

---

# Supplemental Material: Conditional Adversarial Domain Adaptation

---

Mingsheng Long<sup>†</sup>, Zhangjie Cao<sup>†</sup>, Jianmin Wang<sup>†</sup>, and Michael I. Jordan<sup>‡</sup>

<sup>†</sup>School of Software, Tsinghua University, China

<sup>†</sup>KLiss, MOE; BNRist; Research Center for Big Data, Tsinghua University, China

<sup>‡</sup>University of California, Berkeley, Berkeley, USA

{mingsheng, jimwang}@tsinghua.edu.cn caozhangjie14@gmail.com  
jordan@berkeley.edu

## 1 Proof of Theorem 1

This supplemental material provides proof to Theorem 1 in the main paper. To enable better readability, denote by  $\mathbf{z}^1 = \mathbf{f}$ ,  $\mathbf{z}^2 = \mathbf{g}$  and  $\mathbf{R}^1 = \mathbf{R}_f$  and  $\mathbf{R}^2 = \mathbf{R}_g$ . We first rewrite the randomized feature map

$$T_{\odot}(\mathbf{z}) = \frac{1}{\sqrt{d}} (\odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}^{\ell}). \quad (1)$$

**Theorem 1.** *The expectation and variance of the inner products between the randomized feature maps  $T_{\odot}(\mathbf{z})$  (1) generated by random matrices  $\mathbf{R}^{\ell}$ ,  $\ell = 1, 2$  satisfy*

$$\mathbb{E} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle] = \prod_{\ell} \langle \mathbf{z}^{\ell}, \mathbf{z}'^{\ell} \rangle, \quad (2)$$

$$\text{var} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle] = \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left[ \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j'^{\ell})^2 \mathbb{E} [(R_{ij}^{\ell})^4] + C' \right] + C, \quad (3)$$

where  $C$  and  $C'$  are constants that do not depend on the random matrices  $\mathbf{R}^{\ell}$ ,  $\ell = 1, 2$ .

Theorem 1 reveals that the inner product between the randomized feature maps  $T_{\odot}(\mathbf{z})$  is an *unbiased* estimate of the inner product between the original multilinear fusions based on tensor products  $T_{\otimes}(\mathbf{z})$ . The variance of the inner product between the randomized feature maps  $T_{\odot}(\mathbf{z})$  is depending only on the moments  $\mathbb{E}[(R_{ij}^{\ell})^4]$ , which are constants for many symmetric distributions with univariance, i.e.  $\mathbb{E}[R_{ij}^{\ell}] = 0$ ,  $\mathbb{E}[(R_{ij}^{\ell})^2] = 1$ . We can verify that: (1) for Bernoulli distribution,  $\mathbb{E}(R_{ij}^{\ell})^4 = 1$ ; (2) for standard normal distribution,  $\mathbb{E}(R_{ij}^{\ell})^4 = 3$ ; (3) for uniform distribution,  $\mathbb{E}(R_{ij}^{\ell})^4 = 1.8$ . Therefore, for continuous sampling distributions, uniform distribution will yield the lowest estimation variance. The empirical study confirms that uniform distribution leads to the best multilinear fusion accuracy.

*Proof.*

$$\begin{aligned} \mathbb{E} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle] &= \mathbb{E} \left[ \left\langle \frac{1}{\sqrt{d}} (\odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}^{\ell}), \frac{1}{\sqrt{d}} (\odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}'^{\ell}) \right\rangle \right] \\ &= \frac{1}{d} \mathbb{E} \left[ \langle \odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}^{\ell}, \odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}'^{\ell} \rangle \right] \\ &= \frac{1}{d} \mathbb{E} \left[ \sum_{i=1}^d \odot_{\ell} (\mathbf{R}_i^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_i^{\ell} \mathbf{z}'^{\ell}) \right] = \frac{1}{d} \sum_{i=1}^d \mathbb{E} [\odot_{\ell} (\mathbf{R}_i^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_i^{\ell} \mathbf{z}'^{\ell})] \\ &= \frac{1}{d} \sum_{i=1}^d \left[ \prod_{\ell} \mathbf{z}^{\ell T} \mathbb{E} [\mathbf{R}_i^{\ell T} \mathbf{R}_i^{\ell}] \mathbf{z}'^{\ell} \right] = \frac{1}{d} \sum_{i=1}^d \left[ \prod_{\ell} \mathbf{z}^{\ell T} \mathbf{z}'^{\ell} \right] = \prod_{\ell} \langle \mathbf{z}^{\ell}, \mathbf{z}'^{\ell} \rangle. \end{aligned} \quad (4)$$

$$\begin{aligned}
\text{var} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle] &= \mathbb{E} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle^2] - \mathbb{E} [\langle T_{\odot}(\mathbf{z}), T_{\odot}(\mathbf{z}') \rangle]^2 \\
&= \mathbb{E} \left[ \left\langle \frac{1}{\sqrt{d}} (\odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}^{\ell}), \frac{1}{\sqrt{d}} (\odot_{\ell} \mathbf{R}^{\ell} \mathbf{z}'^{\ell}) \right\rangle^2 \right] - \left( \prod_{\ell} \langle \mathbf{z}^{\ell}, \mathbf{z}'^{\ell} \rangle \right)^2 \\
&= \left[ \frac{1}{d^2} \left( \sum_{i=1}^d \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right)^2 \right] - C_1 \\
&= \frac{1}{d^2} \mathbb{E} \left[ \sum_{i=1}^d \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right)^2 + \sum_{i=1}^d \sum_{j \neq i}^d \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \left( \prod_{\ell} (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \right] - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right)^2 \right] \\
&+ \frac{1}{d^2} \sum_{i=1}^d \sum_{j \neq i}^d \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \left( \prod_{\ell} (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \right] - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right)^2 \right] \\
&+ \frac{1}{d^2} \sum_{i=1}^d \sum_{j \neq i}^d \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \right] \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{j \cdot}^{\ell} \mathbf{z}'^{\ell}) \right) \right] - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \mathbb{E} \left[ \left( \prod_{\ell} (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}^{\ell}) (\mathbf{R}_{i \cdot}^{\ell} \mathbf{z}'^{\ell}) \right)^2 \right] + \frac{d(d-1)}{d^2} \left( \prod_{\ell} \langle \mathbf{z}^{\ell}, \mathbf{z}'^{\ell} \rangle \right)^2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \mathbb{E} \left[ \left( \prod_{\ell} \mathbf{z}^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \right)^2 \right] + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \mathbb{E} \left[ \prod_{\ell} \mathbf{z}^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}^{\ell} \right] + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \mathbf{z}^{\ell \top} \mathbb{E} \left[ (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \right] \mathbf{z}^{\ell} + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \sum_{j=1}^{d_{\ell}} \sum_{k=1}^{d_{\ell}} z_j^{\ell \top} \mathbb{E} \left[ (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \right]_{jk} z_k^{\ell} + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} z_j^{\ell \top} \mathbb{E} \left[ (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \right]_{jj} z_j^{\ell} \right. \\
&+ \left. \sum_{k \neq j}^{d_{\ell}} z_j^{\ell \top} \mathbb{E} \left[ (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \right]_{jk} z_k^{\ell} \right) + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} z_j^{\ell \top} \mathbb{E} \left[ (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \mathbf{z}'^{\ell} \mathbf{z}'^{\ell \top} (\mathbf{R}_{i \cdot}^{\ell \top} \mathbf{R}_{i \cdot}^{\ell}) \right]_{jj} z_j^{\ell} + \sum_{j=1}^{d_{\ell}} \sum_{k \neq j}^{d_{\ell}} z_j^{\ell} z_k^{\ell} z_j^{\ell \top} z_k^{\ell \top} \right) + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j^{\ell})^2 \sum_{k=1}^{d_{\ell}} \mathbb{E} \left[ (R_{ij}^{\ell} R_{ij}^{\ell}) \right] \left[ (R_{ik}^{\ell} R_{ik}^{\ell}) \right] + C_3 \right) + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j^{\ell})^2 \mathbb{E} \left[ (R_{ij}^{\ell})^4 \right] + \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j^{\ell})^2 \sum_{k \neq j}^{d_{\ell}} 1 + C_3 \right) + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j^{\ell})^2 \mathbb{E} \left[ (R_{ij}^{\ell})^4 \right] + C_4 + C_3 \right) + C_2 - C_1 \\
&= \frac{1}{d^2} \sum_{i=1}^d \prod_{\ell} \left( \sum_{j=1}^{d_{\ell}} (z_j^{\ell})^2 (z_j^{\ell})^2 \mathbb{E} \left[ (R_{ij}^{\ell})^4 \right] + C' \right) + C.
\end{aligned} \tag{5}$$

Since the equations in this proof look quite lengthy, we simplify the notations by denoting any parts of the equations independent on random matrices  $\mathbf{R}^{\ell}$ ,  $\ell = 1, 2$  as constants, such as  $C_1 \sim C_4$ ,  $C$ , and  $C'$ .  $\square$