Appendices

A. Entry-wise Bound and Proof of Theorem 2

In this section, we will prove the Theorem 2 based on recent results on entry-wise analysis for random matrices (Abbe et al., 2017) and matrix completion with Poisson observation (McRae & Davenport, 2019). The proof idea can be viewed as a generalization from Gaussian noise in the Theorem 3.4 (Abbe et al., 2017) to subexponential noise. In particular, we will proceed the proof in two steps: (i) consider the symmetric scenario where M^*, noises, and anomalies have symmetries; (ii) generalize the results to the asymmetric scenario.

A.1. Symmetric Case

Consider a symmetric scenario. Let $\bar{M}^* \in \mathbb{R}_+^{n\times n}$ be a symmetric matrix. For $1 \le i \le j \le n$, let

$$\begin{cases} \bar{X}_{ij} \sim \text{Poisson}(\bar{M}^*_{ij}) & \text{with prob. } (1-p^*_{\text{A}})p_{\mathcal{O}} \\ \bar{X}_{ij} \sim \text{Anom}(\alpha^*, \bar{M}^*_{ij}) & \text{with prob. } p^*_{\text{A}}p_{\mathcal{O}} \\ \bar{X}_{ij} = 0 & \text{with prob. } 1-p_{\mathcal{O}}. \end{cases} \tag{4}$$

Let $\bar{X}_{ji} = \bar{X}_{ij}$ for $1 \le i \le j \le n$. Let $t = g(\alpha^*)p^*_{\text{A}}p_{\mathcal{O}} + (1-p^*_{\text{A}})p_{\mathcal{O}}$. It is easy to verify $\mathbb{E}\left(\bar{X}/t\right) = \bar{M}^*$. Furthermore, suppose $\max\left(\bar{M}^*_{ij}+1, \left\|\text{Anom}(\alpha^*, \bar{M}^*_{ij})\right\|_{\psi_1}\right) \le L$ for $(i,j) \in [n]\times[n]$.

Denote the eigenvalues of $\bar{M}^*$ by $\lambda^*_1 \ge \lambda^*_2 \ge \ldots \ge \lambda^*_n$ with their associated eigenvectors by $\{\bar{u}^*_j\}_{j=1}^n$. Denote the eigenvalues of $\bar{X}$ by $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$ with their associated eigenvectors by $\{\bar{u}_j\}_{j=1}^n$.

Suppose r is an integer such that $1 \le r < n$. Assume $\bar{M}^*$ satisfies $\lambda^*_1 \ge \lambda^*_2 \ge \ldots \ge \lambda^*_r \ge 0$ and $\lambda^*_{r+1} \le 0$. Let $\bar{U}^* = (u^*_1, u^*_2, \ldots, u^*_r) \in \mathbb{R}^{n\times r}, \bar{U} = (u_1, u_2, \ldots, u_r) \in \mathbb{R}^{n\times r}$. We aim to show that $\bar{U}$ is a good estimation of $\bar{U}^*$ in the entry-wise manner under some proper rotation. In particular, let $\bar{H} := \bar{U}^T\bar{U}^* \in \mathbb{R}^{r\times r}$. Suppose the SVD decomposition of $\bar{H}$ is $\bar{H} = U'\Sigma'V'^T$ where $U', V' \in \mathbb{R}^{r\times r}$ are orthonormal matrices and $\Sigma' \in \mathbb{R}^{r\times r}$ is a diagonal matrix. The matrix sign function of $\bar{H}$ is denoted by $\text{sgn}(\bar{H}) := U'V'^T$. In fact, $\text{sgn}(\bar{H}) = \arg\min_O \left\|\bar{U}O - \bar{U}^*\right\|_{\text{F}}$ subject to $OO^T = I$.[6] We aim to show an upper bound on $\left\|\bar{U}\text{sgn}(\bar{H}) - \bar{U}^*\right\|_{2,\infty}$.

Let $\Delta^* := t\lambda^*_r, \kappa := \frac{\lambda^*_1}{\lambda^*_r}$. We rephrase the Theorem 2.1 in (Abbe et al., 2017) for the above scenario and rewrite it as the following lemma.

Lemma 6 (Theorem 2.1 (Abbe et al., 2017))**.** *Suppose $\gamma \in \mathbb{R}_{\ge 0}$. Let $\phi(x) : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ be a continuous and non-decreasing function with $\phi(0) = 0$ and $\phi(x)/x$ non-increasing in $\mathbb{R}_{>0}$. Let $\delta_0, \delta_1 \in (0,1)$. With the above quantities, consider the following four assumptions:*

A1. $\left\|t\bar{M}^*\right\|_{2,\infty} \le \gamma\Delta^*$.

A2. *For any $m \in [n]$, the entries in the mth row and column of $\bar{X}$ are independent with others.*

A3. $\mathbb{P}\left(\left\|\bar{X} - t\bar{M}^*\right\|_2 \le \gamma\Delta^*\right) \ge 1-\delta_0$.

A4. *For any $m \in [n]$ and any $W \in \mathbb{R}^{n\times r}$,*

$$\mathbb{P}\left(\left\|\left(\bar{X} - t\bar{M}^*\right)_{m,\cdot}W\right\|_2 \le \Delta^*\left\|W\right\|_{2,\infty}\phi\left(\frac{\left\|W\right\|_{\text{F}}}{\sqrt{n}\left\|W\right\|_{2,\infty}}\right)\right) \ge 1 - \frac{\delta_1}{n}$$

If $32\kappa\max(\gamma, \phi(\gamma)) \le 1$, under above Assumptions A1–A4, with probability $1-\delta_0-2\delta_1$, the followings hold,

$$\left\|\bar{U}\right\|_{2,\infty} \lesssim (\kappa + \phi(1))\left\|\bar{U}^*\right\|_{2,\infty} + \gamma\left\|t\bar{M}^*\right\|_{2,\infty}/\Delta^*$$
$$\left\|\bar{U}\text{sgn}(\bar{H}) - \bar{U}^*\right\|_{2,\infty} \lesssim (\kappa(\kappa+\phi(1))(\gamma+\phi(\gamma)) + \phi(1))\left\|\bar{U}^*\right\|_{2,\infty} + \gamma\left\|t\bar{M}^*\right\|_{2,\infty}/\Delta^*.$$

[6] See (Gross, 2011) for more details about the matrix sign function.

To obtain useful results from Lemma 6, one need to find proper γ and $\phi(x)$ and show that the Assumptions A1–A4 hold. We define $\bar{\gamma}$ and $\bar{\phi}(x)$ as the proper form for γ and $\phi(x)$ respectively in the following.

Definition 2. Let $\bar{\gamma} := \frac{\sqrt{np_{\mathcal{O}}}}{\Delta^*}L, \bar{\phi}(x) := \frac{\sqrt{n}}{\Delta^*}L\log(2n^3r)x$.

Under $\bar{\gamma}$ and $\bar{\phi}(x)$, we will show that Assumption A3 holds based on Lemma 11, Assumption A4 holds based on Lemma 13. Note that Assumption A2 naturally holds since each element of $\bar{X}$ is independent of each other. Assumption A1 holds due to that $\left\|t\bar{M}^*\right\|_{2,\infty} = \max_i \sqrt{\sum_j t^2\bar{M}_{ij}^{*2}} \leq t\sqrt{n}L \leq \bar{\gamma}\Delta^*$.

To show that Assumption A3 holds, we introduce a result in (McRae & Davenport, 2019) that helps to control the operator norm of $\bar{X} - t\bar{M}^*$.

Lemma 7 (Lemma 4 in (McRae & Davenport, 2019))**.** *Let Y be a random $n_1 \times n_2$ matrix whose entries are independent and centered, and suppose that for some $v, t_0 > 0$, we have, for all $t_1 \geq t_0$, $\mathbb{P}\left(|Y_{ij}| \geq t_1\right) \leq 2e^{-t_1/v}$. Let $\epsilon \in (0, 1/2)$, and let $K = \max\{t_0, v\log(2mn/\epsilon)\}$. Then,*

$$\mathbb{P}\left(\|Y\|_2 \geq 2\sigma + \frac{\epsilon v}{\sqrt{n_1 n_2}} + t_1\right) \leq \max(n_1, n_2)\exp(-t_1^2/(C_0(2K)^2)) + \epsilon,$$

where C_0 is a constant and $\sigma = \max_i \sqrt{\sum_j \mathbb{E}\left(Y_{ij}^2\right)} + \max_j \sqrt{\sum_i \mathbb{E}\left(Y_{ij}^2\right)}$.

In order to use Lemma 7, we show that every entry of $\bar{X} - t\bar{M}^*$ is a sub-exponential random variable based on Lemmas 8 to 10.

Lemma 8. *Let $Y \sim \mathrm{Poisson}(\lambda)$. Then $\|Y\|_{\psi_1} \leq 4\lambda + 1$.*

Proof. Note that for any $t_1 > 0$,

$$\mathbb{E}\left(e^{|Y|/t_1}\right) = \mathbb{E}\left(e^{Y/t_1}\right) = \sum_{k=0}^{\infty} e^{k/t_1}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^{1/t_1}\lambda)^k}{k!} = e^{-\lambda}e^{e^{1/t_1}\lambda} = e^{\lambda(e^{1/t_1}-1)}.$$

Note that $1/(4\lambda+1) \leq 1$, hence $e^{1/(4\lambda+1)} - 1 = \frac{1}{4\lambda+1}e^s \leq \frac{1}{4\lambda+1}e$ where $s \in [0, 1/(4\lambda+1)]$ by Taylor expansion. Therefore

$$\mathbb{E}\left(e^{|Y|/(4\lambda+1)}\right) \leq e^{\frac{\lambda}{4\lambda+1}e} \leq e^{e/4} \approx 1.973 < 2.$$

By the definition of $\|\cdot\|_{\psi_1}$, we have $\|Y\|_{\psi_1} \leq 4\lambda + 1$. $\square$

Lemma 9. *Let $Y_1, Y_2 \ldots Y_q$ be q subexponential random variables with $\|Y_i\|_{\psi_1} \leq L_{\max}$. Let $c \in \{1, 2, \ldots, q\}$ be a random variable. Then $\|Y_c\|_{\psi_1} \leq L_{\max}$.*

Proof. This is because $\mathbb{E}\left(e^{|Y_c|/L_{\max}}\right) = \sum_{1\leq i\leq q}\mathbb{P}\left(c = i\right)\mathbb{E}\left(e^{|Y_i|/L_{\max}}\right) \leq \sum_{1\leq i\leq q} 2\mathbb{P}\left(c = i\right) = 2$. $\square$

Lemma 10. *For any $(i, j) \in [n] \times [n]$, $\left\|\bar{X}_{ij} - t\bar{M}_{ij}^*\right\|_{\psi_1} \leq 6L$.*

Proof. Note that $\left\|\mathrm{Poisson}(\bar{M}_{ij}^*)\right\|_{\psi_1} \leq 4L$ by Lemma 8 and $\left\|\mathrm{Anom}(\alpha^*, \bar{M}_{ij}^*)\right\| \leq L$ by the definition of L. We have $\left\|\bar{X}_{ij}\right\|_{\psi_1} \leq 4L$ by Eq. (4) and Lemma 9. Then, by the triangle inequality, $\left\|\bar{X}_{ij} - t\bar{M}_{ij}^*\right\|_{\psi_1} \leq \left\|\bar{X}_{ij}\right\|_{\psi_1} + \left\|\bar{M}_{ij}^*\right\|_{\psi_1} \leq 4L + 2L = 6L$. $\square$

Next we show that Assumption A3 holds.

Lemma 11. *Suppose $p_{\mathcal{O}} \geq \frac{\log^3 n}{n}$. $\mathbb{P}\left(\left\|\bar{X} - t\bar{M}^*\right\|_2 \leq C\bar{\gamma}\Delta^*\right) \geq 1 - \frac{1}{n^2}$ for some constant C.*

Proof. Denote $Y \in \mathbb{R}^{n\times n}$ by

$$Y_{ij} = \begin{cases} 2(\bar{X}_{ij} - t\bar{M}_{ij}^*) & i < j \\ (\bar{X}_{ij} - t\bar{M}_{ij}^*) & i = j \\ 0 & i > j \end{cases}.$$

Note that $\|Y_{ij}\|_{\psi_1} \le 2\left\|\bar{X}_{ij} - t\bar{M}^*_{ij}\right\|_{\psi_1} \le 12L$ by Lemma 10. By the property of subexponential random variable, for all $t' \ge 0$, $\mathbb{P}\left(|Y_{ij}| \ge t'\right) \le 2\exp(-t'/C_1 L)$ where C_1 is a constant. By the construction of Y, we also have

$$\mathbb{E}\left(Y_{ij}\right) = 0 \quad \text{and} \quad \mathbb{E}\left(Y_{ij}^2\right) \le 2E[\bar{X}_{ij}^2] \le C_2 p_{\mathcal{O}} L^2 \tag{5}$$

for some constant C_2.

Consider applying Lemma 7 to X with $n_1 = n_2 = n$. Let $\epsilon = \frac{1}{2n^2}$. Then $K = C_1 L \log(4n^2)$. Take $t' = \sqrt{C_0 3 \ln n} 2K + \ln 2$. Then $\max(n_1, n_2)\exp(-t'^2/(C_0(2K)^2)) + \epsilon = \frac{1}{n^2}$. Furthermore, by Eq. (5) and $np_{\mathcal{O}} \ge \log^3(n)$, one can verify that

$$2\sigma + \frac{\epsilon v}{\sqrt{n_1 n_2}} + t' \le C_3\sqrt{np_{\mathcal{O}}}L$$

for some constant C_3.

Therefore, $\|Y\|_2 \le C_3\sqrt{np_{\mathcal{O}}}L$ with probability $1 - \frac{1}{n^2}$. Note that $\bar{X} - t\bar{M}^* = (Y + Y^T)/2$. Hence, with probability $1 - \frac{1}{n^2}$, $\left\|\bar{X} - t\bar{M}^*\right\|_2 \le (\|Y\|_2 + \left\|Y^T\right\|_2)/2 \le C_3\sqrt{np_{\mathcal{O}}}L \le C_3\bar{\gamma}\Delta^*$. □

Next, we will show that Assumption A4 holds based on the matrix Bernstein's inequality to control the tail bound of sum of sub-exponential random variables.

Lemma 12 (Matrix Bernstein's inequality)**.** *Given n independent random $m_1 \times m_2$ matrices $X_1, X_2, \ldots, X_n$ with $E[X_i] = 0$. Let*

$$V \triangleq \max\left(\left\|\sum_{i=1}^n E[X_i X_i^T]\right\|, \left\|\sum_{i=1}^n E[X_i^T X_i]\right\|\right). \tag{6}$$

Suppose $\|\|X_i\|\|_{\psi_1} \le L$ for $i \in [n]$. Then,

$$\|X_1 + X_2 + \ldots + X_n\| \lesssim \sqrt{V\log(n(m_1+m_2))} + L\log(n(m_1+m_2))\log(n) \tag{7}$$

with probability $1 - O(n^{-c})$ for any constant c.

Proof. Let $Y_i = X_i \mathbb{1}\left\{\|X_i\| \le B\right\}$ be the truncated version of X_i. We have,

$$\begin{aligned}
\|\mathbb{E}\left(Y_i\right)\| &\le \left\|\int X_i \mathbb{1}\left\{\|X_i\| > B\right\} df(X_i)\right\| \\
&\overset{(i)}{\le} \int \|X_i\| \, \mathbb{1}\left\{\|X_i\| > B\right\} df(X_i) \\
&\le BP(\|X_i\| > B) + \int_B^\infty P(\|X_i\| > t)dt \\
&\overset{(ii)}{\le} Be^{-B/CL} + CLe^{-B/CL}
\end{aligned} \tag{8}$$

where (i) is due to the convexity of $\|\cdot\|$ and (ii) is due to the subexponential property of $\|X_i\|$ and C is a constant. Meanwhile, we have

$$\begin{aligned}
\left\|\sum_{i=1}^n \mathbb{E}\left((Y_i - \mathbb{E}\left(Y_i\right))(Y_i - \mathbb{E}\left(Y_i\right))^T\right)\right\| &= \left\|\sum_{i=1}^n \mathbb{E}\left(Y_i Y_i^T\right) - \mathbb{E}\left(Y_i\right)\mathbb{E}\left(Y_i\right)^T\right\| \\
&\overset{(i)}{\le} \left\|\sum_{i=1}^n \mathbb{E}\left(Y_i Y_i^T\right)\right\| \\
&= \left\|\sum_{i=1}^n \mathbb{E}\left(X_i X_i^T\right) - \mathbb{E}\left(X_i X_i^T \mathbb{1}\left\{\|X_i\| > B\right\}\right)\right\| \\
&\overset{(ii)}{\le} \left\|\sum_{i=1}^n \mathbb{E}\left(X_i X_i^T\right)\right\| \le V
\end{aligned}$$

where (i) is due to the positive-semidefinite property of $\mathbb{E}\left(Y_i\right)\mathbb{E}\left(Y_i\right)^T$ and $\mathbb{E}\left(Y_iY_i^T\right) - \mathbb{E}\left(Y_i\right)\mathbb{E}\left(Y_i\right)^T$, (ii) is due to the positive-semidefinite property of $\mathbb{E}\left(X_iX_i^T\mathbb{1}\left\{\|X_i\| > B\right\}\right)$ and $\mathbb{E}\left(Y_iY_i^T\right)$. Similarly, $\left\|\sum_{i=1}^n \mathbb{E}\left((Y_i - \mathbb{E}\left(Y_i\right))^T(Y_i - \mathbb{E}\left(Y_i\right))\right)\right\| \le V$.

Then, by Theorem 6.1.1 (Tropp et al., 2015), we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^N (Y_i - \mathbb{E}\left(Y_i\right))\right\| \ge t\right) \le 2\exp\left(-\frac{t^2/2}{V + 2Bt/3}\right).$$

Then, with probability $1 - O(n^{-c})$ for some constant c,

$$\left\|\sum_{i=1}^N (Y_i - \mathbb{E}\left(Y_i\right))\right\| \lesssim \sqrt{V\log(n(m_1+m_2))} + B\log(n(m_1+m_2)).$$

Take $B = L\log(n)C'$ for a proper constant C', by Eq. (8), we have

$$\begin{aligned}\left\|\sum_{i=1}^N Y_i\right\| &\lesssim \sqrt{V\log(n)} + L\log^2(n) + nL\log(n)O(n^{-C'/C}) \\ &\lesssim \sqrt{V\log(n(m_1+m_2))} + L\log(n(m_1+m_2))\log(n).\end{aligned}$$

By the union bound on the event $\|X_i\| \le B$ for all i, we can conclude that, with probability $1 - O(n^{-c'})$ for some constant c',

$$\left\|\sum_{i=1}^N X_i\right\| \lesssim \sqrt{V\log(n(m_1+m_2))} + L\log(n(m_1+m_2))\log(n).$$

□

Consider the Assumption A4.

Lemma 13. *For any $m \in [n]$ and any $W \in \mathbb{R}^{n\times r}$, the following holds*

$$\mathbb{P}\left(\left\|(\bar{X} - t\bar{M}^*)_{m,\cdot}W\right\|_2 \le C\Delta^*\left\|W\right\|_{2,\infty}\bar{\phi}\left(\frac{\|W\|_{\mathrm{F}}}{\sqrt{n}\left\|W\right\|_{2,\infty}}\right)\right) \ge 1 - O(n^{-3})$$

where C is a constant.

Proof. Let $Y_j = \bar{X}_{ij} - t\bar{M}^*_{ij}$ and $Z_j = Y_jW_{j,\cdot} \in \mathbb{R}^{1\times r}$. Note that

$$\left\|\left(\bar{X} - t\bar{M}^*\right)_{m,\cdot}W\right\|_2 = \left\|\sum_{j=1}^n Z_j\right\|_2.$$

We aim to invoke the Lemma 12 for $Z_1, Z_2, \ldots, Z_n$. Note that $\mathbb{E}\left(Z_j\right) = 0$ since $\mathbb{E}\left(Y_j\right) = 0$ and Z_j are independent since Y_j are independent. Also, for the subexponential norm, we have

$$\begin{aligned}\left\|\|Z_j\|_2\right\|_{\psi_1} &\le \left\||Y_j|\right\|_{\psi_1}\|W_{j,\cdot}\|_2 \\ &\le \left\||Y_j|\right\|_{\psi_1}\|W\|_{2,\infty} \\ &\lesssim L\|W\|_{2,\infty}\end{aligned}$$

where the last inequality is due to Lemma 10. Then, one can check

$$
\begin{aligned}
\left\| \sum_{j=1}^{n} \mathbb{E}\left(Z_j^T Z_j\right) \right\| &\leq \sum_{j=1}^{n} \left\| \mathbb{E}\left(Z_j^T Z_j\right) \right\| \\
&\leq \sum_{j=1}^{n} \mathbb{E}\left(Y_j^2\right) \left\| W_{j,\cdot} \right\|_2^2 \\
&\lesssim \sum_{j=1}^{n} L^2 p_{\mathcal{O}} \left\| W_{j,\cdot} \right\|_2^2 \\
&\lesssim L^2 p_{\mathcal{O}} \left\| W \right\|_{\mathrm{F}}^2 .
\end{aligned}
$$

Similarly, one can show that $\left\| \sum_{j=1}^{n} \mathbb{E}\left(Z_j^T Z_j\right) \right\| \lesssim L^2 p_{\mathcal{O}} \left\| W \right\|_{\mathrm{F}}^2$. Then, we can invoke Lemma 12 and obtain, with probability $1 - O(n^{-3})$,

$$
\left\| \left(\bar{X} - t\bar{M}^*\right)_{m,\cdot} W \right\|_2 \lesssim L\sqrt{p_{\mathcal{O}}} \left\| W \right\|_{\mathrm{F}} \sqrt{\log(n)} + L \left\| W \right\|_{2,\infty} \log^2(n).
$$

Since $\bar{\phi}(x) = \sqrt{\log(n)} L \frac{\sqrt{np_{\mathcal{O}}}}{\Delta^*} x + L\frac{\log^2 n}{\Delta^*}$, we have

$$
L\sqrt{p_{\mathcal{O}}} \left\| W \right\|_{\mathrm{F}} \sqrt{\log(n)} + L \left\| W \right\|_{2,\infty} \log^2(n) \lesssim \Delta^* \left\| W \right\|_{2,\infty} \bar{\phi}\left(\frac{\left\| W \right\|_{\mathrm{F}}}{\sqrt{n} \left\| W \right\|_{2,\infty}} \right).
$$

This finishes the proof. □

After showing that Assumptions A1–A4 hold, we can prove the following theorem.

Proposition 2. *Let $t := (g(\alpha^*)p_{\mathrm{A}}^* + (1 - p_{\mathrm{A}}^*))p_{\mathcal{O}}$. Suppose $p_{\mathcal{O}} \geq \frac{\log^3 n}{n}$ and $\sqrt{np_{\mathcal{O}} \log(n)} L\kappa^2 \leq Ct\lambda_1^*$ for some known constant C. Then, with probability $1 - O(n^{-2})$, the following holds*

$$
\begin{aligned}
\left\| \bar{U} \right\|_{2,\infty} &\lesssim \kappa \left\| \bar{U}^* \right\|_{2,\infty} + \frac{\sqrt{np_{\mathcal{O}}}\kappa L}{\lambda_1^* t} \left\| \bar{M}^* \right\|_{2,\infty} / \lambda_1^* \\
\left\| \bar{U}\mathrm{sgn}(H) - \bar{U}^* \right\|_{2,\infty} &\lesssim \frac{\sqrt{np_{\mathcal{O}} \log(n)}}{t\lambda_1^*} \kappa^3 L \left\| \bar{U}^* \right\|_{2,\infty} + \frac{\sqrt{np_{\mathcal{O}}}\kappa L}{\lambda_1^* t} \left\| \bar{M}^* \right\|_{2,\infty} / \lambda_1^*.
\end{aligned}
$$

Proof. Let $\gamma = (C_1 + 1)\bar{\gamma}, \phi(x) = C_2\bar{\phi}(x)$ where C_1, C_2 are constants defined in Lemmas 11 and 13 respectively. One can verify that $\gamma = (C_1 + 1)\frac{\sqrt{np_{\mathcal{O}}}\kappa L}{\lambda_1^* t}, \phi(x) = C_2 \frac{(\sqrt{np_{\mathcal{O}} \log(n)} x + \log^2 n)\kappa L}{\lambda_1^* t}$. In order to apply Lemma 6, we still need to show that $32\kappa \max(\gamma, \phi(\gamma)) \leq 1$. Because $p_{\mathcal{O}} \geq \frac{\log^3 n}{n}$ and $\sqrt{np_{\mathcal{O}} \log(n)} L\kappa^2 \leq Ct\lambda_1^*$, one can verify that $32\kappa \max(\gamma, \phi(\gamma)) \leq 1$ by choosing a sufficient small C. Based on Lemma 11, Lemma 13, we can apply Lemma 6 and obtain that, with probability $1 - O(n^{-2})$,

$$
\begin{aligned}
\left\| \bar{U} \right\|_{2,\infty} &\lesssim (\kappa + \phi(1)) \left\| \bar{U}^* \right\|_{2,\infty} + \gamma \left\| t\bar{M}^* \right\|_{2,\infty} / \Delta^* \\
\left\| \bar{U}\mathrm{sgn}(\bar{H}) - \bar{U}^* \right\|_{2,\infty} &\lesssim (\kappa(\kappa + \phi(1))(\gamma + \phi(\gamma)) + \phi(1)) \left\| \bar{U}^* \right\|_{2,\infty} + \gamma \left\| t\bar{M}^* \right\|_{2,\infty} / \Delta^*.
\end{aligned}
$$

Using the fact $\Delta^* = t\lambda_1^*/\kappa, \phi(1) \leq 1 \leq \kappa$, we have

$$
\begin{aligned}
\left\| \bar{U} \right\|_{2,\infty} &\lesssim \kappa \left\| \bar{U}^* \right\|_{2,\infty} + \gamma \left\| \bar{M}^* \right\|_{2,\infty} / \lambda_1^* \\
\left\| \bar{U}\mathrm{sgn}(\bar{H}) - \bar{U}^* \right\|_{2,\infty} &\lesssim (\kappa^2(\gamma + \phi(\gamma)) + \phi(1)) \left\| \bar{U}^* \right\|_{2,\infty} + \gamma \left\| \bar{M}^* \right\|_{2,\infty} / \lambda_1^*.
\end{aligned}
$$

Plug in the definition of γ and ϕ, we complete the proof. □

A.2. Asymmetric Case

Let X_Ω associated with $M^*, p_{\mathrm{A}}^*, \alpha^*, p_{\mathcal{O}}$ be the observation generated by the model described in the Section 2. Let $t = (p_{\mathrm{A}}^* g(\alpha^*) + (1 - p_{\mathrm{A}}^*))p_{\mathcal{O}}$. Let $M^* = U^*\Sigma^* V^{*T}, M = U\Sigma V^T$ be the singular decomposition of M^* and M, where $M = \arg\min_{\mathrm{rank}(M')\le r} \|X' - M'\|_{\mathrm{F}}$ and X' is obtained from X_Ω by setting unobserved entries to 0. We construct the following: $\bar{M}^* := \begin{pmatrix} 0_{n\times n} & M^* \\ M^{*T} & 0_{m\times m}\end{pmatrix}$. One can verify that the spectral decomposition of $\bar{M}^*$ is

$$\bar{M}^* = \frac{1}{\sqrt{2}}\begin{pmatrix} U^* & U^* \\ V^* & -V^*\end{pmatrix} \cdot \begin{pmatrix} \Sigma^* & \\ & -\Sigma^*\end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} U^* & U^* \\ V^* & -V^*\end{pmatrix}^T.$$

Note that the largest r singular values, $\sigma_1^* \ge \sigma_2^* \ge \ldots \ge \sigma_r^*$, of M^* are the same as the largest r eigenvalues of $\bar{M}^*$. The $(r+1)$-th eigenvalue of $\bar{M}^*$ is non-positive. Let $\bar{U}^* = \frac{1}{\sqrt{2}}\begin{pmatrix} U^* \\ V^*\end{pmatrix}$ be the eigenvectors associated with the largest r singular values of $\bar{M}^*$. Similarly, let $\bar{X} := \begin{pmatrix} 0_{n\times n} & X \\ X^T & 0_{m\times m}\end{pmatrix}$. Let $\bar{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} U \\ V\end{pmatrix}$ be the eigenvectors associated with the largest r singular values of $\bar{X}$.

We can apply Proposition 2 to the $\bar{M}^*$ and $\bar{X}$ constructed in this subsection. This gives us the following result.

Proposition 3. *Let $H = \frac{1}{2}(U^T U^* + V^T V^*), N = n + m, \mu = \max\left(N\|U^*\|_{2,\infty}^2, N\|V^*\|_{2,\infty}^2\right)/r, \kappa = \sigma_1^*/\sigma_r^*, t = (p_{\mathrm{A}}^* g(\alpha^*) + 1 - p_{\mathrm{A}}^*)p_{\mathcal{O}}$. Suppose $p_{\mathcal{O}} \ge \frac{\log^3 m}{m}$ and $\sqrt{mp_{\mathcal{O}}\log(m)}L\kappa^2 \le Ct\sigma_1^*$ for some known small constant C. Then, with probability $1 - O((nm)^{-1})$, the followings hold*

$$(\|U\|_{2,\infty} \vee \|V\|_{2,\infty}) \lesssim \kappa\sqrt{\frac{\mu r}{m}} \tag{9}$$

$$(\|U\mathrm{sgn}(H) - U^*\|_{2,\infty} \vee \|V\mathrm{sgn}(H) - V^*\|_{2,\infty}) \lesssim \frac{\sqrt{p_{\mathcal{O}}\log(m)}\kappa^3 L\sqrt{\mu r}}{t\sigma_1^*} \tag{10}$$

$$\|M' - tM^*\|_{\max} \lesssim \kappa^4 \mu r L\sqrt{\frac{p_{\mathcal{O}}\log(m)}{m}}. \tag{11}$$

Proof. Note that $\sqrt{2}\left\|\bar{U}^*\right\|_{2,\infty} = (\|U^*\|_{2,\infty} \vee \|V^*\|_{2,\infty}) \le \sqrt{\mu r/N}$. Furthermore, we have

$$\left\|\bar{M}^*\right\|_{2,\infty} = \|M^*\|_{2,\infty} \le \|U^*\|_{2,\infty}\|\Sigma^* V^*\|_2 \le \|U^*\|_{2,\infty}\sigma_1^* \le \sqrt{\mu r/N}\sigma_1^*.$$

Apply Proposition 2 on $\bar{M}^*$ and $\bar{X}$ along with the bound on $\left\|\bar{U}^*\right\|_{2,\infty}, \left\|\bar{M}^*\right\|_{2,\infty}$, we can obtain, with probability $1 - O(N^{-2})$,

$$\left\|\bar{U}\right\|_{2,\infty} \lesssim \kappa\sqrt{\frac{\mu r}{N}} \tag{12}$$

$$\left\|\bar{U}\mathrm{sgn}(H) - \bar{U}^*\right\|_{2,\infty} \lesssim \frac{\sqrt{Np_{\mathcal{O}}\log(N)}}{t\sigma_1^*}\kappa^3 L\sqrt{\mu r/N} = \frac{\sqrt{p_{\mathcal{O}}\log(N)}\kappa^3 L\sqrt{\mu r}}{t\sigma_1^*}. \tag{13}$$

This completes the proof of Eq. (9) and Eq. (10). Next we proceed to the proof of Eq. (11).

Let $\tilde{U} = U\mathrm{sgn}(H), \tilde{V} = V\mathrm{sgn}(H), \tilde{\Sigma} = \mathrm{sgn}(H)^T\Sigma\mathrm{sgn}(H)$. Note that $M'_{ij} = U_{i,\cdot}\Sigma V_{j,\cdot}^T = \tilde{U}_{i,\cdot}\tilde{\Sigma}\tilde{V}_{j,\cdot}^T$ and $M_{ij}^* = U_{i,\cdot}^*\Sigma^* V_{j,\cdot}^{*T}$. Then,

$$\begin{aligned}
|M'_{ij} - tM_{ij}^*| &= |\mathrm{tr}(\tilde{U}_{i,\cdot}\tilde{\Sigma}\tilde{V}_{j,\cdot}^T) - t\mathrm{tr}(U_{i,\cdot}^*\Sigma^* V_{j,\cdot}^{*T})| \\
&= |\mathrm{tr}(\tilde{\Sigma}\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot}) - t\mathrm{tr}(\Sigma^* V_{j,\cdot}^{*T}U_{i,\cdot}^*)| \\
&= |\mathrm{tr}((\tilde{\Sigma} - t\Sigma^*)(\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot})) + \mathrm{tr}(t\Sigma^*(\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot} - V_{j,\cdot}^{*T}U_{i,\cdot}^*))| \\
&\le \left\|\tilde{\Sigma} - t\Sigma^*\right\|_2\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot}\right\|_* + t\|\Sigma^*\|_2\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot} - V_{j,\cdot}^{*T}U_{i,\cdot}^*\right\|_*
\end{aligned} \tag{14}$$

where Eq. (14) is due to the triangle inequality and $|\mathrm{tr}(AB)| \leq \|A\|_2 \|B\|_*$ by the Von Neumann's trace inequality. We derive the bound on the term $\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot} - V_{j,\cdot}^{*T}U_{i,\cdot}^*\right\|_*$. Let $\hat{\gamma} = \kappa\sqrt{np_{\mathcal{O}}}L/(\sigma_1^* t)$. Note that

$$\begin{aligned}
\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot} - V_{j,\cdot}^{*T}U_{i,\cdot}^*\right\|_* &= \left\|(\tilde{V}_{j,\cdot}^T - V_{j,\cdot}^{*T})\tilde{U}_{i,\cdot} + V_{j,\cdot}^{*T}(\tilde{U}_{i,\cdot} - U_{i,\cdot}^*)\right\|_* \\
&\leq \left\|(\tilde{V}_{j,\cdot}^T - V_{j,\cdot}^{*T})\tilde{U}_{i,\cdot}\right\|_* + \left\|V_{j,\cdot}^{*T}(\tilde{U}_{i,\cdot} - U_{i,\cdot}^*)\right\|_* \\
&\leq \left\|\tilde{V}_{j,\cdot}^T - V_{j,\cdot}^{*T}\right\|_2 \left\|\tilde{U}_{i,\cdot}\right\|_2 + \|V_{j,\cdot}^{*T}\|_2 \left\|\tilde{U}_{i,\cdot} - U_{i,\cdot}^*\right\|_2 &(15)\\
&\lesssim \kappa^2\sqrt{\log(N)}\hat{\gamma}\sqrt{\mu r/N}\left(\left\|\tilde{U}_{i,\cdot}\right\|_2 + \|V_{j,\cdot}^*\|_2\right) &(16)\\
&\lesssim \kappa^3\sqrt{\log(N)}\hat{\gamma}\mu r/N &(17)
\end{aligned}$$

where Eq. (15) is due to $\left\|ab^T\right\|_* = \left\|ab^T\right\|_2 \leq \|a\|_2 \|b\|_2$ for any vector a, b, Eq. (16) is due to Eq. (13), and Eq. (17) is due to Eq. (12). We then bound $\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot}\right\|_*$,

$$\begin{aligned}
\left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot}\right\|_* &\leq \|V_{j,\cdot}^{*T}U_{i,\cdot}^*\|_* + \left\|\tilde{V}_{j,\cdot}^T\tilde{U}_{i,\cdot} - V_{j,\cdot}^{*T}U_{i,\cdot}^*\right\|_* \\
&\lesssim \|V_{j,\cdot}^{*T}U_{i,\cdot}^*\|_* + \kappa^3\sqrt{\log(N)}\hat{\gamma}\frac{\mu r}{N} &(18)\\
&\lesssim \|V^*\|_{2,\infty}\|U^*\|_{2,\infty} + \kappa^3\sqrt{\log(N)}\hat{\gamma}\frac{\mu r}{N} \\
&\lesssim \frac{\mu r}{N} + \kappa^3\sqrt{\log(N)}\hat{\gamma}\frac{\mu r}{N} \\
&\lesssim \kappa^2\frac{\mu r}{N} &(19)
\end{aligned}$$

where Eq. (18) is due to Eq. (17) and Eq. (19) is due to $\kappa\sqrt{\log(N)}\hat{\gamma} \lesssim 1$. Next we bound $\left\|\tilde{\Sigma} - t\Sigma^*\right\|_2$. Note that

$$\begin{aligned}
\left\|\tilde{\Sigma} - \Sigma\right\|_2 &= \left\|\mathrm{sgn}(H)^T(\Sigma\mathrm{sgn}(H) - \mathrm{sgn}(H)\Sigma)\right\|_2 \\
&\leq \|\Sigma\mathrm{sgn}(H) - \mathrm{sgn}(H)\Sigma\|_2 \\
&= \|(\Sigma H - H\Sigma) + \Sigma(\mathrm{sgn}(H) - H) + (H - \mathrm{sgn}(H))\Sigma\|_2 \\
&\leq \|\Sigma H - H\Sigma\|_2 + 2\|\Sigma\|_2 \|\mathrm{sgn}(H) - H\|_2 . &(20)
\end{aligned}$$

By Lemma 2 in (Abbe et al., 2017), we have

$$\|\mathrm{sgn}(H) - H\|_2 \lesssim (\left\|\bar{X} - t\bar{M}^*\right\|_2/(t\sigma_r^*))^2 \lesssim \hat{\gamma}^2 \tag{21}$$

$$\|\Sigma H - H\Sigma\|_2 \leq 2\left\|\bar{X} - t\bar{M}^*\right\|_2 \lesssim t\hat{\gamma}\sigma_r^* \tag{22}$$

where $\left\|\bar{X} - t\bar{M}^*\right\|_2 \leq \hat{\gamma}t\sigma_r^*$ by Lemma 11. By Weyl's inequality, we also have $\|\Sigma - t\Sigma^*\|_2 \leq \left\|\bar{X} - t\bar{M}^*\right\|_2 \lesssim t\hat{\gamma}\sigma_r^*$. Hence,

$$\|\Sigma\|_2 \leq \|t\Sigma^*\|_2 + \|\Sigma - t\Sigma^*\|_2 \lesssim t\sigma_1^* + t\hat{\gamma}\sigma_r^* \lesssim t\sigma_1^*. \tag{23}$$

Plugging Eqs. (21) to (23) into Eq. (20), we have $\left\|\tilde{\Sigma} - \Sigma\right\|_2 \lesssim t\hat{\gamma}\sigma_r^* + t\hat{\gamma}^2\sigma_1^* \lesssim t\hat{\gamma}\sigma_1^*$. Therefore,

$$\left\|\tilde{\Sigma} - t\Sigma^*\right\|_2 \leq \left\|\tilde{\Sigma} - \Sigma\right\|_2 + \|\Sigma - t\Sigma^*\|_2 \lesssim t\hat{\gamma}\sigma_1^*. \tag{24}$$

Plugging Eqs. (17), (19) and (24) into Eq. (14), we arrive at

$$\left\|M' - tM^*\right\|_{\max} \lesssim t\kappa^3\sqrt{\log(N)}\hat{\gamma}\frac{\mu r}{N}\sigma_1^* \lesssim t\kappa^3\sqrt{\log(N)}\frac{\mu r}{N}\sigma_1^*\frac{\kappa L\sqrt{Np_{\mathcal{O}}}}{t\sigma_1^*} \lesssim \kappa^4\mu r L\sqrt{\frac{p_{\mathcal{O}}\log(N)}{N}}.$$

□

Next, we provide a lemma for the concentration bound of the sum over Ω.

Lemma 14. *Let $\Omega = \{(i,j)|O_{ij} = 1\} \subset [n] \times [m]$ where $O_{ij} \sim \mathrm{Ber}(p_{\mathcal{O}})$ are i.i.d random variables. Let $\{T_{ij}|(i,j) \in [n] \times [m]\}$ be independent random variables with $\mathbb{E}(T_{ij}) = p_{ij}$. Let $S = \sum_{(i,j)\in[n]\times[m]} p_{\mathcal{O}} p_{ij}$. Then, with probability $1 - 1/(nm)$,*

$$\left| \sum_{(i,j)\in\Omega} T_{ij} - S \right| \le C\left(\sqrt{S\log(mn)} + \log(mn)\right)$$

where C is a constant. In particular, if $S \gtrsim \log(nm)$, then

$$\left| \sum_{(i,j)\in\Omega} T_{ij} - S \right| \le C_1 S$$

where C_1 is a constant.

Proof. Let $Z_{ij} = T_{ij} O_{ij}$. Then $Z_{ij} \in [0,1], \mathbb{E}(Z_{ij}) = p_{\mathcal{O}} p_{ij}$. By the Bernstein's inequality (Bernstein, 1946), we have

$$\mathbb{P}\left(\left| \sum_{ij} Z_{ij} - \sum_{ij} p_{\mathcal{O}} p_{ij} \right| > t \right) \le 2e^{-\frac{t^2/2}{\sum_{ij}\mathbb{E}((Z_{ij}-p_{\mathcal{O}}p_{ij})^2)+\frac{t}{3}}} \le 2e^{-\frac{t^2/2}{S+\frac{t}{3}}}$$

due to

$$\mathbb{E}\left((Z_{ij} - p_{\mathcal{O}} p_{ij})^2\right) \le \mathbb{E}\left(Z_{ij}^2\right) \le \mathbb{E}(Z_{ij}) = p_{\mathcal{O}} p_{ij}.$$

Take $t = C_2\left(\sqrt{S\log(nm)} + \log(nm)\right)$ where C_2 is a constant. Then we have

$$\mathbb{P}\left(\left| \sum_{(i,j)\in\Omega} T_{ij} - \sum_{(i,j)\in[n]\times[m]} S \right| > t \right) \le \frac{1}{nm}$$

for a proper C_2. □

Proof of Theorem 2. Next we proceed the proof of Theorem 2 based on Proposition 3. By the assumption in Section 2, $\log^{1.5}(m)\mu r L\kappa^2/(\|M^*\|_{\max}\sqrt{m}) \lesssim \sqrt{p_{\mathcal{O}}}$ and $1 - p_{\mathrm{A}}^* \gtrsim 1$. Note that $\|M^*\|_{\max} \lesssim \sigma_1^* \mu r/m$, this implies that $\sqrt{\log(m)}\sqrt{m}L\kappa^2 \lesssim \sqrt{p_{\mathcal{O}}}\sigma_1^*$ and $\sqrt{p_{\mathcal{O}}} \gtrsim \frac{\log^{1.5}(m)}{\sqrt{m}}$, which is the condition required by Proposition 3. Also, by taking $T_{ij} = 1$ in Lemma 14 and noting that $p_{\mathcal{O}} \gtrsim \frac{\log^3(m)}{n}$, we have, with probability $1 - O(\frac{1}{nm})$,

$$|nmp_{\mathcal{O}} - |\Omega|| < C\sqrt{\log(nm)p_{\mathcal{O}}nm}$$

where C is a constant. Then

$$\begin{aligned}
\left| \frac{nm}{|\Omega|} - \frac{1}{p_{\mathcal{O}}} \right| &= \frac{|nmp_{\mathcal{O}} - |\Omega||}{|\Omega|p_{\mathcal{O}}} \\
&\le \frac{C\sqrt{\log(nm)p_{\mathcal{O}}nm}}{|\Omega|p_{\mathcal{O}}} \\
&\le \frac{C'\sqrt{\log(nm)}}{\sqrt{p_{\mathcal{O}}nm}p_{\mathcal{O}}}
\end{aligned}$$

where C' is a constant. Finally, we can obtain

$$\begin{aligned}\left\|M' \frac{nm}{|\Omega|} - \frac{t}{p_{\mathcal{O}}} M^*\right\|_{\max} &= \left\|M' \frac{nm}{|\Omega|} - M' \frac{1}{p_{\mathcal{O}}} + M' \frac{1}{p_{\mathcal{O}}} - \frac{t}{p_{\mathcal{O}}} M^*\right\|_{\max} \\ &\lesssim \left\|\frac{1}{p_{\mathcal{O}}}(M' - tM^*)\right\|_{\max} + \left\|M'\right\|_{\max} \frac{\sqrt{\log(nm)/p_{\mathcal{O}} nm}}{p_{\mathcal{O}}} \\ &\lesssim \frac{\kappa^4 \mu r L}{p_{\mathcal{O}}} \sqrt{\frac{\log(m) p_{\mathcal{O}}}{m}} + L \sqrt{\frac{\log(nm)}{p_{\mathcal{O}} nm}} \\ &\lesssim \kappa^4 \mu r L \sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}. \end{aligned} \tag{25}$$

This completes the proof. □

B. Analysis of π^{EW} and Proof of Theorem 1

B.1. Moment Matching Estimator

B.1.1. Proof of Lemma 1

Recall that

$$g_t(\theta, M) = \frac{1}{nm} \sum_{(i,j) \in [n] \times [m]} \left(p_{\mathrm{A}} \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} \le t | \alpha, M_{ij}\right) + (1 - p_{\mathrm{A}}) \mathbb{P}_{\mathrm{Poisson}}\left(X_{ij} \le t | M_{ij}\right)\right).$$

Let $\delta' = \kappa^4 \mu r L \sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}$ and

$$h(\theta) = \sum_{t=0}^{T-1} \left(g_t(\theta, \hat{M}/e(\theta)) - \frac{|X_{ij} = t, (i,j) \in \Omega|}{|\Omega|}\right)^2.$$

We have the following result.

Lemma 15. *With probability* $1 - O(\frac{1}{nm})$*, for any* $\theta \in \Theta$ *and* $t = 0, 1, \ldots, T$*,*

$$|g_t(\theta, M^* e(\theta^*)/e(\theta)) - g_t(\theta, \hat{M}/e(\theta))| \lesssim (K + L)\delta'.$$

Proof. Note that $\mathbb{P}_{\mathrm{Anom}}\left(X_{ij} = t | \alpha, M\right)$ is K-lipschitz on M. One also can verify that $\mathbb{P}_{\mathrm{Poisson}}\left(X_{ij} = t | M\right)$ is L-Lipschitz on M. Hence

$$\left(p_{\mathrm{A}} \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} \le t | \alpha, M_{ij}\right) + (1 - p_{\mathrm{A}}) \mathbb{P}_{\mathrm{Poisson}}\left(X_{ij} \le t | M_{ij}\right)\right)$$

is $(K + L)$-Lipschitz on M_{ij}. Let C_1, C_2 be two constants. By Theorem 2, with probability $1 - O((nm)^{-1})$, $|\hat{M}_{ij}/e(\theta^*) - M^*_{ij}| \le C_1 \delta'$. This implies that

$$\left|\frac{M^*_{ij} e(\theta^*)}{e(\theta)} - \frac{\hat{M}_{ij}}{e(\theta)}\right| \le \frac{C_1 \delta'}{e(\theta)} \le C_2 \delta'$$

where we use that $e(\theta) \ge (1 - p_A) \ge c$ for some constant c. This implies that

$$\begin{aligned}\left|g_t\left(\theta, \frac{M^* e(\theta^*)}{e(\theta)}\right) - g_t\left(\theta, \frac{\hat{M}}{e(\theta)}\right)\right| &\le \frac{1}{nm} \sum_{ij} \left|\frac{M^*_{ij} e(\theta^*)}{e(\theta)} - \frac{\hat{M}_{ij}}{e(\theta)}\right| (K + L) \\ &\lesssim (K + L)\delta'.\end{aligned}$$

□

Lemma 16. *With probability* $1-O((nm)^{-1})$, $h(\theta^*) \lesssim (K+L)^2(\delta')^2$.

Proof. Set C_1, C_2, C_3, C_4, C_5 be proper constants.

Note that by Lemma 14, with probability $1-O((nm)^{-1})$,

$$||X_{ij} \le t, (i,j) \in \Omega| - p_{\mathcal{O}} nm g_t(\theta^*, M^*)| \le C_2\sqrt{p_{\mathcal{O}} nm g_t(\theta^*, M^*)\log(nm)} + C_2\log(nm)$$

Also, we can similarly obtain $||\Omega| - p_{\mathcal{O}} nm| \le C_3\sqrt{p_{\mathcal{O}} nm \log(nm)}$ by Lemma 14. Then, one can verify that

$$\begin{aligned}
&\left|\frac{|X_{ij} = t, (i,j) \in \Omega|}{|\Omega|} - g_t(\theta^*, M^*)\right| \\
&= \left|\frac{|X_{ij} = t, (i,j) \in \Omega| - |\Omega| g_t(\theta^*, M^*)}{|\Omega|}\right| \\
&\le \frac{1}{|\Omega|}\left(C_2\sqrt{p_{\mathcal{O}} nm g_t(\theta^*, M^*)\log(nm)} + C_3\sqrt{p_{\mathcal{O}} nm\log(nm)}g_t(\theta^*, M^*) + C_2\log(nm)\right) \\
&\le C_4\frac{\sqrt{p_{\mathcal{O}} nm}\log(nm)}{nmp_{\mathcal{O}}} \\
&\le C_4\frac{\log(nm)}{\sqrt{nmp_{\mathcal{O}}}}.
\end{aligned}$$

Then, taking $\theta = \theta^*$ in Lemma 15, we have

$$\begin{aligned}
h(\theta^*) &= \sum_{t=0}^{T}\left(g_t(\theta^*, \hat{M}/e(\theta^*)) - |X_{ij} = t, (i,j) \in \Omega|/|\Omega|\right)^2 \\
&\le \sum_{t=0}^{T}\left(|g_t(\theta^*, \hat{M}/e(\theta^*)) - g_t(\theta^*, M^*)| + C_5\log(nm)/\sqrt{nmp_{\mathcal{O}}}\right)^2 \\
&\lesssim (K+L)^2(\delta')^2 + \log(nm)\log(nm)/(nmp_{\mathcal{O}}) \\
&\lesssim (K+L)^2(\delta')^2
\end{aligned}$$

due to the fact that $\delta' \gtrsim \sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}$ and $p_{\mathcal{O}} \gtrsim \frac{\log^3(m)}{m}$. □

Proof of Lemma 1. By Lemma 16, with probability $1-O((nm)^{-1})$, $h(\hat{\theta}) \le h(\theta^*) \lesssim (K+L)^2(\delta')^2$. This implies, for each $t < T$, $|g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - \frac{|X_{ij}=t,(i,j)\in\Omega|}{|\Omega|}| \lesssim (K+L)\delta'$. Combining with Lemma 16, we have, for each $t < T$,

$$|g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - g_t(\theta^*, \hat{M}/e(\theta^*))| \lesssim (K+L)\delta'. \tag{26}$$

Note that

$$\begin{aligned}
&|g_t(\theta^*, M^*) - g_t(\hat{\theta}, M^* e(\theta^*)/e(\hat{\theta}))| \\
&\le |g_t(\theta^*, M^*) - g_t(\hat{\theta}, \hat{M}/e(\hat{\theta}))| + |g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - g_t(\hat{\theta}, M^* e(\theta^*)/e(\hat{\theta}))|.
\end{aligned}$$

By Lemma 15, $|g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - g_t(\hat{\theta}, M^* e(\theta^*)/e(\hat{\theta}))| \lesssim (K+L)\delta'$. Also we have

$$|g_t(\theta^*, M^*) - g_t(\hat{\theta}, \hat{M}/e(\hat{\theta}))| \le |g_t(\theta^*, M^*) - g_t(\theta^*, \hat{M}/e(\theta^*))| + |g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - g_t(\theta^*, \hat{M}/e(\theta^*))|.$$

By Lemma 15 agian, we have $|g_t(\theta^*, M^*) - g_t(\theta^*, \hat{M}/e(\theta^*))| \lesssim (K+L)\delta'$. By Eq. (26), we have $|g_t(\hat{\theta}, \hat{M}/e(\hat{\theta})) - g_t(\theta^*, \hat{M}/e(\theta^*))| \lesssim (K+L)\delta'$. In conclusion,

$$|g_t(\theta^*, M^*) - g_t(\hat{\theta}, \hat{M}/e(\hat{\theta}))| \lesssim (K+L)\delta'.$$

Therefore, $\left\|F(\hat{\theta}) - F(\theta^*)\right\| \lesssim (K+L)\delta'$ since T is a constant. □

B.1.2. Proof of Lemma 2

Lemma 17. *Suppose F satisfies the following condition:*

- $F : \Theta \subset \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ *is continuously differentiable and injective.*
- $B_{2C_2\delta}(\theta^*) \subset \Theta$ *where* $B_r(\theta^*) = \{\theta : \|\theta - \theta^*\| \le r\}$.
- $\|J_F(\theta) - J_F(\theta^*)\|_{\max} \le C_1 \|\theta - \theta^*\|$ *for* $\theta \in B_{2C_2\delta}(\theta^*)$.
- $\left\|J_F(\theta^*)^{-1}\right\|_2 \le C_2$.

Suppose $2\sqrt{d_1 d_2} C_1 (C_2)^2 \delta < 1/2$. *For any* $\theta \in \Theta$,

$$\|F(\theta) - F(\theta^*)\| \le \delta \implies \|\theta - \theta^*\| \le 2C_2\delta. \tag{27}$$

Proof. Suppose $\|F(\theta) - F(\theta^*)\| \le \delta$. We construct a sequence of θ_i such that $\lim_{i\to\infty} F(\theta_i) = F(\theta)$ while $\|\theta_i - \theta^*\|$ is well bounded for every i. Let $\theta_1 - \theta^* = J_F^{-1}(\theta^*)(F(\theta) - F(\theta^*))$. Note that

$$\|\theta_1 - \theta^*\| \le \left\|J_F^{-1}(\theta^*)\right\|_2 \|F(\theta) - F(\theta^*)\| \le C_2\delta.$$

Furthermore, by multivariate Taylor theorem,

$$F(\theta_1) = F(\theta^*) + A(\theta_1 - \theta^*)^T$$

where the i-th row $A_i = (\nabla F_i(x_i))^T$ such that $x_i = \theta^* + c(\theta_1 - \theta^*)$ for some $c \in [0, 1]$. Hence, $F(\theta_1) = F(\theta^*) + J_F(\theta^*)(\theta_1 - \theta^*)^T + (A - J_F(\theta^*))(\theta_1 - \theta^*)^T$. Note that $F(\theta^*) + J_F(\theta^*)(\theta_1 - \theta^*)^T = F(\theta)$ by the definition of θ_1. Therefore,

$$\begin{aligned}
&F(\theta_1) = F(\theta) + (A - J_F(\theta^*))(\theta_1 - \theta^*)^T \\
\implies &\|F(\theta_1) - F(\theta)\| \le \|A - J_F(\theta^*)\|_2 \|\theta_1 - \theta^*\| \\
\implies &\|F(\theta_1) - F(\theta)\| \le \|A - J_F(\theta^*)\|_{\max} \sqrt{d_1 d_2} \|\theta_1 - \theta^*\| \\
\implies &\|F(\theta_1) - F(\theta)\| \le C_1 \sqrt{d_1 d_2} \|\theta_1 - \theta^*\|^2 \\
\implies &\|F(\theta_1) - F(\theta)\| \le C_1 \sqrt{d_1 d_2} C_2^2 \delta^2.
\end{aligned}$$

We can use the similar idea to the successive construction. In particular, let $t = 2\sqrt{d_1 d_2}(C_1)(C_2)^2\delta < 1/2, a = 2C_1C_2\sqrt{d_1 d_2}, \theta_0 = \theta^*$. Suppose

$$\|\theta_k - \theta_{k-1}\| \le \frac{1}{a} t^k, \|\theta_k - \theta^*\| \le \frac{1}{a}(2t - t^k), \|F(\theta_k) - F(\theta)\| \le \frac{1}{aC_2} t^{k+1}.$$

It is easy to verify that the above conditions are satisfied for $k = 1$. Then let $\theta_{k+1} - \theta_k = J_F^{-1}(\theta^*)(F(\theta) - F(\theta_k))$ for $k > 1$.

Then, we have $\|\theta_{k+1} - \theta_k\| \le C_2 \frac{t^{k+1}}{aC_2} \le \frac{t^{k+1}}{a}$. Also, $\|\theta_{k+1} - \theta^*\| \le \|\theta_{k+1} - \theta_k\| + \|\theta_k - \theta^*\| \le \frac{2t - t^k + t^{k+1}}{a} \le \frac{2t - t^{k+1}}{a}$. Furthermore,

$$\begin{aligned}
&F(\theta_{k+1}) = F(\theta_k) + J_F(\theta^*)(\theta_{k+1} - \theta_k)^T + (A - J_F(\theta^*))(\theta_{k+1} - \theta_k)^T \\
\implies &F(\theta_{k+1}) = F(\theta) + (A - J_F(\theta^*))(\theta_{k+1} - \theta_k)^T \\
\implies &\|F(\theta_{k+1}) - F(\theta)\| \le \|A - J_F(\theta^*)\|_2 \|\theta_{k+1} - \theta_k\| \\
\implies &\|F(\theta_{k+1}) - F(\theta)\| \le \|A - J_F(\theta^*)\|_{\max} \sqrt{d_1 d_2} \|\theta_{k+1} - \theta_k\| \\
\implies &\|F(\theta_{k+1}) - F(\theta)\| \le C_1 \sqrt{d_1 d_2} (\|\theta_k - \theta_*\| + \|\theta_{k+1} - \theta_k\|) \|\theta_{k+1} - \theta_k\| \\
\implies &\|F(\theta_{k+1}) - F(\theta)\| \le C_1 \sqrt{d_1 d_2} \frac{2(t)}{a} \frac{t^{k+1}}{a} \\
\implies &\|F(\theta_{k+1}) - F(\theta)\| \le \frac{C_1 C_2 \sqrt{d_1 d_2} 2}{a} \frac{t^{k+2}}{aC_2} \le \frac{t^{k+2}}{aC_2}.
\end{aligned}$$

Note that $\|\theta_k - \theta^*\| \le \frac{2t}{a}$. Therefore $\theta_k \in \Theta$ is well-defined. Furthermore, we can conclude for any $\epsilon > 0$, there exists N, if $k_1, k_2 > N$, $\|\theta_{k_1} - \theta_{k_2}\| \le \epsilon$. Therefore, the sequence converges. Suppose $\lim_k \theta_k = \theta'$. Note that $\|\theta' - \theta^*\| \le \frac{2t}{a}$ due to $\|\theta_k - \theta^*\| \le \frac{2t}{a}$. Also note that $\|J_F(\theta)\|$ is bounded for $\|\theta - \theta^*\| \le \frac{2t}{a}$. This implies that $\lim_k F(\theta_k) = F(\theta')$. On the other hand, due to the convergence of $F(\theta_k)$, $\lim_k F(\theta_k) = F(\theta)$. By injectivity, $\theta = \theta'$ and $\|\theta - \theta^*\| \le 2t/a$. This completes the proof. $\square$

Proof of Lemma 2. We proceed the proof of Lemma 2. By the regularity conditions (RC) stated in Section 4.2 and Lemma 17, we have $\|F(\theta) - F(\theta^*)\| \lesssim (K+L)\delta' \implies \|\theta - \theta^*\| \lesssim (K+L)\delta'$. By Lemma 1, we complete the proof. $\square$

B.2. Confidence Interval Estimator and Proof of Lemma 3

Let $\delta' = (K+L)\kappa^4 \mu r L \sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}$. Let

$$\hat{x}_{ij} := [\hat{p}_{\mathrm{A}} \mathbb{P}_{\mathrm{Anom}}(X_{ij} | \hat{\alpha}, \hat{M}_{ij}/e(\hat{\theta}))]$$
$$\hat{y}_{ij} := [(1 - \hat{p}_{\mathrm{A}}) \mathbb{P}_{\mathrm{Poisson}}(X_{ij} | \hat{M}_{ij}/e(\hat{\theta}))].$$

Let $x_{ij} = p_{\mathrm{A}}^* \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \alpha^*, M_{ij}^*\right), y_{ij} = (1 - p_{\mathrm{A}}^*) \mathbb{P}_{\mathrm{Poisson}}\left(X_{ij} | M_{ij}^*\right)$. We have the following result.

Lemma 18. *With probability* $1 - O(1/(nm))$, $\max(|\hat{x}_{ij} - x_{ij}|, |\hat{y}_{ij} - y_{ij}|) \le C(L+K)^2 L\delta'$ *for any* $(i,j) \in \Omega$.

Proof. By Lemma 2, with probability $1 - O(1/(nm))$, we have $\left\|\hat{\theta} - \theta^*\right\| \lesssim \delta'$.

Note that $g(\theta)$ is K-Lipschitz in θ and $e(\theta) = p_{\mathrm{A}} g(\theta) + (1 - p_{\mathrm{A}})$. Hence

$$\begin{aligned}
|e(\hat{\theta}) - e(\theta^*)| &\le |\hat{p}_{\mathrm{A}} - p_{\mathrm{A}}^*|(1 - g(\hat{\theta})) + p_{\mathrm{A}}^* |g(\hat{\theta}) - g(\theta^*)| \\
&\lesssim (K+1)\delta'.
\end{aligned}$$

Furthermore

$$\begin{aligned}
\left|\frac{\hat{M}_{ij}}{e(\hat{\theta})} - M^*\right| &= \frac{1}{e(\hat{\theta})} |\hat{M} - M^* e(\hat{\theta})| \\
&\le \frac{1}{e(\hat{\theta})} \left(|\hat{M} - M^* e(\theta^*)| + M^* |e(\theta^*) - e(\hat{\theta})|\right) \\
&\lesssim \frac{\delta'}{K+L} + L(K+1)\delta' \\
&\lesssim L(K+1)\delta'.
\end{aligned}$$

Note that $\mathbb{P}_{\mathrm{Anom}}(\alpha, M)$ is K-Lipschitz in α and M. The implies that

$$\begin{aligned}
|\hat{x}_{ij} - x_{ij}| &\le |\hat{p}_{\mathrm{A}} \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \hat{\alpha}, \hat{M}_{ij}\right) - p_{\mathrm{A}}^* \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \alpha^*, M_{ij}^*\right)| \\
&\le |\hat{p}_{\mathrm{A}} - p_{\mathrm{A}}^*| \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \alpha^*, M_{ij}^*\right) + |\mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \alpha^*, M_{ij}^*\right) - \mathbb{P}_{\mathrm{Anom}}\left(X_{ij} | \hat{\alpha}, \hat{M}_{ij}\right)| \hat{p}_{\mathrm{A}} \\
&\lesssim \delta' + KL(K+1)\delta' \\
&\lesssim KL(K+1)\delta'.
\end{aligned}$$

Similarly, one can obtain $|\hat{y}_{ij} - y_{ij}| \lesssim L^2(K+1)\delta'$. In conclusion,

$$\max(|\hat{x}_{ij} - x_{ij}|, |\hat{y}_{ij} - y_{ij}|) \lesssim (L+K)^2 L.$$

$\square$

Lemma 19. *Suppose* $|\hat{x}-x| \le \delta, |\hat{y}-y| \le \delta$ *where* $x, y, \hat{x}, \hat{y} \in [0,1], x+y > 0$. *Let* $\hat{s} = \frac{\hat{x}}{\hat{x}+\hat{y}}$ *if* $\hat{x}+\hat{y} > 0$ *otherwise* $\hat{s} = 0$. *Then,*

$$\left|\hat{s} - \frac{x}{x+y}\right| \le \min\left(\frac{\delta}{x+y}, \frac{\delta}{\hat{x}+\hat{y}}, 1\right). \tag{28}$$

Proof. Note that $\left|\hat{s} - \frac{x}{x+y}\right| \le 1$ is trivial since $\hat{s} \in [0,1]$ and $\frac{x}{x+y} \in [0,1]$.

When $\hat{x} = \hat{y} = 0$, $\frac{x}{x+y} \le \frac{\delta}{x+y}$ due to $x \le \delta$.

When $\hat{x}+\hat{y} > 0$,

$$\begin{aligned}
\left|\hat{s} - \frac{x}{x+y}\right| &= \left|\frac{\hat{x}(x+y) - x(\hat{x}+\hat{y})}{(\hat{x}+\hat{y})(x+y)}\right| \\
&= \left|\frac{\hat{x}y - x\hat{y}}{(\hat{x}+\hat{y})(x+y)}\right| \\
&= \left|\frac{\hat{x}(\hat{y}-(\hat{y}-y)) - (\hat{x}-(\hat{x}-x))\hat{y}}{(\hat{x}+\hat{y})(x+y)}\right| \\
&= \left|\frac{-\hat{x}(\hat{y}-y) + (\hat{x}-x)\hat{y}}{(\hat{x}+\hat{y})(x+y)}\right| \\
&\le \frac{\hat{x}}{\hat{x}+\hat{y}}\frac{|\hat{y}-y|}{x+y} + \frac{\hat{y}}{\hat{x}+\hat{y}}\frac{|\hat{x}-x|}{x+y} \\
&\le \frac{\hat{x}}{\hat{x}+\hat{y}}\frac{\delta}{x+y} + \frac{\hat{y}}{\hat{x}+\hat{y}}\frac{\delta}{x+y} \\
&= \frac{\delta}{x+y}.
\end{aligned}$$

By symmetry, $\left|\hat{s} - \frac{x}{x+y}\right| \le \frac{\delta}{\hat{x}+\hat{y}}$, which completes the proof. $\square$

Proof of Lemma 3. Next, we proceed the proof of Lemma 3. For notation simplification, write $\hat{x}_{ij}, \hat{y}_{ij}, x_{ij}, y_{ij}$ as $\hat{x}, \hat{y}, x, y$.

Let $\delta = (K+L)^3\kappa^4\mu r L^2\sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}$. and C be the constant denoted in Lemma 18. Then by Lemma 18, with probability $1 - O(1/(nm))$,

$$|x - \hat{x}| \le C\delta, |y - \hat{y}| \le C\delta.$$

By Lemma 19, we have

$$\frac{\hat{x} - C\delta}{\hat{x}+\hat{y}} \le \frac{x}{x+y}.$$

Therefore $f_{ij}^{\mathrm{L}} \le f_{ij}^*$. Next we show that $f_{ij}^* \le f_{ij}^{\mathrm{L}} + 4C\delta$.

If $4C\delta \ge (x+y)$, then $f_{ij}^{\mathrm{L}} + \frac{4C\delta}{x+y} \ge 1 \ge \frac{x}{x+y}$. On the other hand, suppose $4C\delta < (x+y)$. Note that

$$\begin{aligned}
&(x+y) > 4C\delta \\
\implies &4(x+y-2C\delta) > 2(x+y) \\
\implies &\frac{4C\delta}{x+y} > \frac{2C\delta}{x+y-2C\delta} \\
\implies &\frac{4C\delta}{x+y} > \frac{2C\delta}{\hat{x}+\hat{y}}.
\end{aligned}$$

Then,

$$\frac{\hat{x}-C\delta}{\hat{x}+\hat{y}}+\frac{4C\delta}{x+y}\geq\left(\frac{x}{x+y}-\frac{2C\delta}{\hat{x}+\hat{y}}\right)+\frac{4C\delta}{x+y}\geq\frac{x}{x+y}.$$

This implies that $f_{ij}^{\mathrm{L}}+\frac{4C\delta}{x+y}\geq f_{ij}^{*}$. Similar result can be shown for f_{ij}^{R}. This completes the proof. □

B.3. Analysis of the optimization problem $\mathcal{P}^{\mathrm{EW}}$

Note that $\mathcal{P}^{\mathrm{EW}}$ is obtained from $\mathcal{P}^{*}$ by replacing f_{ij}^{*} with the confidence interval estimators f_{ij}^{L} and f_{ij}^{R}. Intuitively, we could expect that $\mathcal{P}^{\mathrm{EW}}\approx\mathcal{P}^{*}$, and therefore the algorithm π^{EW} should achieve the desired performance. We first have the following lemma to show that $\mathrm{FPR}_{\pi^{\mathrm{EW}}(\gamma)}(X)\leq\gamma$ since $f_{ij}^{\mathrm{L}}\leq f_{ij}^{*}\leq f_{ij}^{\mathrm{R}}$ and so $\{t_{ij}^{\mathrm{EW}}\}$ is a feasible solution of $\mathcal{P}^{*}$.

Lemma 20. *With probability $1-O(1/(nm))$, for any $0<\gamma\leq 1$*

$$\mathrm{FPR}_{\pi^{\mathrm{EW}}(\gamma)}(X_{\Omega})\leq\gamma.$$

Proof. This is because

$$\sum_{(i,j)\in\Omega}t_{ij}^{\mathrm{EW}}f_{ij}^{*}\leq\sum_{(i,j)\in\Omega}t_{ij}^{\mathrm{EW}}f_{ij}^{\mathrm{R}}\leq\gamma\sum_{(i,j)\in\Omega}f_{ij}^{\mathrm{L}}\leq\gamma\sum_{(i,j)\in\Omega}f_{ij}^{*}.$$

due to that $f_{ij}^{\mathrm{L}}\leq f_{ij}^{*}\leq f_{ij}^{\mathrm{R}}$ and the constraint of t_{ij}^{EW}. □

To show the desired performance guarantee for $\mathrm{TPR}_{\pi^{\mathrm{EW}}}(X)$, we provide the following Lemma that characterizes how f_{ij}^{L} and f_{ij}^{R} are close to f_{ij}^{*} in an accumulated manner (the proof is shown momentarily):

Lemma 21. *Let $\delta=(K+L)^{3}\kappa^{4}\mu rL^{2}\sqrt{\frac{\log(m)}{p_{\mathcal{O}}m}}$. With probability $1-O(\frac{1}{nm})$,*

$$\sum_{(i,j)\in\Omega}\left(|f_{ij}^{\mathrm{L}}-f_{ij}^{*}|+|f_{ij}^{\mathrm{R}}-f_{ij}^{*}|\right)\leq CL\log(m)\delta p_{\mathcal{O}}nm.$$

Next we proceed to the analysis of $\mathcal{P}^{\mathrm{EW}}$. For a fixed η, let $\{t'_{ij}\}$ be the optimal solution of $\pi^{*}(\gamma')$ for some γ' such that $\frac{\sum_{(i,j)\in\Omega}t'_{ij}}{\sum_{(i,j)\in\Omega}t^{*}_{ij}}=\eta<1$. The key idea is to find some η such that $\{t'_{ij}\}$ is a feasible solution of $\mathcal{P}^{\mathrm{EW}}$, while maintaining good TPR performance compared to $\pi^{*}(\gamma)$. Indeed, a sufficiently large η can be achieved by Lemma 21. In particular, we have (the proof is shown momentarily):

Lemma 22. *Let $\delta=(K+L)^{3}\kappa^{4}\mu rL^{2}\sqrt{\frac{\log(m)}{p_{\mathcal{O}}m}},\eta=1-CL\delta\log(m)/\gamma$. Then $\{t'_{ij}\}$ is a feasible solution of $\mathcal{P}^{\mathrm{EW}}$. Furthermore,* $\min\left(1,\frac{\sum_{(i,j)\in\Omega}t^{*}_{ij}-\sum_{(i,j)\in\Omega}t'_{ij}}{\sum_{(i,j)\in\Omega}(1-f^{*}_{ij})}\right)\leq C_{1}\frac{L\delta\log(m)}{\gamma p_{\mathrm{A}}^{*}}$ *for a constant C_1.*

B.3.1. Proof of Lemma 21

Next, we prove Lemma 21, i.e., show that the accumulated error induced by the approximation of f_{ij}^{*} by f_{ij}^{L} and f_{ij}^{R} has the desired bound.

Proof of Lemma 21. Let $x_{ij}:=p_{\mathrm{A}}^{*}\mathbb{P}_{\mathrm{Anom}}\left(X_{ij}|\alpha^{*},M_{ij}^{*}\right),y_{ij}:=(1-p_{\mathrm{A}}^{*})\mathbb{P}_{\mathrm{Poisson}}\left(X_{ij}|M_{ij}^{*}\right)$. By Lemma 3,

$$\max(|f_{ij}^{\mathrm{L}}-f_{ij}^{*}|,|f_{ij}^{\mathrm{R}}-f_{ij}^{*}|)\leq\epsilon_{ij}$$

where $\epsilon_{ij}:=\min\left(\frac{4C\delta}{x_{ij}+y_{ij}},1\right)$ for some constant C and $\delta=(K+L)^{3}\kappa^{4}\mu rL^{2}\sqrt{\frac{\log(m)}{p_{\mathcal{O}}m}}$.

Note that when $X_{ij}=t$,

$$x_{ij}+y_{ij}=\mathbb{P}\left(X_{ij}=t\right).$$

Note that $\|X_{ij}\|_{\psi_1} \lesssim L$ is a sub-exponential random variable by Lemmas 8 and 9. Then, we have

$$\begin{aligned} &\mathbb{P}\left(X_{ij} > t\right) \leq \exp^{-t/C'L} \\ \Longrightarrow &\mathbb{P}\left(X'_{ij} > C'L\log(1/\delta)\right) \leq \delta \end{aligned}$$

where C' is a proper constant. Let $z_{ij} = \min\left(\frac{\delta}{\mathbb{P}(X_{ij}=t)}, 1\right)$. Then,

$$\begin{aligned} \mathbb{E}\left(z_{ij}\right) &= \sum_{t=0}^{\infty} \min\left(1, \delta/\mathbb{P}\left(X_{ij}=t\right)\right)\mathbb{P}\left(X_{ij}=t\right) \\ &\leq \sum_{t=0}^{C'L\log(1/\delta)} \delta + \sum_{t=C'L\log(1/\delta)+1}^{\infty} \mathbb{P}\left(X_{ij}=t\right) \\ &\leq C'L\log(1/\delta)\delta + \delta. \end{aligned}$$

Note that $z_{ij} \in [0,1]$ are independent random variables. Then, by Lemma 14, with probability $1 - O(\frac{1}{nm})$,

$$\begin{aligned} \sum_{(i,j)\in\Omega} z_{ij} &\lesssim L\log(1/\delta)\delta p_{\mathcal{O}} nm + \sqrt{p_{\mathcal{O}} nm \log(nm)} \\ &\lesssim L\log(m)\delta p_{\mathcal{O}} nm \end{aligned}$$

given that $\delta \gtrsim \sqrt{\frac{\log(m)}{p_{\mathcal{O}} m}}$.

Therefore,

$$\sum_{(i,j)\in\Omega} \max(|f^{\mathrm{L}}_{ij} - f^*_{ij}|, |f^{\mathrm{R}}_{ij} - f^*_{ij}|) \leq \sum_{(i,j)\in\Omega} \epsilon_{ij} \lesssim \sum_{(i,j)\in\Omega} z_{ij} \lesssim L\log(m)\delta p_{\mathcal{O}} nm.$$

$\square$

B.3.2. Proof of Lemma 22

Consider a concentration bound

Lemma 23. *Let C_1, C_2, C_3 be constants. With probability $1 - O(\frac{1}{nm})$,*

$$\begin{aligned} \sum_{(i,j)\in\Omega} f^*_{ij} &\geq C_1 nmp_{\mathcal{O}} \\ |\Omega| &\leq C_2 nmp_{\mathcal{O}}. \end{aligned}$$

Furthermore, if $p^*_{\mathrm{A}} p_{\mathcal{O}} nm \gtrsim \log(nm)$,

$$\sum_{(i,j)\in\Omega} 1 - f^*_{ij} \geq C_3 p^*_{\mathrm{A}} p_{\mathcal{O}} nm.$$

Proof. Let $Z_{ij} = \mathbb{P}\left(B_{ij} = 1 | X_{ij}\right)$. Then $\sum_{(i,j)\in\Omega} 1 - f^*_{ij} = \sum_{(i,j)\in\Omega} Z_{ij}$. Note that $\mathbb{E}\left(Z_{ij}\right) = p^*_{\mathrm{A}}$ and $Z_{ij} \in [0,1]$ are independent. Hence, by Lemma 14, with probability $1 - O(\frac{1}{nm})$, $\sum_{(i,j)\in\Omega} 1 - f^*_{ij} \geq Cp^*_{\mathrm{A}} p_{\mathcal{O}} nm$ where C is a constant given that $p^*_{\mathrm{A}} p_{\mathcal{O}} nm \gtrsim \log(nm)$ Similar results for $\sum_{(i,j)\in\Omega} f^*_{ij}$ (with $1 - p^*_{\mathrm{A}} \geq c$ for some constant c) and $|\Omega|$ can also be obtained. $\square$

Proof of Lemma 22. Let $\{t'_{ij}, (i,j) \in \Omega\}$ be the optimal solution of the algorithm $\pi^*(\gamma')$. Let $\{t^*_{ij}, (i,j) \in \Omega\}$ be the optimal solution of $\pi^*(\gamma)$. Suppose

$$\frac{\sum_{(i,j)\in\Omega} t'_{ij}}{\sum_{(i,j)\in\Omega} t^*_{ij}} = \eta < 1.$$

Order f^*_{ij} by $f^*_{a_1b_1} \le f^*_{a_2b_2} \le \ldots \le f^*_{a_{|\Omega|}b_{|\Omega|}}$. One can easily verify that $t'_{a_1b_1} \le t^*_{a_1b_1}, t'_{a_2b_2} \le t^*_{a_2b_2}, \ldots, t'_{a_{|\Omega|}b_{|\Omega|}} \le t^*_{a_{|\Omega|}b_{|\Omega|}}$. Furthermore, for any k and l such that $t^{'}_{a_kb_k} > 0$ and $t^*_{a_lb_l} - t'_{a_lb_l} > 0$, we have $f^*_{a_kb_k} \le f^*_{a_lb_l}$. Let $A = \sum_{ij} t'_{ij}, B = \sum_{ij} t^*_{ij} - t'_{ij}, C = \sum_{ij} t'_{ij} f^*_{ij}, D = \sum_{ij}(t^*_{ij} - t'_{ij}) f^*_{ij}$. Then the following weighted average inequality holds: $\frac{C}{A} \le \frac{D}{B}$. This implies that $\frac{C}{A} \le \frac{C+D}{A+B}$, i.e.,

$$\frac{1}{\sum_{(i,j)\in\Omega} t'_{ij}} \sum_{(i,j)\in\Omega} t'_{ij} f^*_{ij} \le \frac{1}{\sum_{(i,j)\in\Omega} t^*_{ij}} \sum_{(i,j)\in\Omega} t^*_{ij} f^*_{ij}. \tag{29}$$

This implies that $\sum_{(i,j)\in\Omega} t'_{ij} f^*_{ij} \le \eta \sum_{(i,j)\in\Omega} t^*_{ij} f^*_{ij}$. Then, we have,

$$\begin{aligned}
\sum_{(i,j)\in\Omega} t'_{ij} f^{\mathrm{R}}_{ij} &\le \sum_{(i,j)\in\Omega} t'_{ij}(f^*_{ij} + |f^{\mathrm{R}}_{ij} - f^*_{ij}|) \\
&\le \left(\eta \sum_{(i,j)\in\Omega} t^*_{ij} f^*_{ij}\right) + \sum_{(i,j)\in\Omega} |f^{\mathrm{R}}_{ij} - f^*_{ij}| && \text{by Eq. (29) and } 0 \le t'_{ij} \le 1 \\
&\le \left(\gamma\eta \sum_{(i,j)\in\Omega} f^*_{ij}\right) + \sum_{(i,j)\in\Omega} |f^{\mathrm{R}}_{ij} - f^*_{ij}| && \sum_{(i,j)\in\Omega} t^*_{ij} f^*_{ij} \le \gamma \sum_{(i,j)\in\Omega} f^*_{ij} \\
&\le \gamma \sum_{(i,j)\in\Omega} f^*_{ij} + \sum_{(i,j)\in\Omega} |f^{\mathrm{R}}_{ij} - f^*_{ij}| - \gamma(1-\eta) \sum_{(i,j)\in\Omega} f^*_{ij}.
\end{aligned}$$

Note that

$$\gamma \sum_{(i,j)\in\Omega} f^*_{ij} \le \gamma \sum_{(i,j)\in\Omega} f^{\mathrm{L}}_{ij} + \sum_{(i,j)\in\Omega} |f^*_{ij} - f^{\mathrm{L}}_{ij}|.$$

Therefore, we have

$$\sum_{(i,j)\in\Omega} t'_{ij} f^{\mathrm{R}}_{ij} \le \gamma \sum_{(i,j)\in\Omega} f^{\mathrm{L}}_{ij} + \left(\sum_{(i,j)\in\Omega} \left(|f^*_{ij} - f^{\mathrm{L}}_{ij}| + |f^*_{ij} - f^{\mathrm{R}}_{ij}|\right)\right) - \gamma(1-\eta) \sum_{(i,j)\in\Omega} f^*_{ij}.$$

By Lemma 21, we have $\left(\sum_{(i,j)\in\Omega} \left(|f^*_{ij} - f^{\mathrm{L}}_{ij}| + |f^*_{ij} - f^{\mathrm{R}}_{ij}|\right)\right) \le C_1 L \log(m) \delta p_{\mathcal{O}} nm$. By Lemma 23, we have $\gamma(1 - \eta) \sum_{(i,j)\in\Omega} f^*_{ij} \ge C_2 \gamma (1-\eta) p_{\mathcal{O}} nm$. Take $\eta = 1 - \frac{C_1}{C_2\gamma} L \log(m)\delta$. We then have $\{t'_{ij}\}$ is a feasible solution of $\mathcal{P}^{\mathrm{EW}}$:

$$\sum_{(i,j)\in\Omega} t'_{ij} f^{\mathrm{R}}_{ij} \le \gamma \sum_{(i,j)\in\Omega} f^{\mathrm{L}}_{ij}.$$

Furthermore, for any $0 < \gamma \le 1$, we can get

$$\frac{\sum_{(i,j)\in\Omega}(t^*_{ij} - t'_{ij})}{\sum_{(i,j)\in\Omega}(1 - f^*_{ij})} = \frac{(1-\eta)\sum_{(i,j)\in\Omega} t^*_{ij}}{\sum_{(i,j)\in\Omega}(1 - f^*_{ij})}.$$

By Lemma 23, $\sum_{(i,j)\in\Omega} t^*_{ij} \le |\Omega| \lesssim nmp_{\mathcal{O}}$. Suppose $p^*_{\mathrm{A}} p_{\mathcal{O}} nm \gtrsim \log(nm)$, then by Lemma 23, $\sum_{(i,j)\in\Omega}(1 - f^*_{ij}) \gtrsim nmp_{\mathcal{O}} p^*_{\mathrm{A}}$. This leads to

$$\frac{\sum_{(i,j)\in\Omega}(t^*_{ij} - t'_{ij})}{\sum_{(i,j)\in\Omega}(1 - f^*_{ij})} \lesssim \frac{(1-\eta) p_{\mathcal{O}} nm}{p_{\mathcal{O}} p^*_{\mathrm{A}} nm} \lesssim \frac{L\log(m)\delta}{\gamma p^*_{\mathrm{A}}}. \tag{30}$$

Note that $\delta \gtrsim \frac{1}{\sqrt{p_{\mathcal{O}} m}}$. Suppose $p^*_{\mathrm{A}} p_{\mathcal{O}} nm \lesssim \log(nm)$, then

$$\frac{L\log(n)\delta}{\gamma p^*_{\mathrm{A}}} \gtrsim \frac{1}{p^*_{\mathrm{A}}\sqrt{p_{\mathcal{O}} m}} \gtrsim \frac{nm\sqrt{p_{\mathcal{O}}}}{\log(nm)\sqrt{m}} \gtrsim 1.$$

This completes the proof. $\square$

B.4. Proof of Theorem 1

Proof of Theorem 1. Finally, we proceed the proof of Theorem 1. Note that

$$\begin{aligned}
&\mathrm{TPR}_{\pi^*(\gamma)}(X_\Omega) - \mathrm{TPR}_{\pi^{\mathrm{EW}}(\gamma)}(X_\Omega) \\
&= \frac{\sum_{(i,j)\in\Omega} t_{ij}^*(1-f_{ij}^*) - \sum_{(i,j)\in\Omega} t_{ij}^{\mathrm{EW}}(1-f_{ij}^*)}{\sum_{(i,j)\in\Omega}(1-f_{ij}^*)} \\
&\le \frac{\sum_{(i,j)\in\Omega}(t_{ij}^* - t_{ij}^{\mathrm{EW}}) + (\sum_{(i,j)\in\Omega} t_{ij}^{\mathrm{EW}} f_{ij}^* - \sum_{(i,j)\in\Omega} t_{ij}^* f_{ij}^*)}{\sum_{(i,j)\in\Omega}(1-f_{ij}^*)}.
\end{aligned}$$

Note that $\sum_{(i,j)\in\Omega} t_{ij}^* f_{ij}^* = \gamma \sum_{(i,j)\in\Omega} f_{ij}^*$ and $\sum_{(i,j)\in\Omega} t_{ij}^{\mathrm{EW}} f_{ij}^* \le \gamma \sum_{(i,j)\in\Omega} f_{ij}^*$ by Lemma 20. Furthermore, $\sum_{(i,j)\in\Omega} t_{ij}^{\mathrm{EW}} \ge \sum_{(i,j)\in\Omega} t'_{ij}$ since $\{t'_{ij}\}$ is a feasible solution of $\mathcal{P}^{\mathrm{EW}}$ and the objective function of $\mathcal{P}^{\mathrm{EW}}$ maximizes $\sum_{(i,j)\in\Omega} t_{ij}^{\mathrm{EW}}$ given the constraint. Hence,

$$\mathrm{TPR}_{\pi^*(\gamma)}(X_\Omega) - \mathrm{TPR}_{\pi^{\mathrm{EW}}(\gamma)}(X_\Omega) \le \frac{\sum_{(i,j)\in\Omega}(t_{ij}^* - t_{ij}^{\mathrm{EW}})}{\sum_{(i,j)\in\Omega}(1-f_{ij}^*)} \le \frac{\sum_{(i,j)\in\Omega}(t_{ij}^* - t'_{ij})}{\sum_{(i,j)\in\Omega}(1-f_{ij}^*)}.$$

Also, note that $\mathrm{TPR}_{\pi^*(\gamma)}(X_\Omega) - \mathrm{TPR}_{\pi^{\mathrm{EW}}(\gamma)}(X_\Omega) \le 1$ since $\mathrm{TPR} \le 1$ by definition. By Lemma 22,

$$\mathrm{TPR}_{\pi^*(\gamma)}(X_\Omega) - \mathrm{TPR}_{\pi^{\mathrm{EW}}(\gamma)}(X_\Omega) \lesssim \frac{L\log(m)\delta}{\gamma p_{\mathrm{A}}^*},$$

which completes the proof. □

C. Minimax Lower Bound

C.1. Proof of Proposition 1

Consider the model $X \sim \mathrm{H}(p_{\mathrm{A}}^*, M^*)$. Recall that the construction of $\mathcal{M}_n = \{M^b \in \mathbb{R}^{n\times n}, b \in \{0,1\}^{n/2}\}$ is: for the i-th and $(i+1)$-th rows, set $M_{ij} = 1$ and $M_{i+1j} = 1 - \frac{c^*}{\sqrt{n}}$ if $b_{i/2} = 0$; otherwise set $M_{ij} = 1 - \frac{c^*}{\sqrt{n}}$ and $M_{i+1j} = 1$. Here $c^* < \frac{1}{2}$ is some sufficient small constant.

For a constant C_0, let Π'_γ denote the set of all policies such that

$$\mathbb{P}_{X\sim\mathrm{H}(p_{\mathrm{A}}^*, M)}\left(\mathrm{FPR}_\pi(X) \le \gamma\right) \ge 1 - C_0/n^2 \quad \text{for all } M \in \mathcal{M}_n. \tag{31}$$

Set $\gamma = \frac{1}{2e}, p_{\mathrm{A}}^* = \frac{1}{2}$. We write $\sum_{(i,j)\in[n]\times[n]}$ as $\sum_{ij}$ if there is no ambiguity. Note that anomaly can only occur when $X_{ij} = 0$. One can verify that

$$\begin{aligned}
f_{ij}^* &= \frac{1/2 e^{-M_{ij}^*}}{1/2 + 1/2 e^{-M_{ij}^*}} \mathbb{1}\{X_{ij} = 0\} + \mathbb{1}\{X_{ij} > 0\} \\
&= \frac{e^{-M_{ij}^*}}{1 + e^{-M_{ij}^*}} \mathbb{1}\{X_{ij} = 0\} + \mathbb{1}\{X_{ij} > 0\}.
\end{aligned}$$

Since the anomaly only occurs when $X_{ij} = 0$, a “rational” algorithm should not claim anomalies for those entries with $X_{ij} > 0$. We have the following result:

Lemma 24. *For any $\pi' \in \Pi'_\gamma$, there exists π such that for any X,*

$$\mathrm{FPR}_\pi(X) \le \mathrm{FPR}_{\pi'}(X), \mathrm{TPR}_\pi(X) = \mathrm{TPR}_{\pi'}(X).$$

Furthermore, $\mathbb{P}\left(A_{ij}^\pi = 1, X_{ij} > 0\right) = 0.$

Proof. For any algorithm π', we can construct π as the following: let $A_{ij}^{\pi} = A_{ij}^{\pi'}$ if $X_{ij} = 0$; otherwise $A_{ij}^{\pi} = 0$. Then it is easy to see that

$$\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X\right) f_{ij}^* \leq \sum_{ij} \mathbb{P}\left(A_{ij}^{\pi'} = 1 | X\right) f_{ij}^*.$$

This implies $\mathrm{FPR}_{\pi}(X) \leq \mathrm{FPR}_{\pi'}(X)$. Furthermore,

$$\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X\right) (1 - f_{ij}^*) = \sum_{ij} \mathbb{P}\left(A_{ij}^{\pi'} = 1 | X\right) (1 - f_{ij}^*).$$

This implies $\mathrm{TPR}_{\pi}(X) = \mathrm{TPR}_{\pi'}(X)$. $\square$

Hence, it is sufficient to only consider π that does not claim anomalies for entries with $X_{ij} > 0$. Let $\Pi_\gamma = \{\pi \in \Pi'_\gamma \mid \mathbb{P}\left(A_{ij}^{\pi} = 1, X_{ij} > 0\right) = 0\}$. Note that the FPR constraint is a high probability statement, hence it is possible that different $M \in \mathcal{M}_n$ satisfies the constraint on different sets of X and makes the problem hard to analyze. To address this issue, we consider the "expectation" of the FPR constraint and have the following lemma.

Lemma 25. *For any $\pi \in \Pi_\gamma$ and any $M \in \mathcal{M}$,*

$$\sum_{ij} a_{ij}^{\pi}(M) e^{-M_{ij}} \leq \gamma n^2 + 2C_0. \tag{32}$$

where $a_{ij}^{\pi}(M) = \mathbb{P}_{X \sim \mathrm{H}(p_{\mathrm{A}}^*, M)}\left(A_{ij}^{\pi} = 1 | X_{ij} = 0\right)$.

Proof. By Eq. (31), with probability $1 - \frac{C_0}{n^2}$, rewrite Eq. (31)

$$\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X\right) f_{ij}^* \leq \gamma \sum_{ij} f_{ij}^*. \tag{33}$$

Take the expectation on the left hand side of Eq. (33), we have

$$\begin{aligned}
&\mathbb{E}_{X \sim \mathrm{H}(p_A^*, M)}\left(\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X\right) f_{ij}^*\right) \\
&= \sum_{ij} \frac{e^{-M_{ij}}}{1 + e^{-M_{ij}}} \mathbb{E}\left(\mathbb{P}\left(A_{ij}^{\pi} = 1, X_{ij} = 0 | X\right)\right) + \mathbb{E}\left(\mathbb{P}\left(A_{ij}^{\pi} = 1, X_{ij} > 0 | X\right)\right) \\
&= \sum_{ij} \frac{e^{-M_{ij}}}{1 + e^{-M_{ij}}} \mathbb{P}\left(A_{ij}^{\pi} = 1, X_{ij} = 0\right) \\
&= \sum_{ij} \frac{1}{2} e^{-M_{ij}} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X_{ij} = 0\right) = \sum_{ij} \frac{1}{2} a_{ij}^{\pi}(M) e^{-M_{ij}}.
\end{aligned}$$

where we use $\mathbb{P}\left(A_{ij}^{\pi} = 1, X_{ij} > 0\right) = 0$ and $\mathbb{P}\left(X_{ij} = 0\right) = \frac{1}{2} + \frac{1}{2} e^{-M_{ij}}$. Take the expectation on the right hand side of Eq. (33), we have

$$\mathbb{E}_{X \sim \mathrm{H}(p_A^*, M)}\left(\gamma \sum_{ij} f_{ij}^*\right) = \gamma \sum_{ij} \left(\frac{1}{2} e^{-M_{ij}} + \frac{1}{2}(1 - e^{-M_{ij}})\right) = \frac{\gamma n^2}{2}.$$

Let $G(t) = t$ in the event that Eq. (33) holds; otherwise $G(t) = n^2$. Then it is easy to verify that $\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi} = 1 | X\right) f_{ij}^* \leq G(\gamma \sum_{ij} f_{ij}^*)$ with probability 1. Take expectation on both sides, we have

$$\sum_{ij} \frac{1}{2} a_{ij}^{\pi}(M) e^{-M_{ij}} \leq \frac{\gamma n^2}{2} + \frac{C_0}{n^2} n^2 \leq \frac{\gamma n^2}{2} + C_0,$$

which completes the proof. $\square$

Next, we consider the expectation of TPR. In particular, note that

$$\begin{aligned}
&\mathbb{E}_{X\sim \mathrm{H}(p_A^*,M)}\left(\sum_{ij}\mathbb{P}\left(A_{ij}^{\pi}=1|X\right)(1-f_{ij}^*)\right)\\
&=\sum_{ij}\frac{1}{1+e^{-M_{ij}}}\mathbb{E}\left(\mathbb{P}\left(A_{ij}^{\pi}=1,X_{ij}=0|X\right)\right)\\
&=\sum_{ij}\frac{1}{1+e^{-M_{ij}}}\mathbb{P}\left(A_{ij}^{\pi}=1,X_{ij}=0\right)\\
&=\sum_{ij}\frac{1}{2}\mathbb{P}\left(A_{ij}^{\pi}=1|X_{ij}=0\right)=\sum_{ij}\frac{1}{2}a_{ij}^{\pi}(M). \qquad (34)
\end{aligned}$$

Let $M^+=1, M^-=1-\frac{c^*}{\sqrt{n}}$. For any $M\in\mathcal{M}_n$, there are one-half M^+ and one-half M^- entries in M. Note that, when observing $X_{ij}=0$, π^* would claim anomaly on the entry M^+ with priority than M^-, because it is more possible to observe 0 for M^- in the normal situation. Indeed, we choose γ and p_{A}^* in a way that π^* roughly claims anomalies for all M^+ entries with $X_{ij}=0$.

Intuitively speaking, if an algorithm π achieves the similar performance as π^*, it must be able to distinguish M^+ and M^- from the observation X. However, the construction of $\mathcal{M}_n$ prevent this distinguishability. We next provide a lemma to connect TPR and the ability of recognizing M^+.

Lemma 26. *For any $\pi\in\Pi_\gamma$ and any $M\in\mathcal{M}$,*

$$\frac{e-e^{M^-}}{e+e^{M^-}}\sum_{ij}\left|\mathbb{1}\left\{M_{ij}=M^+\right\}-a_{ij}^{\pi}(M)\right|\le\frac{n^2}{2}-\sum_{ij}a_{ij}^{\pi}(M)+4C_0e.$$

Proof. Let

$$\begin{aligned}
x &:= \sum_{ij,M_{ij}=M^+}a_{ij}^{\pi}(M)\\
y &:= \sum_{ij,M_{ij}=M^-}a_{ij}^{\pi}(M).
\end{aligned}$$

By Eq. (32), we have $xe^{-M^+}+ye^{-M^-}\le\gamma n^2+2C_0=\frac{n^2}{2e}+2C_0$. Hence

$$y\le\left(\frac{n^2}{2e}+2C_0-xe^{-M^+}\right)e^{M^-}\le\left(\frac{n^2}{2}-x\right)\frac{e^{M^-}}{e}+2C_0e. \qquad (35)$$

Furthermore,

$$\begin{aligned}
\sum_{ij}\left|\mathbb{1}\left\{M_{ij}=M^+\right\}-a_{ij}^{\pi}(M)\right| &= \sum_{ij,M_{ij}=M^+}(1-a_{ij}^{\pi}(M))+\sum_{ij,M_{ij}=M^-}a_{ij}^{\pi}(M)\\
&=\frac{n^2}{2}-x+y\\
&\le\left(\frac{n^2}{2}-x\right)+\left(\frac{n^2}{2}-x\right)\frac{e^{M^-}}{e}+2C_0e. \qquad \text{By Eq. (35).}
\end{aligned}$$

Further algebra provides us

$$
\begin{aligned}
(\frac{n^2}{2} - x) + (\frac{n^2}{2} - x)\frac{e^{M^-}}{e} &\leq \left((\frac{n^2}{2} - x)\frac{e - e^{M^-}}{e} \right) \frac{e + e^{M^-}}{e - e^{M^-}} \\
&= \left(\frac{n^2}{2} - x - (\frac{n^2}{2} - x)\frac{e^{M-}}{e} \right) \frac{e + e^{M^-}}{e - e^{M^-}} \\
&= \left(\frac{n^2}{2} - x - y + (y - (\frac{n^2}{2} - x)\frac{e^{M-}}{e}) \right) \frac{e + e^{M^-}}{e - e^{M^-}} \\
&\leq \left(\frac{n^2}{2} - x - y + 2C_0 e \right) \frac{e + e^{M^-}}{e - e^{M^-}} \\
&= \left(\frac{n^2}{2} - \sum_{ij} a_{ij}^{\pi}(M) + 2C_0 e \right) \frac{e + e^{M^-}}{e - e^{M^-}}.
\end{aligned}
$$

This implies that

$$
\sum_{ij} |\mathbb{1}\left\{M_{ij} = M^+\right\} - a_{ij}^{\pi}(M)| + 2C_0 e \leq \left(\frac{n^2}{2} - \sum_{ij} a_{ij}^{\pi}(M) + 2C_0 e \right) \frac{e + e^{M^-}}{e - e^{M^-}}.
$$

Then, we can conclude

$$
\frac{e - e^{M^-}}{e + e^{M^-}} \sum_{ij} |\mathbb{1}\left\{M_{ij} = M^+\right\} - a_{ij}^{\pi}(M)| \leq \frac{n^2}{2} - \sum_{ij} a_{ij}^{\pi}(M) + 4C_0 e
$$

which completes the proof. □

Next, we show that $a_{ij}^{\pi}(M^a) \approx a_{ij}^{\pi}(M^b)$ if $M^a \approx M^b$.

Lemma 27. *Let $M^a \in \mathbb{R}^{n \times n}$ and $M^b \in \mathbb{R}^{n \times n}$ only differ on two rows (WOLG, the first row and the second row). In particular, $M^a_{ij} = M^b_{ij}$ for any $j \in [n]$ and $i = 3, 4, \ldots, n$. Furthermore, $M^a_{1j} = 1$ and $M^b_{1j} = 1 - \frac{c^*}{\sqrt{n}}$ for $j \in [n]$; $M^a_{2j} = 1 - \frac{c^*}{\sqrt{n}}$ and $M^b_{2j} = 1$ for $j \in [n]$. Here $c^* < \frac{1}{2}$. Then for any $(i, j) \in [n] \times [n]$,*

$$
|a_{ij}^{\pi}(M_1) - a_{ij}^{\pi}(M_2)| \leq c^*.
$$

Proof. Consider some set S such that

$$
a_{ij}^{\pi}(M) = \mathbb{P}_{X \sim \mathrm{H}(p_{\mathrm{A}}^*, M)}\left(A_{ij}^{\pi} = 1 | X_{ij} = 0\right) = \mathbb{P}_{X \sim \mathrm{H}(p_{\mathrm{A}}^*, M)}\left(X \in S\right).
$$

Let $X(M)$ be $X \sim \mathrm{H}(p_{\mathrm{A}}^*, M)$, $\delta(X||Y)$ be the total variation distance between X and Y, $D_{\mathrm{KL}}(X||Y)$ be the KL-divergence between X and Y. Then,

$$
\begin{aligned}
|a_{ij}^{\pi}(M^a) - a_{ij}^{\pi}(M^b)| &= |\mathbb{P}_{X(M^a)}\left(X \in S\right) - \mathbb{P}_{X(M^b)}\left(X \in S\right)| \\
&\leq \delta(X(M^a)||X(M^b)) && \text{total variation distance} \\
&\leq \sqrt{\frac{1}{2} D_{KL}(X(M^a)||X(M^b))} && \text{Pinsker's inequality} \\
&= \sqrt{\frac{1}{2} \sum_{ij} D_{KL}(X(M^a)_{ij}||X(M^b)_{ij})} && X_{ij} \text{ are independent.}
\end{aligned}
$$

Note that there are only two rows that are different between M^a and M^b. Let X^+ be the observation of the entry with value M^+ and X^- be the observation of the entry with the value M^-. Then we have

$$
\sum_{ij} D_{KL}(X(M^a)_{ij}||X(M^b)_{ij}) = nD_{KL}(X^+||X^-) + nD_{KL}(X^-||X^+).
$$

Note that $X^+ = Y^+ b, X^- = Y^- b$ where $Y^+ = \text{Poisson}(M^+), Y^- = \text{Poisson}(M^-)$, and b indicates whether the anomaly occurs. Hence by the data processing inequality and formula of KL-divergence of Poisson random variables,

$$\begin{aligned}
D_{KL}(X^+||X^-) &\leq D_{KL}(Y^+||Y^-) \\
&= (M^+ \log(M^+/M^-) + M^- - M^+) \\
&= -\log(1 - \frac{c^*}{\sqrt{n}}) - \frac{c^*}{\sqrt{n}} \\
&= \frac{c^*}{\sqrt{n}} + \frac{(c^*)^2}{2n} + \sum_{k=3}^{\infty} \frac{1}{k}(\frac{c^*}{\sqrt{n}})^k - \frac{C^*}{\sqrt{n}} \\
&\leq \frac{(c^*)^2}{2n} + \frac{1}{3}(\frac{c^*}{\sqrt{n}})^3 \sum_{k=0}^{\infty}(\frac{c^*}{\sqrt{n}})^k \\
&\leq \frac{(c^*)^2}{2n} + \frac{2c^*}{3}\frac{(c^*)^2}{n^2} \leq \frac{(c^*)^2}{n}.
\end{aligned}$$

where $c^* < \frac{1}{2}$. Similarly,

$$\begin{aligned}
D_{KL}(X^-||X^+) &\leq D_{KL}(Y^+||Y^-) \\
&= (M^- \log(M^-/M^+) + M^+ - M^-) \\
&= (1 - \frac{c^*}{\sqrt{n}}) \log(1 - \frac{c^*}{\sqrt{n}}) + \frac{c^*}{\sqrt{n}} \\
&\leq (1 - \frac{c^*}{\sqrt{n}})(-\frac{c^*}{\sqrt{n}}) + \frac{c^*}{\sqrt{n}} \\
&\leq \frac{(c^*)^2}{n}.
\end{aligned}$$

Hence,

$$|a^{\pi}_{ij}(M_1) - a^{\pi}_{ij}(M_2)| \leq c^*.$$

□

Next, we show a bound related to the “aggregated TPR” of all $M \in \mathcal{M}_n$.

Lemma 28.

$$\frac{n^2}{2} - \frac{1}{2^{n/2}} \sum_{M \in \mathcal{M}_n} \sum_{ij} a^{\pi}_{ij}(M) \geq -4C_0 e + \frac{c^*(1-c^*)}{4} n\sqrt{n}. \tag{36}$$

where c is a constant.

Proof. Recall that $|\mathcal{M}_n| = \frac{1}{2^{n/2}}$. In order to use the Lemma 26, we derive a lower bound on

$$\frac{1}{2^{n/2}} \sum_{M \in \mathcal{M}_n} \sum_{ij} |\mathbb{1}\left\{M_{ij} = M^+\right\} - a^{\pi}_{ij}(M)| = \frac{1}{2^{n/2}} \sum_{ij} \sum_{M \in \mathcal{M}_n} |\mathbb{1}\left\{M_{ij} = M^+\right\} - a^{\pi}_{ij}(M)|.$$

Consider fixed (i, j). Let $M^a, M^b \in \mathcal{M}_n$ be a pair of matrices such that the only different rows between M^a, M^b are the i-th row and the $i+1$-th row (or the $i-1$-th row). Without loss of generality, suppose $M^a_{ij} = M^+$. Note that there are $2^{n/2-1}$ such pairs. Consider

$$\begin{aligned}
&|\mathbb{1}\left\{M^a_{ij} = M^+\right\} - a^{\pi}_{ij}(M^a)| + |\mathbb{1}\left\{M^b_{ij} = M^+\right\} - a^{\pi}_{ij}(M^b)| \\
&= |1 - a^{\pi}_{ij}(M^a)| + |a^{\pi}_{ij}(M^b)| \\
&= 1 - a^{\pi}_{ij}(M^a) + a^{\pi}_{ij}(M^b) \\
&\geq 1 - |a^{\pi}_{ij}(M^a) - a^{\pi}_{ij}(M^b)| \\
&\geq 1 - c^*.
\end{aligned}$$

The last inequality is due to Lemma 27. Hence,

$$\begin{aligned}
&\frac{1}{2^{n/2}}\sum_{ij}\sum_{M\in\mathcal{M}_n}\left|\mathbb{1}\left\{M_{ij}=M^+\right\}-a^{\pi}_{ij}(M)\right| \\
&\geq \frac{1}{2^{n/2}}\sum_{ij}(2^{n/2-1})(1-c^*) \\
&= \frac{n^2}{2}(1-c^*)
\end{aligned}$$

Also note that (Recall $c^*\leq\frac{1}{2}$)

$$\frac{e-e^{M-}}{e+e^{M-}}\geq\frac{e(1-e^{-\frac{c^*}{\sqrt{n}}})}{e}\geq 1-e^{-\frac{c^*}{\sqrt{n}}}\geq\frac{c^*}{2\sqrt{n}}.$$

Then by Lemma 26, one can obtain

$$\frac{c^*}{2\sqrt{n}}\sum_{ij}\left|\mathbb{1}\left\{M_{ij}=M^+\right\}-a^{\pi}_{ij}(M)\right|\leq\frac{n^2}{2}-\sum_{ij}a^{\pi}_{ij}(M)+4C_0e.$$

Sum over $M\in\mathcal{M}_n$ on both sides, we have

$$\begin{aligned}
&\frac{n^2}{2}-\frac{1}{2^{n/2}}\sum_{M\in\mathcal{M}_n}\sum_{ij}a^{\pi}_{ij}(M)+4C_0e \\
&\geq\frac{c^*}{2\sqrt{n}}\frac{1}{2^{n/2}}\sum_{M\in\mathcal{M}_n}\left|\mathbb{1}\left\{M_{ij}=M^+\right\}-a^{\pi}_{ij}(M)\right| \\
&\geq\frac{c^*}{2\sqrt{n}}\frac{n^2}{2}(1-c^*) \\
&\geq\frac{c^*(1-c^*)}{4}n\sqrt{n}.
\end{aligned}$$

□

Next, we consider the ideal policy $\pi^*(\gamma)$. Write $\pi^*(\gamma)$ as π^* if there is no ambiguity.

Lemma 29. *For any $M\in\mathcal{M}_n$,*

$$\sum_{ij}a^{\pi^*}_{ij}(M)\geq\frac{n^2}{2}-Cn\log n-2.$$

where C is a constant.

Proof. By Lemma 14, we have, with probability $1-\frac{1}{n^2}$,

$$\left|\sum_{ij}f^*_{ij}-n^2(1-p^*_{\mathrm{A}})\right|\leq C_1 n\log n.$$

where C_1 is a constant. Consider a policy π' that knows the true rate matrix M. Without loss of generality, let first $\frac{n}{2}$ rows of M be M^+. Suppose π' claims anomalies for (i,j) with $X_{ij}=0$ and $i\leq(n-k_1)/2$ with $k_1=8eC_1n\log n$. Then, with probability $1-\frac{1}{n^2}$,

$$\begin{aligned}
\sum_{ij}\mathbb{P}\left(A^{\pi'}_{ij}=1|X\right)f^*_{ij}&=\sum_{i\leq(n-k_1)/2}\mathbb{1}\left\{X_{ij}=0\right\}\frac{e^{-M^+}}{1+e^{-M^+}} \\
&\leq\frac{n(n-k_1)(1-p^*_{\mathrm{A}})}{2e}
\end{aligned}$$

Then we have

$$\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi'} = 1|X\right) f_{ij}^* \leq \frac{n^2(1-p_{\mathrm{A}}^*)}{2e} - C_1 n \log n \leq \gamma \sum_{ij} f_{ij}^*.$$

Therefore, with probability $1 - \frac{2}{n^2}$,

$$\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi'} = 1|X\right)(1 - f_{ij}^*) \leq \sum_{ij} \mathbb{P}\left(A_{ij}^{\pi^*} = 1|X\right)(1 - f_{ij}^*).$$

Finally, we have

$$\begin{aligned}
\mathbb{E}_{X(M)}\left(\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi^*} = 1|X\right)(1 - f_{ij}^*)\right) &\geq \mathbb{E}_{X(M)}\left(\sum_{ij} \mathbb{P}\left(A_{ij}^{\pi'} = 1|X\right)(1 - f_{ij}^*)\right) - (n^2)/\frac{2}{n^2} \\
&= \sum_{i \leq (n-k_1)/2} \mathbb{1}\left\{X_{ij} = 0\right\} \frac{1}{1+e^{-1}} - 2 \\
&\geq \frac{n(n-k_1)}{4} - n \log n - 2 \\
&= \frac{n^2}{4} - Cn \log n - 2.
\end{aligned}$$

□

Proof of Proposition 1. Next, we finish the proof of Proposition 1. Combining Lemma 28 and Lemma 29, we have

$$\begin{aligned}
&\frac{1}{2^{n/2}} \sum_{M \in \mathcal{M}_n} \sum_{ij} a_{ij}^{\pi^*}(M) - \frac{1}{2^{n/2}} \sum_{M \in \mathcal{M}_n} \sum_{ij} a_{ij}^{\pi}(M) \\
&\geq -Cn \log n - 2 - 4C_0 e + \frac{c^*(1-c^*)}{4} n\sqrt{n}.
\end{aligned}$$

Therefore, there exists a $M' \in \mathcal{M}_n$, such that

$$\sum_{ij} a_{ij}^{\pi^*}(M') - \sum_{ij} a_{ij}^{\pi}(M') \geq C_1 n\sqrt{n}.$$

Finally, we have

$$\begin{aligned}
&\mathbb{E}_{X \sim \mathrm{H}(p_{\mathrm{A}}^*, M')} \left(\mathrm{TPR}_{\pi^*}(X) - \mathrm{TPR}_{\pi}(X)\right) \\
&= \mathbb{E}_{X \sim \mathrm{H}(p_{\mathrm{A}}^*, M')} \left(\frac{\sum_{ij} \left(\mathbb{P}\left(A_{ij}^{\pi^*} = 1|X\right) - \mathbb{P}\left(A_{ij}^{\pi} = 1|X\right)\right)(1 - f_{ij})^*}{\sum_{ij}(1 - f_{ij}^*)}\right) \\
&\geq \frac{\sum_{ij} a_{ij}^{\pi^*}(M') - \sum_{ij} a_{ij}^{\pi}(M')}{2n^2} \\
&\geq \frac{C_1}{2\sqrt{n}}
\end{aligned}$$

which completes the proof. □

D. Experiments

In this section, we provide further implementation details of the experiments.

Computing Infrastructure. all experiments are done in a personal laptop equipped with 2.6 GHz 6-Core Intel Core i7 and 16 GB 2667 MHz DDR4. The operating system is macOS Catalina. For each instance, the running time is within seconds for our algorithm.

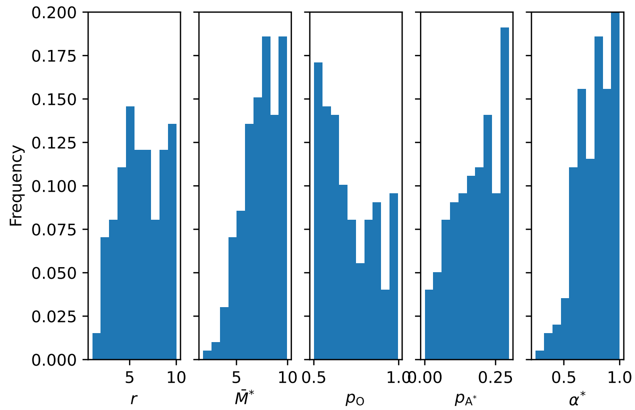

Figure 3: Synthetic data. Histograms shows problem characteristics where EW performs worst relative to ideal algorithm (20th percentile).

Synthetic Data. We present the implementation details of our algorithm and three state-of-the-arts. For practical consideration, we implemented a slight variant of the EW algorithm where (i) the matrix completion step used the typical soft impute algorithm (Mazumder et al., 2010); (ii) the anomaly model estimation used MLE; and (iii) solving $\mathcal{P}^{\mathrm{EW}}$ by replacing f^*_{ij} directly by $\frac{\hat{y}_{ij}}{\hat{x}_{ij}+\hat{y}_{ij}}$. Given the observation X_Ω, the soft impute algorithm solves the optimization problem $\min_M \|P_\Omega(X-M)\|_{\mathrm{F}}^2 + \lambda \|M\|_*$ where λ is a hyper-parameter. To tune λ, we start with a small λ and gradually increase it until the rank of the solution fits the true rank of M^* (all other algorithms also use the knowledge of the true rank). In order to generate the AUC curve for each instance, we vary γ in our algorithm.

In the implementation of Stable-PCP, we solve the following optimization problem $(\hat{M}, \hat{A}) = \arg\min_{M,A} \|M\|_* + \lambda \|A\|_1 + \mu \|P_\Omega(M+A-X)\|_{\mathrm{F}}^2$ by alternating optimization (Ma & Aybat, 2018). The set of anomalies is identified from $\{(i,j) \mid \hat{A}_{ij} \neq 0\}$. In order to choose suitable (λ, μ) and generate the AUC curve, note that when $\hat{M}$ fixed, the ratio of λ/μ decides the portion that will be classified as anomalies (i.e., different points on the AUC curve). Hence, we iterate the ratio λ/μ and then tune λ (accordingly, μ) such that the solution $\hat{M}$ fits the true rank of M^*. This provides an AUC curve.

In the DRMF algorithm, we implement the Algorithm 1 in (Xiong et al., 2011)[7] to solve the following optimization problem $(\hat{M}, \hat{A}) = \arg\min \|P_\Omega(X-A-M)\|_{\mathrm{F}}$ with the constraints rank$(\hat{M}) \leq r, \|A\|_0 \leq e$. The set of anomalies is identified from $\{(i,j) \mid \hat{A}_{ij} \neq 0\}$. Here, we provide the true rank r and vary e for the DRMF algorithm to generate the AUC curve.

For the RMC algorithm (Klopp et al., 2017), the authors propose the following optimization problem $(\hat{M}, \hat{A}) = \arg\min_{M,A} \|M\|_* + \lambda \|A\|_1 + \mu \|P_\Omega(M+A-X)\|_{\mathrm{F}}^2$ with constraints $\|M\|_{\max} \leq a, \|A\|_{\max} \leq a$. This is effectively the Stable-PCP algorithm with the max norm constraints. We choose $a = k\|M^*\|_{\max}$ for some constant scale $k > 1$. Then we implement RMC based on Stable-PCP and a projection of (M, A) into the set with max norm constraints in every iteration during the alternating optimization.

We also study the limitation of our algorithm, in which the performance starts to degrade. Figure 3 shows that the problem instances (in the experiment of the synthetic data) where the AUC of EW was furthest away from the ideal AUC (20th percentile). The results show largely intuitive characteristics: higher α^* (so anomalies look similar to non-anomalous entires), lower $p_{\mathcal{O}}$, higher p^*_{A} and higher r (so that M^* is harder to estimate). The behavior with respect to $\bar{M}^*$ is surprising but was consistently observed across other ensembles as well.

Real Data. We estimate the rank of M^* to obtain $r \sim 30$ via cross-validation. The observation X is generated in a same way described in the synthetic data. The EW and Stable-PCP algorithms are implemented in a same way as in the synthetic data with rank information r.

[7]Although (Xiong et al., 2011) does not consider the partial observation scenario, but the generalization to address missing entries is straightforward.