Supplementary Material

A. Proof of Theorem 4.1

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

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

Since $w_l,\ 1 \le l \le L$ is a random vector, the above quantity is the empirical mean of L random variables, $z_1, ..., z_L$, where $z_l = (\mathbf{x}^{\mathrm{T}}Aw_l)_+(\mathbf{y}^{\mathrm{T}}Aw_l)_+$. 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 $L \to +\infty$, the empirical mean of z_l's converges to the true mean, that is,

$$\lim_{L\to\infty}\mathbf{h}_{\mathbf{x}}^{\mathrm{T}}\mathbf{h}_{\mathbf{y}} \tag{10}$$

$$= \lim_{L\to+\infty}\frac{1}{L}\sum_{l=1}^{L}(\mathbf{x}^{\mathrm{T}}Aw_l)_+(\mathbf{y}^{\mathrm{T}}Aw_l)_+ \tag{11}$$

$$= \int_{w\in\mathbb{R}^d}(\mathbf{x}^{\mathrm{T}}Aw)_+(\mathbf{y}^{\mathrm{T}}Aw)_+p(w)\mathrm{d}w \tag{12}$$

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

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

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

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

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

Furthermore, the mean and variance of the approximate kernel after T iterations is given by

$$\begin{aligned}m(\mathbf{x},\mathbf{y}) &= \frac{1}{TLI^2}\sum_{t,l}\sum_{i,j}\mathbb{E}[(p_i^{\mathrm{T}}Aw_{t\ell})_+(q_j^{\mathrm{T}}Aw_{t\ell})_+] \\ &= \frac{1}{TLI^2}\sum_{t,l}\sum_{i,j}k_\Sigma(p_i,q_j) \\ &= \frac{1}{I^2}\sum_{i,j}k_\Sigma(p_i,q_j) \\ &= k(x,y). \\ v(\mathbf{x},\mathbf{y}) &= \frac{1}{T^2L^2}\sum_{t,l}var(\frac{1}{I^2}\sum_{i,j}(p_i^{\mathrm{T}}Aw_{t\ell})_+(q_j^{\mathrm{T}}Aw_{t\ell})_+) \\ &\le \frac{1}{T^2L^2I^2}\sum_{t,l}\sum_{i,j}var((p_i^{\mathrm{T}}Aw_{t\ell})_+(q_j^{\mathrm{T}}Aw_{t\ell})_+) \\ &= \frac{1}{TLI^2}\sum_{i,j}var((p_i^{\mathrm{T}}Aw)_+(q_j^{\mathrm{T}}Aw)_+)\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}[(p_i^{\mathrm{T}}Aw)_+^2(q_j^{\mathrm{T}}Aw)_+^2] \\ &= \|Ap_i\|^2\|Aq_j\|^2\frac{(3\sin\theta\cos\theta + (\pi-\theta)(1+2\cos^2\theta))}{2\pi} \\ &\le 2\|Ap_i\|^2\|Aq_j\|^2 \quad .\end{aligned}$$

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

$$\begin{aligned}&P(|k_{ap}(\mathbf{x},\mathbf{y}) - k(\mathbf{x},\mathbf{y})| \ge \epsilon) \\ &\le 2\frac{1}{TLI^2\epsilon^2}\sum_{i=1}^{I}\sum_{j=1}^{I}\|Ap_i\|^2\|Aq_j\|^2 \,.\end{aligned}$$

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

$$\begin{aligned}&|k_{ap}(\mathbf{x},\mathbf{y}) - k(\mathbf{x},\mathbf{y})| \\ &\le \sqrt{\frac{2}{TL\delta}}\sqrt{\frac{1}{I^2}\sum_{i=1}^{I}\sum_{j=1}^{I}\|Ap_i\|^2\|Aq_j\|^2}\end{aligned}$$

□