Supplemental material - Improving training of deep neural networks via Singular Value Bounding

Lemma 1 *For a matrix $\boldsymbol{W} \in \mathbb{R}^{M \times N}$ with singular values of all 1, and a diagonal matrix $\boldsymbol{G} \in \mathbb{R}^{M \times M}$ with nonzero entries $\{g_i\}_{i=1}^{M}$, let $g_{\max} = \max(|g_1|, \dots, |g_M|)$ and $g_{\min} = \min(|g_1|, \dots, |g_M|)$, the singular values of $\widetilde{\boldsymbol{W}} = \boldsymbol{G}\boldsymbol{W}$ is bounded in $[g_{min}, g_{\max}]$. When $\boldsymbol{W}$ is fat, i.e., $M \leq N$, and rank$(\boldsymbol{W}) = M$, singular values of $\widetilde{\boldsymbol{W}}$ are exactly $\{|g_i|\}_{i=1}^{M}$.*

Proof. We first consider the general case, and let $P = \min(M, N)$. Denote singular values of $\boldsymbol{W}$ as $\sigma_1 = \cdots = \sigma_P = 1$, and singular values of $\widetilde{\boldsymbol{W}}$ as $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_P$. Based on the properties of matrix extreme singular values, we have

$$\sigma_1 = \|\boldsymbol{W}\|_2 = \max_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{W}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \min_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{W}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_P = 1.$$

Let $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x} \neq 0} \frac{\|\widetilde{\boldsymbol{W}}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$, we have

$$\tilde{\sigma}_1 = \frac{\|\widetilde{\boldsymbol{W}}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2} = \frac{\|\boldsymbol{G}\boldsymbol{W}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2} \leq \frac{\|\boldsymbol{G}\|_2 \|\boldsymbol{W}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2},$$

where we have used the fact that $\|\boldsymbol{A}\boldsymbol{b}\|_2 \leq \|\boldsymbol{A}\|_2 \|\boldsymbol{b}\|_2$ for any $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$. We thus have

$$\tilde{\sigma}_1 \leq \|\boldsymbol{G}\|_2 \frac{\|\boldsymbol{W}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2} \leq \|\boldsymbol{G}\|_2 \max_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{W}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = |g_{\max}|.$$

Since $\boldsymbol{G}$ has nonzero entries, we have $\boldsymbol{W} = \boldsymbol{G}^{-1}\widetilde{\boldsymbol{G}}$. Let $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \neq 0} \frac{\|\widetilde{\boldsymbol{W}}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$, the properties of matrix extreme singular values give $\tilde{\sigma}_P = \frac{\|\widetilde{\boldsymbol{G}}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2}$, and $\sigma_P = \min_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{W}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = 1$. We thus have

$$1 = \min_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{G}^{-1}\widetilde{\boldsymbol{G}}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \frac{\|\boldsymbol{G}^{-1}\widetilde{\boldsymbol{G}}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2} \leq \|\boldsymbol{G}^{-1}\|_2 \frac{\|\widetilde{\boldsymbol{G}}\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2},$$

which gives $\tilde{\sigma}_P \geq |g_{\min}|$. Overall, we have

$$|g_{\max}| \geq \tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_P \geq |g_{\min}|.$$

We next consider the special case of $M \leq N$ and rank$(\boldsymbol{W}) = M$. Without loss of generality, we assume diagonal entries $\{g_i\}_{i=1}^{M}$ of $\boldsymbol{G}$ are all positive and ordered. By definition we have $\widetilde{\boldsymbol{W}} = \boldsymbol{I}\boldsymbol{G}\boldsymbol{W}$, where $\boldsymbol{I}$ is an identity matrix of size $M \times M$. Let $\boldsymbol{V} = \left[\boldsymbol{W}^\top, \boldsymbol{W}^{\perp\top}\right]$, where $\boldsymbol{W}^\perp$ denotes the orthogonal complement of $\boldsymbol{W}$, we thus have the SVD of $\widetilde{\boldsymbol{W}}$ by construction as $\widetilde{\boldsymbol{W}} = \boldsymbol{I}\,[\boldsymbol{G}, \boldsymbol{0}]\,\boldsymbol{V}^\top$. When some values of $\{g_i\}_{i=1}^{M}$ are not positive, the SVD can be constructed by changing the signs of the corresponding columns of either $\boldsymbol{I}$ or $\boldsymbol{V}$. Since matrix singular values are uniquely determined (while singular vectors are not), singular values of $\widetilde{\boldsymbol{W}}$ are thus exactly $\{|g_i|\}_{i=1}^{M}$. □