Recitation #3: HW #2 from 2017 Review

TA: Rajesh Jayaram

September 27, 2019

Spectral Norm Low Rank Approximation

Recall Low rank approximation: Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, $n \gg d$, we would like to find a rank k matrix $\mathbf{B}'$ such that

$$\|\mathbf{A} - \mathbf{B}'\| \leq (1+\epsilon) \min_{\mathbf{B},\ \text{rank}-k} \|\mathbf{A} - \mathbf{B}\|$$

where $\|\cdot\|$ is some matrix norm.

Spectral Norm Low Rank Approximation

Recall Low rank approximation: Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, $n \gg d$, we would like to find a rank k matrix $\mathbf{B}'$ such that

$$\|\mathbf{A} - \mathbf{B}'\| \leq (1+\epsilon) \min_{\mathbf{B},\ \text{rank}-k} \|\mathbf{A} - \mathbf{B}\|$$

where $\|\cdot\|$ is some matrix norm.

- In class we saw Frobenius norm LRA $\|\cdot\| = \|\cdot\|_F$

Spectral Norm Low Rank Approximation

Recall Low rank approximation: Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, $n \gg d$, we would like to find a rank k matrix $\mathbf{B}'$ such that

$$\|\mathbf{A} - \mathbf{B}'\| \leq (1+\epsilon) \min_{\mathbf{B},\, \text{rank}-k} \|\mathbf{A} - \mathbf{B}\|$$

where $\|\cdot\|$ is some matrix norm.

- In class we saw Frobenius norm LRA $\|\cdot\| = \|\cdot\|_F$
- Spectral norm LRA can be much stronger:

$$\min_{\mathbf{B},\, \text{rank}-k} \|\mathbf{A} - \mathbf{B}\|_2 = \min_{\mathbf{B},\, \text{rank}-k} \max_{x \in \mathbb{R}^n} \|(\mathbf{A} - \mathbf{B})x\|_2$$

Spectral Norm Low Rank Approximation

First observe that the optimum solution is the same as in Frobenius LRA:

$$\min_{\mathbf{B},\ \text{rank}-k} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}(\mathbf{A})$$

Where $\mathbf{A}_k$ is the truncated SVD, and $\sigma_{k+1}(\mathbf{A})$ the $k+1$ largest singular value of $\mathbf{A}$.

First: A recap of Frobenius LRA

Recall how we solved Frobenius LRA:

- We first generated a random matrix $\mathbf{S} \in \mathbb{R}^{r \times n}$ where $r = \text{poly}(k/\epsilon)$.

First: A recap of Frobenius LRA

Recall how we solved Frobenius LRA:

- We first generated a random matrix $\mathbf{S} \in \mathbb{R}^{r \times n}$ where $r = \text{poly}(k/\epsilon)$.
- We computed $\mathbf{SA}$, and argued that there is a good rank-k approximation to $\mathbf{A}$ in the row span of $\mathbf{SA}$.

First: A recap of Frobenius LRA

Recall how we solved Frobenius LRA:

- We first generated a random matrix $\mathbf{S} \in \mathbb{R}^{r \times n}$ where $r = \text{poly}(k/\epsilon)$.
- We computed $\mathbf{SA}$, and argued that there is a good rank-k approximation to $\mathbf{A}$ in the row span of $\mathbf{SA}$.
- By Pythagorean theorem, the best LRA in the row span of $\mathbf{SA}$ comes from first projecting the rows of $\mathbf{A}$ onto $\mathbf{SA}$, then computing the SVD of these projected rows.

First: A recap of Frobenius LRA

Recall how we solved Frobenius LRA:

- We first generated a random matrix $\mathbf{S} \in \mathbb{R}^{r \times n}$ where $r = \text{poly}(k/\epsilon)$.
- We computed $\mathbf{SA}$, and argued that there is a good rank-k approximation to $\mathbf{A}$ in the row span of $\mathbf{SA}$.
- By Pythagorean theorem, the best LRA in the row span of $\mathbf{SA}$ comes from first projecting the rows of $\mathbf{A}$ onto $\mathbf{SA}$, then computing the SVD of these projected rows.
- Formally, we could project $\mathbf{AP_{SA}} = (\mathbf{SA})^{-}\mathbf{SA}$, and then compute

$$[\mathbf{AP_{SA}}]_k = [\mathbf{A}(\mathbf{SA})^{-}\mathbf{SA}]_k$$

Spectral Norm Low Rank Approximation

Part 1: Let $\tilde{\mathbf{A}}$ be the best Spectral rank-k approx to $\mathbf{A}$ in the row span of some matrix $\mathbf{B}$:

$$\tilde{\mathbf{A}} = \mathbf{Y}\mathbf{B}, \quad \text{where } \mathbf{Y} = \arg \min_{\text{rank-k } \mathbf{Y}'} \|\mathbf{Y}'\mathbf{B} - \mathbf{A}\|_2$$

Show that for any $\mathbf{A}, \mathbf{B}$, we have

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Spectral Norm Low Rank Approximation

We have $\tilde{\mathbf{A}} = \mathbf{YB}$, where $\mathbf{Y} = \arg\min_{\text{rank-k } \mathbf{Y}'} \|\mathbf{Y}'\mathbf{B} - \mathbf{A}\|_2$
Show that for any $\mathbf{A}, \mathbf{B}$, we have

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Spectral Norm Low Rank Approximation

We have $\tilde{\mathbf{A}} = \mathbf{YB}$, where $\mathbf{Y} = \arg\min_{\text{rank-k } \mathbf{Y}'} \|\mathbf{Y}'\mathbf{B} - \mathbf{A}\|_2$
Show that for any $\mathbf{A}, \mathbf{B}$, we have

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Use def of spectral norm: fix any vector x

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A}(\mathbb{I} - \mathbf{P_B}) + x\mathbf{A}\mathbf{P_B} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \quad (1)$$

Spectral Norm Low Rank Approximation

We have $\tilde{\mathbf{A}} = \mathbf{YB}$, where $\mathbf{Y} = \arg\min_{\text{rank-k } \mathbf{Y}'} \|\mathbf{Y}'\mathbf{B} - \mathbf{A}\|_2$
Show that for any $\mathbf{A}, \mathbf{B}$, we have

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Use def of spectral norm: fix any vector x

$$\begin{aligned} \|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 &= \|x\mathbf{A}(\mathbb{I} - \mathbf{P_B}) + x\mathbf{AP_B} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\ &= \|x\mathbf{A}(\mathbb{I} - \mathbf{P_B})\|_2^2 + \|x\mathbf{AP_B} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\ &= \|x\mathbf{A} - x\mathbf{AP_B}\|_2^2 + \|x\mathbf{AP_B} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \end{aligned} \tag{2}$$

Spectral Norm Low Rank Approximation

We have $\tilde{\mathbf{A}} = \mathbf{Y}\mathbf{B}$, where $\mathbf{Y} = \arg\min_{\text{rank-k } \mathbf{Y}'} \|\mathbf{Y}'\mathbf{B} - \mathbf{A}\|_2$
Show that for any $\mathbf{A}, \mathbf{B}$, we have

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Use def of spectral norm: fix any vector x

$$\begin{aligned}
\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 &= \|x\mathbf{A}(\mathbb{I} - \mathbf{P}_{\mathbf{B}}) + x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\
&= \|x\mathbf{A}(\mathbb{I} - \mathbf{P}_{\mathbf{B}})\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\
&= \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2
\end{aligned} \tag{3}$$

Hint: Just show each of the last two terms are $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$

Spectral Norm Low Rank Approximation

We have:

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2$$

W.T.S each term on RHS is $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$.

Spectral Norm Low Rank Approximation

We have:

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2$$

W.T.S each term on RHS is $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$.

First Term:

$$\|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 \leq \|x\mathbf{A} - x\tilde{\mathbf{A}}\|_2^2 \quad (4)$$

Why?

Spectral Norm Low Rank Approximation

We have:

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2$$

W.T.S each term on RHS is $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$.

First Term:

$$\|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 \leq \|x\mathbf{A} - x\tilde{\mathbf{A}}\|_2^2 \tag{5}$$

Why?

$x\mathbf{A}\mathbf{P}_{\mathbf{B}}$ is the closest point to $x\mathbf{A}$ in the row span of $\mathbf{B}$, but $x\tilde{\mathbf{A}} = x\mathbf{Y}'\mathbf{B}$ is also point in the row span of $\mathbf{B}$, so $x\mathbf{A}\mathbf{P}_{\mathbf{B}}$ is at least as close to $x\mathbf{A}$ as $x\tilde{\mathbf{A}}$ is to $x\mathbf{A}$.

Spectral Norm Low Rank Approximation

We have:

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2$$

W.T.S each term on RHS is $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$.

Second Term:

$$\begin{aligned} \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 &\leq \|\mathbf{A}\mathbf{P}_{\mathbf{B}} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\ &\leq \|\mathbf{A}\mathbf{P}_{\mathbf{B}} - \tilde{\mathbf{A}}\|_2^2 \end{aligned} \tag{6}$$

Spectral Norm Low Rank Approximation

We have:

$$\|x\mathbf{A} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 = \|x\mathbf{A} - x\mathbf{A}\mathbf{P}_{\mathbf{B}}\|_2^2 + \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2$$

W.T.S each term on RHS is $\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$.

Second Term:

$$\begin{aligned} \|x\mathbf{A}\mathbf{P}_{\mathbf{B}} - x[\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 &\leq \|\mathbf{A}\mathbf{P}_{\mathbf{B}} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \\ &\leq \|\mathbf{A}\mathbf{P}_{\mathbf{B}} - \tilde{\mathbf{A}}\|_2^2 \\ &\leq \|(\mathbf{A} - \tilde{\mathbf{A}})\mathbf{P}_{\mathbf{B}}\|_2^2 \\ &\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2 \end{aligned} \tag{7}$$

Spectral Norm Low Rank Approximation

Summary: we have shown:

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Spectral Norm Low Rank Approximation

Summary: we have shown:

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Unfortunately, the above can be tight, so we cannot first find a matrix $\mathbf{B} = \mathbf{S}\mathbf{A}$ whose span contains a good rank-k approximation and then approximately compute $[\mathbf{A}P_{\mathbf{B}}]_k$.

Spectral Norm Low Rank Approximation

Summary: we have shown:

$$\|\mathbf{A} - [\mathbf{A}P_{\mathbf{B}}]_k\|_2^2 \leq 2\|\mathbf{A} - \tilde{\mathbf{A}}\|_2^2$$

Unfortunately, the above can be tight, so we cannot first find a matrix $\mathbf{B} = \mathbf{S}\mathbf{A}$ whose span contains a good rank-k approximation and then approximately compute $[\mathbf{A}P_{\mathbf{B}}]_k$.

Instead: we do the following. For simplicity assume $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric matrix, let $\mathbf{G} \in \mathbb{R}^{n \times k}$ be a matrix of i.i.d. Gaussians. We compute $\mathbf{A}^r\mathbf{G}$ where $r = O(\epsilon^{-1} \log n)$. One can then prove that

$$\|\mathbf{A} - \mathbf{P}_{\mathbf{A}^r\mathbf{G}}\mathbf{A}\|_2 \leq \|\mathbf{A}^r - \mathbf{P}_{\mathbf{A}^r\mathbf{G}}\mathbf{A}^r\|_2^{1/r}$$

Spectral Norm Low Rank Approximation

Instead: we do the following. For simplicity assume $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric matrix, let $\mathbf{G} \in \mathbb{R}^{n \times k}$ be a matrix of i.i.d. Gaussians. We compute $\mathbf{A}^r \mathbf{G}$ where $r = O(\epsilon^{-1} \log n)$. Let $\mathbf{P} = \mathbf{P}_{\mathbf{A}^r \mathbf{G}}$. One can then prove that

$$\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq \|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^{1/r}$$

Now we show that if

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 \text{poly}(n)$$

for some poly(n) that does not depend on r, then

$$\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq (1 + \epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$$

Spectral Norm Low Rank Approximation

In other words, we know that $\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq \|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^{1/r}$ and that$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$poly$(n)$ for some poly(n) that does not depend on $r = O(\epsilon^{-1} \log n)$, and we want to show that $\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq (1 + \epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$.

Spectral Norm Low Rank Approximation

In other words, we know that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r}$ and that$\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$poly$(n)$ for some poly(n) that does not depend on $r = O(\epsilon^{-1} \log n)$, and we want to show that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$.

$$\begin{aligned} \|\mathbf{A} - \mathbf{PA}\|_2 &\leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r} \\ &\leq \left(\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2\right)^{1/(2r)} \\ &\leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} \left(\text{poly}(n)\right)^{1/(2r)} \end{aligned} \tag{8}$$

Spectral Norm Low Rank Approximation

In other words, we know that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r}$ and that$\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$poly$(n)$ for some poly(n) that does not depend on $r = O(\epsilon^{-1} \log n)$, and we want to show that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$.

$$\begin{aligned}
\|\mathbf{A} - \mathbf{PA}\|_2 &\leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r} \\
&\leq \left(\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2\right)^{1/(2r)} \\
&\leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} \left(\text{poly}(n)\right)^{1/(2r)} \\
&\leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} 2^{\Theta(\log(n)/r)} \\
&\leq (1+\epsilon)\|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r}
\end{aligned} \tag{9}$$

Spectral Norm Low Rank Approximation

In other words, we know that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r}$ and that $\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 \text{poly}(n)$ for some poly(n) that does not depend on $r = O(\epsilon^{-1} \log n)$, and we want to show that $\|\mathbf{A} - \mathbf{PA}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$.

$$
\begin{aligned}
\|\mathbf{A} - \mathbf{PA}\|_2 &\leq \|\mathbf{A}^r - \mathbf{PA}^r\|_2^{1/r} \\
&\leq \left(\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2\right)^{1/(2r)} \\
&\leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} \left(\text{poly}(n)\right)^{1/(2r)} \\
&\leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} 2^{\Theta(\log(n)/r)} \\
&\leq (1+\epsilon)\|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^{1/r} \\
&= (1+\epsilon)(\sigma_{k+1}(\mathbf{A}^r))^{1/r} \\
&= (1+\epsilon)((\sigma_{k+1}(\mathbf{A}))^r)^{1/r} \\
&= (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2
\end{aligned}
\tag{10}
$$

Spectral Norm Low Rank Approximation

Thus, we have shown that if $\mathbf{P} = \mathbf{P}_{\mathbf{A}^r\mathbf{G}}$, then:

$$\|\mathbf{A} - \mathbf{PA}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$$

We only assumed that $\|\mathbf{A}^r - \mathbf{PA}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 \text{poly}(n)$, so lets prove this now.

Spectral Norm Low Rank Approximation

Thus, we have shown that if $\mathbf{P} = \mathbf{P}_{\mathbf{A}^r\mathbf{G}}$, then:

$$\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$$

We only assumed that $\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 \text{poly}(n)$, so lets prove this now.

Let $\mathbf{A} {=} \mathbf{U}\Sigma\mathbf{V}^T$. Let $\mathbf{V}_k^T$ be top k rows of $\mathbf{V}^T$, and $\mathbf{V}_{n-k}^T$ the bottom $n-k$. Then for any S with $\text{rank}(\mathbf{V}_k^T S) = k$, if $\mathbf{P}$ is the projection onto columns of $\mathbf{A}^r\mathbf{S}$, then:

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 + \|(\mathbf{A}^r - [\mathbf{A}^r]_k)\mathbf{S}(\mathbf{V}_k^T\mathbf{S})^-\|_2^2$$

Spectral Norm Low Rank Approximation

We have the fact that:

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 + \|(\mathbf{A}^r - [\mathbf{A}^r]_k)\mathbf{S}(\mathbf{V}_k^T\mathbf{S})^-\|_2^2 \quad (11)$$

where $\mathbf{P}$ is the projection onto columns of $\mathbf{A}^r\mathbf{S}$, then: write $(\mathbf{A}^r - [\mathbf{A}^r]_k) = \mathbf{U}_{n-k}\Sigma'\mathbf{V}_{n-k}^T$, then:

$$\begin{aligned}\|(\mathbf{A}^r - [\mathbf{A}^r]_k)G(\mathbf{V}_k^T\mathbf{G})^-\|_2^2 &\leq \|\mathbf{U}_{n-k}\Sigma'\mathbf{V}_{n-k}^T G(\mathbf{V}_k^T\mathbf{G})^-\|_2^2 \\ &\leq \|\mathbf{U}_{n-k}\Sigma'\|_2^2\|\mathbf{V}_{n-k}^T G\|_2^2\|(\mathbf{V}_k^T\mathbf{G})^-\|_2^2 \\ &= \|(\mathbf{A}^r - [\mathbf{A}^r]_k)\|_2^2\|\mathbf{V}_{n-k}^T G\|_2^2\|(\mathbf{V}_k^T\mathbf{G})^-\|_2^2\end{aligned} \quad (12)$$

Putting this together with eqn 11, we have

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2(1 + \|\mathbf{V}_{n-k}^T G\|_2^2\|(\mathbf{V}_k^T\mathbf{G})^-\|_2^2)$$

Spectral Norm Low Rank Approximation

We have

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 (1 + \|\mathbf{V}_{n-k}^T G\|_2^2 \|(\mathbf{V}_k^T \mathbf{G})^- \|_2^2)$$

By rotational invariance: $\mathbf{V}_{n-k}^T G$ distributed as $n - k \times k$ matrix of i.i.d. gaussians, and $\mathbf{V}_k^T G$ is a $k \times k$ matrix of Gaussians.

Spectral Norm Low Rank Approximation

We have

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 (1 + \|\mathbf{V}_{n-k}^T G\|_2^2 \|(\mathbf{V}_k^T \mathbf{G})^{-}\|_2^2)$$

By rotational invariance: $\mathbf{V}_{n-k}^T G$ distributed as $n-k \times k$ matrix of i.i.d. gaussians, and $\mathbf{V}_k^T G$ is a $k \times k$ matrix of Gaussians.

$$\|\mathbf{V}_{n-k}^T G\|_2^2 \leq O(n)$$

$$\|(\mathbf{V}_k^T G)^{-}\|_2^2 \leq \frac{1}{\sigma_{\min}(\mathbf{V}_k^T G)} \leq O(k)$$

Spectral Norm Low Rank Approximation

We have

$$\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2 (1 + \|\mathbf{V}_{n-k}^T G\|_2^2 \|(\mathbf{V}_k^T \mathbf{G})^{-}\|_2^2)$$

By rotational invariance: $\mathbf{V}_{n-k}^T G$ distributed as $n - k \times k$ matrix of i.i.d. gaussians, and $\mathbf{V}_k^T G$ is a $k \times k$ matrix of Gaussians.

$$\|\mathbf{V}_{n-k}^T G\|_2^2 \leq O(n)$$

$$\|(\mathbf{V}_k^T G)^{-}\|_2^2 \leq \frac{1}{\sigma_{\min}(\mathbf{V}_k^T G)} \leq O(k)$$

So $\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \text{poly}(n) \|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$!

Spectral Norm Low Rank Approximation

Summary: We proved $\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \text{poly}(n)\|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$, and this is all we needed to prove that

$$\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$$

So we can output $\mathbf{P}\mathbf{A}$, where $\mathbf{P} \in \mathbb{R}^{k\times k}$ is the projection onto the column span of $\mathbf{A}^r\mathbf{G}$.

Spectral Norm Low Rank Approximation

Summary: We proved $\|\mathbf{A}^r - \mathbf{P}\mathbf{A}^r\|_2^2 \leq \text{poly}(n)\|\mathbf{A}^r - [\mathbf{A}^r]_k\|_2^2$, and this is all we needed to prove that

$$\|\mathbf{A} - \mathbf{P}\mathbf{A}\|_2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$$

So we can output $\mathbf{P}\mathbf{A}$, where $\mathbf{P} \in \mathbb{R}^{k \times k}$ is the projection onto the column span of $\mathbf{A}^r\mathbf{G}$.

- To compute $\mathbf{A}^r\mathbf{G}$, can compute $\mathbf{A}\mathbf{G} \in \mathbb{R}^{n \times k}$ in $O(k\text{nnz}(\mathbf{A}))$ time, then $\mathbf{A}^2\mathbf{G} = \mathbf{A}(\mathbf{A}\mathbf{G})$ in $O(k\text{nnz}(\mathbf{A}))$ time, and so on.
- Thus, can compute $\mathbf{A}^r\mathbf{G}$ in $O(\frac{k\log(n)}{\epsilon}\text{nnz}(\mathbf{A}))$ time.
- Then we can output $\mathbf{P}\mathbf{A} = (\mathbf{A}^r\mathbf{G})(\mathbf{A}^r\mathbf{G})^{-}\mathbf{A}$ in factored form in $O(\frac{k\log(n)}{\epsilon}\text{nnz}(\mathbf{A}) + \text{nnz}(\mathbf{A})k^2)$ time for pseudoinverse.