Supplementary material for k-means clustering using random matrix sparsification

Kaushik Sinha [1]

1. Proof of auxiliary lemmas

1.1. Proof of Lemma 1

Proof. Let $\tilde{\mathbf{A}}_{r-m} = \tilde{\mathbf{A}} - \tilde{\mathbf{A}}_m$ (we have assumed that rank of $\tilde{\mathbf{A}}$ is r and $r > m$). First note that $\tilde{\mathbf{A}}_m$ and $\tilde{\mathbf{A}}_{r-m}$ are orthogonal to each other. To see this write $\tilde{\mathbf{A}}$ using singular value decomposition as $\tilde{\mathbf{A}} = \sum_{i=1}^r \sigma_i(\tilde{\mathbf{A}})\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^\top$, where $\tilde{\mathbf{u}}_i$ and $\tilde{\mathbf{v}}_i$ are left and right singular vectors of $\tilde{\mathbf{A}}$ respectively, and $\sigma_1(\tilde{\mathbf{A}}) \geq \sigma_2(\tilde{\mathbf{A}}) \geq \cdots \geq \sigma_r(\tilde{\mathbf{A}})$ are singular values of $\tilde{\mathbf{A}}$. Clearly, $\tilde{\mathbf{A}}_m = \sum_{i=1}^m \sigma_i(\tilde{\mathbf{A}})\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^\top$ and $\tilde{\mathbf{A}}_{r-m} = \sum_{i=m+1}^r \sigma_i(\tilde{\mathbf{A}})\tilde{\mathbf{u}}_i\tilde{\mathbf{v}}_i^\top$. Therefore, $\tilde{\mathbf{A}}_m\tilde{\mathbf{A}}_{r-m}^\top = \sum_{i=1}^m\sum_{j=m+1}^r \sigma_i(\tilde{\mathbf{A}})\sigma_j(\tilde{\mathbf{A}})\tilde{\mathbf{u}}_i(\tilde{\mathbf{v}}_i^\top\tilde{\mathbf{v}}_j)\tilde{\mathbf{u}}_j^\top = \mathbf{0}$. Similarly, $\tilde{\mathbf{A}}_{r-m}\tilde{\mathbf{A}}_m^\top = 0$.

Due to orthogonality of $\tilde{\mathbf{A}}_m$ and $\tilde{\mathbf{A}}_{r-m}$, using Pythagorean theorem, it holds that $\|\tilde{\mathbf{A}}-\mathbf{P}\tilde{\mathbf{A}}\|_F^2 = \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}\|_F^2 = \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}_m\|_F^2 + \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}_{r-m}\|_F^2$ for any rank k projection matrix $\mathbf{P}$. To see this, let $\mathbf{Y} = \mathbf{I} - \mathbf{P}$. Then,

$$
\begin{aligned}
&\|\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}}\|_F^2 \\
&= \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}\|_F^2 = \|\mathbf{Y}\tilde{\mathbf{A}}\|_F^2 = \|\mathbf{Y}(\tilde{\mathbf{A}}_m + \tilde{\mathbf{A}}_{r-m})\|_F^2 \\
&\stackrel{a}{=} \text{trace}\left(\mathbf{Y}(\tilde{\mathbf{A}}_m + \tilde{\mathbf{A}}_{r-m})\left(\mathbf{Y}(\tilde{\mathbf{A}}_m + \tilde{\mathbf{A}}_{r-m})\right)^\top\right) \\
&= \text{trace}\left(\mathbf{Y}(\tilde{\mathbf{A}}_m + \tilde{\mathbf{A}}_{r-m})(\tilde{\mathbf{A}}_m + \tilde{\mathbf{A}}_{r-m})^\top\mathbf{Y}^\top\right) \\
&= \text{trace}\Big(\mathbf{Y}\tilde{\mathbf{A}}_m\tilde{\mathbf{A}}_m^\top\mathbf{Y}^\top + \mathbf{Y}\tilde{\mathbf{A}}_{r-m}\tilde{\mathbf{A}}_{r-m}^\top\mathbf{Y}^\top + \\
&\qquad \mathbf{Y}\tilde{\mathbf{A}}_m\tilde{\mathbf{A}}_{r-m}^\top\mathbf{Y}^\top + \mathbf{Y}\tilde{\mathbf{A}}_{r-m}\tilde{\mathbf{A}}_m^\top\mathbf{Y}^\top\Big) \\
&\stackrel{b}{=} \text{trace}\left(\mathbf{Y}\tilde{\mathbf{A}}_m\tilde{\mathbf{A}}_m^\top\mathbf{Y}^\top\right) + \text{trace}\left(\mathbf{Y}\tilde{\mathbf{A}}_{r-m}\tilde{\mathbf{A}}_{r-m}^\top\mathbf{Y}^\top\right) \\
&= \|\mathbf{Y}\tilde{\mathbf{A}}_m\|_F^2 + \|\mathbf{Y}\tilde{\mathbf{A}}_{r-m}\|_F^2 \\
&= \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}_m\|_F^2 + \|(\mathbf{I}-\mathbf{P})\tilde{\mathbf{A}}_{r-m}\|_F^2
\end{aligned}
$$

where, equality a follows from the fact that for any matrix $\mathbf{B}$, $\|\mathbf{B}\|_F^2 = \text{trace}(\mathbf{B}\mathbf{B}^\top)$ and equality b follows from linearity of trace and orthogonality of $\tilde{\mathbf{A}}_m$ and $\tilde{\mathbf{A}}_{r-m}$. Since any rank k projection matrix $\mathbf{P}$ can be written as $\mathbf{P} = \mathbf{Q}\mathbf{Q}^\top$, where $\mathbf{Q}$ is a matrix of k orthonormal columns, $\mathbf{P}\tilde{\mathbf{A}}$ is a rank k matrix spanned by the columns of $\mathbf{Q}$. Clearly, $(\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}})$ is spanned by $(r-k)$ orthonormal columns (which are orthogonal to the columns of $\mathbf{Q}$ since $\tilde{\mathbf{A}}$ has rank r). Therefore, $\mathbf{P}\tilde{\mathbf{A}}$ and $(\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}})$ are orthogonal to each other and using similar argument as above we see that, $\|\tilde{\mathbf{A}}\|_F^2 = \|\mathbf{P}\tilde{\mathbf{A}} + (\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}})\|_F^2 = \text{trace}\left((\mathbf{P}\tilde{\mathbf{A}} + (\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}}))(\mathbf{P}\tilde{\mathbf{A}} + (\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}}))^\top\right) = \text{trace}\left(\mathbf{P}\tilde{\mathbf{A}}(\mathbf{P}\tilde{\mathbf{A}})^\top\right) + \text{trace}\left((\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}})(\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}})^\top\right) = \|\mathbf{P}\tilde{\mathbf{A}}\|_F^2 + \|\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}}\|_F^2$. In other words, $\|\tilde{\mathbf{A}} - \mathbf{P}\tilde{\mathbf{A}}\|_F^2 = \|\tilde{\mathbf{A}}\|_F^2 - \|\mathbf{P}\tilde{\mathbf{A}}\|_F^2$.

Now using optimality of $\tilde{\mathbf{P}}^*$ and definition of $\hat{\mathbf{P}}$, we get, $\|(\mathbf{I} - \hat{\mathbf{P}})\tilde{\mathbf{A}}\|_F^2 \leq \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}^*)\tilde{\mathbf{A}}\|_F^2 \leq \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}\|_F^2$ Expanding on both side, we get,

$$
\begin{aligned}
&\|(\mathbf{I} - \hat{\mathbf{P}})\tilde{\mathbf{A}}_m\|_F^2 + \|(\mathbf{I} - \hat{\mathbf{P}})\tilde{\mathbf{A}}_{r-m}\|_F^2 \\
&\leq \gamma\left(\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 + \|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_{r-m}\|_F^2\right)
\end{aligned}
$$

Rearranging the terms,

$$
\begin{aligned}
&\|(\mathbf{I} - \hat{\mathbf{P}})\tilde{\mathbf{A}}_m\|_F^2 \\
&\leq \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 + \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_{r-m}\|_F^2 \\
&\quad - \|(\mathbf{I} - \hat{\mathbf{P}})\tilde{\mathbf{A}}_{r-m}\|_F^2 \\
&= \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 + \gamma\|\tilde{\mathbf{A}}_{r-m}\|_F^2 - \gamma\|\tilde{\mathbf{P}}_m^*\tilde{\mathbf{A}}_{r-m}\|_F^2 \\
&\quad - \left(\|\tilde{\mathbf{A}}_{r-m}\|_F^2 - \|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2\right) \\
&\leq \gamma\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 + (\gamma - 1)\|\tilde{\mathbf{A}}_{r-m}\|_F^2 + \|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2
\end{aligned}
\tag{1}
$$

This establishes the first result. When $\gamma = 1$, $\hat{\mathbf{P}} = \tilde{\mathbf{P}}^*$ and the above inequality becomes

$$
\|(\mathbf{I} - \tilde{\mathbf{P}}^*)\tilde{\mathbf{A}}_m\|_F^2 \leq (\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 + \|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2 \tag{2}
$$

Next we show how to bound $\|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2$. First note that, $\|\tilde{\mathbf{A}}_{r-m}\|_F^2 = \sum_{i=1}^{r-m}\sigma_i^2(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_m) = \sum_{i=m+1}^r \sigma_i^2(\tilde{\mathbf{A}})$. Since $\hat{\mathbf{P}}$ is a rank k projection matrix, assuming $r - m > k$, $\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}$ has rank k, and therefore, $\|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2$ has value no more than the top k singular values of $\tilde{\mathbf{A}}_{r-m}$. In other

[1] Department of Electrical Engineering & Computer Science, Wichita State University, KS, USA. Correspondence to: Kaushik Sinha <kaushik.sinha@wichita.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

words,

$$\|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2 \leq \sum_{i=1}^{k} \sigma_i^2(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_m) = \sum_{i=m+1}^{m+k} \sigma_i^2(\tilde{\mathbf{A}}) \quad (3)$$

Using singular value decomposition it is easy to see that $\|\tilde{\mathbf{A}}_m - \tilde{\mathbf{A}}_k\|_F^2 = \sum_{i=k+1}^{m} \sigma_i^2(\tilde{\mathbf{A}})$. From this we can write

$$\begin{aligned} & 2\|\tilde{\mathbf{A}}_m - \tilde{\mathbf{A}}_k\|_F^2 \\ &= \sum_{i=k+1}^{m}\sigma_i^2(\tilde{\mathbf{A}}) + \sum_{i=k+1}^{m}\sigma_i^2(\tilde{\mathbf{A}}) \overset{a}{\geq} \sum_{i=k+1}^{m}\sigma_i^2(\tilde{\mathbf{A}}) + \sum_{i=m+1}^{m+k}\sigma_i^2(\tilde{\mathbf{A}}) \\ &= \sum_{i=k+1}^{m+k}\sigma_i^2(\tilde{\mathbf{A}}) \overset{b}{\geq} m\left(\frac{1}{k}\sum_{i=m+1}^{m+k}\sigma_i^2(\tilde{\mathbf{A}})\right) \overset{c}{\geq} \frac{1}{\epsilon}\sum_{i=m+1}^{m+k}\sigma_i^2(\tilde{\mathbf{A}}) \end{aligned} \quad (4)$$

where, inequality a follows from the fact that sum of $(m-k)$ singular values of $\tilde{\mathbf{A}}$ is being bounded from below by sum of k (smaller) singular values of $\tilde{\mathbf{A}}$ (in the worst case if all these singular values are of same value then inequality a will hold if $(m-k) \geq k$ or $m \geq 2k$, which will always hold as long as $\epsilon \leq 1/2$). Inequality b follows from the fact that sum of m consecutive singular values of $\tilde{\mathbf{A}}$ is bounded from below by m times average of the smallest k of those m singular values of $\tilde{\mathbf{A}}$. Finally, inequality c follows from our choice of m. Combining equation 3 and 4 we get,

$$\begin{aligned} \|\hat{\mathbf{P}}\tilde{\mathbf{A}}_{r-m}\|_F^2 &\leq 2\epsilon\|\tilde{\mathbf{A}}_m - \tilde{\mathbf{A}}_k\|_F^2 \leq 2\epsilon\|\tilde{\mathbf{A}}_m - \tilde{\mathbf{P}}_m^*\tilde{\mathbf{A}}_m\|_F^2 \\ &= 2\epsilon\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2 \end{aligned} \quad (5)$$

where, the second inequality follows from the fact that $\tilde{\mathbf{A}}_k$ is the best rank k approximation of $\tilde{\mathbf{A}}_m$ since $m > k$ and $\tilde{\mathbf{P}}_m^*\tilde{\mathbf{A}}_m$ is a rank k matrix having same size that of $\tilde{\mathbf{A}}_m$. Combining equation 2 and 5 yields part (i) of the Lemma.

To prove part (ii) of the lemma note that $(\gamma-1)\|\tilde{\mathbf{A}}_{r-m}\|_F^2 = \epsilon_1 \sum_{i=m+1}^{r} \sigma_i^2(\tilde{\mathbf{A}}) \leq \sum_{i=m+1}^{m+k} \sigma_i^2(\tilde{\mathbf{A}}) \leq 2\epsilon\|\tilde{\mathbf{A}}_m - \tilde{\mathbf{A}}_k\|_F^2 \leq 2\epsilon\|(\mathbf{I} - \tilde{\mathbf{P}}_m^*)\tilde{\mathbf{A}}_m\|_F^2$. Combining this with equation 1, 5 and $\gamma = 1 + \epsilon_1$, yields the desired result. □

1.2. Proof of Lemma 2

Proof. We observe, using Lemma 1 of (Achlioptas & Mcsherry, 2007), that

$$\|\mathbf{A}-\tilde{\mathbf{A}}_m\|_F \leq \|\mathbf{A}-\mathbf{A}_m\|_F + \|\mathbf{N}_m\|_F + 2\sqrt{\|\mathbf{N}_m\|_F\|\mathbf{A}_m\|_F} \quad (6)$$

For the choice of p and using Theorem **??**, we get $\|N_m\|_F \leq \sqrt{m}\epsilon_2\|\mathbf{A}\|_F \leq m^{1/4}\sqrt{\epsilon_2}\|\mathbf{A}\|_F$, where the second inequality follows from the restriction of ϵ_2. Next, $\sqrt{\|\mathbf{N}_m\|_F\|\mathbf{A}_m\|_F} \leq \sqrt{m^{1/2}\epsilon_2\|\mathbf{A}\|_F\|\mathbf{A}_m\|_F} \leq m^{1/4}\sqrt{\epsilon_2}\|\mathbf{A}\|_F$. Plugging in these values in equation 6 we get the desired result. □

1.3. Proof of Lemma 3

Proof. Note that the choice of m satisfies, $\|\mathbf{A}_m\|_F^2 = \sum_{i=1}^{m} \sigma_i^2(\mathbf{A}) \leq \frac{1}{2}\sum_{i=1}^{\rho} \sigma_i^2(\mathbf{A}) = \frac{1}{2}\|\mathbf{A}\|_F^2 = \frac{1}{2}\left(\|\mathbf{A}_m\|_F^2 + \|\mathbf{A} - \mathbf{A}_m\|_F^2\right) \Rightarrow 2\|\mathbf{A}_m\|_F^2 \leq \|\mathbf{A}_m\|_F^2 + \|\mathbf{A} - \mathbf{A}_m\|_F^2 \Rightarrow \|\mathbf{A}_m\|_F^2 \leq \|\mathbf{A} - \mathbf{A}_m\|_F^2 \Rightarrow \|\mathbf{A}_m\|_F^2 + \|\mathbf{A} - \mathbf{A}_m\|_F^2 \leq 2\|\mathbf{A} - \mathbf{A}_m\|_F^2 \Rightarrow \|\mathbf{A}\|_F^2 \leq 2\|\mathbf{A} - \mathbf{A}_m\|_F^2 \Rightarrow \|\mathbf{A}\|_F \leq \sqrt{2}\|\mathbf{A} - \mathbf{A}_m\|_F$. Now invoking Lemma **??** ensures that , $\|\mathbf{A} - \tilde{\mathbf{A}}_m\|_F \leq \left(1 + 3\sqrt{2\epsilon_2}(k/\epsilon_3)^{1/4}\right)\|\mathbf{A} - \mathbf{A}_m\|_F$. Setting $\epsilon_2 = \frac{1}{18}\sqrt{\frac{\epsilon_3^9}{k}}$ yields the desired result. □

1.4. Proof of Corollary 1

Proof. Since $\gamma = 1 + \epsilon_1$, it is easy to see that $\frac{\gamma_1(1+11\epsilon)}{1-4\epsilon} \leq \frac{(1+\epsilon_1+4\epsilon)(1+11\epsilon)}{1-4\epsilon} = \frac{(\gamma+4\epsilon)(1+11\epsilon)}{1-4\epsilon} \leq \frac{\gamma(1+4\epsilon)(1+11\epsilon)}{1-4\epsilon} \leq \frac{\gamma(1+15\epsilon+44\epsilon^2)}{1-4\epsilon} \leq \frac{\gamma(1+59\epsilon)}{1-4\epsilon}$. We require $\frac{(1+59\epsilon)}{1-4\epsilon} \leq 1 + \epsilon' \Rightarrow (1 + 59\epsilon) \leq (1 - 4\epsilon + \epsilon'(1 - 4\epsilon)) \Rightarrow 63\epsilon \leq (\epsilon' - 4\epsilon\epsilon') \Rightarrow \epsilon \leq \frac{\epsilon'}{63+4\epsilon'}$. Setting $\epsilon = \epsilon'/67$ and plugging in Theorem **??** yields the desired result. □

References

Achlioptas, D. and Mcsherry, F. Fast computation of low-rank matrix approximations. *J. ACM*, 54(2), April 2007. ISSN 0004-5411. doi: 10.1145/1219092.1219097. URL http://doi.acm.org/10.1145/1219092.1219097.