Supplement to "Optimal Deterministic Coresets for Ridge Regression"

Praneeth Kacham
Carnegie Mellon University

David P. Woodruff
Carnegie Mellon University

This supplementary document contains the omitted proofs of Lemmas in Section 1 of the main paper. We restate the lemmas and prove them here.

Lemma 0.1. *If $\hat{A}$ is a matrix with orthonormal columns such that $\texttt{range}(\hat{A}) = \texttt{range}(\begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix})$ and if U_1 comprises the first n rows of $\hat{A}$, then $\|U_1\|_F^2 = \textit{sd}_\lambda(A)$ and $\|U_1\|_2^2 = 1/(1 + \lambda/\sigma_1^2) \leq 1$.*

Proof. See Lemma 12 of (Avron et al., 2017) for a proof. $\qquad\square$

Lemma 0.2. *If A' is the sub-matrix of A formed by taking rows of A, then $\texttt{sd}_\lambda(A') \leq \texttt{sd}_\lambda(A)$.*

Proof. The minimax characterization of singular values is as follows:

$$\sigma_i(A) = \max_{U:\dim(U)=i} \min_{x \in U: \|x\|_2=1} \|Ax\|_2 \tag{0.1}$$

For any vector x, we have $\|A'x\|_2 \leq \|Ax\|_2$ (Since, vector Ax has all the elements of $A'x$ and more). Hence, for any subspace U, $\min_{x \in U: \|x\|_2=1} \|A'x\| \leq \min_{x \in U: \|x\|_2=1} \|Ax\|_2$. Now, it is easy to see that

$$\max_{U:\dim(U)=i} \min_{x \in U: \|x\|_2=1} \|A'x\|_2 \leq$$
$$\max_{U:\dim(U)=i} \min_{x \in U: \|x\|_2=1} \|Ax\|_2$$
$$\implies \sigma_i(A') \leq \sigma_i(A) \tag{0.2}$$

Now,

$$\begin{aligned}
\texttt{sd}_\lambda(A') &= \sum_{i=1}^{\texttt{rank}(A')} \frac{1}{1 + \lambda/\sigma_i(A')^2} \\
&\leq \sum_{i=1}^{\texttt{rank}(A')} \frac{1}{1 + \lambda/\sigma_i(A)^2} \quad \text{(Using (0.2))} \\
&\leq \sum_{i=1}^{\texttt{rank}(A)} \frac{1}{1 + \lambda/\sigma_i(A)^2} \\
&= \texttt{sd}_\lambda(A) \qquad\qquad\qquad\qquad\square
\end{aligned}$$

Lemma 0.3. *For any $r \geq 1$, $\texttt{sd}_{\lambda/r}(A) \leq \min(r \cdot \texttt{sd}_\lambda(A), \texttt{rank}(A))$.*

Proof. By definition of statistical dimension, $\texttt{sd}_\lambda(A) \leq \texttt{rank}(A)$. We also have for all $r \geq 1$, $1/(1 + \lambda/r\sigma_i^2) \leq r/(1 + \lambda/\sigma_i^2)$. By summing the inequality for all i, we get $\texttt{sd}_{\lambda/r}(A) \leq r \cdot \texttt{sd}_\lambda(A)$. $\qquad\square$

References

Haim Avron, Kenneth L. Clarkson, and David P. Woodruff. Sharper bounds for regularized data fitting. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2017, August 16-18, 2017, Berkeley, CA, USA*, pages 27:1–27:22, 2017. doi: 10.4230/LIPIcs.APPROX-RANDOM.2017.27. URL https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2017.27.