ACsleuth: Domain Adaptive and Fine-grained Anomalous Cell Detection for Single-cell Multiomics

Supplementary Material

1 Proof of Theorem

Proof. [1] provides an unbiased empirical MMD for samples:

$$\begin{aligned} & MMD^2\left(\boldsymbol{\delta}_m^x, \boldsymbol{\delta}_n^y\right) \\ & = \frac{1}{m(m-1)} \sum_i^m \sum_{j \neq i}^m k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^x) + \frac{1}{n(n-1)} \sum_i^n \sum_{j \neq i}^n k(\boldsymbol{\delta}_i^y, \boldsymbol{\delta}_j^y) - \frac{2}{mn} \sum_i^m \sum_j^n k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^y) \\ & = \sum_i^{m+n} \sum_{j \neq i}^{m+n} k(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \gamma(s_i, s_j) \end{aligned} \tag{1}$$

where $\boldsymbol{\delta}_i$ denotes unlabeled total samples, and the adjustment coefficients γ are defined as:

$$\gamma(s_i, s_j) = \begin{cases} \frac{1}{m(m-1)}, & s_i = s_j = 0 \\ \frac{1}{n(n-1)}, & s_i = s_j = 1 \\ \frac{-1}{mn}, & s_i \neq s_j \end{cases} \tag{2}$$

If $s_i = 1$, sample i will be classified as anomaly; otherwise, it will be classified as normal. By the way, the theorem has been proved. □

2 Proof of Theorem

Proof. We define the two-dimensional sequence $\{\gamma_{mn}\}_{m,n \in \mathbb{Z}}$ as:

$$\gamma_{00} = \frac{1}{m(m-1)}, \quad \gamma_{01} = \gamma_{10} = \frac{-1}{mn}, \quad \gamma_{11} = \frac{1}{n(n-1)}, \quad \forall i, j \geq 2, \gamma_{ij} = 0 \tag{3}$$

The ordinary generating function $H(x, y)$ for γ_{mn} is:

$$H(x, y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \gamma_{ij} x^i y^j \tag{4}$$

where $(x, y) \in \mathbb{D} \coloneqq [0, 1] \times [0, 1]$. In this case, $H(x, y)$ is a formal power series, and γ_{ij} corresponds to the coefficients in front of each term $x^i y^j$.

According to [2, 3], the extension of Ramanujan's master theorem in the k-dimensional case have been proposed as Lemma 2.1.

Lemma 2.1. *If a complex-valued function $f(x_1, \cdots, x_k)$ has an expansion:*

$$f(x_1, \cdots, x_k) = \sum_{n_1, \cdots, n_k}^{\infty} g(n_1, \cdots, n_k) \prod_{i=1}^{k} \frac{(-1)^{n_i}}{n_i!} x_i^{n_i} \tag{5}$$

where $g(n_1, \cdots, n_k)$ is a continuously analytic function everywhere, then the k-dimensional Mellin transform satisfies a multivariate version of Ramanujan's master theorem as follows:

$$\begin{aligned} mathcal M[f(x_1, \cdots, x_k)](s_1, \cdots, s_k) &\coloneqq \int_{\mathbb{R}_+^k} \prod_{i=1}^{k} x_i^{s_i - 1} f(x_1, \cdots, x_k) dx_1 \cdots dx_k \\ &= \prod_{i=1}^{k} \Gamma(s_i) g(-s_1, \cdots, -s_k) \end{aligned} \tag{6}$$

The integral is convergent when $0 < Re(s_i) < 1, \forall i \in \{1, \cdots, k\}$.

Note that $H(-x, -y)$ satisfies:

$$H(-x, -y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \gamma_{ij} \Gamma(i+1) \Gamma(j+1) \frac{(-1)^i (-1)^j}{i! j!} x^i y^j \tag{7}$$

By Lemma 2.1, the k-dimensional Mellin transform follows:

$$\begin{aligned} \mathcal{M}[H(-x, -y)](s, t) &\coloneqq \int_{\mathbb{D}} x^{s-1} y^{t-1} H(-x, -y) dx dy \\ &= \Gamma(s) \Gamma(t) \Gamma(1-s) \Gamma(1-t) \gamma_c(-s, -t) \\ &= \frac{\pi^2}{\sin \pi s \sin \pi t} \gamma_c(-s, -t) \end{aligned} \tag{8}$$

where $\gamma_c(s, t)$ is exactly the extension of sequence γ_{ij} in the continuous scenario. Ultimately, by solving the definite integral, we can obtain:

$$\begin{aligned} \gamma_c(-s, -t) &= \frac{\sin \pi s \sin \pi t}{\pi^2} \int_{\mathbb{D}} x^{s-1} y^{t-1} H(-x, -y) dx dy \\ &= \frac{\sin \pi s \sin \pi t}{\pi^2} \int_{\mathbb{D}} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \gamma_{ij} x^{i+s-1} y^{j+t-1} dx dy \\ &= \frac{\sin \pi s \sin \pi t}{\pi^2} \int_{\mathbb{D}} (\gamma_{00} x^{s-1} y^{t-1} + \gamma_{10} x^s y^{t-1} + \gamma_{01} x^{s-1} y^t + \gamma_{11} x^s y^t) dx dy \\ &= \frac{\sin \pi s \sin \pi t}{\pi^2} \left(\gamma_{00} \frac{x^s}{s} \Big|_0^1 \frac{y^t}{t} \Big|_0^1 + \gamma_{10} \frac{x^{s+1}}{s+1} \Big|_0^1 \frac{y^t}{t} \Big|_0^1 + \gamma_{01} \frac{x^s}{s} \Big|_0^1 \frac{y^{t+1}}{t+1} \Big|_0^1 + \gamma_{11} \frac{x^{s+1}}{s+1} \Big|_0^1 \frac{y^{t+1}}{t+1} \Big|_0^1 \right) \\ &= \frac{\sin \pi s \sin \pi t}{\pi^2} \left(\frac{[m(m-1)]^{-1}}{st} + \frac{(mn)^{-1}}{(s+1)t} + \frac{(mn)^{-1}}{s(t+1)} + \frac{[n(n-1)]^{-1}}{(s+1)(t+1)} \right) \end{aligned} \tag{9}$$

By replacing $-s$ and $-t$, the original theorem is thereby proven. $\square$

3 Proof of Theorem

Proof. Let the reference samples belong to batch k and the target samples belong to batch k'. If our assumption holds, the reference samples $\boldsymbol{\zeta}_l$ satisfy:

$$\begin{cases} \boldsymbol{\zeta}_l = \boldsymbol{\zeta}_l^* + \boldsymbol{b}_l^k + \boldsymbol{\epsilon}_l \\ \widehat{\boldsymbol{\zeta}}_l \coloneqq G(\boldsymbol{\zeta}_l) = \widehat{\boldsymbol{\zeta}}_l^* + \boldsymbol{b}_l^k \end{cases} \tag{10}$$

where $\widehat{\boldsymbol{\zeta}}_m^* \sim P_{\widehat{\boldsymbol{\zeta}}^*}, \boldsymbol{b}_l^k \sim P_{\boldsymbol{b}^k}$, and l is the reference sample size. Because the generator doesn't learn the distribution of random noise, $\widehat{\boldsymbol{\zeta}}_l$ excludes $\boldsymbol{\epsilon}_l$.

Considering that the target samples $\widehat{\boldsymbol{y}}_n, \widehat{\boldsymbol{x}}_m$ are reconstructed with the reference information,

$$\begin{cases} \boldsymbol{x}_m = \boldsymbol{x}_m^* + \boldsymbol{b}_m^{k'} + \boldsymbol{\epsilon}_m \\ \widehat{\boldsymbol{x}}_m = \widehat{\boldsymbol{\zeta}}_m^* + \boldsymbol{b}_m^k \\ \boldsymbol{y}_n = \boldsymbol{y}_n^* + \boldsymbol{b}_n^{k'} + \boldsymbol{\epsilon}_n \\ \widehat{\boldsymbol{y}}_n = \widehat{\boldsymbol{\zeta}}_n^* + \boldsymbol{b}_n^k \end{cases} \tag{11}$$

where $\widehat{\boldsymbol{\zeta}}_m^*, \widehat{\boldsymbol{\zeta}}_n^* \overset{i.i.d.}{\sim} P_{\widehat{\boldsymbol{\zeta}}^*}$, $\boldsymbol{b}_m^k, \boldsymbol{b}_n^k \overset{i.i.d.}{\sim} P_{\boldsymbol{b}^k}$ and $\boldsymbol{b}_m^{k'}, \boldsymbol{b}_n^{k'} \overset{i.i.d.}{\sim} P_{\boldsymbol{b}^{k'}}$. Subsequently, the reconstruction errors satisfy:

$$\begin{cases} \boldsymbol{\delta}_m^x = \boldsymbol{x}_m - \widehat{\boldsymbol{x}}_m = \boldsymbol{x}_m^* - \widehat{\boldsymbol{\zeta}}_m^* + \boldsymbol{b}_m^{k'} - \boldsymbol{b}_m^k + \boldsymbol{\epsilon}_m \\ \boldsymbol{\delta}_n^y = \boldsymbol{y}_n - \widehat{\boldsymbol{y}}_n = \boldsymbol{y}_n^* - \widehat{\boldsymbol{\zeta}}_n^* + \boldsymbol{b}_n^{k'} - \boldsymbol{b}_n^k + \boldsymbol{\epsilon}_n \end{cases} \tag{12}$$

For a more concise representation, we define:

$$\begin{cases} \boldsymbol{\delta}_m^{x*} = \boldsymbol{x}_m^* - \widehat{\boldsymbol{\zeta}}_m^*, \quad \boldsymbol{\delta}_m^b = \boldsymbol{b}_m^{k'} - \boldsymbol{b}_m^k + \boldsymbol{\epsilon}_m \\ \boldsymbol{\delta}_m^{y*} = \boldsymbol{y}_n^* - \widehat{\boldsymbol{\zeta}}_n^*, \quad \boldsymbol{\delta}_n^b = \boldsymbol{b}_n^{k'} - \boldsymbol{b}_n^k + \boldsymbol{\epsilon}_n \end{cases} \tag{13}$$

Under these symbol representations, if MMD is induced by linear kernel, the kernel is expandable as follows:

$$\begin{aligned} k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^x) &= k(\boldsymbol{\delta}_i^{x*}, \boldsymbol{\delta}_j^{x*}) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^{x*}, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^{x*}) \\ k(\boldsymbol{\delta}_i^y, \boldsymbol{\delta}_j^y) &= k(\boldsymbol{\delta}_i^{y*}, \boldsymbol{\delta}_j^{y*}) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^{y*}, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^{y*}) \\ k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^y) &= k(\boldsymbol{\delta}_i^{x*}, \boldsymbol{\delta}_j^{y*}) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^{x*}, \boldsymbol{\delta}_j^b) + k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^{y*}) \end{aligned} \tag{14}$$

Subsequently, considering (1), therefore it follows that:

$$\begin{aligned} &MMD^2\left(\boldsymbol{\delta}_m^x, \boldsymbol{\delta}_n^y\right) \\ &= \frac{1}{m(m-1)} \sum_i^m \sum_{j \neq i}^m k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^x) + \frac{1}{n(n-1)} \sum_i^n \sum_{j \neq i}^n k(\boldsymbol{\delta}_i^y, \boldsymbol{\delta}_j^y) - \frac{2}{mn} \sum_i^m \sum_j^n k(\boldsymbol{\delta}_i^x, \boldsymbol{\delta}_j^y) \\ &= MMD^2\left(\boldsymbol{\delta}_m^{x*}, \boldsymbol{\delta}_n^{y*}\right) + MMD^2\left(\boldsymbol{\delta}_m^b, \boldsymbol{\delta}_n^b\right) + 2R_{mn}^x + 2R_{mn}^y \end{aligned} \tag{15}$$

where the remainder term R_{mn}^x, R_{mn}^y are defined as:

$$\begin{aligned} R_{mn}^x &\coloneqq \frac{1}{m(m-1)} \sum_i^m \sum_{j \neq i}^m {\boldsymbol{\delta}_i^b}^T \boldsymbol{\delta}_j^{x*} - \frac{1}{n^2} \sum_i^n \sum_j^n {\boldsymbol{\delta}_i^b}^T \boldsymbol{\delta}_j^{x*} \\ R_{mn}^y &\coloneqq \frac{1}{n(n-1)} \sum_i^m \sum_{j \neq i}^m {\boldsymbol{\delta}_i^b}^T \boldsymbol{\delta}_j^{y*} - \frac{1}{m^2} \sum_i^m \sum_j^m {\boldsymbol{\delta}_i^b}^T \boldsymbol{\delta}_j^{y*} \end{aligned} \tag{16}$$

Before analyzing the transfer error of $MMD^2\left(\boldsymbol{\delta}_m^x, \boldsymbol{\delta}_n^y\right)$, let's first discuss the convergence of each term in (15). For the first and second terms, we analyze them using the following Lemma 3.1 [1].

Lemma 3.1. *Assume $0 \leq k(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq K$. Then:*

$$\mathbb{P}\left(\left|MMD^2(\boldsymbol{X}, \boldsymbol{Y}) - MMD^2(p, q)\right| \geq \varepsilon\right) \leq 2 \exp\left(\frac{-\varepsilon^2 mn}{8K^2(m+n)}\right) \tag{17}$$

where $\boldsymbol{x}_i \sim p, \boldsymbol{y}_j \sim q$.

Then, we obtain:

$$\begin{gathered} \mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^{x*}, \boldsymbol{\delta}_n^{y*}) - MMD^2(P_{\boldsymbol{\delta}^{x*}}, P_{\boldsymbol{\delta}^{y*}})\right| \geq \varepsilon\right) \leq 2 \exp\left(\frac{-Cn\varepsilon^2}{8(1+C){K_+^x}^2}\right) \\ \mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^b, \boldsymbol{\delta}_n^b) - 0\right| \geq \varepsilon\right) \leq 2 \exp\left(\frac{-Cn\varepsilon^2}{8(1+C){K_+^b}^2}\right) \end{gathered} \tag{18}$$

where $K_+^x \coloneqq \sup_{ij} k(\boldsymbol{\delta}_i^{x*}, \boldsymbol{\delta}_j^{x*})$, $K_+^b \coloneqq \sup_{ij} k(\boldsymbol{\delta}_i^b, \boldsymbol{\delta}_j^b)$.

For the third term R_{mn}^x in (15), we employ the following Lemma 3.2 to discuss the convergence, and the proof is available at Section 4.

Lemma 3.2. *For any random variables $x_1, x_2, \cdots, x_k$, they always satisfy:*

$$\mathbb{P}\left(\left|\sum_{i=1}^k x_i\right| \geq \varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^k |x_i| \geq \varepsilon\right) \leq \sum_{i=1}^k \mathbb{P}\left(|x_i| \geq \frac{\varepsilon}{k}\right) \tag{19}$$

where $\varepsilon \geq 0$, $k \in \mathbb{Z}_+$

Considering that $\boldsymbol{\delta}^b$ and $\boldsymbol{\delta}^{x*}$ are mutually independent, we define the random variable $\xi \coloneqq {\boldsymbol{\delta}^b}^T \boldsymbol{\delta}^{x*} \in \mathbb{R}$. According to Lemma 3.2 and Hoeffding's Inequality [4], R_{mn}^x can be bounded as follows:

$$\begin{aligned}
\mathbb{P}(\left|R_{mn}^x\right| \geq \varepsilon) &= \mathbb{P}\left(\left|\frac{1}{m(m-1)}\sum_{i=1}^{m(m-1)} \xi_i - \frac{1}{n^2}\sum_{j=1}^{n^2}\xi_j\right| \geq \varepsilon\right) \\
&= \mathbb{P}\left(\left|\frac{1}{m(m-1)}\sum_{i=1}^{m(m-1)} \xi_i - \mathbb{E}(\xi) + \mathbb{E}(\xi) - \frac{1}{n^2}\sum_{j=1}^{n^2}\xi_j\right| \geq \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{1}{m(m-1)}\sum_{i=1}^{m(m-1)} \xi_i - \mathbb{E}(\xi)\right| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\left|\frac{1}{n^2}\sum_{j=1}^{n^2}\xi_j - \mathbb{E}(\xi)\right| \geq \frac{\varepsilon}{2}\right) \\
&\leq 2\exp\left(\frac{-2m(m-1)(\varepsilon/2)^2}{{K_+^\xi}^2}\right) + 2\exp\left(\frac{-2n^2(\varepsilon/2)^2}{{K_+^\xi}^2}\right) \\
&\leq 4\exp\left(\frac{-m(m-1)\varepsilon^2}{2{K_+^\xi}^2}\right)
\end{aligned} \tag{20}$$

where $K_+^\xi \coloneqq \sup_{i,j}(\xi_i - \xi_j)$

Similarly, for the forth term R_{mn}^y in (15), we define the random variable $\theta \coloneqq {\boldsymbol{\delta}^b}^T \boldsymbol{\delta}^{y*} \in \mathbb{R}$, and $K_+^\theta \coloneqq \sup_{i,j}(\theta_i - \theta_j)$. Thus, R_{mn}^y can be bounded as follows:

$$\mathbb{P}(\left|R_{mn}^y\right| \geq \varepsilon) \leq 4\exp\left(\frac{-m^2\varepsilon^2}{2{K_+^\theta}^2}\right) \tag{21}$$

Combined (18), (20) and (21), we also employ Lemma 3.2 to obtain:

$$\begin{aligned}
&\mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^x, \boldsymbol{\delta}_n^y) - MMD^2(P_{\boldsymbol{\delta}^{x*}}, P_{\boldsymbol{\delta}^{y*}})\right| \geq \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^{x*}, \boldsymbol{\delta}_n^{y*}) - MMD^2(P_{\boldsymbol{\delta}^{x*}}, P_{\boldsymbol{\delta}^{y*}})\right| \geq \frac{\varepsilon}{4}\right) \\
&\quad + \mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^b, \boldsymbol{\delta}_n^b)\right| \geq \frac{\varepsilon}{4}\right) + \mathbb{P}(\left|2R_{mn}^x\right| \geq \frac{\varepsilon}{4}) + \mathbb{P}(\left|2R_{mn}^y\right| \geq \frac{\varepsilon}{4}) \\
&\leq 4\exp\left(\frac{-Cn\varepsilon^2}{128(1+C)K_+^2}\right) + 4\exp\left(\frac{-m(m-1)\varepsilon^2}{128K_+^2}\right) + 4\exp\left(\frac{-m^2\varepsilon^2}{128K_+^2}\right)
\end{aligned} \tag{22}$$

where $K_+ \coloneqq \max\{K_+^x, K_+^b, K_+^\xi, K_+^\theta\}$. When $m > 1 + (1+C)^{-1}$, the following inequality always holds[1]:

$$\frac{Cn}{1+C} < m(m-1) < m^2 \tag{23}$$

[1] In fact, this condition is always satisfied as long as there is more than only two normal sample.

Finally, we can obtain:

$$\mathbb{P}\left(\left|MMD^2(\boldsymbol{\delta}_m^x, \boldsymbol{\delta}_n^y) - MMD^2(P_{\boldsymbol{\delta}^{x*}}, P_{\boldsymbol{\delta}^{y*}})\right| \geq \varepsilon\right) \leq 12 \exp\left(\frac{-Cn\varepsilon^2}{128(1+C)K_+^2}\right) \tag{24}$$

If $\alpha \coloneqq 12, \beta \coloneqq (128K_+^2)^{-1}$, the original theorem will be proven. □

4 Proof of Lemma 3.2

Proof. On one hand, according to the triangle inequality, we have:

$$\left|\sum_{i=1}^{k} x_i\right| \leq \sum_{i=1}^{k} |x_i| \tag{25}$$

which also indicates

$$\left\{\left|\sum_{i=1}^{k} x_i\right| \geq \varepsilon\right\} \subset \left\{\sum_{i=1}^{k} |x_i| \geq \varepsilon\right\} \tag{26}$$

On the other hand, the following relationship always holds:

$$\left\{\sum_{i=1}^{k} |x_i| \geq \varepsilon\right\} \subset \bigcup_{i=1}^{k} \left\{|x_i| \geq \frac{\varepsilon}{k}\right\} \tag{27}$$

Thus, we have:

$$\mathbb{P}\left(\left|\sum_{i=1}^{k} x_i\right| \geq \varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} |x_i| \geq \varepsilon\right) \leq \sum_{i=1}^{k} \mathbb{P}\left(|x_i| \geq \frac{\varepsilon}{k}\right) \tag{28}$$

□

References

[1] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[2] Zachary P Bradshaw and Christophe Vignat. An operational calculus generalization of ramanujan's master theorem. *Journal of Mathematical Analysis and Applications*, 523(2):127029, 2023.

[3] Tewodros Amdeberhan, Olivier Espinosa, Ivan Gonzalez, Marshall Harrison, Victor H Moll, and Armin Straub. Ramanujan's master theorem. *The Ramanujan Journal*, 29:103–120, 2012.

[4] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.