Fast and Simple Spectral Clustering in Theory and Practice Appendix

Anonymous Author(s)
Affiliation
Address
email

A Omitted detail from Section 3

In this section, we prove the main theoretical result of the paper. In order that this section is self-contained, we repeat some of the steps included in the main paper. We first show that the lengths of the random vectors $\boldsymbol{x}_i$ generated in Algorithm 2 are close to their expected value. Notice that $\mathrm{E}\left[\|\boldsymbol{x}_i\|_2\right] = \sqrt{n}$ and $\mathrm{E}\left[\|\mathbf{P}\boldsymbol{x}_i\|_2\right] = \sqrt{k}$. We use Chebyshev's inequality to show the following.

Lemma A.1. *Let $\boldsymbol{x} \in \mathbb{R}^n$ be drawn from the n-dimensional Gaussian distribution $\mathrm{N}(\mathbf{0}, \mathbf{I})$. Let $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k \in \mathbb{R}^n$ be orthogonal vectors and let $\mathbf{P} = \sum_{i=1}^{k} \boldsymbol{f}_i \boldsymbol{f}_i^{\intercal}$ be the projection onto the space spanned by $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$. With probability at least $1 - (1/10k)$,*

- $\|\mathbf{P}\boldsymbol{x}\|_2 \leq \sqrt{6k}$, *and*
- $\|\boldsymbol{x}\|_2 \leq \sqrt{6n}$.

Proof of Lemma A.1. Since $\boldsymbol{x}$ is drawn from a symmetric n-dimensional Gaussian distribution, $\|\boldsymbol{x}\|_2^2$ is distributed according to a χ^2 distribution with n degrees of freedom. Similarly, since $\mathbf{P}$ is a projection matrix, $\|\mathbf{P}\boldsymbol{x}\|_2^2$ is distributed according to a χ^2 distribution with k degrees of freedom. By the Chebyshev inequality, we have that

$$\Pr\left[\|\mathbf{P}\boldsymbol{x}_i\|_2^2 \geq 6k\right] \leq \frac{k}{(5k)^2} = \frac{1}{25k},$$

and

$$\Pr\left[\|\boldsymbol{x}_i\|_2^2 \geq 6n\right] \leq \frac{n}{(5n)^2} = \frac{1}{25n}.$$

The lemma follows by the union bound and since $k \leq n$. ∎

We now show that the output of the PowerMethod algorithm is close to a random vector in the space spanned by $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$.

Lemma 3.1. *Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be a matrix with eigenvalues $1 \geq \gamma_1 \geq \ldots \geq \gamma_n \geq 0$ and corresponding eigenvectors $\boldsymbol{f}_1, \ldots \boldsymbol{f}_n$. Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ be drawn from the n-dimensional Gaussian distribution $N(\mathbf{0}, \mathbf{I})$. Let $\boldsymbol{x}_t = \textsc{PowerMethod}(\mathbf{M}, \boldsymbol{x}_0, t)$ for $t = \Theta\big(\log(n/\epsilon^2 k)\big)$. If $\gamma_k \geq 1 - O\big(\epsilon \cdot \log(n/\epsilon^2 k)^{-1}\big)$ and $\gamma_{k+1} \leq 1 - \Omega(1)$, then with probability at least $1 - (1/10k)$,*

$$\|\boldsymbol{x}_t - \mathbf{P}\boldsymbol{x}_0\|_2 \leq \epsilon\sqrt{k},$$

where $P = \sum_{i=1}^{k} \boldsymbol{f}_i \boldsymbol{f}_i^{\intercal}$ is the projection onto the space spanned by the first k eigenvectors of $\mathbf{M}$.

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

Proof of Lemma 3.1. By the assumptions of the Lemma, we can assume that

- $\gamma_{k+1} \le c_1 < 1$,
- $\gamma_k \ge 1 - c_2 \epsilon \log(24n/\epsilon^2 k)^{-1}$, and
- $t = c_3 \log(24n/\epsilon^2 k)$,

for constants c_1, c_2, and c_3. Fixing $c_1 < 1$, we will set

$$c_3 = \frac{1}{2\log\left(\frac{1}{c_1}\right)}$$

and

$$c_2 = \frac{1}{c_3 \cdot 2\sqrt{6}}.$$

Furthermore, by Lemma A.1, with probability at least $1 - (1/10k)$ it holds that

$$\|\mathbf{P}\boldsymbol{x}_0\|_2 \le \sqrt{6k}$$

and

$$\|\boldsymbol{x}_0\|_2 \le \sqrt{6n},$$

and we assume that this holds in the remainder of the proof.

Now, we write $\boldsymbol{x}_0$ in terms of its expansion in the basis given by the eigenvectors $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$:

$$\boldsymbol{x}_0 = \sum_{j=1}^{n} a_j \boldsymbol{f}_j,$$

where $a_j = \langle \boldsymbol{x}_0, \boldsymbol{f}_j \rangle$. Similarly, we have

$$\mathbf{P}\boldsymbol{x}_0 = \sum_{j=1}^{k} a_j \boldsymbol{f}_j$$

and

$$\boldsymbol{x}_t = \sum_{j=1}^{n} a_j \gamma_j^t \boldsymbol{f}_j.$$

Then,

$$\begin{aligned}
\|\boldsymbol{y}_i - \mathbf{P}\boldsymbol{x}_0\|_2 &= \left\| \sum_{j=1}^{k} \left(a_j \gamma_j^t - a_j\right) \boldsymbol{f}_j + \sum_{j=k+1}^{n} a_j \gamma_j^t \boldsymbol{f}_j \right\|_2 \\
&\le \left\| \sum_{j=1}^{k} \left(a_j \gamma_j^t - a_j\right) \boldsymbol{f}_j \right\|_2 + \left\| \sum_{j=k+1}^{n} a_j \gamma_j^t \boldsymbol{f}_j \right\|_2 \\
&\le \left\| \left(1 - \gamma_k^t\right) \sum_{j=1}^{k} a_j \boldsymbol{f}_j \right\|_2 + \left\| \gamma_{k+1}^t \sum_{j=k+1}^{n} a_j \boldsymbol{f}_j \right\|_2 \\
&= \left(1 - \gamma_k^t\right) \|\mathbf{P}\boldsymbol{x}_0\|_2 + \gamma_{k+1}^t \|(\mathbf{I} - \mathbf{P})\, \boldsymbol{x}_0\|_2 \\
&\le \left(1 - \gamma_k^t\right) \|\mathbf{P}\boldsymbol{x}_0\|_2 + \gamma_{k+1}^t \|\boldsymbol{x}_0\|_2
\end{aligned}$$

where we used the fact that $1 \ge \gamma_1 \ge \ldots \ge \gamma_n$. Now, we have

$$\begin{aligned}
\gamma_k^t &\ge \left(1 - c_2 \epsilon \log(24n/\epsilon^2 k)^{-1}\right)^{c_3 \log(24n/\epsilon^2 k)} \\
&\ge 1 - c_2 c_3 \epsilon \\
&= 1 - \frac{\epsilon}{2\sqrt{6}}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\gamma_{k+1}^{t} &\leq c_1^{c_3 \log(24n/\epsilon^2 k)} \\
&= \left(\frac{1}{c_1}\right)^{c_3 \log(\epsilon^2 k/24n)} \\
&= \left(\frac{\epsilon^2 k}{24n}\right)^{c_3 \log(1/c_1)} \\
&= \epsilon\sqrt{\frac{k}{24n}}.
\end{aligned}$$

Combining everything together, we have

$$\begin{aligned}
\|\boldsymbol{y}_i - \mathbf{P}\boldsymbol{x}_0\|_2 &\leq \frac{\epsilon}{2\sqrt{6}}\,\|\mathbf{P}\boldsymbol{x}_0\|_2 + \epsilon\sqrt{\frac{k}{24n}}\,\|\boldsymbol{x}_0\|_2 \\
&\leq \frac{\epsilon}{2\sqrt{6}}\sqrt{6k} + \epsilon\sqrt{\frac{6kn}{24n}} \\
&\leq \epsilon\sqrt{k},
\end{aligned}$$

which completes the proof. ∎

It remains to prove that the k-means cost is preserved in the embedding produced by the power method. Recall that $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$ are the eigenvectors of $\mathbf{M}$ corresponding to the eigenvalues $\gamma_1, \ldots, \gamma_k$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l$ are the vectors computed in Algorithm 2. We will also consider the vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_l$ given by $\boldsymbol{z}_i = \mathbf{P}\boldsymbol{x}_i$, where $\{\boldsymbol{x}_i\}_{i=1}^{k}$ are the random vectors sampled in Algorithm 2, and $\mathbf{P} = \sum_{i=1}^{k} \boldsymbol{f}_i \boldsymbol{f}_i^{\intercal}$ is the projection onto the space spanned by $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$. We also define

$$\mathbf{F} = \begin{bmatrix} \vert & & \vert \\ \boldsymbol{f}_1 & \ldots & \boldsymbol{f}_k \\ \vert & & \vert \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \vert & & \vert \\ \boldsymbol{y}_1 & \ldots & \boldsymbol{y}_l \\ \vert & & \vert \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} \vert & & \vert \\ \boldsymbol{z}_1 & \ldots & \boldsymbol{z}_l \\ \vert & & \vert \end{bmatrix}.$$

We will use the following result shown by Makarychev et al. [22].

Lemma 3.2 ([22], Theorem 1.3)**.** *Given data $\mathbf{X} \in \mathbb{R}^{n \times k}$, let $\mathbf{\Pi} \in \mathbb{R}^{k \times l}$ be a random matrix with each column sampled from the k-dimensional Gaussian distribution $\mathrm{N}(\mathbf{0}, \mathbf{I}_k)$ and*

$$l = O\left(\frac{\log(k) + \log(1/\epsilon)}{\epsilon^2}\right).$$

Then, with probability at least $1 - \epsilon$, it holds for all partitions $\{A_i\}_{i=1}^{k}$ of $[n]$ that

$$\mathrm{COST}_{\mathbf{X}}(A_1, \ldots, A_k) \in (1 \pm \epsilon)\mathrm{COST}_{\mathbf{X\Pi}}(A_1, \ldots, A_k).$$

Applying this lemma with $\mathbf{X} = \mathbf{D}^{-\frac{1}{2}}\mathbf{F}$ and $\mathbf{\Pi} = \mathbf{F}^{\intercal}\mathbf{Z}$ shows that the k-means cost is approximately equal in the embeddings given by $\mathbf{D}^{-\frac{1}{2}}\mathbf{F}$ and $\mathbf{D}^{-\frac{1}{2}}\mathbf{Z}$, since $\mathbf{F}\mathbf{F}^{\intercal}\mathbf{Z} = \mathbf{Z}$ and each of the entries of $\mathbf{F}^{\intercal}\mathbf{Z}$ is distributed according to the Gaussian distribution $\mathrm{N}(0, 1)$. By Lemma 3.1, we can also show that the k-means objective in $\mathbf{D}^{-\frac{1}{2}}\mathbf{Y}$ is within an additive error of $\mathbf{D}^{-\frac{1}{2}}\mathbf{Z}$. This allows us to prove the following lemma.

Lemma 3.3. *With probability at least $0.9 - \epsilon$, for any partitioning $\{A_i\}_{i=1}^{k}$ of the vertex set V, we have*

$$\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Y}}(A_1, \ldots, A_k) \geq (1 - \epsilon)\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{F}}(A_1, \ldots, A_k) - \epsilon k$$

and

$$\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Y}}(A_1, \ldots, A_k) \leq (1 + \epsilon)\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{F}}(A_1, \ldots, A_k) + \epsilon k.$$

In order to prove this, we will use the fact shown by Boutsidis and Magdon-Ismail [5] that we can write the k-means cost as

$$\mathrm{COST}_{\mathbf{B}}(A_i, \ldots, A_k) = \|\mathbf{B} - \mathbf{X}\mathbf{X}^{\intercal}\mathbf{B}\|_F^2 \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{n \times k}$ is the indicator matrix of the partition, defined by

$$\mathbf{X}(u,i) = \begin{cases} \frac{1}{\sqrt{|A_i|}} & \text{if } u \in A_i \\ 0 & \text{otherwise} \end{cases},$$

and $\|\mathbf{B}\|_F \triangleq (\sum_{i,j} \mathbf{B}_{i,j}^2)^{1/2}$ is the Frobenius norm.

Proof of Lemma 3.3. Notice that $\mathbf{F}^\intercal \mathbf{Z} \in \mathbb{R}^{k \times l}$ is a random matrix with columns drawn from the standard k-dimensional Gaussian distribution. Then, by Lemma 3.2, with probability at least $1-\epsilon$, we have for any partition $\{A_i\}_{i=1}^k$ that

$$\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{F}}(A_1, \ldots, A_k) \in (1 \pm \epsilon)\, \mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Z}}(A_1, \ldots, A_k) \tag{2}$$

since $\mathbf{D}^{-1/2}\mathbf{F}\mathbf{F}^\intercal\mathbf{Z} = \mathbf{D}^{-1/2}\mathbf{Z}$, where we use the fact that the columns of $\mathbf{Z}$ are in the span of $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$.

Furthermore, by the union bound, we can assume with probability at least 0.9 that the conclusion of Lemma 3.1 holds for every vector $\boldsymbol{y}_i$ computed by Algorithm 2.

Now, we will establish that $\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Y}}(\cdot)$ is close to $\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Z}}(\cdot)$ which will complete the proof. For some arbitrary partition $\{A_i\}_{i=1}^k$, let $\mathbf{X}$ be the indicator matrix of the partition. Then, we have

$$\begin{aligned}
&\left\|\mathbf{D}^{-1/2}\mathbf{Y} - \mathbf{X}\mathbf{X}^\intercal\mathbf{D}^{-1/2}\mathbf{Y}\right\|_F - \left\|\mathbf{D}^{-1/2}\mathbf{Z} - \mathbf{X}\mathbf{X}^\intercal\mathbf{D}^{-1/2}\mathbf{Z}\right\|_F \\
&= \left\|(\mathbf{I} - \mathbf{X}\mathbf{X}^\intercal)\,\mathbf{D}^{-1/2}\mathbf{Y}\right\|_F - \left\|(\mathbf{I} - \mathbf{X}\mathbf{X}^\intercal)\,\mathbf{D}^{-1/2}\mathbf{Z}\right\|_F \\
&\leq \left\|(\mathbf{I} - \mathbf{X}\mathbf{X}^\intercal)\,\mathbf{D}^{-1/2}\,(\mathbf{Y} - \mathbf{Z})\right\|_F \\
&\leq \left\|(\mathbf{I} - \mathbf{X}\mathbf{X}^\intercal)\,(\mathbf{Y} - \mathbf{Z})\right\|_F \\
&\leq \|\mathbf{Y} - \mathbf{Z}\|_F \\
&= \sqrt{\sum_{i=1}^{l} \|\boldsymbol{y}_i - \boldsymbol{z}_i\|_2^2} \\
&\leq \sqrt{l\epsilon^2 k} \\
&\leq \epsilon k,
\end{aligned}$$

Where we use Lemma 3.1, and the fact that $l \leq k$. Combining this with (2) completes the proof. ■

Now we come to the proof of the main theorem.

Theorem 3.1. *Let G be a graph with $\lambda_{k+1} = \Omega(1)$ and $\rho(k) = O\big(\epsilon \cdot \log(n/\epsilon)^{-1}\big)$. Additionally, let $\{S_i\}_{i=1}^k$ be the k-way partition corresponding to $\rho(k)$ and suppose that $\{S_i\}_{i=1}^k$ are almost balanced. Let $\{A_i\}_{i=1}^k$ be the output of Algorithm 2. With probability at least $0.9-\epsilon$, there exists a permutation $\sigma : [k] \to [k]$ such that*

$$\sum_{i=1}^{k} \mathrm{vol}(A_i \triangle S_{\sigma(i)}) = O(\epsilon \cdot \mathrm{vol}(V_G))\,.$$

Moreover, the running time of Algorithm 2 is

$$\widetilde{O}\big(m \cdot \epsilon^{-2}\big) + T_{\mathrm{KM}}(n, k, l),$$

where m is the number of edges in G and $T_{\mathrm{KM}}(n,k,l)$ is the running time of the k-means approximation algorithm on n points in l dimensions.

To complete the proof, we will make use of the following results proved by Macgregor and Sun [21], which hold under the same assumptions as Theorem 3.1.

Lemma 3.4 ([21], Lemma 4.1)**.** *There exists a partition $\{A_i\}_{i=1}^k$ of the vertex set V such that*

$$\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{F}}(A_1, \ldots A_k) < \epsilon \cdot k.$$

Lemma 3.5 ([21], Theorem 2)**.** *Given some partition of the vertices,* $\{A_i\}_{i=1}^k$*, such that*

$$\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{F}}(A_1, \ldots A_k) \leq ck,$$

then there exists a permutation $\sigma : [k] \to [k]$ *such that*

$$\sum_{i=1}^{k} \mathrm{vol}(A_i \triangle S_{\sigma(i)}) = O(c \cdot \mathrm{vol}(V))\,.$$

Proof of Theorem 3.1. By Lemma 3.4 and Lemma 3.3, with probability at least $0.9 - \epsilon$, there exists some partition $\{\widehat{A}_i\}_{i=1}^k$ of the vertex set V_G such that $\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Y}}(\widehat{A}_1, \ldots, \widehat{A}_k) \leq (1+\epsilon)\epsilon k + \epsilon k \leq 3k\epsilon$. Since we use a constant-factor approximation algorithm for k-means, the partition $\{A_i\}_{i=1}^k$ returned by Algorithm 2 satisfies $\mathrm{COST}_{\mathbf{D}^{-1/2}\mathbf{Y}}(A_1, \ldots, A_k) = O(\epsilon k)$. Then, by Lemma 3.5, for some permutation $\sigma : [k] \to [k]$, we have

$$\sum_{i=1}^{k} \mathrm{vol}(A_i \triangle S_{\sigma(i)}) = O(\epsilon \cdot \mathrm{vol}(V_G))\,.$$

To bound the running time, notice that the number of non-zero entries in $\mathbf{M}$ is $2m$, and the time complexity of matrix multiplication is proportional to the number of non-zero entries. Therefore, the running time of $\textsc{PowerMethod}(\mathbf{M}, \boldsymbol{x}_0, t)$ is $\widetilde{O}(m)$. Since the loop in Algorithm 2 is executed $\Theta\big(\log(k) \cdot \epsilon^{-2}\big)$ times, the total running time of Algorithm 2 is $\widetilde{O}\big(m \cdot \epsilon^{-2}\big) + T_{\mathrm{KM}}(n, k, l)$. ∎