

## Supplementary Material

### A. Proof of Theorem 4.1

*Proof.* When  $L$  is finite, the inner product between the feature representation is given by

$$\begin{aligned} \mathbf{h}_{\mathbf{x}}^T \mathbf{h}_{\mathbf{y}} &= \frac{1}{L} (W^T A^T \mathbf{x})_+^T (W^T A^T \mathbf{y})_+ \\ &= \frac{1}{L} (\mathbf{x}^T A W)_+ (W^T A^T \mathbf{y})_+ \\ &= \frac{1}{L} \sum_{l=1}^L (\mathbf{x}^T A w_l)_+ (\mathbf{y}^T A w_l)_+ \end{aligned}$$

Since  $w_l$ ,  $1 \leq l \leq L$  is a random vector, the above quantity is the empirical mean of  $L$  random variables,  $z_1, \dots, z_L$ , where  $z_l = (\mathbf{x}^T A w_l)_+ (\mathbf{y}^T A w_l)_+$ . Furthermore since  $w_l$ 's are independent and identically distributed,  $z_l$ 's being functions of  $w_l$ 's are also independent and identically distributed. Hence, by law of large numbers, as  $L \rightarrow +\infty$ , the empirical mean of  $z_l$ 's converges to the true mean, that is,

$$\lim_{L \rightarrow \infty} \mathbf{h}_{\mathbf{x}}^T \mathbf{h}_{\mathbf{y}} \quad (10)$$

$$= \lim_{L \rightarrow +\infty} \frac{1}{L} \sum_{l=1}^L (\mathbf{x}^T A w_l)_+ (\mathbf{y}^T A w_l)_+ \quad (11)$$

$$= \int_{w \in \mathbb{R}^d} (\mathbf{x}^T A w)_+ (\mathbf{y}^T A w)_+ p(w) dw \quad (12)$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{w \in \mathbb{R}^d} (\mathbf{x}^T A w)_+ (\mathbf{y}^T A w)_+ e^{-\frac{\|w\|^2}{2}} dw \quad (13)$$

$$= \frac{1}{2\pi} \|A^T \mathbf{x}\| \|A^T \mathbf{y}\| (\sin \theta_A + (\pi - \theta_A) \cos \theta_A), \quad (14)$$

where  $\theta_A = \cos^{-1} \frac{\mathbf{x}^T A A^T \mathbf{y}}{\|A \mathbf{x}\| \|A \mathbf{y}\|}$ . Here, the last equation follows from the derivation of arc-cosine kernel by [Cho & Saul \(2010\)](#).  $\square$

### B. Proof of Theorem 5.1

*Proof.* After  $T$  iteration of Algorithm 1, the approximate kernel matrix  $k_{ap}(\mathbf{x}, \mathbf{y})$  is given by

$$\begin{aligned} k_{ap}(\mathbf{x}, \mathbf{y}) &= \frac{1}{T} \sum_{t=1}^T \frac{1}{L I^2} \sum_{i=1}^I \sum_{j=1}^I \sum_{\ell=1}^L (p_i^T A w_{t\ell})_+ (q_j^T A w_{t\ell})_+, \\ &= \frac{1}{T L} \sum_{t,l} \sum_{i,j} (p_i^T A w_{t\ell})_+ (q_j^T A w_{t\ell})_+. \end{aligned}$$

Furthermore, the mean and variance of the approximate

kernel after  $T$  iterations is given by

$$\begin{aligned} m(\mathbf{x}, \mathbf{y}) &= \frac{1}{T L I^2} \sum_{t,l} \sum_{i,j} \mathbb{E}[(p_i^T A w_{t\ell})_+ (q_j^T A w_{t\ell})_+] \\ &= \frac{1}{T L I^2} \sum_{t,l} \sum_{i,j} k_{\Sigma}(p_i, q_j) \\ &= \frac{1}{I^2} \sum_{i,j} k_{\Sigma}(p_i, q_j) \\ &= k(x, y). \\ v(\mathbf{x}, \mathbf{y}) &= \frac{1}{T^2 L^2 I^2} \sum_{t,l} \text{var}\left(\frac{1}{I^2} \sum_{i,j} (p_i^T A w_{t\ell})_+ (q_j^T A w_{t\ell})_+\right) \\ &\leq \frac{1}{T^2 L^2 I^2} \sum_{t,l} \sum_{i,j} \text{var}((p_i^T A w_{t\ell})_+ (q_j^T A w_{t\ell})_+) \\ &= \frac{1}{T L I^2} \sum_{i,j} \text{var}((p_i^T A w)_+ (q_j^T A w)_+) \end{aligned}$$

Since  $\text{var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 \leq \mathbb{E}X^2$ ,  $\text{var}((p_i^T A w)_+ (q_j^T A w)_+) \leq \mathbb{E}[(p_i^T A w)_+^2 (q_j^T A w)_+^2]$ . From equation (7) of [Cho et al. \(?\)](#), the same quantity can be rewritten as

$$\begin{aligned} &\mathbb{E}[(p_i^T A w)_+^2 (q_j^T A w)_+^2] \\ &= \|A p_i\|^2 \|A q_j\|^2 \frac{(3 \sin \theta \cos \theta + (\pi - \theta)(1 + 2 \cos^2 \theta))}{2\pi} \\ &\leq 2 \|A p_i\|^2 \|A q_j\|^2. \end{aligned}$$

Hence, by applying Chebyshev's inequality to the quantity  $k_{ap}(\mathbf{x}, \mathbf{y})$ , we get

$$\begin{aligned} &P(|k_{ap}(\mathbf{x}, \mathbf{y}) - k(\mathbf{x}, \mathbf{y})| \geq \epsilon) \\ &\leq 2 \frac{1}{T L I^2 \epsilon^2} \sum_{i=1}^I \sum_{j=1}^I \|A p_i\|^2 \|A q_j\|^2. \end{aligned}$$

Hence, as  $T \rightarrow \infty$ , approximate kernel entry converges in probability to the exact kernel entry. Rewriting the above statement, we can say that with probability at least  $1 - \delta$ .

$$\begin{aligned} &|k_{ap}(\mathbf{x}, \mathbf{y}) - k(\mathbf{x}, \mathbf{y})| \\ &\leq \sqrt{\frac{2}{T L \delta}} \sqrt{\frac{1}{I^2} \sum_{i=1}^I \sum_{j=1}^I \|A p_i\|^2 \|A q_j\|^2} \end{aligned}$$

$\square$

990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044

1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099