A Deterministic Streaming Sketch for Ridge Regression
Supplementary Materials

A OTHER VARIANCE BOUNDS FOR RISK

We provide two different bounds for variance $\mathcal{V}(\hat{\mathbf{x}}_\gamma)$ that are not strictly comparable with the one provided in Lemma 3.

Lemma 6. *Considering the data generation model and the risk described in Lemma 3. The variance of the approximate solution* $\hat{\mathbf{x}}_\gamma = (\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{b}$ *satisfy*

$$\mathcal{V}(\hat{\mathbf{x}}_\gamma) \leq \left(1 + \frac{1}{\gamma}\|\mathbf{A}\|_2^2\|\mathbf{A}^\top\mathbf{A} - \mathbf{C}^\top\mathbf{C}\|_2^2\|\mathbf{A}^\dagger\|_2^2\right)\mathcal{V}(\mathbf{x}_\gamma)$$

Proof.

$$\begin{aligned}
\mathcal{V}(\hat{\mathbf{x}}_\gamma) =& \mathbb{E}_{\boldsymbol{Z}}\left[\|\mathbf{A}\left(\hat{\mathbf{x}}_\gamma - \mathbb{E}_{\boldsymbol{Z}}\left[\hat{\mathbf{x}}_\gamma\right]\right)\|^2\right] \\
=& \mathbb{E}_{\boldsymbol{Z}}\left[\|\mathbf{A}\left((\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top s\boldsymbol{Z}\right)\|^2\right] \\
=& s^2\|\mathbf{A}(\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
=& s^2\|\mathbf{A}\left((\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1} - (\mathbf{K} + \gamma\mathbf{I})^{-1} + (\mathbf{K} + \gamma\mathbf{I})^{-1}\right)\mathbf{A}^\top\|_F^2 \\
=& s^2\|\mathbf{A}\left((\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\mathbf{K} - \hat{\mathbf{K}})(\mathbf{K} + \gamma\mathbf{I})^{-1} + (\mathbf{K} + \gamma\mathbf{I})^{-1}\right)\mathbf{A}^\top\|_F^2 \\
=& s^2\|\mathbf{A}\left((\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\mathbf{K} - \hat{\mathbf{K}}) + \mathbf{I}\right)(\mathbf{K} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
=& s^2\|\mathbf{A}\left((\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\mathbf{K} - \hat{\mathbf{K}}) + \mathbf{I}\right)\mathbf{A}^+\mathbf{A}(\mathbf{K} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
\leq& s^2\|\mathbf{A}\left((\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\mathbf{K} - \hat{\mathbf{K}}) + \mathbf{I}\right)\mathbf{A}^+\|_2^2\|\mathbf{A}(\mathbf{K} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
=& \|\mathbf{A}(\hat{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\mathbf{A}^\top\mathbf{A} - \mathbf{C}^\top\mathbf{C})\mathbf{A}^+ + \mathbf{I}\|_2^2\mathcal{V}(\mathbf{x}_\gamma) \\
\leq& \left(1 + \frac{1}{\gamma}\|\mathbf{A}\|_2^2\|\mathbf{A}^\top\mathbf{A} - \mathbf{C}^\top\mathbf{C}\|_2^2\|\mathbf{A}^\dagger\|_2^2\right)\mathcal{V}(\mathbf{x}_\gamma)
\end{aligned}$$

□

Lemma 7. *Considering the data generation model and the risk described in Lemma 3. The variance of the approximate solution* $\hat{\mathbf{x}}_\gamma = (\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{b}$ *satisfy*

$$\mathcal{V}(\hat{\mathbf{x}}_\gamma) \leq \frac{1}{1 - \|\mathbf{A}^+\|^2\|(\mathbf{C}^\top\mathbf{C} - \mathbf{A}^\top\mathbf{A})\|_2}\mathcal{V}(\mathbf{x}_\gamma)$$

Proof. The proof Follows the strategy used by Wang et al. (2018) for the Hessian Sketch variance bound.

$$
\begin{aligned}
\mathcal{V}(\hat{\mathbf{x}}_\gamma) =& \mathbb{E}_{\boldsymbol{Z}}\left[\|\mathbf{A}\left(\hat{\mathbf{x}}_\gamma - \mathbb{E}_{\boldsymbol{Z}}\left[\hat{\mathbf{x}}_\gamma\right]\right)\|^2\right] \\
=& \mathbb{E}_{\boldsymbol{Z}}\left[\|\mathbf{A}\left((\mathbf{C}^\top\mathbf{C}+\gamma\mathbf{I})^{-1}\mathbf{A}^\top s\boldsymbol{Z}\right)\|^2\right] \\
=& s^2\|\mathbf{A}(\mathbf{C}^\top\mathbf{C}+\gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
=& s^2\|({\mathbf{A}^+}^\top\mathbf{C}^\top\mathbf{C}\mathbf{A}^+ + \gamma(\mathbf{A}^\top\mathbf{A})^{-1})^{-1}\|_F^2 \\
\leq& \frac{1}{1-\|\mathbf{A}^+\|^2\|(\mathbf{C}^\top\mathbf{C}-\mathbf{A}^\top\mathbf{A})\|_2} s^2\|(\mathbf{I}+\gamma(\mathbf{A}^\top\mathbf{A})^{-1})^{-1}\|_F^2 \\
=& \frac{1}{1-\|\mathbf{A}^+\|^2\|(\mathbf{C}^\top\mathbf{C}-\mathbf{A}^\top\mathbf{A})\|_2} s^2\|\mathbf{A}(\mathbf{A}^\top\mathbf{A}+\gamma\mathbf{I})^{-1}\mathbf{A}^\top\|_F^2 \\
=& \frac{1}{1-\|\mathbf{A}^+\|^2\|(\mathbf{C}^\top\mathbf{C}-\mathbf{A}^\top\mathbf{A})\|_2} \mathcal{V}(\mathbf{x}_\gamma).
\end{aligned}
$$

The inequality follows

$$
\begin{aligned}
&\|{\mathbf{A}^+}^\top\mathbf{C}^\top\mathbf{C}\mathbf{A}^+ - \mathbf{I}\|_2 \\
=&\|{\mathbf{A}^+}^\top\mathbf{C}^\top\mathbf{C}\mathbf{A}^+ - {\mathbf{A}^+}^\top\mathbf{A}^\top\mathbf{A}\mathbf{A}^+\|_2 \\
=&\|{\mathbf{A}^+}^\top(\mathbf{C}^\top\mathbf{C}-\mathbf{A}^\top\mathbf{A})\mathbf{A}^+\|_2 \\
\leq&\|\mathbf{A}^+\|^2\|(\mathbf{C}^\top\mathbf{C}-\mathbf{A}^\top\mathbf{A})\|_2.
\end{aligned}
$$

□