Supplementary Materials

The proofs use the following two results, i.e., Lemma 1 and Lemma 2, of subsampled randomized Hadamard transform (SRHT).

Lemma 1 ((Tropp, 2011))**.** *Suppose that $\boldsymbol{V}$ is an $n \times k$ matrix with orthonormal columns and $\boldsymbol{\Pi}$ is an $n \times n$ SRHT matrix, it satisfies that*

$$\Pr\left(\max_j \left\|\boldsymbol{e}_j^\top \boldsymbol{\Pi} \boldsymbol{V}\right\| \geq \sqrt{\frac{k}{n}} + \sqrt{\frac{8\log(n/\delta)}{n}}\right) \leq \delta.$$

Lemma 2 ((Tropp, 2011))**.** *Let $\boldsymbol{V}$ be an $n \times k$ matrix with orthonormal columns, and denote the maximum squared row norm by $\gamma = \max_j \left\|\boldsymbol{e}_j^\top \boldsymbol{V}\right\|^2$. Sample uniformly without replacement m rows of $\boldsymbol{V}$ to obtain a reduced matrix $\boldsymbol{V}'$. For any $t > 0$, the extreme singular values satisfy*

$$\sigma_1(\boldsymbol{V}') \leq \sqrt{\frac{(1+\alpha)m}{n}} \text{ and } \sigma_k(\boldsymbol{V}') \geq \sqrt{\frac{(1-\beta)m}{n}}$$

with failure probability at most

$$k\left[\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right]^{\frac{m}{n\gamma}} + k\left[\frac{e^{-\beta}}{(1-\beta)^{1-\beta}}\right]^{\frac{m}{n\gamma}}.$$

The following theorem is a consequence of Lemma 1 and Lemma 2. The theorem basically states that kernel approximation in Algorithm 1 is close to the true kernel up to some scaling factor $1 \pm \epsilon$.

Theorem 3 (Approximate matrix multiplication)**.** *Let $\boldsymbol{A}$ be an $n \times p$ matrix with rank r. Let $\boldsymbol{\Pi}$ be an $m \times p$ SRHT matrix with*

$$m \geq \frac{6\left[\sqrt{r} + \sqrt{8\log(rp)}\right]^2 \log r}{\epsilon^2}.$$

Suppose that $p > m$ and compute $\widehat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}^\top$, then the inequality

$$(1-\epsilon)\boldsymbol{A}\boldsymbol{A}^\top \preceq \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top \preceq \left(1 + \sqrt{\frac{2}{3}}\epsilon\right)\boldsymbol{A}\boldsymbol{A}^\top.$$

fails with probability at most 3/n.

Proof. Note that the failure probability in Lemma 2 is no more than

$$k\exp\left(-\frac{\alpha^2 m}{3n\gamma}\right) + k\exp\left(-\frac{\beta^2 m}{2n\gamma}\right).$$

To make the failure probability no more than $2k^{-1}$, it suffices to set

$$\alpha \geq \sqrt{\frac{6n\gamma\log k}{m}} \quad \text{and} \quad \beta \geq \sqrt{\frac{4n\gamma\log k}{m}}.$$

Incorporating the scaling factors of the SRHT, the extreme singular values of the transformed $\boldsymbol{V}$ satisfy

$$\sigma_1\left(\sqrt{\frac{n}{m}}\boldsymbol{V}'\right) \leq \sqrt{1 + \sqrt{\frac{6n\gamma\log k}{m}}} \quad \text{and}$$
$$\sigma_k\left(\sqrt{\frac{n}{m}}\boldsymbol{V}'\right) \geq \sqrt{1 - \sqrt{\frac{4n\gamma\log k}{m}}}.$$

From Lemma 1, with failure probability at most k^{-1} that

$$\gamma \leq \left[\sqrt{\frac{k}{n}} + \sqrt{\frac{8\log(nk)}{n}}\right]^2.$$

Combined with the singular value bounds, this result establishes the connection between m and the desired singular value bounds. One may choose $m = tn\gamma\log k$, for some $t \geq 4$.

The rest of the proof is now straightforward. We consider the singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where $\boldsymbol{V}$ is $p \times r$ orthonormal $\boldsymbol{V}$ and has orthonormal columns. Let $\boldsymbol{\Pi}$ be the SRHT, we have that $\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top = \boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{V}^\top\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\boldsymbol{V}\right)\boldsymbol{\Sigma}\boldsymbol{U}^\top$. The desired result follows by invoking Lemma 2 to bound the extreme singular values of $\boldsymbol{V}^\top\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\boldsymbol{V}$. □

Proof of Theorem 1. The idea is to simplify the analysis by dealing with the equivalent primal form of (16), involving only one $\boldsymbol{\Pi}$ term. We then perform a perturbation analysis of the inverse component. In addition, Weyl's inequalities as well the exponentiated version of Horn's inequalities are used for eigenvalue manipulations.

First, (16) can be equivalently expressed in the primal form (3):

$$\begin{aligned}\widehat{\boldsymbol{\beta}}' &= \sqrt{\boldsymbol{\Phi}}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\sqrt{\boldsymbol{\Phi}}\boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1} \\ &\quad \left(\boldsymbol{I} + \boldsymbol{X}\sqrt{\boldsymbol{\Phi}}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\sqrt{\boldsymbol{\Phi}}\boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1}\right)^{-1}\boldsymbol{y} \\ &= \left[\left(\sqrt{\boldsymbol{\Phi}}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\sqrt{\boldsymbol{\Phi}}\right)^{-1} + \boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right]^{-1} \\ &\quad \boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y}.\end{aligned} \tag{20}$$

Let $\boldsymbol{\Gamma} = \boldsymbol{\Phi}^{-1} + \boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X}$, the idea is to bound the error norm using the perturbation of the singular values of $\boldsymbol{\Gamma}^{-1}$. Denote by $\boldsymbol{\Phi}' = \sqrt{\boldsymbol{\Phi}}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\sqrt{\boldsymbol{\Phi}}$ and $\boldsymbol{\Delta} =$

$\mathbf{\Phi}'^{-1} - \mathbf{\Phi}^{-1}$, a basic result from matrix perturbation theory (Stewart and Sun, 1990) gives

$$\left\|\mathbf{\Gamma}^{-1} - (\mathbf{\Gamma} + \mathbf{\Delta})^{-1}\right\|_2 \leq \|\mathbf{\Delta}\|_2 \left\|\mathbf{\Gamma}^{-1}\right\|_2 \left\|(\mathbf{\Gamma} + \mathbf{\Delta})^{-1}\right\|_2.$$

From Weyl's inequalities, one further obtains

$$\begin{aligned} &\left\|(\mathbf{\Gamma} + \mathbf{\Delta})^{-1}\right\|_2 \\ &\leq \left[\lambda_{\text{min}}\left(\mathbf{\Phi}'^{-1}\right) + \lambda_{\text{min}}\left(\boldsymbol{X}^\top \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X}\right)\right]^{-1}. \end{aligned}$$

We now provide a bound for $\|\mathbf{\Delta}\|_2$. Observe that $\mathbf{\Delta} = \sqrt{\mathbf{\Phi}}^{-1}\left(\left(\mathbf{\Pi}^\top \mathbf{\Pi}\right)^{-1} - \boldsymbol{I}\right)\sqrt{\mathbf{\Phi}}^{-1}$ in which the extreme singular values of the parenthesized difference are bounded via Theorem 3. Thus, we have

$$\begin{aligned} \|\mathbf{\Delta}\|_2 &\leq \max\left\{\frac{\epsilon}{1-\epsilon}, \frac{\sqrt{2/3\epsilon}}{1+\sqrt{2/3\epsilon}}\right\}\left\|\mathbf{\Phi}^{-1}\right\|_2 \\ &= \frac{\epsilon}{1-\epsilon}\left\|\mathbf{\Phi}^{-1}\right\|_2. \end{aligned}$$

It remains to give a lower bound for $\lambda_{\text{min}}\left(\mathbf{\Phi}'^{-1}\right)$. From Theorem 3 and Horn's inequalities, one has

$$\begin{aligned} \lambda_{\text{min}}\left(\mathbf{\Phi}'^{-1}\right) &\geq \lambda_{\text{min}}\left(\left(\mathbf{\Pi}^\top \mathbf{\Pi}\right)^{-1}\right)\lambda_{\text{min}}\left(\mathbf{\Phi}^{-1/2}\right)^2 \\ &= \frac{\|\mathbf{\Phi}\|_2^{-1}}{1+\sqrt{2/3\epsilon}}. \end{aligned}$$

Finally, the desired estimation bound satisfies

$$\begin{aligned} \frac{\left\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'\right\|}{\left\|\widehat{\boldsymbol{\beta}}\right\|} &\leq \frac{\left\|\mathbf{\Gamma}^{-1} - (\mathbf{\Gamma} + \mathbf{\Delta})^{-1}\right\|_2}{\sigma_{\text{min}}\left(\mathbf{\Gamma}^{-1}\right)} \\ &\leq \frac{\epsilon}{1-\epsilon}\,\frac{\left\|\mathbf{\Phi}^{-1}\right\|_2 \kappa(\mathbf{\Gamma})}{\frac{\|\mathbf{\Phi}\|_2^{-1}}{1+\sqrt{2/3\epsilon}} + \lambda_{\text{min}}\left(\boldsymbol{X}^\top \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X}\right)}. \end{aligned}$$

Setting $\lambda_{\text{min}}\left(\boldsymbol{X}^\top \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X}\right) = 0$ yields the simplified worst-case bound

$$\left\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'\right\| \leq \frac{\epsilon\left(1+\sqrt{2/3\epsilon}\right)}{1-\epsilon}\kappa(\mathbf{\Phi})\,\kappa(\mathbf{\Gamma})\left\|\widehat{\boldsymbol{\beta}}\right\|.$$

□

Proof of Theorem 2. Let $\boldsymbol{P} = \boldsymbol{Z}\boldsymbol{Z}^\dagger$, then $\boldsymbol{I} - \boldsymbol{P}$ is idempotent. Thus, the noise AVC $\widehat{\sigma}^2_{\text{AVC}}$ in (12) can be expressed as

$$\widehat{\sigma}^2_{\text{AVC}} = \frac{\operatorname{tr}\left[(\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{S}\,(\boldsymbol{I} - \boldsymbol{P})\right]}{n-q}.$$

The SRHT version $\widehat{\sigma}'^2_{\text{AVC}}$ using Algorithm 1 satisfies

$$\begin{aligned} &\left|\widehat{\sigma}^2_{\text{AVC}} - \widehat{\sigma}'^2_{\text{AVC}}\right| \\ &= \left|\frac{\operatorname{tr}\left[(\boldsymbol{I} - \boldsymbol{P})\left(\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top - \boldsymbol{A}\boldsymbol{A}^\top\right)(\boldsymbol{I} - \boldsymbol{P})\right]}{n-q}\right|, \end{aligned}$$

where $\boldsymbol{A}$ is given in Algorithm 1. One then invokes Theorem 3 to bound the singular values of $\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top - \boldsymbol{A}\boldsymbol{A}^\top$:

$$\begin{aligned} \left|\widehat{\sigma}^2_{\text{AVC}} - \widehat{\sigma}'^2_{\text{AVC}}\right| &\leq \epsilon \cdot \frac{\operatorname{tr}\left[(\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{P})\right]}{n-q} \\ &\leq \frac{\epsilon\sum_{i=1}^{n-q}\lambda_i\left(\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top\right)}{n-q} \end{aligned}$$

fails with probability at most $3/n$. The second line follows from the exponentiated Horn's inequalities and the fact that $\boldsymbol{I} - \boldsymbol{P}$ is an idempotent projection matrix of rank $n-q$. The sum in the fraction equals to the Ky Fan $(n-q)$-norm of $\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top$.

To show the bound for $\widehat{\mathbf{\Lambda}}_{\text{AVC}}$, it follows from (15) that

$$\begin{aligned} &\left\|\widehat{\mathbf{\Lambda}}_{\text{AVC}} - \widehat{\mathbf{\Lambda}}'_{\text{AVC}}\right\|_2 \\ &\leq \left\|\boldsymbol{Z}^\dagger\left(\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top - \boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{Z}^{\dagger\top}\right\|_2 \\ &\quad + \left|\widehat{\sigma}^2_{\text{AVC}} - \widehat{\sigma}'^2_{\text{AVC}}\right|\left\|\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\right\|_2 \\ &\leq \frac{\epsilon}{\sigma_{\text{min}}(\boldsymbol{Z})^2}\left(\left\|\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top\right\|_2 + \frac{\left|\left|\left|\boldsymbol{X}\mathbf{\Phi}\boldsymbol{X}^\top\right|\right|\right|_{n-q}}{n-q}\right), \end{aligned}$$

where we used Theorem 3 and the earlier bound on $\widehat{\sigma}^2_{\text{AVC}}$.

□