A. Appendix

A.1. Proof of Lemma 4

Recall that $\mathcal{V} = \{\Theta : \mathcal{R}(\Theta_{\overline{\mathcal{M}}^\perp}) \le 3\mathcal{R}(\Theta_{\overline{\mathcal{M}}})\}$. To prove Lemma 4, consider the nuclear norm ball $S_\mathcal{R}(t) = \{\Delta : \mathcal{R}(\Delta) \le t\}$. We

1. Show that, $P\left(\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \left|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\right| > \frac{8t}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + \frac{k_1}{2}t\sqrt{\frac{n\log n}{|\Omega|}}\right) \le C_1' e^{-C_2' t^2 n\log n}$, where $\kappa(n, |\Omega|)$ is a quantity that depends only on the dimensions n and $|\Omega|$. This is done by:
   (a) Bounding the expectation, $\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|\Big]$
   (b) Showing an exponential decay of the tail.
2. Then use a peeling argument (Raskutti et al., 2010) to derive at the result in Lemma 4.

A.1.1. Bounding $\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|\Big]$

Note that $\forall\ \Delta \in \mathcal{E}$, $\mathbb{E}[\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2] = \|\Delta\|_F^2 = 1$. Thus, by using standard symmetrization argument (Lemma 6.3 of (Ledoux & Talagrand, 1991), with a Rademacher sequence, $\{\epsilon_{ij}, \forall\ ij\in\Omega\}$, we have:

$$\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|\Big] \le \frac{2mn}{|\Omega|}\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)}\Big|\sum_{ij\in\Omega}\epsilon_{ij}\Delta_{ij}^2\Big|\Big] \tag{1}$$

Also, $\forall\Delta\in\mathcal{E}$, $\phi_{ij}(\Delta) \triangleq \frac{\Delta_{ij}^2}{2\sup_{\Delta\in\mathcal{E}}\|\Delta\|_{\max}}$ is a contraction, and $\forall\Delta\in\mathcal{E}$, $\|\Delta\|_{\max} = \frac{\alpha_{\text{sp}}(\Delta)}{\sqrt{mn}} \le \frac{1}{c_0\Psi(\overline{\mathcal{M}})\sqrt{mn}}\sqrt{\frac{|\Omega|}{n\log n}}$. Thus, using Theorem 4.12 of (Ledoux & Talagrand, 1991) in Equation 1, we have:

$$\begin{aligned}\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|\Big] &\le \frac{8}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|}{n\log n}}\mathbb{E}\left[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)}\Big|\frac{\sqrt{mn}}{|\Omega|}\Big\langle\sum_{ij\in\Omega}\epsilon_{ij}e_ie_j^*, \Delta\Big\rangle\Big|\right] \\ &\overset{(a)}{\le} \frac{8t}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|}{n\log n}}\mathbb{E}\left[\frac{\sqrt{mn}}{|\Omega|}\mathcal{R}^*\Big(\sum_{ij\in\Omega}\epsilon_{ij}e_ie_j^*\Big)\right]\end{aligned} \tag{2}$$

where (a) follows from Cauchy–Schwartz and as $\mathcal{R}(\Delta)\le t$. Note that $\mathcal{R}^*\Big(\sum_{ij\in\Omega}\epsilon_{ij}e_ie_j^*\Big)$ is independent of Δ and depends only on n and $|\Omega|$. Let $\kappa(n,|\Omega|) \ge \mathbb{E}\left[\frac{\sqrt{mn}}{|\Omega|}\mathcal{R}^*\Big(\sum_{ij\in\Omega}\epsilon_{ij}e_ie_j^*\Big)\right]$ be a suitable upper bound.

$$\mathbb{E}\Big[\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|\Big] \le \frac{8t}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} \tag{3}$$

A.1.2. Tail Behavior

Let $G_t(\Omega) \triangleq \sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big|$. Let $\Omega' \subset [m]\times[n]$ be another set of indices that differ from Ω in exactly one element. We then have:

$$\begin{aligned}G_t(\Omega) - G_t(\Omega') &= \sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big| - \sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)} \Big|\frac{mn}{|\Omega|}\sum_{kl\in\Omega'}\Delta_{kl}^2 - 1\Big| \\ &\le \frac{mn}{|\Omega|}\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)}\left(\Big|\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big| - \Big|\sum_{kl\in\Omega'}\Delta_{kl}^2 - 1\Big|\right) \le \frac{mn}{|\Omega|}\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)}\left(\Big|\sum_{ij\in\Omega}\Delta_{ij}^2 - \sum_{kl\in\Omega'}\Delta_{kl}^2\Big|\right) \\ &\le \frac{2mn}{|\Omega|}\sup_{\Delta\in\mathcal{E}\cap S_\mathcal{R}(t)}\|\Delta\|_{\max}^2 \le \frac{2}{c_0^2\Psi^2(\overline{\mathcal{M}})n\log n}\end{aligned} \tag{4}$$

By similar arguments on $G_t(\Omega')-G_t(\Omega)$, we conclude that $|G_t(\Omega)-G_t(\Omega')| \leq \frac{2}{c_0^2\Psi^2(\overline{\mathcal{M}})n\log n}$. Therefore, using Mc Diarmid's inequality, we have $P(|G_t(\Omega)-\mathbb{E}[G_t(\Omega)]|>\delta) \leq 2\exp\left(-\frac{c_0^4\delta^2\Psi^4(\overline{\mathcal{M}})n^2\log^2 n}{2|\Omega|}\right)$. Using $\delta = \frac{k_1}{2}t\sqrt{\frac{n\log n}{|\Omega|}}$ for appropriate constant $\frac{k_1}{2}$,

$$P\Big(|G_t(\Omega)-\mathbb{E}[G_t(\Omega)]| > \frac{k_1}{2}t\sqrt{\frac{n\log n}{|\Omega|}}\Big) \leq 2\exp\left(-c_1' t^2 n\log n\left(\frac{\Psi^2(\overline{\mathcal{M}})n\log n}{|\Omega|}\right)^2\right)$$

$$P\left(G_t(\Omega) > \frac{8t}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + \frac{k_1}{2}t\sqrt{\frac{n\log n}{|\Omega|}}\right) \leq 2\exp(-c_1'' t^2 n\log n)$$

A.1.3. Peeling Argument

Consider the following sets, $S_\ell = \{\Delta \in \mathcal{E} : 2^{\ell-1}\Psi_{\min} \leq \mathcal{R}(\Delta) \leq 2^\ell\Psi_{\min}\}$, for all (integers) $\ell \geq 1$. Recall that $\Psi_{\min} = \inf_{X\setminus\{0\}} \frac{\mathcal{R}(X)}{\|X\|_F}$. Since, $\forall \Delta \in \mathcal{E}$, $\mathcal{R}(\Delta) \geq \Psi_{\min}\|\Delta\|_F = \Psi_{\min}$, for each $\Delta \in \mathcal{E}$, $\Delta \in S_\ell$ for some $\ell \geq 1$. Further, if for some $\Delta \in \mathcal{E}$, $\left|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\right| > \frac{16\mathcal{R}(\Delta)}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + k_1\mathcal{R}(\Delta)\sqrt{\frac{n\log n}{|\Omega|}}$, then for some ℓ:

$$\begin{aligned}\Big|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\Big| &> \frac{16(2^{\ell-1})\Psi_{\min}}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + k_1 2^{\ell-1}\Psi_{\min}\sqrt{\frac{n\log n}{|\Omega|}} \\ &= \frac{8(2^\ell\Psi_{\min})}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + \frac{k_1}{2}2^\ell\Psi_{\min}\sqrt{\frac{n\log n}{|\Omega|}}\end{aligned} \tag{5}$$

Thus,

$$\begin{aligned}&P\left(\sup_{\Delta\in\mathcal{E}}\left|\frac{mn}{|\Omega|}\sum_{ij\in\Omega}\Delta_{ij}^2 - 1\right| > \frac{16\mathcal{R}(\Delta)}{c_0\Psi(\overline{\mathcal{M}})}\sqrt{\frac{|\Omega|\kappa^2(n,|\Omega|)}{n\log n}} + k_1\|\Delta\|_*\sqrt{\frac{n\log n}{|\Omega|}}\right) \\ &\leq \sum_{\ell=1}^\infty P\left(G_{2^\ell}(\Omega) > \frac{8(2^\ell\Psi_{\min})}{c_0\Psi(\overline{\mathcal{M}})} + \frac{k_1}{2}2^\ell\Psi_{\min}\sqrt{\frac{n\log n}{|\Omega|}}\right) \leq \sum_{\ell=1}^\infty 2\exp(-c_1'' 2^{2\ell}\Psi_{\min}^2 n\log n) \\ &\overset{(a)}{\leq} \sum_{\ell=1}^\infty 2\exp(-2\log(2)\, c_1''\ell\Psi_{\min}n\log n) \leq \frac{2e^{-c_2''\Psi_{\min}n\log n}}{1-e^{-c_2''\Psi_{\min}n\log n}} \approx C_1 e^{-C_2\Psi_{\min}n\log n}\end{aligned} \tag{6}$$

where (a) follows as $x \geq \log x$ for $x > 1$, and the last step holds for $1 \gg e^{-c_2''\Psi_{\min}n\log n}$.

A.2. Proof of Lemma 2

Let $\widehat{\Delta} = \widehat{\Theta} - \Theta^*$.

$$\mathcal{R}(\widehat{\Theta}) = \mathcal{R}(\Theta^* + \widehat{\Delta}_{\overline{\mathcal{M}}} + \widehat{\Delta}_{\overline{\mathcal{M}}^\perp}) \geq \mathcal{R}(\Theta^* + \widehat{\Delta}_{\overline{\mathcal{M}}^\perp}) - \mathcal{R}(\widehat{\Delta}_{\overline{\mathcal{M}}}) = \mathcal{R}(\Theta^*) + \mathcal{R}(\widehat{\Delta}_{\overline{\mathcal{M}}^\perp}) - \mathcal{R}(\widehat{\Delta}_{\overline{\mathcal{M}}}) \tag{7}$$

The above inequalities hold due to triangle inequality, and decomposability of $\mathcal{R}$ over $\Theta^* \in \mathcal{M}$ and $\Delta_{\overline{\mathcal{M}}^\perp} \in \overline{\mathcal{M}}^\perp$.

$$\begin{aligned}&\frac{mn}{|\Omega|}\sum_{(i,j)\in\Omega} B_G(\widehat{\Theta}_{ij}, \Theta_{ij}^*) = \frac{mn}{|\Omega|}\left(\sum_{(i,j)\in\Omega} G(\widehat{\Theta}_{ij}) - X_{ij}\widehat{\Theta}_{ij} - G(\Theta_{ij}^*) + X_{ij}\Theta_{ij}^*\right) + \frac{mn}{|\Omega|}\langle\mathcal{P}_\Omega(X - g(\Theta^*), \widehat{\Delta}\rangle \\ &\overset{(a)}{\leq} \lambda\mathcal{R}(\Theta^*) - \lambda\mathcal{R}(\widehat{\Theta}) + \frac{mn}{|\Omega|}\mathcal{R}^*(\mathcal{P}_\Omega(X - g(\Theta^*)))\mathcal{R}(\widehat{\Delta}) \overset{(b)}{\leq} \lambda\mathcal{R}(\Delta_{\overline{\mathcal{M}}}) - \lambda\mathcal{R}(\Delta_{\overline{\mathcal{M}}^\perp}) + \frac{\lambda}{2}\mathcal{R}(\Delta_{\overline{\mathcal{M}}} + \Delta_{\overline{\mathcal{M}}^\perp}) \\ &\overset{(c)}{\leq} \frac{3\lambda}{2}\mathcal{R}(\Delta_{\overline{\mathcal{M}}}) - \frac{\lambda}{2}\mathcal{R}(\Delta_{\overline{\mathcal{M}}^\perp}) \leq \frac{3\lambda\Psi(\overline{\mathcal{M}})}{2}\|\Theta^* - \widehat{\Theta}_{\overline{\mathcal{M}}}\|_F \leq \frac{3\lambda\Psi(\overline{\mathcal{M}})}{2}\|\Theta^* - \widehat{\Theta}\|_F\end{aligned} \tag{8}$$

where (a) follows as $\widehat{\Theta}$ is the minimizer of Equation (6) and using Cauchy Schwartz, (b) follows from Equation 7 and using $\frac{mn}{|\Omega|}\mathcal{R}^*(\mathcal{P}_\Omega(X - g(\Theta^*)) \leq \frac{\lambda}{2}$, and (c) follows from triangle inequality. □

A.3. Ahlswede–Winter Matrix Bound

The Orlicz norm of a random matrix $Z \in \mathbb{R}^{m \times n}$ w.r.t to a convex, differentiable and monotonically increasing function, $\phi(x) : \mathbb{R}^{+} \to \mathbb{R}$ as follows:

$$\|Z\|_{\phi} \triangleq \inf\{t \geq 0 : \mathbb{E}\left[\phi\left(|\langle Z, Z' \rangle|/t\right)\right] \leq 1, \\ \forall\, Z' \in \mathbb{R}^{m \times n},\ and\ Z'_{ij} \in [0,1]\}$$

Lemma 1 (Ahlswede-Winter Matrix Bound). *Let $Z^{(1)}, Z^{(2)}, \ldots, Z^{(K)}$ be random matrices of dimensions $m \times n$. Let $\|Z^{(i)}\|_{\phi} \leq M$, $\forall i$. Further, $\sigma_i^2 = \max\{\|\mathbb{E}[Z^{(i)^T} Z^{(i)}]\|_2, \|\mathbb{E}[Z^{(i)} Z^{(i)^T}]\|_2\}$, and $\sigma^2 = \sum_{i=1}^{K} \sigma_i^2$, then:*

$$P\left(\|\sum_{i=1}^{K} Z^{(i)}\|_2 \geq t\right) \leq mn \max\left\{e^{-\frac{t^2}{4\sigma^2}}, e^{-\frac{t}{2M}}\right\}$$

The above lemma is an extension noted by (Vershynin, 2009) (Theorem 1 and a later remark) for the matrix bounds resulting from (Ahlswede & Winter, 2002).

B. Additional Experimental Results

We provide the additional experimental results where we compare the error of the estimate in the parameter space. We plot the results first against the proportion of the total entries sampled, $\frac{|\Omega|}{mn}$ (Figure on left), and then against the "normalized" sample size, $\frac{|\Omega|}{rn \log n}$ (Figures on right). We observe trends similar to those observed in Section 5. Again, we find that the curves (for different n) given the "normalized" sample size, align and converge (left), corroborating the theoretical results. Note that, the curves do not align when plotted against, unnormalized sample size (right). Further, as with errors in observation space, with $|\Omega| > 1.5rn \log n$ samples, the errors parameter space also decay to a sufficiently small value.

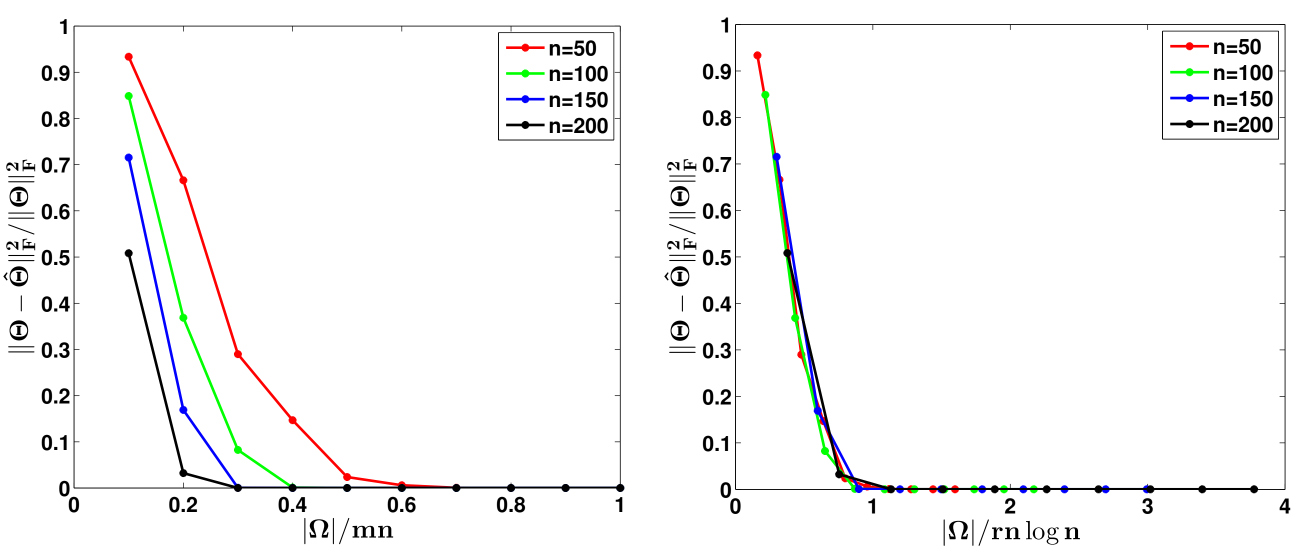

Figure 1. Parameter Error when measured (a) against proportion of the sampled values, and (b) against the 'normalized" sample size, when the distribution of the observations $P(X|\Theta^*)$, is Gaussian

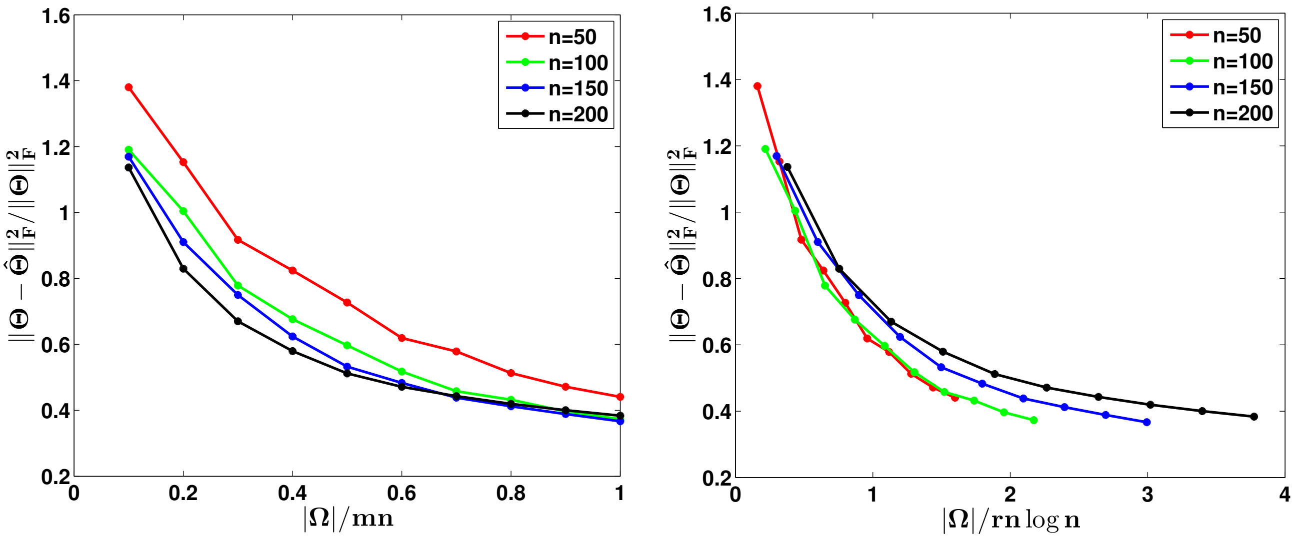

Figure 2. Parameter Error when measured (a) against proportion of the sampled values, and (b) against the 'normalized" sample size, when the distribution of the observations $P(X|\Theta^*)$, is Bernoulli

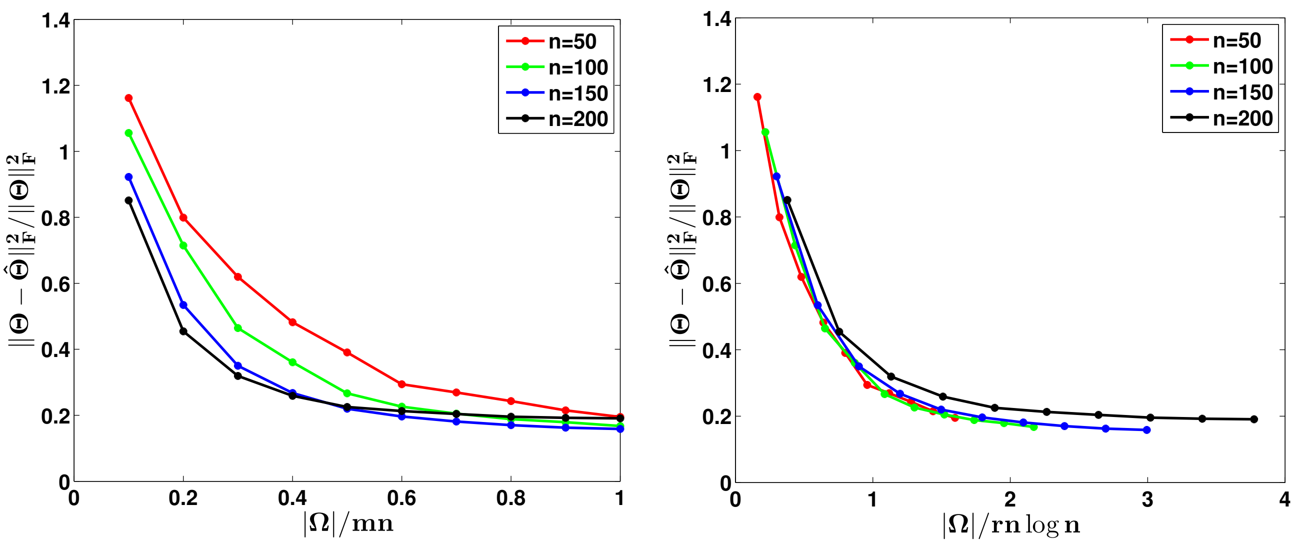

Figure 3. Parameter Error when measured (a) against proportion of the sampled values, and (b) against the 'normalized" sample size, when the distribution of the observations $P(X|\Theta^*)$, is Binomial

References

Ahlswede, R. and Winter, A. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

Ledoux, M. and Talagrand, M. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 1991.

Raskutti, G., Wainwright, M. J, and Yu, B. Restricted eigenvalue properties for correlated gaussian designs. *The Journal of Machine Learning Research*, 99:2241–2259, 2010.

Vershynin, R. A note on sums of independent random matrices after ahlswede-winter. *Lecture Notes*, 2009.