Learning One Convolutional Layer with Overlapping Patches

Surbhi Goel [1] **Adam Klivans** [1] **Raghu Meka** [2]

1. Omitted Proofs

1.1. Proof of Lemma 1

We will follow the following notation:

$$T_1 = \mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}}[\sigma(a^T x)(b^T x)]$$
$$T_2 = \mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}}[(a^T x)(b^T x)].$$

Since x is drawn from a symmetric distribution we have $E_{x \sim \mathcal{D}_{\mathcal{X}}}[F(x)] = E_{x \sim \mathcal{D}_{\mathcal{X}}}[F(-x)]$ for any function F. Thus, we have

$$T_1 = \mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}}[\sigma(-a^T x)(-b^T x)]$$
$$\implies 2T_1 = \mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}}[(\sigma(-a^T x) - \sigma(-a^T x))(b^T x)]$$

Observe that $\sigma(c) - \sigma(-c) = \frac{(1-\alpha)|c| + (1+\alpha)c}{2} - \frac{(1-\alpha)|a| - (1+\alpha)c}{2} = (1+\alpha)c$. Substituting this in the above, we get the required result $2T_1 = (1+\alpha)T_2$.

1.2. Proof of Lemma 2

We have,

$$\begin{aligned}
&\frac{1}{k} \sum_{1 \le i \le k} \mathbb{E}_x[(\sigma(w_* P_i x) - \sigma(w P_i x))(w_* - w)^T P_i x)] \\
&= \frac{1+\alpha}{2k} \sum_{1 \le i \le k} \mathbb{E}_x[((w_* - w)^T P_i x))^2] \\
&= \frac{1+\alpha}{2k} (w_* - w_t)^T \left(\sum_{1 \le i \le k} P_i \mathbb{E}_x[x x^T] P_i^T \right) (w_* - w_t) \\
&= \frac{1+\alpha}{2k} (w_* - w_t)^T \left(\sum_{1 \le i \le k} P_i \Sigma P_i^T \right) (w_* - w_t) \\
&\le \frac{1+\alpha}{2k} \lambda_{\max}(\Sigma) \left(\sum_{1 \le i \le k} \lambda_{\max}(P_i P_i^T) \right) ||w_* - w_t||^2 \\
&= \frac{1+\alpha}{2} \lambda_{\max}(\Sigma) ||w_* - w||^2
\end{aligned}$$

The first equality follows from using Lemma 1 and the last follows since for all i, $P_i P_i^T$ is a permutation of the identity matrix by definition.

Using monotonicity of σ and Jensen's inequality, we also have,

$$\begin{aligned}
&\frac{1}{k} \sum_{1 \le i \le k} \mathbb{E}_x[(\sigma(w_* P_i x) - \sigma(w P_i x))(w_* - w)^T P_i x] \\
&\ge \frac{1}{k} \sum_{1 \le i \le k} \mathbb{E}_x[(\sigma(w_* P_i x) - \sigma(w P_i x))^2] \\
&\ge \mathbb{E}_x \left[\left(\frac{1}{k} \sum_{1 \le i \le k} (\sigma(w_* P_i x) - \sigma(w P_i x)) \right)^2 \right] \\
&= L(w).
\end{aligned}$$

Combining the two above lemmas, we get the required result.

1.3. Proof of Lemma 3

We have,

$$\begin{aligned}
&(f_{w_*}(x) - f_{w_t}(x))^2 \\
&= \left(\frac{1}{k} \sum_{i=1}^{k} (\sigma(w_*^T P_i x) - \sigma(w^T P_i x)) \right)^2 \\
&\le \frac{1}{k} \sum_{i=1}^{k} (\sigma(w_*^T P_i x) - \sigma(w^T P_i x))^2 \\
&\le \frac{1}{k} \sum_{i=1}^{k} (w_*^T P_i x - w^T P_i x)^2 \\
&\le \frac{1}{k} \sum_{i=1}^{k} ||w_* - w||^2 \lambda_{\max}(P_i P_i^T) ||||x||^2 \\
&\le ||w_* - w||^2 ||x||^2
\end{aligned}$$

The first inequality follows from using Jensen's, the second inequality follows from the 1-Lipschitz property of σ, the third follows from observing that $P_i P_i^T$ is a PSD matrix and the last inequality follows since for all i, $\lambda_{\max}(P_i P_i^T) = 1$ since $P_i P_i^T$ is a permutation of the identity matrix.

[*]Equal contribution [1]Department of Computer Science, University of Texas at Austin [2]Department of Computer Science, UCLA. Correspondence to: Surbhi Goel <surbhi@cs.utexas.edu>.

Proceedings of the 35^{th} *International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80, 2018.

2. Properties of Patch Matrix P

Let $r = pd + q$ for some $p \geq 0$ and $1 \leq q \leq d$.

Lemma A. *For $d < r/3$, P^{-1} has the following form:*

$$P^{-1}_{i,j} = \begin{cases} \alpha_0 \text{ if } i = j \in \{1, \dots, q\} \cup \{r - q + 1, \dots, r\} \\ \alpha_1 \text{ if } i = j \in \{q + 1, \dots, d\} \cup \{r - d + 1, r - q\} \\ 1 \text{ if } i = j \in \{d + 1, \dots, r - d\} \\ -0.5 \text{ if } |i - j| = d + 1 \\ \phi \text{ if } |i - j| = (p - 1)d + 1 \text{ and } i \text{ or } j \in \{q + 1, \dots, d\} \\ \beta \text{ if } |i - j| = pd + 1 \\ 0 \text{ otherwise} \end{cases}$$

where $\alpha_0 = \beta + 0.5$, $\alpha_1 = \phi + 0.5$, $\beta = \frac{0.5}{2k-p}$ and $\phi = \frac{0.5}{2k-p-1}$. Also, $\lambda_{\max}(P^{-1}) \leq 2$.

Proof. We need to show that $A = PP^{-1} = I$. Observe that P and P^{-1} are bisymmetric, thus A is centrosymmetric implying $A_{i,j} = A_{r-1-i,r-1-j}$. Hence, we need to only prove that the lower triangular matrix matches I. We show the result for $p > 2$, as the same ideas apply for the other case.

To verify this, consider each diagonal entry,

- $d \leq i \leq \lceil d/2 \rceil$: $A_{i,i} = -0.5(k-1) + k - 0.5(k-1) = 1$.
- $i \in \{1, \dots, q\}$: $A_{i,i} = \alpha_0 k - 0.5(k-1) + \beta\,(k - p) = 1$.
- $i \in \{q + 1, \dots, d\}$: $A_{i,i} = \alpha_1 k - 0.5(k - 1) + \phi\,(k - p - 1) = 1$.

For non-diagonal entries, that is, $j \neq i$,

- $d \leq j \leq \lceil d/2 \rceil$: $A_{i,j} = -0.5P_{i,j-d} + P_{i,j} - 0.5P_{i,j+d}$. If $|i-j| = ad$ then $A_{i,j} = -0.5\,(k - a - 1) + k - a - 0.5\,(k - a + 1) = 0$, else $P_{i,j} = P_{i,j-d} = P_{i,j+d} = 0 \implies A_{i,j} = 0$.
- $j \in \{1, \dots, q\}$: $A_{i,j} = \alpha_0 P_{i,j} - 0.5P_{i,j+d} + \beta P_{i,j+pd}$. Now if $i - j = ad$, then $A_{i,j} = \alpha_0(k - a) - 0.5(k - a + 1) + \beta(k - p + a) = 0$ else $P_{i,j} = P_{i,j+d} = P_{i,j+pd} = 0 \implies A_{i,j} = 0$.
- $j \in \{q + 1, \dots, d\}$: $A_{i,j} = \alpha_1 P_{i,j} - 0.5P_{i,j+d} + \beta P_{i,j+pd}$. Now if $i - j = ad$, then $A_{i,j} = \alpha_1(k - a) - 0.5(k - a + 1) + \phi(k - p + a + 1) = 0$ else $P_{i,j} = P_{i,j+d} = P_{i,j+pd} = 0 \implies A_{i,j} = 0$.

Hence $A = I$.

Using Theorem 1, we have $\lambda_{\max}(P^{-1}) = \max_i \left(P^{-1}_{i,i} + \sum_{j \neq i} |P^{-1}_{i,j}|\right)$. If $q < d$, then $\lambda_{\max}(P^{-1}) = \max(\alpha_0 + 0.5 + \beta, \alpha_1 + 0.5 + \phi, 1 + 0.5 + 0.5) = \max(2\beta + 1, 2\phi + 1, 2) = 2$ as $\beta, \phi \leq 0.5$ which follows from $2k - p - 1 \geq 1$. Similarly, when $q = d$, $\lambda_{\max}(P^{-1}) = \max(\alpha_0 + 0.5 + \beta, 1 + 0.5 + 0.5) = \max(2\beta + 1, 2) = 2$. ☐

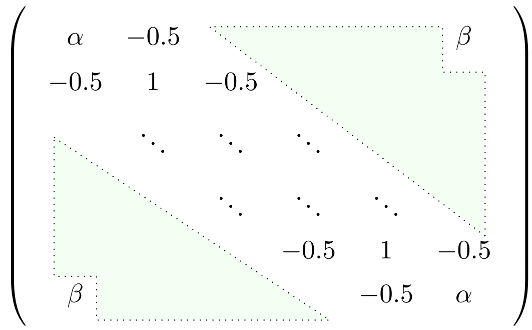

Figure 1. P^{-1} for $d = 1$. Here $\alpha = \beta + 0.5$ and $\beta = \frac{0.5}{2k-p} = \frac{0.5}{2n-3r+3}$. The shaded area is all 0s.

Lemma B. *For $r/3 \leq d < r/2$, P^{-1} has the following form:*

$$P^{-1}_{i,j} = \begin{cases} \alpha_0 \text{ if } i = j \in \{1, \dots, q\} \cup \{r - q + 1, \dots, r\} \\ \alpha_1 \text{ if } i = j \in \{q + 1, \dots, d\} \cup \{r - d + 1, r - q\} \\ 1 \text{ if } i = j \in \{d + 1, \dots, r - d\} \\ -0.5 \text{ if } |i - j| = d + 1 \text{ and } i \text{ or } j \in \{d + 1, \dots, r - d\} \\ \phi \text{ if } |i - j| = d + 1 \text{ and } i \text{ or } j \in \{q + 1, \dots, d\} \\ \beta \text{ if } |i - j| = 2d + 1 \\ 0 \text{ otherwise} \end{cases}$$

where $\alpha_0 = \beta + 0.5$, $\alpha_1 = \frac{k}{2k-1}$, $\beta = \frac{0.5}{2k-2}$ and $\phi = -\frac{k-1}{2k-1}$. Also, $\lambda_{\max}(P^{-1}) \leq 2$.

Proof. Similar to the previous lemma, to verify this, consider each diagonal entry,

- $d \leq i \leq \lceil d/2 \rceil$: $A_{i,i} = -0.5(k-1) + k - 0.5(k-1) = 1$.
- $i \in \{1, \dots, q\}$: $A_{i,i} = \alpha_0 k - 0.5(k-1) + \beta\,(k - p) = 1$.
- $i \in \{q + 1, \dots, d\}$: $A_{i,i} = \alpha_1 k + \phi(k - 1) = 1$.

For non-diagonal entries, that is, $j \neq i$,

- $d \leq j \leq \lceil d/2 \rceil$: $A_{i,j} = -0.5P_{i,j-d} + P_{i,j} - 0.5P_{i,j+d}$. If $|i-j| = ad$ then $A_{i,j} = -0.5\,(k - a - 1) + k - a - 0.5\,(k - a + 1) = 0$, else $P_{i,j} = P_{i,j-d} = P_{i,j+d} = 0 \implies A_{i,j} = 0$.

- $j \in \{1, \ldots, q\}$: $A_{i,j} = \alpha_0 P_{i,j} - 0.5 P_{i,j+d} + \beta P_{i,j+2d}$. Now if $i - j = ad$, then $A_{i,j} = \alpha_0(k-a) - 0.5(k-a+1) + \beta(k-2+a) = 0$ else $P_{i,j} = P_{i,j+d} = P_{i,j+pd} = 0 \implies A_{i,j} = 0$.
- $j \in \{q+1, \ldots, d\}$: $A_{i,j} = \alpha_1 P_{i,j} + \phi P_{i,j+d}$. Now if $i - j = ad$, then $a = 1$, implying $A_{i,j} = \alpha_1(k-1) + \phi k = 0$ else $P_{i,j} = P_{i,j+d} = 0 \implies A_{i,j} = 0$.

Hence $A = I$.

Similar to the previous lemma, we have $\lambda_{\max}(P^{-1}) = \max(\alpha_0 + 0.5 + \beta, \alpha_1 + |\phi|, 1 + 0.5 + 0.5) = \max(2\beta + 1, 1, 2) = 2$ as $\alpha_1 + |\phi| = 1$ and $\beta \leq 0.5$ which follows from $2k - p - 1 \geq 1$. $\square$