A Technical Lemmas

**Lemma 4** (Theorem 1.1 in [10]). There exists a constant $K$ such that, for any $n$, $m$ any $h \leq 2 \log \max \{m, n\}$ and any $m \times n$ matrix $A = (a_{ij})$ where $a_{ij}$ are i.i.d. symmetric random variables, the following inequality holds:

$$
\max \left\{ \mathbb{E} \max_{1 \leq i \leq m} \|a_i\|_2^h, \mathbb{E} \max_{1 \leq j \leq n} \|a_j^h\|_2 \right\} \leq \mathbb{E} \|A\|_h^2 \leq K \left( \mathbb{E} \max_{1 \leq i \leq m} \|a_i\|_2^h + \mathbb{E} \max_{1 \leq j \leq n} \|a_j^h\|_2 \right).
$$

**Lemma 5** (Symmetrization, Lemma 6.3 in [30]). Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex. Then, for any finite sequence $\{t_i\}$ of independent mean zero random variables in $B$ such that for every $i \mathbb{E} \left[ F(\|t_i\|_2) \right] < \infty$, then

$$
\mathbb{E} \left[ F \left( \frac{1}{2} \sum \xi_i t_i \right) \right] \leq \mathbb{E} \left[ F \left( \sum t_i \right) \right] \leq \mathbb{E} \left[ F \left( 2 \sum \xi_i t_i \right) \right],
$$

where $\{\xi_i\}$ are i.i.d. Rademacher random variables.

**Lemma 6** (Contraction, Theorem 4.12 in [30]). Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing. Let $\psi_i : \mathbb{R} \to \mathbb{R}$ be contraction such that $\psi_i(0) = 0$. Then it holds that

$$
\mathbb{E} \left[ F \left( \frac{1}{2} \sup_{t_1, \ldots, t_N} \left| \sum_{i=1}^N \xi_i \psi_i(t_i) \right| \right) \right] \leq \mathbb{E} \left[ F \left( \sup_{t_1, \ldots, t_N} \left| \sum_{i=1}^N \xi_i t_i \right| \right) \right],
$$

where $\{\xi_i\}$ are i.i.d. Rademacher random variables.

**Lemma 7** (Lemma 2 in [17]). Let $f$ be a differentiable function and assume $\max \left\{ \|M\|_\infty, \|\hat{M}\|_\infty \right\} \leq \alpha$. Then

$$
d_H^2 \left( f(M), f(\hat{M}) \right) \geq \inf_{|x| \leq \alpha} \frac{(f'(x))^2}{8f(x)(1-f(x))} \frac{\|M - \hat{M}\|_F^2}{d_1d_2}.\]

**Lemma 8** (Lemma 4 in [17]). Suppose that $x, y \in (0, 1)$. Then

$$
D(x|y) \leq \frac{(x-y)^2}{y(1-y)}.
$$

**Lemma 9** (Lemma 3 in [17]). Let $\mathcal{K}$ be the set of matrices that satisfy (A2) and (A3). Let $0 < \nu \leq 1$ be a scalar such that $\nu r^{-2}$ is an integer that is not larger than $d_1$. Then there exists a subset $\mathcal{X} \subset \mathcal{K}$ with the following properties:

1. $|\mathcal{X}| \geq \exp \left( \frac{\nu}{16 \log \nu} \right)$.
2. $\forall X \in \mathcal{X}$, $|X_{ij}| = \alpha \nu$.
3. $\forall X, \tilde{X} \in \mathcal{X}$ with $X \neq \tilde{X}$, $\|X - \tilde{X}\|_F^2 > \frac{1}{4} \alpha^2 \nu^2 d_1d_2$.

B Proof for Main Results

Recall the observation model: $M \in \mathbb{R}^{d_1 \times d_2}$ is the true low-rank matrix and $\Omega \subset [d_1] \times [d_2]$ is the index set of entries we observed. $Y \in \mathbb{R}^{d_1 \times d_2}$ is the binary matrix determined by $M$: for all $(i, j) \in \Omega$,

$$
Y_{ij} = \begin{cases} +1, \text{ with probability } f(M_{ij}), \\ -1, \text{ with probability } 1 - f(M_{ij}). \end{cases}
$$

In the setting of symmetric noise, the observation $Y'_{ij} = \delta_{ij} Y_{ij}$ where $\delta_{ij}$ are i.i.d. and

$$
\delta_{ij} = \begin{cases} +1, \text{ with probability } 1 - \tau, \\ -1, \text{ with probability } \tau, \end{cases}
$$
where \( \tau \in (0, 1/2) \) itself can be a random variable. Therefore, conditioning on \( \tau \), we observe
\[
\Pr(Y_{ij}' = 1 \mid \tau) = (1 - \tau)f(M_{ij}) + \tau(1 - f(M_{ij})).
\]

**Case 1.** If \( \tau \) is a discrete random variable, say
\[
\Pr(\tau = \tau_k) = p_k, \quad 1 \leq k \leq s,
\]
then it is easy to see that
\[
\Pr(Y_{ij}' = 1) = \sum_{k=1}^{s} \Pr(Y_{ij}' = 1, \ \tau = \tau_k)
\]
\[
= \sum_{k=1}^{s} \Pr(Y_{ij}' = 1 \mid \tau = \tau_k) \cdot \Pr(\tau = \tau_k)
\]
\[
= \sum_{k=1}^{s} p_k \left[ (1 - \tau_k)f(M_{ij}) + \tau_k(1 - f(M_{ij})) \right].
\]

Denote
\[
g(x) = \sum_{k=1}^{s} p_k \left[ (1 - \tau_k)f(x) + \tau_k(1 - f(x)) \right] = (1 - 2E[\tau])f(x) + E[\tau].
\]

We have
\[
Y_{ij}' = \begin{cases} 
  +1, \text{ with probability } g(M_{ij}), \\
  -1, \text{ with probability } 1 - g(M_{ij}).
\end{cases}
\]

**Case 2.** If \( \tau \) is a continuous random variable with probability density function (pdf) \( h_\tau(t) \), then we have
\[
\Pr(Y_{ij}' = 1) = \int_t h_{Y,\tau}(Y_{ij}' = 1, t)dt
\]
\[
= \int_t h_{Y\mid\tau}(Y_{ij}' = 1 \mid t)h_\tau(t)dt
\]
\[
= \int_t h_\tau(t) \left[ (1 - t)f(M_{ij}) + t(1 - f(M_{ij})) \right] dt,
\]
where \( h_{Y,\tau}(y, t) \) is the joint pdf of \( Y_{ij} \) and \( \tau \), and \( h_{Y\mid\tau}(y \mid t) \) is the conditional pdf. Thus, define
\[
g(x) = \int_t h_\tau(t) \left[ (1 - t)f(x) + t(1 - f(x)) \right] dt = (1 - 2E[\tau])f(x) + E[\tau].
\]

We again have
\[
Y_{ij}' = \begin{cases} 
  +1, \text{ with probability } g(M_{ij}), \\
  -1, \text{ with probability } 1 - g(M_{ij}).
\end{cases}
\]

Hence, the maximum likelihood estimator is given as follows:
\[
\hat{M} = \arg\max_X L_{\Omega,Y'}(X), \quad \text{s.t. } \|X\|_* \leq \alpha \sqrt{rd_1d_2}, \|X\|_\infty \leq \gamma,
\]
where
\[
L_{\Omega,Y'}(X) := \sum_{(i,j) \in \Omega} (1_{(Y_{ij} = 1)} \log g(X_{ij}) + 1_{(Y_{ij} = -1)} \log(1 - g(X_{ij}))).
\]
For the sake of a principled analysis, we will treat \( g(x) \) as a general function at this point. Associated with the function \( g(x) \) are two quantities:

\[
\rho_\gamma^+ := \sup_{|x| \leq \gamma} \frac{|g'(x)|}{g(x)(1-g(x))}, \quad \rho_\gamma^- := \sup_{|x| \leq \gamma} \frac{g(x)(1-g(x))}{(g'(x))^2}.
\]

We will use several kinds of distances in the proof. The first one is Hellinger distance that is given by

\[
d_H^2(p, q) := (\sqrt{p} - \sqrt{q})^2 + (\sqrt{1-p} + \sqrt{1-q})^2, \quad \forall \ 0 \leq p, q \leq 1.
\]

Extending it to the matrix, we write

\[
d_H^2(P, Q) := \frac{1}{d_1d_2} \sum_{i,j} d_H^2(P_{ij}, Q_{ij}),
\]

where \( P, Q \in \mathbb{R}^{d_1 \times d_2} \) and the entries therein are between 0 and 1.

For two probability distributions \( P \) and \( Q \) on a finite set \( A \), the Kullback-Leibler (KL) divergence is defined as

\[
D(P || Q) = \sum_{x \in A} P(x) \log \frac{P(x)}{Q(x)}.
\]

With a slight abuse, we write for two scalars \( p, q \in [0, 1] \)

\[
D(p || q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q},
\]

and for two matrices \( P, Q \in [0, 1]^{d_1 \times d_2} \),

\[
D(P || Q) = \frac{1}{d_1d_2} \sum_{i,j} D(P_{ij} || Q_{ij}).
\]

Throughout the proof, we will work with a shifted MLE, i.e.

\[
\bar{L}_{\Omega, Y'}(X) := L_{\Omega, Y'}(X) - L_{\Omega, Y'}(0) = \sum_{(i,j) \in \Omega} \left( 1_{\{Y_{ij} = 1\}} \log \frac{g(X_{ij})}{g(0)} + 1_{\{Y_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(0)} \right),
\]

\[
= \sum_{i,j} 1_{\{(i,j) \in \Omega\}} \left( 1_{\{Y_{ij} = 1\}} \log \frac{g(X_{ij})}{g(0)} + 1_{\{Y_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(0)} \right). \quad (12)
\]

**B.1 Proof for Lemma 3**

*Proof.* Using the Markov’s inequality, we have for any \( \theta > 0 \),

\[
\Pr \left( \sup_{X \in S} \left| \bar{L}_{\Omega, Y'}(X) - \mathbb{E} \bar{L}_{\Omega, Y'}(X) \right| \geq C_0 \alpha \rho_\gamma^+ \sqrt{\theta} \sqrt{n(d_1 + d_2)} + d_1d_2 \log(d_1d_2) \right) \\
\leq \frac{\mathbb{E} \left[ \sup_{X \in S} \left| \bar{L}_{\Omega, Y'}(X) - \mathbb{E} \bar{L}_{\Omega, Y'}(X) \right|^\theta \right]}{(C_0 \alpha \rho_\gamma^+ \sqrt{\theta} \sqrt{n(d_1 + d_2)} + d_1d_2 \log(d_1d_2))^{\frac{\theta}{2}}} \quad (13)
\]

We bound the numerator above. Recall that

\[
\bar{L}_{\Omega, Y'}(X) = \sum_{i,j} 1_{\{(i,j) \in \Omega\}} \left( 1_{\{Y'_{ij} = 1\}} \log \frac{g(X_{ij})}{g(0)} + 1_{\{Y'_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(0)} \right).
\]

Let the random variable

\[
\bar{e}_{ij} = 1_{\{(i,j) \in \Omega\}} \left( 1_{\{Y'_{ij} = 1\}} \log \frac{g(X_{ij})}{g(0)} + 1_{\{Y'_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(0)} \right),
\]
and let
\[ t_{ij} = \tilde{t}_{ij} - \mathbb{E} \tilde{t}_{ij}. \]

Then
\[ \bar{L}_{\Omega,Y'}(X) - \mathbb{E} \bar{L}_{\Omega,Y'}(X) = \sum_{i,j} t_{ij}. \]

Note that \{t_{ij}\} are i.i.d. random variables with zero mean. The function \( F(t) = \sup t^\theta \) is convex for \( \theta \geq 1 \), and \( \mathbb{E} F(|t_{ij}|) \) is finite for all \((i,j) \in [d_1] \times [d_2]\). Hence, we can apply Lemma 5 to obtain
\[
\mathbb{E} \left[ \sup_{X \in S} \left| \bar{L}_{\Omega,Y'}(X) - \mathbb{E} \bar{L}_{\Omega,Y'}(X) \right|^\theta \right] \leq 2^\theta \mathbb{E} \left[ \sup_{X \in S} \sum_{i,j} \xi_{ij} 1_{\{(i,j) \in \Omega\}} \left( 1_{\{Y'_{ij} = 1\}} \log \frac{g(X_{ij})}{g(0)} + 1_{\{Y'_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(0)} \right) \right]^\theta,
\]

where \{\xi_{ij}\} are i.i.d. Rademacher random variables. Now observe that due to the construction of \( \rho^+_\alpha \), both \( \frac{1}{p_\alpha} \log \frac{g(x)}{g(0)} \) and \( \frac{1}{p_\alpha} \log \frac{1 - g(x)}{1 - g(0)} \) are contractions and vanish at \( x = 0 \). Thereby, using Lemma 6 we have
\[
\mathbb{E} \left[ \sup_{X \in S} \left| \bar{L}_{\Omega,Y'}(X) - \mathbb{E} \bar{L}_{\Omega,Y'}(X) \right|^\theta \right] \leq (4 \rho^+_\alpha)^\theta \mathbb{E} \left[ \sup_{X \in S} \sum_{i,j} \xi_{ij} 1_{\{(i,j) \in \Omega\}} \left( 1_{\{Y'_{ij} = 1\}} X_{ij} - 1_{\{Y'_{ij} = -1\}} X_{ij} \right) \right]^\theta.
\]

With a simple algebra, we have
\[ \Pr(\xi_{ij} Y'_{ij} = 1) = \Pr(\xi_{ij} = 1, Y'_{ij} = 1) + \Pr(\xi_{ij} = -1, Y'_{ij} = -1) = \frac{1}{2} \left( \Pr(Y'_{ij} = 1) + \Pr(Y'_{ij} = -1) \right) = \frac{1}{2}, \]

which implies that the distribution of \( \xi_{ij} Y'_{ij} \) is the same as that of \( \xi_{ij} \) for all \((i,j) \in [d_1] \times [d_2]\). Thus, by denoting \( \Delta_\Omega \) the matrix such that its \((i,j)\)-th element is 1 if \((i,j) \in \Omega\) and 0 otherwise, and \( \Xi = (\xi_{ij}) \), it follows that
\[
\mathbb{E} \left[ \sup_{X \in S} \left| \bar{L}_{\Omega,Y'}(X) - \mathbb{E} \bar{L}_{\Omega,Y'}(X) \right|^\theta \right] \leq (4 \rho^+_\alpha)^\theta \mathbb{E} \left[ \sup_{X \in S} \sum_{i,j} \xi_{ij} 1_{\{(i,j) \in \Omega\}} X_{ij} \right]^\theta = (4 \rho^+_\alpha)^\theta \mathbb{E} \left[ \sup_{X \in S} \|\Delta_\Omega \circ \Xi, X\|^\theta \right] \leq (4 \rho^+_\alpha)^\theta \mathbb{E} \left[ \sup_{X \in S} \|\Delta_\Omega \circ \Xi\|^\theta \right] \leq (\alpha \sqrt{rd_1 d_2})^\theta \mathbb{E} \left[ \|\Delta_\Omega \circ \Xi\|^\theta \right].
\]

Above, the last inequality follows from the nuclear norm constraint we imposed in the MLE estimator. Note that the \((i,j)\)-th entry of the matrix \( \Delta_\Omega \circ \Xi \) is given by \( 1_{\{(i,j) \in \Omega\}} \xi_{ij} \), which are i.i.d. symmetric random variables. Thus, Lemma 7 implies that
\[
\mathbb{E} \left[ \|\Delta_\Omega \circ \Xi\|^\theta \right] \leq C \left( \mathbb{E} \max_{1 \leq i \leq d_1} \left( \sum_{j=1}^{d_2} (\xi_{ij} \Delta_{ij})^2 \right)^{\theta/2} + \mathbb{E} \max_{1 \leq j \leq d_2} \left( \sum_{i=1}^{d_1} (\xi_{ij} \Delta_{ij})^2 \right)^{\theta/2} \right) \leq C \left( \mathbb{E} \max_{1 \leq i \leq d_1} \left( \sum_{j=1}^{d_2} \Delta_{ij} \right)^{\theta/2} + \mathbb{E} \max_{1 \leq j \leq d_2} \left( \sum_{i=1}^{d_1} \Delta_{ij} \right)^{\theta/2} \right). \]
Fix $i$. By Bernstein’s inequality, for all $t > 0$,

$$\Pr \left( \sum_{j=1}^{d_2} \left| \Delta_{ij} - \frac{n}{d_1 d_2} \right| > t \right) \leq 2 \exp \left( \frac{-t^2/2}{n/d_1 + t/3} \right).$$

When $t \geq \frac{6n}{d_1}$, the above reduces to

$$\Pr \left( \sum_{j=1}^{d_2} \left| \Delta_{ij} - \frac{n}{d_1 d_2} \right| > t \right) \leq 2 \exp(-t).$$

Suppose that $W_1, \ldots, W_{d_1}$ are i.i.d. exponential random variables with pdf $\exp(-t)$. Then it follows that

$$\Pr \left( \sum_{j=1}^{d_2} \left| \Delta_{ij} - \frac{n}{d_1 d_2} \right| > t \right) \leq 2 \Pr(W_i > t).$$

On the other hand, we have

$$\left( \mathbb{E} \max_{1 \leq i \leq d_1} \left( \sum_{j=1}^{d_2} \Delta_{ij} \right)^{\theta/2} \right)^{1/\theta} \leq \sqrt{\frac{n}{d_1}} + \left( \frac{6n}{d_1} \right)^{\theta/2} + \int_0^{+\infty} \Pr \left( \max_{1 \leq i \leq d_1} \left( \sum_{j=1}^{d_2} \left| \Delta_{ij} - \frac{n}{d_1 d_2} \right| \right)^{\theta/2} \geq t \right) dt \right)^{1/\theta} \leq \sqrt{\frac{n}{d_1}} + \left( \frac{6n}{d_1} \right)^{\theta/2} + \int_0^{+\infty} \Pr \left( \max_{1 \leq i \leq d_1} W_i^{\theta/2} \geq t \right) dt \right)^{1/\theta} \leq \sqrt{\frac{n}{d_1}} + \left( \frac{6n}{d_1} \right)^{\theta/2} + 2 \mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2} \right)^{1/\theta} \leq (1 + \sqrt{6}) \sqrt{\frac{n}{d_1}} + 2^{1/\theta} \left( \mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2} \right)^{1/\theta}.$$

Here, $\zeta_1$ and $\zeta_2$ use the identity $\mathbb{E} x = \int_0^{+\infty} \Pr(x \geq t) dt$ for any positive random variable $x$. It remains to bound $\mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2}$. Using the fact that $W_i$ is exponential, we have

$$\mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2} \leq \max_{1 \leq i \leq d_1} W_i - \log d_1 \left( \frac{\theta/2}{\log^{\theta/2} d_1 + \log^{\theta/2} d_1 \leq 2((\theta/2)!)} + \log^{\theta/2} d_1 \leq 2(\theta/2)^{\theta/2} + \log^{\theta/2} d_1,\right.$$ 

where we apply Stirling’s approximation in the last inequality. Thus,

$$2^{1/\theta} \left( \mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2} \right)^{1/\theta} \leq 2^{1/\theta} \left( \sqrt{\log d_1} + 2^{1/\theta} \sqrt{\theta/2} \right).$$

Picking $\theta = 2 \log(d_1 + d_2)$ gives

$$2^{1/\theta} \left( \mathbb{E} \max_{1 \leq i \leq d_1} W_i^{\theta/2} \right)^{1/\theta} \leq (2 + \sqrt{2}) \sqrt{\log(d_1 + d_2)}.$$
Putting pieces together, we obtain

\[
\left( \mathbb{E} \max_{1 \leq i \leq d_1} \left( \sum_{j=1}^{d_2} \Delta_{ij} \right)^{\theta/2} \right)^{1/\theta} \leq (1 + \sqrt{6}) \sqrt{\frac{n}{d_1}} + (2 + \sqrt{2}) \sqrt{\log(d_1 + d_2)}.
\]

Likewise, we can show that

\[
\left( \mathbb{E} \max_{1 \leq i \leq d_2} \left( \sum_{i=1}^{d_1} \Delta_{ij} \right)^{\theta/2} \right)^{1/\theta} \leq (1 + \sqrt{6}) \sqrt{\frac{n}{d_2}} + (2 + \sqrt{2}) \sqrt{\log(d_1 + d_2)}.
\]

Note that \(\sqrt{\cdot}\) is a concave function. Hence, Jensen’s inequality implies that (13) can be bounded as follows:

\[
\left( \mathbb{E} \left[ \|\Delta_\Omega \circ \Theta \|^\theta \right] \right)^{1/\theta} \leq C^{1/\theta} \left( (1 + \sqrt{6}) \sqrt{\frac{2n(d_1 + d_2)}{d_1 d_2}} + (2 + \sqrt{2}) \sqrt{\log(d_1 + d_2)} \right)
\]

\[
\leq C^{1/\theta} 2(1 + \sqrt{6}) \sqrt{\frac{n(d_1 + d_2) + d_1 d_2 \log(d_1 + d_2)}{d_1 d_2}}.
\]

Plugging this back to (14), we have

\[
\left( \mathbb{E} \left[ \sup_{X \in \mathcal{S}} |\hat{L}_{\Omega,Y'}(X) - \mathbb{E} \hat{L}_{\Omega,Y'}(X)|^\theta \right] \right)^{1/\theta} \leq C^{1/\theta} 8(1 + \sqrt{6}) \alpha \rho_1^\gamma \sqrt{\frac{n(d_1 + d_2) + d_1 d_2 \log(d_1 + d_2)}{n}}.
\]

Therefore, (13) is upper bounded by

\[
C \left( \frac{8(1 + \sqrt{6})}{C_0} \right)^{2\log(d_1 + d_2)} \leq \frac{C}{d_1 + d_2},
\]

as soon as we choose \(C_0 \geq 8(1 + \sqrt{6})\sqrt{e}\). \(\square\)

### B.2 Proof for Theorem 11

We need the following result in our proof.

**Proposition 10.** Assume same conditions as in Theorem 11 but with a slightly more general assumption that \(\|M\|_\infty \leq \gamma\) in place of \(\|M\|_\infty \leq \alpha\). Then, with probability at least \(1 - C_1/(d_1 + d_2)\), the follow holds:

\[
d^2_{TV}(g(M), g(M)) \leq C_2 \rho_1^\gamma \alpha \sqrt{\frac{r(d_1 + d_2)}{n}} \sqrt{1 + \frac{(d_1 + d_2) \log(d_1 d_2)}{n}},
\]

where \(C_1\) and \(C_2\) are absolute constants.

**Proof.** For any matrix \(X \in \mathbb{R}^{d_1 \times d_2}\), we have

\[
\mathbb{E} \left[ \hat{L}_{\Omega,Y'}(X) - \hat{L}_{\Omega,Y'}(M) \right] = \mathbb{E} \left[ L_{\Omega,Y'}(X) - L_{\Omega,Y'}(M) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i,j} 1_{\{i,j\} \in \Omega} \left( 1_{\{y'_{ij} = 1\}} \log \frac{g(X_{ij})}{g(M_{ij})} + 1_{\{y'_{ij} = -1\}} \log \frac{1 - g(X_{ij})}{1 - g(M_{ij})} \right) \right]
\]

\[
= \mathbb{E} \left[ \sum_{i,j} \frac{n}{d_1 d_2} \left( g(M_{ij}) \log \frac{g(X_{ij})}{g(M_{ij})} + (1 - g(M_{ij})) \log \frac{1 - g(X_{ij})}{1 - g(M_{ij})} \right) \right]
\]

\[
= -nD(g(M)||g(X)).
\]
On the other hand, for the optimum $\hat{M}$, it holds that
\[
\hat{L}_{\Omega,Y'}(\hat{M}) - L_{\Omega,Y'}(M) = \mathbb{E} [L_{\Omega,Y'}(\hat{M}) - L_{\Omega,Y'}(M)] + \left(\hat{L}_{\Omega,Y'}(\hat{M}) - \mathbb{E} [L_{\Omega,Y'}(\hat{M})]\right)
+ \left(\mathbb{E} [L_{\Omega,Y'}(M) - L_{\Omega,Y'}(M)]\right)
\leq \mathbb{E} [\hat{L}_{\Omega,Y'}(X) - L_{\Omega,Y'}(M)] + 2 \sup_{X \in \mathcal{S}} \left|\hat{L}_{\Omega,Y'}(X) - \mathbb{E} [L_{\Omega,Y'}(X)]\right|,
\]
where we recall that $\mathcal{S}$ was defined in Lemma 3. Since $\hat{M}$ also maximizes $\hat{L}_{\Omega,Y'}(X)$, we obtain
\[
- \mathbb{E} [\hat{L}_{\Omega,Y'}(X) - L_{\Omega,Y'}(M)] \leq 2 \sup_{X \in \mathcal{S}} \mathbb{E} [\hat{L}_{\Omega,Y'}(X) - \mathbb{E} [L_{\Omega,Y'}(X)]].
\]

This together with (16) and Lemma 3 imply that
\[
D(g(M)||g(\hat{M})) \leq 2C_0C_0^\alpha \rho_\alpha^+ \sqrt{\frac{r(d_1 + d_2)}{n}} \sqrt{1 + \frac{(d_1 + d_2) \log(d_1d_2)}{n}}
\]
holds with probability at least $1 - C_1/(d_1 + d_2)$. Since the Hellinger distance is upper bounded by the KL divergence, we complete the proof. \hfill \square

Now we are in the position to prove Theorem 1. In fact, Theorem 1 follows immediately from Prop. 10 and Lemma 7.

B.3 Proof for Theorem 2

Proof. Without loss of generality, suppose that $d_1 \leq d_2$. Let
\[
e^2 = \min \left\{ \frac{1}{1024}, C \alpha \sqrt{\frac{\rho_\alpha^+ r d_2}{n}} \right\}.
\]
Pick
\[
\frac{4\sqrt{2}}{\alpha} \leq \nu \leq \frac{8e}{\alpha}.
\]
It is easy to see that
\[
\frac{r \alpha^2}{64e^2} \leq \frac{\nu}{\nu^2} \leq \frac{r \alpha^2}{32e^2}.
\]
The length of this interval is $\frac{r \alpha^2}{64e^2}$, which is larger than 1 since $\alpha \geq 1$, $r \geq 16$ and $e^2 \leq 1/1024$. Hence, it is possible to pick a proper $\nu$ such that $\frac{\nu}{\nu^2}$ is an integer. Also, the assumption that $e^2 \geq O(r \alpha^2/d_1)$ suggests $r/\nu^2 \leq d_1$. Hence we have found an appropriate $\nu$ for Lemma 9.

Let $\mathcal{X}_\alpha/2,\nu$ be a set that satisfies all the properties in Lemma 9 with parameter $\alpha/2$. Let
\[
\mathcal{X} = \left\{ X' + \alpha \left(1 - \frac{\nu}{2}\right) U : X' \in \mathcal{X}_\alpha/2,\nu \right\},
\]
where all the entries of $U$ equal one.

First, we verify that each component in $\mathcal{X}$ satisfies (A2) and (A3). It is easy to see that for any $X \in \mathcal{X}$, $|X_{ij}|$ either equals $\alpha$ or $(1 - \nu)\alpha$, i.e., $\|X\|_\infty \leq \alpha$ since $\nu < 1$. In addition,
\[
\left\|X' + \alpha \left(1 - \frac{\nu}{2}\right) U\right\|_* \leq \|X'\|_* + \alpha \left(1 - \frac{\nu}{2}\right) \|U\|_* \leq \frac{\alpha}{2} \sqrt{rd_1d_2} + \alpha \left(1 - \frac{\nu}{2}\right) \|U\|_*.
\]
Since $\nu \in (0,1)$ and $r \geq 16$, we have $2 - \nu \leq \sqrt{r}$, which together with $\|U\|_* = \sqrt{d_1d_2}$ imply that $\|X\|_* \leq \alpha \sqrt{rd_1d_2}$ for all $X \in \mathcal{X}$. 

Robust Matrix Completion from Quantized Observations
We prove the theorem by showing that its converse is false. That is, suppose that there exists an algorithm such that for any \( M \in \mathcal{X} \) (which satisfies (A2) and (A3)), with probability at least 1/4, its output \( \hat{X} \) satisfies
\[
\frac{1}{d_1d_2} \left\| \hat{X} - M \right\|_F^2 < \epsilon^2. \tag{17}
\]
Let \( X^* \in \mathcal{X} \) be the closest member to \( \hat{X} \). For any \( \tilde{X} \notin M \in \mathcal{X} \), it follows that
\[
\left\| \tilde{X} - \hat{X} \right\|_F \geq \left\| \tilde{X} - M \right\|_F - \left\| \hat{X} - M \right\|_F > 2\epsilon \sqrt{d_1d_2} - \epsilon \sqrt{d_1d_2} = \epsilon \sqrt{d_1d_2}, \tag{18}
\]
where the last inequality follows from (17) and the fact that for any \( X, \tilde{X} \in \mathcal{X} \),
\[
\left\| X - \tilde{X} \right\|_F^2 \geq \frac{\alpha^2 \nu^2 d_1d_2}{8} \geq 4d_1d_2 \nu^2.
\]
The first inequality above uses the third property in Lemma 8 and the second inequality follows from our choice of \( \nu \).

On the other hand, since \( X^* \) is the closest one to \( \hat{X} \), we have
\[
\left\| X^* - \hat{X} \right\|_F \leq \left\| M - \hat{X} \right\|_F \leq \epsilon \sqrt{d_1d_2}. \tag{19}
\]
Combining (18) and (19), we obtain
\[
\left\| X^* - \hat{X} \right\|_F < \left\| \tilde{X} - \hat{X} \right\|_F, \forall \tilde{X} \neq M,
\]
which implies \( X^* = M \). Since (17) holds with probability at least 1/4,
\[
\Pr(X^* \neq M) \leq \frac{3}{4}. \tag{20}
\]
From a variant of Fano’s inequality,
\[
\Pr(X^* \neq M) \geq 1 - \frac{1 + d_1d_2 \max_{X \neq \tilde{X}} D(Y_{ij} | X || Y_{ij} | \tilde{X})}{\log |\mathcal{X}|} \tag{21}
\]
Denote
\[
D = d_1d_2D(Y_{ij} | X || Y_{ij} | \tilde{X}) = \sum_{(i,j) \in \Omega} D(Y_{ij} | X_{ij} || Y_{ij} | \tilde{X}_{ij}).
\]
For each \((i, j) \in \Omega, D(Y_{ij} | X_{ij} || Y_{ij} | \tilde{X}_{ij})\) is either 0, \( D(g(\alpha)||g(\alpha')) \) or \( D(g(\alpha)||g(\alpha')) \), where \( \alpha' = (1 - \nu)\alpha \) and we recall that \( X_{ij}, \tilde{X}_{ij} \) can only take value from \{\alpha, \alpha'\}. It thus follows from Lemma 8 that
\[
D(Y_{ij} | X_{ij} || Y_{ij} | \tilde{X}_{ij}) \leq \frac{(g(\alpha) - g(\alpha'))^2}{g(\alpha')(1 - g(\alpha'))},
\]
since \( \alpha' < \alpha \). Now using the mean value theorem, we obtain
\[
D \leq n(g'(\theta))^2 \frac{(\alpha - \alpha')^2}{g(\alpha')(1 - g(\alpha'))}, \text{ for some } \theta \in [\alpha', \alpha].
\]
As we assumed that \( \nabla g(x) \) is decreasing in \((0, +\infty)\), we get
\[
D \leq \frac{n(\nu \alpha)^2}{\rho_{\alpha'}} \leq 64n\epsilon^2 \frac{\rho}{\rho_{\alpha'}}.
\]
Due to the construction, the cardinality of $\mathcal{X}$ equals to that of $\mathcal{X}'_{\alpha/2,\nu}$. Hence, combining (20) and (21), we can show

$$\frac{1}{4} \leq \frac{D + 1}{\log |\mathcal{X}|} \leq \frac{16\nu^2}{rd_2} \left( \frac{64n\epsilon^2}{\rho_{\alpha'}} + 1 \right) \leq \frac{1024\epsilon^2}{\alpha^2rd_2} \left( \frac{64n\epsilon^2}{\rho_{\alpha'}} + 1 \right).$$

(22)

Note that when $64n\epsilon^2 \leq \rho_{\alpha'}$, we have

$$\frac{1}{4} \leq 1024 \times \frac{2048\epsilon^2}{\alpha^2rd_2},$$

implying $\alpha^2rd_2 \leq 8$ due to the definition of $\epsilon$. This contradicts our assumption that $\alpha^2rd_2 \geq C_0$ if we specify $C_0 > 8$.

When $64n\epsilon^2 > \rho_{\alpha'}$, then (22) suggests

$$\frac{1}{4} \leq \frac{1024 \times 128 \times n\epsilon^4}{\rho_{\alpha'}^2 \alpha^2rd_2},$$

which gives

$$\epsilon^2 > \frac{\alpha \sqrt{\rho_{\alpha'}} \sqrt{\frac{rd_2}{n}}}{1024}.$$ 

Picking $C_2 = 1/1024$ in the definition of $\epsilon$ and noting $\rho_{\alpha'} \geq \bar{\rho}_{0.75\alpha}$ yields a contradiction. Therefore, (17) fails to hold with probability at least $3/4$. 

$\Box$