# A Provable Generalized Tensor Spectral Method for Uniform Hypergraph Partitioning (Supplementary material)

**Debarghya Ghoshdastidar** DEBARGHYA.G@CSA.IISC.ERNET.IN
**Ambedkar Dukkipati** AD@CSA.IISC.ERNET.IN
Departiment of Computer Science & Automation,
Indian Institute of Science, Bangalore – 560012, India

## A. Proofs of Technical Results

### Proof of Theorem 3

Since, the output clusters (say, $\mathcal{V}_1, \ldots, \mathcal{V}_k$) are disjoint, one can immediately see that the columns of matrix $H$ are orthonormal. The objective function in (4) is

$$\begin{aligned}&\text{Trace}(\mathbf{W} \times_1 H^T \times_2 \ldots \times_m H^T)\\ &= \sum_{j=1}^{k} (\mathbf{W} \times_1 H^T \times_2 \ldots \times_m H^T)_{j\ldots j},\end{aligned}$$

where each term in the summation can be expressed using Definition 2 as

$$\begin{aligned}&(\mathbf{W} \times_1 H^T \times_2 \ldots \times_m H^T)_{j\ldots j}\\ &= \sum_{i_1\ldots i_m=1}^{n} \mathbf{W}_{i_1\ldots i_m} H_{i_1 j} \ldots H_{i_m j}\\ &= \frac{1}{|\mathcal{V}_j|^{m/2}} \sum_{i_1\ldots i_m=1}^{n} \mathbf{W}_{i_1\ldots i_m} \mathbf{1}\{v_{i_1}, \ldots, v_{i_m} \in \mathcal{V}_j\}\\ &= \frac{1}{|\mathcal{V}_j|^{m/2}} \sum_{v_{i_1},\ldots,v_{i_m} \in \mathcal{V}_j} \mathbf{W}_{i_1\ldots i_m} = \frac{\text{Assoc}(\mathcal{V}_j)}{|\mathcal{V}_j|^{m/2}}.\end{aligned}$$

Thus, summing over $j$, we can see that the objective is simply N-associativity$(\mathcal{V}_1, \ldots, \mathcal{V}_k)$.

### Proof of Corollary 4

The relation in (6) follows from the derivation in above proof. The fact that, at stationary, each column $z_j$ is a $\ell_2$-eigenvector can be argued as below. For each $j$,

$$z_j = \underset{\|z\|_2=1}{\text{argmax}} \mathbf{W}(z, \ldots, z)$$

and hence, at $z_j$, the derivative of the Lagrangian is zero, which gives the defining equation of a $\ell_2$-eigenvector

$$\mathbf{W}(\cdot, z_j, \ldots, z_j) = \frac{2\lambda}{m} z_j.$$

### Proof of Corollary 6

Note that $n_{\max} = \frac{n}{k}$. So, $|M_n| \leq \frac{16n^2 \log n}{\delta_n^2}$. We now compute $\delta_n$ in this case, which is simply

$$\delta_n = \frac{p(\frac{n}{k})^{m-1}}{n^{(0.5m-1)}} = \frac{pn^{0.5m}}{k^{m-1}}$$

since $g = 0$ in (15). Substituting $\delta_n$ in the error bound, we have

$$|M_n| \leq \frac{16k^{(2m-2)} n^2 \log n}{p^2 n^m} = O\left(\frac{\log n}{p^2 n^{m-3+\frac{1}{m}}}\right)$$

when $k = O(n^{1/2m})$.

### Proof of Lemma 7

Let

$$\mathcal{A} = \mathcal{W}\left(\cdot, \cdot, \frac{1_n}{\sqrt{n}}, \ldots, \frac{1_n}{\sqrt{n}}\right)$$

be the matrix computed in first step of Algorithm 1. From the structure of $\mathcal{W}$ given in (14), one can write

$$\begin{aligned}\mathcal{A} &= \sum_{j=1}^{k} \frac{p_j (c_j^T 1_n)^{m-2}}{n^{(m-2)/2}} c_j c_j^T + \frac{q(1_n^T 1_n)^{m-2}}{n^{(m-2)/2}} 1_n 1_n^T\\ &= \sum_{j=1}^{k} \frac{p_j n_j^{m-2}}{n^{(m-2)/2}} c_j c_j^T + qn^{(m-2)/2} 1_n 1_n^T,\end{aligned}$$

where the last step follows from the fact that $c_j \in \{0,1\}^n$ with exactly $n_j$ ones. Further noting that $1_n = \sum_j c_j$, we can write $\mathcal{A} = CS^{-1/2}BS^{-1/2}C^T$, where $C = [c_1 \ldots c_k] \in \mathbb{R}^{n\times k}$, and $S = \text{diag}(n_j) \in \mathbb{R}^{k\times k}$ is diagonal with entries being $n_1, \ldots, n_k$. The matrix $B \in \mathbb{R}^{k\times k}$ is

$$B = \text{diag}\left(\frac{p_j n_j^{m-1}}{n^{(0.5m-1)}}\right) + vv^T,$$

where $v \in \mathbb{R}^k$ with $j^{th}$ component being $\sqrt{qn_j n^{(0.5m-1)}}$. We observe that $CS^{-1/2}$ has orthonormal columns, and so, $\mathcal{A}$ is rank-$k$ with non-zero eigenvalues same as that of $B$.

Now, the above representation of $B$ as a rank-one perturbation of a diagonal matrix allows one to bound the smallest eigenvalue of $B$ using (Ipsen & Nadler, 2009). Note that this bound is tighter than Weyl's inequality, which does not provide useful bounds in certain situations, for instance, when $p_k = 0$. From Theorem 2.1 in (Ipsen & Nadler, 2009), one can claim that the smallest eigenvalue of $B$,

$$\begin{aligned}\lambda_k(B) \geq \frac{p_{j^*} n_{j^*}^{m-1}}{n^{(0.5m-1)}} + \\ \frac{g + qn^{0.5m} - \sqrt{(g+qn^{0.5m})^2 - 4gqn_j n^{(0.5m-1)}}}{2} \\ \geq \frac{p_{j^*} n_{j^*}^{m-1}}{n^{(0.5m-1)}} + \frac{gqn_{j^*} n^{(0.5m-1)}}{g + qn^{0.5m}},\end{aligned}$$

where $j^*$ and $g$ are as in (15). The second inequality is obtained by observing that $(a - \sqrt{a^2 - b}) = \frac{b}{a + \sqrt{a^2 - b}} \geq \frac{b}{2a}$ for $a, b \geq 0$. The above lower bound is defined as $2\delta_n$ (15).

Thus, whenever $\lambda_k(B) \geq 2\delta_n > 0$, the largest $k$ eigenvalues of $\mathcal{A}$ are strictly positive with eigen-gap between non-zero spectrum and the eigenvalue 0 being at least $2\delta_n$. Further, the corresponding eigenspace is spanned by the columns of $CS^{-1/2}$, which implies that two rows of the eigenvector matrix $\mathcal{Z}$ are equal only when corresponding rows of $C$ are identical, *i.e.*, the nodes lie in same partition.

**Proof of Lemma 9**

When Algorithm 1 is run on a random hypergraph, one computes eigen-decomposition of a random matrix $A$ instead of $\mathcal{A}$. Considering $\mathbf{W} = \mathcal{W} + \mathcal{E}$, the perturbation of $A$ from $\mathcal{A}$ in terms of the spectral norm can be bounded as

$$\begin{aligned}\|A - \mathcal{A}\|_2 &\leq \left\| \mathcal{E}\left(\cdot, \cdot, \frac{1_n}{\sqrt{n}}, ..., \frac{1_n}{\sqrt{n}}\right)\right\|_2 \\ &= \max_{\|x_1\|_2 = \|x_2\|_2 = 1} \left| \mathcal{E}\left(x_1, x_2, \frac{1_n}{\sqrt{n}}, ..., \frac{1_n}{\sqrt{n}}\right)\right| \\ &\leq \|\mathcal{E}\|_{op}\end{aligned}$$

since the other argument above is $\frac{1_n}{\sqrt{n}}$ that has unit $\ell_2$-norm.

We can combine Davis-Kahan $\sin\Theta$ theorem (see Proposition 2.1 in (Rohe et al., 2011)) with above bound to obtain

$$\|\sin\Theta(Z, \mathcal{Z})\|_2 \leq \frac{2\|A - \mathcal{A}\|_2}{\lambda_k(B)} \leq \frac{\|\mathcal{E}\|_{op}}{\delta_n}, \quad (19)$$

when $\|\mathcal{E}\|_{op} < \delta_n$. Here $\sin\Theta(Z, \mathcal{Z})$ is a matrix of the sines of the canonical angles between the subspaces spanned by columns of $Z$ and $\mathcal{Z}$, and $\lambda_k(B) \geq 2\delta_n$ is the eigen-gap between largest $k$ eigenvalues of $\mathcal{A}$ and the remaining spectrum.

A more convenient form of the above bound is required for our purpose. Let $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_k$ denote the canonical angles, all of which lie in $[0, \pi/2]$. Then $\sin\Theta(Z, \mathcal{Z})\|_2 = \sin\theta_1$. Now, the singular values of the matrix $Z^T\mathcal{Z}$ are cosines of the above angles. Let $Z^T\mathcal{Z} = U\Sigma V^T$ be the svd of $Z^T\mathcal{Z}$, *i.e.*, the diagonal entries of $\Sigma$ are $\cos\theta_j$, $j = 1, \ldots, k$. Then

$$\begin{aligned}&\|Z - \mathcal{Z}VU^T\|_F^2 \\ &= \mathrm{Trace}\left((Z - \mathcal{Z}VU^T)^T(Z - \mathcal{Z}VU^T)\right) \\ &= 2\sum_{j=1}^k (1 - \cos\theta_j) \leq 2\sum_{j=1}^k (1 - \cos^2\theta_j)\end{aligned}$$

since $\cos^2\theta_j \leq \cos\theta_j$. Thus,

$$\|Z - \mathcal{Z}VU^T\|_F^2 \leq 2\sum_{j=1}^k \sin^2\theta_j \leq 2k\sin^2\theta_1.$$

Note that $Q = VU^T \in \mathbb{R}^{k\times k}$ is an orthonormal (rotation) matrix. The result follows by combining above inequality with the Davis-Kahan perturbation bound (19).

**Proof of Lemma 10**

From Definition 8, we have

$$\|\mathcal{E}\|_{op} = \max_{\|x\|_2 = 1} |\mathcal{E}(x, \ldots, x)| .$$

Thus, we need to find a bound on

$$\mathsf{P}\left(\max_{\|x\|_2 = 1} |\mathcal{E}(x, \ldots, x)| > \lambda\right),$$

where the maximum is taken over the unit ball in $\mathbb{R}^n$. Since, the maximum is over an uncountable set, a direct use of union bound does not yield reasonable bound. This is taken care of by using an $\epsilon$-net argument.

Let $S_\epsilon$ be a maximal $\epsilon$-net on the unit ball in $\mathbb{R}^n$, *i.e.*, for any $x, y \in S_\epsilon$, $\|x - y\|_2 > \epsilon$ and for any $x \notin S_\epsilon$ with $\|x\|_2$, $S_\epsilon \cup \{x\}$ is not an $\epsilon$-net. It is easy to see that such a maximal $\epsilon$-net always exists, and its size $|S_\epsilon| \leq (\frac{2}{\epsilon} + 1)^n$.

We claim that, choosing $\epsilon = \frac{1}{m}$,

$$\mathsf{P}(\|\mathcal{E}\|_{op} > \lambda) \leq \sum_{x \in S_\epsilon} \mathsf{P}(|\mathcal{E}(x, \ldots, x)| > \lambda/m). \quad (20)$$

To prove (20), it suffices to show that whenever $\|\mathcal{E}\|_{op} > \lambda$, there exists some $x \in S_\epsilon$ such that $|\mathcal{E}(x, \ldots, x)| > \frac{\lambda}{m}$. If this holds, then (20) follows from union bound.

Note that there exists some $y$ in the unit ball which achieves the maximum, *i.e.*, $|\mathcal{E}(y, \ldots, y)| = \|\mathcal{E}\|_{op}$. Given $\|\mathcal{E}\|_{op} > \lambda$, if $y \in S_\epsilon$, then the above condition trivially holds. If

$y \notin S_\epsilon$, then by maximality of $S_\epsilon$, there exists $x \in S_\epsilon$ such that $\|y-x\|_2 \leq \epsilon = \frac{1}{m}$. Now, we can write

$$\mathcal{E}(y,\ldots,y) = \mathcal{E}(x,\ldots,x) + \sum_{i=1}^{m-1} \mathcal{E}(x,..,x,y\text{-}x,y,..,y),$$

where the $i^{th}$ $m$-linear functional in the sum contains $(i-1)$ arguments as $x$, one argument as $(y-x)$, and $(m-i)$ arguments as $y$. One can verify that for the $i^{th}$ term

$$\begin{aligned}\mathcal{E}(x,..,x,y-x,y,..,y) &\leq \|\mathcal{E}\|_{op}\|x\|_2^{i-1}\|y-x\|_2\|y\|_2^{m-i} \\ &\leq \frac{\|\mathcal{E}\|_{op}}{m}\end{aligned}$$

since $\|x\|_2 = \|y\|_2 = 1$. Thus,

$$\|\mathcal{E}\|_{op} = |\mathcal{E}(y,\ldots,y)| \leq |\mathcal{E}(x,\ldots,x)| + \frac{(m-1)\|\mathcal{E}\|_{op}}{m}.$$

Hence, we have $|\mathcal{E}(x,\ldots,x)| \geq \frac{1}{m}\|\mathcal{E}\|_{op} > \frac{\lambda}{m}$.

Next, we derive an upper bound for

$$\mathsf{P}(|\mathcal{E}(x,\ldots,x)| > \lambda/m),$$

where $x \in \mathbb{R}^n$ is arbitrary with $\|x\|_2 = 1$. Recall that $\mathcal{E} = \mathbf{W} - \mathsf{E}[\mathbf{W}]$. Hence, each entry of $\mathcal{E}$ has mean zero and varies over an unit length interval. But all the entries are not independent due to the symmetric structure. However, the random variables $\{\mathcal{E}_{i_1 i_2 \ldots i_m} : i_1 \leq i_2 \leq \ldots \leq i_m\}$ are independent. So for any $i_1 \leq i_2 \leq \ldots \leq i_m$, we define

$$\overline{\mathcal{E}}_{\{i_1,i_2,\ldots,i_m\}} = N_{\{i_1,i_2,\ldots,i_m\}}\mathcal{E}_{i_1 i_2 \ldots i_m},$$

where $N_{\{i_1,i_2,\ldots,i_m\}}$ is the number of possible permutations of $(i_1,i_2,\ldots,i_m)$ that are distinct in the sense of $n$-tuples. Note that $N_{\{i_1,i_2,\ldots,i_m\}} \leq m!$ for any $i_1,i_2,\ldots,i_m$. One can exploit the symmetry of $\mathcal{E}$ to write

$$\begin{aligned}\mathcal{E}(x,\ldots,x) &= \sum_{i_1,i_2,\ldots,i_m=1}^{n} \mathcal{E}_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \ldots x_{i_m} \\ &= \sum_{i_1 \leq i_2 \leq \ldots \leq i_m} \overline{\mathcal{E}}_{\{i_1,i_2,\ldots,i_m\}} x_{i_1} x_{i_2} \ldots x_{i_m},\end{aligned}$$

where $\{(\overline{\mathcal{E}}_{\{i_1,\ldots,i_m\}} x_{i_1} \ldots x_{i_m}) : i_1 \leq \ldots \leq i_m\}$ is a collection of mean zero independent random variables, each varying over an interval of length $|N_{\{i_1,\ldots,i_m\}} x_{i_1} \ldots x_{i_m}|$.

The above representation shows that one can directly use Hoeffding's inequality to find a bound on the tail probability for $|\mathcal{E}(x,\ldots,x)|$ as

$$\begin{aligned}&\mathsf{P}(|\mathcal{E}(x,\ldots,x)| > \lambda/m) \\ &\leq 2\exp\left(-\frac{2(\lambda/m)^2}{\sum\limits_{i_1 \leq \ldots \leq i_m} N^2_{\{i_1,\ldots,i_m\}} x^2_{i_1} \ldots x^2_{i_m}}\right) \\ &\leq 2\exp\left(-\frac{2(\lambda/m)^2}{m! \sum\limits_{i_1 \leq \ldots \leq i_m} N_{\{i_1,\ldots,i_m\}} x^2_{i_1} \ldots x^2_{i_m}}\right) \\ &= 2\exp\left(-\frac{2(\lambda/m)^2}{m! \sum\limits_{i_1,\ldots,i_m=1}^{n} x^2_{i_1} \ldots x^2_{i_m}}\right)\end{aligned}$$

since $N_{\{i_1,\ldots,i_m\}}$ accounts for all the repetitions. Observe that the sum in the denominator is $\|x\|_2^{2m} = 1$ since $x \in S_\epsilon$. Combining above bound with (20) and the fact that $S_\epsilon = (2m+1)^n$, we obtain the result.

## B. Experiments on Motion Segmentation

In this section, we discuss an application of subspace clustering in the problem of motion segmentation. We conduct experiments on the Hopkins 155 motion segmentation database (Tron & Vidal, 2007). We consider 120 videos that contain two independent affine motions. The videos contain feature trajectories, each of which can be viewed as a high-dimensional vector. It is known that the trajectories associated with same rigid body lie in an affine space of dimension at most 3. Thus, each dataset is a union of vectors from two affine subspaces. We try to fit 2-dimensional affine spaces through the data, and use $4^{th}$-order affinity tensors. The similarity among 4 vectors is $e^{-\beta f(\cdot)}$, where $\beta > 0$ is a tuning parameter and $f(\cdot)$ is the error of fitting a 2-dimensional affine space through the 4 vectors, given by the sum of smallest two squared singular values of the centered data matrix for the 4 vectors. To reduce the complexity, we use only sample $1000N$ entries of the tensor for Algorithm 1, where $N$ is the number of vectors in each dataset. This is done by the sampling suggested in Section 5.2. We approximate $A$ in Algorithm 1 as a sum of 80 matrices, each obtained by randomly selecting a pairs of vectors. We further compute only 25 columns for each matrix, thus maintaining the sampling rate. We also tune $\beta$ and average the result over 20 trials, and report best result over a set of values of parameter $\beta$ in the range $[0.001, 1]$, as considered in (Ghoshdastidar & Dukkipati, 2014; 2015).

Table 3 reports the performance of some subspace clustering algorithms that are popular in the vision community. These include local subspace affinity (LSA), spectral curvature clustering (SCC), low rank representation with heuristic post-processing (LRR-H), low rank sub-

| Algorithm | Mean error (%) |
|---|---|
| LSA | 4.23 |
| SCC | 2.89 |
| LRR-H | 2.13 |
| LRSC | 3.69 |
| SSC | 1.52 |
| SGC | 1.03 |
| HOSVD | 1.83 |
| $\text{HOSVD}_{\text{sampled}}$ | 1.05 |
| Algorithm 1 | 1.54 |

*Table 3.* Error for different algorithms. Results for the other algorithms have been taken from (Ghoshdastidar & Dukkipati, 2015).

space clustering (LRSC), sparse subspace clustering (SSC), sparse Grassmann clustering (SGC), tensor HOSVD based clustering (HOSVD) and its column sampled variant ($\text{HOSVD}_{\text{sampled}}$). Table 3 shows that Algorithm 1 performs better than most approaches, while the best results are obtained by SGC (Jain & Govindu, 2013) and $\text{HOSVD}_{\text{sampled}}$ (Ghoshdastidar & Dukkipati, 2015). which consider improved sampling techniques for the HOSVD algorithm. Thus, improving the sampling for Algorithm 1 requires further study.