7. Appendix

7.1. Noiseless GRC

Proof of lemma 3

Proof. For any two matrices $A \in \mathbb{R}_{n_1 \times n_2}$ and $B \in \mathbb{R}_{m_1 \times m_2}$ we define the Kronecker product as a matrix in $\mathbb{R}_{n_1 m_1 \times n_2 m_2}$:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & . & a_{1n_2}B \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{n_1 1}B & a_{n_1 2}B & . & a_{n_1 n_2}B \end{pmatrix} \tag{19}$$

The form of the transformation $V^T A^{(C)}$ in vectors is $(I_n \otimes V^T)vec(A^{(C)})$ where the distribution of $(I_n \otimes V^T)vec(A^{(C)})$ is also Gaussian with covariance matrix

$$\begin{aligned} &(I_n \otimes V^T)cov(vec(A^{(C)}))(I_n \otimes V^T)^T = \\ &\sigma^2(I_n \otimes V^T)(I_n \otimes V^T)^T = \sigma^2 I_r \otimes I_n \end{aligned} \tag{20}$$

□

7.2. Gradient Descent

The gradient descent stage is performed directly in the space of rank r matrices, using the decomposition $\hat{X}$=WS where $W \in \mathbb{R}_{n_1 \times r}$ and $S \in \mathbb{R}_{r \times n_2}$ and computing the gradient of the loss as a function of W, S (see Appendix in (21,22).

$$L(W,S) \ = ||A^{(R)}WS - B^{(R)}||_F^2 + ||WSA^{(C)} - B^{(C)}||_F^2 \tag{21}$$

We want to find the minimum of (21) using gradient descent our problem is that the loss L isn't convex and therefore we can't promise the gradient descent will converge to a global optimum. but if we got $\hat{X}$ (the output of SVLS) as starting point we might get that the gradient descent converges to the global minimum since we start close to it.

The derivative of L is (using chain rule)

$$\begin{aligned} \frac{\partial L}{\partial W} &= A^{(R)^T}(A^{(R)}WS - B^{(R)})S^T + (WSA^{(C)} - B^{(C)})A^{(C)^T}S^T \\ \frac{\partial L}{\partial S} &= W^T A^{(R)^T}(A^{(R)}WS - B^{(R)}) + W^T(WSA^{(C)} - B^{(C)})A^{(C)^T} \end{aligned} \tag{22}$$

7.3. Proof of RCMC

We give here some useful lemmas to prove lemma 4 we start with lemma from (Candès & Romberg, 2007).

Lemma 5. *If y_i is a family of vectors in R^d and r_i is a $0/1$ Bernoulli sequence or random variables with $P(r_i = 1) = p$, then*

$$E(p^{-1}||\Sigma_i(r_i - p)y_i \otimes y_i||) < C\sqrt{\frac{log(d)}{p}}max_i||y_i|| \tag{23}$$

for some numerical C provided that the right hand side is less than 1.

For lemma 4 we uses a result from large deviations theory that proved by Talagrand (Talagrand, 1996).

Theorem 4. *Let $Y_1 ... Y_n$ be a sequence of independent random variables taking values in a Banach space and define*

$$Z = sup_{f \in F}\sum_{i=1}^{n} f(Y_i) \tag{24}$$

where F is a real countable set of functions such that if $f \in F$ than $-f \in F$.

Assume that $|f| \leq B$ and $E(f(Y_i)) = 0$ for every $f \in F$ and $i \in [n]$. Then there exists a constant C such that for every $t \geq 0$

$$P(|Z - E(Z)| \geq t) \leq 3exp\left(\frac{-t}{CB}log(1 + \frac{t}{\sigma + Br})\right) \tag{25}$$

where $\sigma = sup_{f \in F} \sum_{i=1}^{n} E(f^2(Y_i))$.

Theorem (4) helps us in the next lemma which taken from Theorem 4.2 in (Candès & Recht, 2009). We bring here the proof in our notations for convenience.

Lemma 6. *Let $Y_i = p^{-1}(r_i - p)P_U(e_i) \otimes P_u(e_i)$, $Y = \sum_{i=1}^{n} Y_i$* and $Z = ||Y||_2$. Suppose $E(Z) \leq 1$. Then for every $\lambda > 0$ we have

$$p(|Z - E(Z)| \geq \lambda\sqrt{\frac{\mu r log(n)}{pn}}) \leq 3exp(-\gamma min(\lambda^2 log(n), \lambda\sqrt{\frac{pnlog(n)}{\mu r}}) \tag{26}$$

for some positive constant γ .

Proof. We know that $Z = ||Y||_2 = sup_{f_1,f_2} < f_1, Y f_2 >= sup_{f_1,f_2} \sum_i < f_1, Y_i f_2 >$, where the supremum is taken over a countable set of unit vectors $f_1, f_2 \in F_V$. Let F be the set of all functions f such that $f(Y) =< f_1, Y f_2 >$ for some unit vectors $f_1, f_2 \in F_V$. For every $f \in F$ and $i \in [n]$ we have$E(f(Y_i)) = 0$. From the incoherence of U we conclude that

$$|f(Y_i)| = p^{-1}|r_i - p|| < f_1, P_U(e_i) > || < P_U(e_i), f_2 > | \leq p^{-1}||P_U(e_i)||^2 \leq p^{-1}\frac{r}{n}\mu. \tag{27}$$

In addition

$$\begin{aligned} Ef^2(Y_i) = p^{-1}(1 - p) < f_1, P_U(e_i) >^2 < P_U(e_i), f_2 >^2 \leq \\ p^{-1}||P_U(e_i)||^2| < P_U(e_i), f_2 > | \leq p^{-1}\frac{r}{n}\mu| < P_U(e_i), f_2 > |^2. \end{aligned} \tag{28}$$

Since $\sum_i | < P_U(e_i), f_2 > |^2 = \sum_i | < e_i, P_U(f_2) > |^2 = ||P_U(f_2)||^2 \leq 1$. we get $\sum_i Ef^2(Y_i) \leq p^{-1}\frac{r}{n}\mu$.

We can take $B = 2p^{-1}\frac{r}{n}\mu$ and $t = \lambda\sqrt{\frac{\mu r log(n)}{pn}}$ and from Theorem (4)

$$p(|Z - E(Z)| \geq t) \leq 3exp(\frac{-t}{KB}log(1 + \frac{t}{2})) \leq 3exp\left(\frac{-tlog(2)}{KB}min(1, \frac{t}{2})\right) \tag{29}$$

where the last inequality is due to the fact that for every $u > 0$ we have $log(1 + u) \geq log(2)min(1, u)$. Taking $\gamma = -log(2)/K$ finish our proof. □

We now prove lemma 4

Proof. 4 Decompose any vector $w \in R^n$ as $w = \sum_{i=1}^{n} < w, e_i > e_i$. Therefore $P_U(w) = \sum_i < P_U(w), e_i > e_i = \sum_i < w, P_U(e_i) > e_i$. Hence

$$P_{A(R)^T}P_U(w) = \sum_i r_i < w, P_U(e_i) > e_i \Longrightarrow P_U P_{A(R)^T}P_U(w) = \sum_i r_i < w, P_U(e_i) > P_U(e_i) \tag{30}$$

In other words the matrix $P_U P_{A(R)^T} P_U$ is

$$P_U P_{A(R)^T} P_U = \sum r_i P_U(e_i) \otimes P_U(e_i) \tag{31}$$

U is $\mu-$incoherent, thus $max_{i\in[n]}||P_U(e_i)|| \leq \sqrt{\frac{r\mu}{n}}$, hence from 5 we have $E(p^{-1}||P_U P_{A(R)^T}P_U - pP_U||_2) < C\sqrt{\frac{log(n)r\mu}{pn}} \leq 1$ for p large enough.

Take $\lambda = \sqrt{\frac{\beta}{\gamma}}$ where γ as in Theorem 4 and get that if $p > \frac{\mu log(n)r\beta}{n\gamma}$ then from lemma 6 with probability of at least $1 - 3n^{-\beta}$ we have $Z \leq C\sqrt{\frac{log(n)r\mu}{pn}} + \frac{1}{\sqrt{\gamma}}\sqrt{\frac{log(n)r\mu\beta}{pn}}$. Taking $C_R = C + \frac{1}{\sqrt{\gamma}}$ finishes our proof. □

7.4. Proof for Noisy GRC

The proof of Theorem 3 is using strong concentration results on the largest and smallest singular values of $n \times k$ matrix with i.i.d Gaussian entries

Theorem 5. *(Szarek, 1991)Let $A \in \mathbb{R}_{n\times k}$ be a random matrix $A \overset{i.i.d.}{\sim} N(0, \frac{1}{n})$. Then, its largest and smallest singular values obey:*

$$P(\sigma_1(A) > 1 + \frac{\sqrt{k}}{\sqrt{n}} + t) \leq e^{-nt^2/2},$$

$$P(\sigma_k(A) \leq 1 - \frac{\sqrt{k}}{\sqrt{n}} - t) \leq e^{-nt^2/2} \tag{32}$$

Corrolary 2. *Let $A \in \mathbb{R}_{n\times k}$ where $n \geq 4k$ be random matrices with i.i.d $N(0,1)$ entries, and let$A^\dagger$ be the pseudoinverse of A. Then*

$$P\left(||A^\dagger||_2 \leq \frac{6}{\sqrt{n}}\right) > 1 - e^{-n/18} \tag{33}$$

Proof. Since $A^\dagger$ is the pseudoinverse of A, $||A^\dagger||_2 = \frac{1}{\sigma_k(A)}$, from Theorem (5) $\sigma_k(A) \geq \sqrt{n} - \sqrt{k} - t\sqrt{n}$ with probability $1 - e^{nt^2/2}$. Therefore, if we take $n \geq 4k$ and $t = \frac{1}{3}$ we get

$$P\left(||A^\dagger||_2 \leq \frac{6}{\sqrt{n}}\right) = P\left(\sigma_k(A) \geq \frac{\sqrt{n}}{6}\right) > 1 - e^{-n/18}. \tag{34}$$

□

We also use the following lemma from (Shalev-Shwartz & Ben-David, 2014):

Lemma 7. *Let Q to be a finite set of vectors in $\mathbb{R}^n$, let $\delta \in (0,1)$ and k be an integer such that*

$$\epsilon = \sqrt{\frac{6log(2|Q|/\delta)}{k}} \leq 3 \tag{35}$$

Let $A \in \mathbb{R}_{k\times n}$ be a random matrix with $A \overset{i.i.d.}{\sim} N(0, \frac{1}{k})$. Then,

$$P\left(max_{x\in Q}\left|\frac{||Ax||^2}{||x||^2} - 1\right| \leq \epsilon\right) > 1 - \delta \tag{36}$$

Lemma 7 is a direct result of Johnson-Lindenstrauss lemma (Dasgupta & Gupta, 2003) applied to each vector in Q and using the union bound . Representing the vectors in Q as a matrix, the lemma shows that $A^{(R)}, A^{(C)}$ preserve matrix Frobenius norm in high probability, which is a weaker property than the RIP which holds for *any* low-rank matrix.

To prove Theorem 3, we first represent $||X - \hat{X}||_F$ as a sum three parts (lemma 8), and then give probabilistic upper bounds to each of the parts. We define $A^{(R)}_{\hat{U}} = A^{(R)}\hat{U}$ and $A^{(C)}_{V^T} = V^T A^{(C)}$. From lemma 3

$A^{(R)}_{\hat{U}}, A^{(C)}_{V^T} \overset{i.i.d.}{\sim} N(0,1)$, hence $rank(A^{(R)}_{\hat{U}}) = rank(A^{(C)}_{V^T}) = r$. We assume w.l.o.g that $\hat{X} = \hat{X}^{(R)}$ (see Algorithm 2). Therefore, from eq.(8) we have $\hat{X} = \hat{U}(A^{(R)^T}_{\hat{U}} A^{(R)}_{\hat{U}})^{-1} A^{(R)^T}_{\hat{U}} B^{(R)}$.

We denote by $A^{(R)^\dagger}_{\hat{U}} = (A^{(R)^T}_{\hat{U}} A^{(R)}_{\hat{U}})^{-1} A^{(R)^T}_{\hat{U}}$ and $A^{(C)^\dagger}_{V^T} = A^{(C)^T}_{V^T}(A^{(C)}_{V^T} A^{(C)^T}_{V^T})^{-1}$ the Moore-Penrose pseudo-inverse of $A^{(R)}_{\hat{U}}$ and $A^{(C)}_{V^T}$, respectively. We next prove the following lemma

Lemma 8. *Let $A^{(R)}$ and $A^{(C)}$ be as in the GRC model and $Z^{(R)}, Z^{(C)}$ be some noise. Let $\hat{X}$ be the output of SVLS. Then:*

$$||X - \hat{X}||_F \leq \mathbf{I} + \mathbf{II} + \mathbf{III}$$

where:

$$\mathbf{I} \equiv ||(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F \tag{37}$$

$$\mathbf{II} \equiv ||\hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F \tag{38}$$

$$\mathbf{III} \equiv ||\hat{U}A^{(R)\dagger}_{\hat{U}}Z^{(R)}||_F \tag{39}$$

Proof. We represent $||X - \hat{X}||_F$ as follows

$$\begin{aligned} &||X - \hat{X}||_F = \\ &||X - \hat{U}(A^{(R)^T}_{\hat{U}}A^{(R)}_{\hat{U}})^{-1}A^{(R)^T}_{\hat{U}}(A^{(R)}X + Z^{(R)})||_F = \\ &||X - \hat{U}(A^{(R)^T}_{\hat{U}}A^{(R)}_{\hat{U}})^{-1}A^{(R)^T}_{\hat{U}}A^{(R)}X \\ &-\hat{U}(A^{(R)^T}_{\hat{U}}A^{(R)}_{\hat{U}})^{-1}A^{(R)}_{\hat{U}}Z^{(R)}||_F \leq \\ &||X - \hat{U}(A^{(R)^T}_{\hat{U}}A^{(R)}_{\hat{U}})^{-1}A^{(R)^T}_{\hat{U}}A^{(R)}X||_F + \mathbf{III} \end{aligned} \tag{40}$$

where we have used the triangle inequality. We next use the following equality

$$XA^{(C)}A^{(C)\dagger}_{V^T}V^T = U\Sigma V^T A^{(C)}A^{(C)\dagger}_{V^T}V^T = U\Sigma V^T = X \tag{41}$$

to obtain:

$$\begin{aligned} &||X - \hat{U}(A^{(R)^T}_{\hat{U}}A^{(R)}_{\hat{U}})^{-1}A^{(R)^T}_{\hat{U}}A^{(R)}X||_F = \\ &||(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})X||_F = \\ &||(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})XA^{(C)}A^{(C)\dagger}_{V^T}V^T||_F = \\ &||(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})B^{(C,0)}A^{(C)\dagger}_{V^T}||_F \end{aligned} \tag{42}$$

where the last equality is true because V is orthogonal.

Since $\hat{U}$ is a basis for $span(B^{(C)}_{(r)})$ there exist a matrix L such that $\hat{U}L = B^{(C)}_{(r)}$ and we get:

$$\begin{aligned} &(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})B^{(C)}_{(r)} = \\ &B^{(C)}_{(r)} - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)}\hat{U}L = \\ &B^{(C)}_{(r)} - \hat{U}L = 0 \end{aligned} \tag{43}$$

Therefore

$$\begin{aligned} &||(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})B^{(C,0)}A^{(C)\dagger}_{V^T}||_F = \\ &||(I_n - \hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)})(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F \leq \\ &||(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F + \\ &||\hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F = \mathbf{I} + \mathbf{II} \end{aligned} \tag{44}$$

Combining eq. (44) and eq. (40) gives the required result. □

We want to bound each of the three parts in the formula of lemma 8. We use the following claim:

Claim 1. $||B^{(C,0)} - B^{(C)}_{(r)}||_2 \leq 2||Z^{(C)}||_2$

Proof. We know that $||B^{(C)} - B^{(C)}_{(r)}||_2 \leq ||B^{(C)} - B^{(C,0)}||_2$ since $rank(B^{(C)}_{(r)}) = rank(B^{(C,0)})$, and $B^{(C)}_{(r)}$ is the closest rank r matrix to $B^{(C)}$ by definition. Therefore from the triangle inequality

$$\begin{aligned} ||(B^{(C,0)} - B^{(C)}_{(r)})||_2 \leq \\ ||B^{(C)} - B^{(C)}_{(r)}||_2 + ||B^{(C)} - B^{(C,0)}||_2 \leq \\ 2||B^{(C,0)} - B^{(C)}||_2 = 2||Z^{(C)}||_2 \end{aligned} \tag{45}$$

□

Now we are ready to prove Theorem 3. The proof uses the following inequalities for matrix norms: for any two matrices A, B (i) $||AB||_2 \leq ||A||_2||B||_2$, (ii) $||AB||_F \leq ||A||_F||B||_2$ and (iii) if $rank(A) \leqslant r$ then $||A||_F \leq \sqrt{r}||A||_2$.

Proof. We prove (probabilistic) upper bounds on the three terms appearing in lemma 8.

1. We have

$$rank\left((B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}\right) \leqslant rank\left(A^{(C)\dagger}_{V^T}\right) \leqslant r \tag{46}$$

Therefore

$$\mathbf{I} = ||(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F \leq \sqrt{r}||(B^{(C,0)} - B^{(C)}_{(r)})||_2||A^{(C)\dagger}_{V^T}||_2 \tag{47}$$

Since $A^{(C)}_{V^T} \overset{i.i.d.}{\sim} N(0,1)$, from Corollary 2 $||A^{(C)\dagger}_{V^T}||_2 \leq \frac{6}{\sqrt{k}}$ for $k \geq 4r$ with probability $1 - e^{-k/18}$, hence

$$\mathbf{I} \leq 4\sqrt{\frac{r}{k}}||(B^{(C,0)} - B^{(C)}_{(r)})||_2. \tag{48}$$

From Claim 1 we have bound on (37)

$$\mathbf{I} \leq C_1\sqrt{\frac{r}{k}}||Z^{(C)}||_2 \tag{49}$$

with probability $1 - e^{-c_1 k}$ for absolute constants C_1, c_1.

2. $\hat{U}$ is orthogonal and can be omitted from $\mathbf{II}$ without changing the norm. Applying inequality (ii) above twice, we get the inequality:

$$\mathbf{II} = ||\hat{U}A^{(R)\dagger}_{\hat{U}}A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})A^{(C)\dagger}_{V^T}||_F \leq ||A^{(R)\dagger}_{\hat{U}}||_2||A^{(R)\dagger}_{\hat{U}}A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})||_F||A^{(C)\dagger}_{V^T}||_2. \tag{50}$$

From Corollary 2 we know that for $k > 4r$ we have $||A^{(R)\dagger}_{\hat{U}}||_2 \leq \frac{4}{\sqrt{k}}$ and $||A^{(C)\dagger}_{V^T}||_2 \leq \frac{4}{\sqrt{k}}$, each with probability $> 1 - e^{-k/18}$. Therefore, with probability $> 1 - 2e^{-k/18}$

$$\mathbf{II} \leq \frac{16}{k}||A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})||_F. \tag{51}$$

$A^{(R)}$ and $B^{(C,0)} - B^{(C)}_{(r)}$ are independent and $rank(B^{(C,0)} - B^{(C)}_{(r)}) \leq 2r$. Therefore we can apply lemma 7 with k such that $\frac{k}{6} > log(2k) + \frac{k}{18}$ (this holds for $k \geq 40$) to get with probability $> 1 - 2e^{-k/18}$:

$$\mathbf{II} \leq \frac{16}{k}||A^{(R)}(B^{(C,0)} - B^{(C)}_{(r)})||_F \leq \frac{16\sqrt{2k}}{k}||(B^{(C,0)} - B^{(C)}_{(r)})||_F \leq 16\sqrt{4\frac{r}{k}}||(B^{(C,0)} - B^{(C)}_{(r)})||_2. \tag{52}$$

From eq. (51) and (52) together with Claim 1 we have constants C_2 and c_2 such that with probability $1 - 3e^{-ck}$

$$\mathbf{II} \leq C_2||Z^{(C)}||_2 \tag{53}$$

3. $rank(A_{\hat{U}}^{(R)^\dagger}) \leq r$ and $||A_{\hat{U}}^{(R)^\dagger}||_2 \leq \frac{4}{\sqrt{k}}$ for $k > 4r$ from corollary (2) with probability $> 1 - e^{-k/18}$: Hence with probability$> 1 - e^{-k/18}$:

$$\textbf{III} = ||\hat{U} A_{\hat{U}}^{(R)^\dagger} Z^{(R)}||_F = ||A_{\hat{U}}^{(R)^\dagger} Z^{(R)}||_F \leq \sqrt{r}||A_{\hat{U}}^{(R)^\dagger} Z^{(R)}||_2 \leq \sqrt{r}||A_{\hat{U}}^{(R)^\dagger}||_2||Z^{(R)}||_2 \leq \frac{4\sqrt{r}}{\sqrt{k}}||Z^{(R)}||_2 \quad (54)$$

hence we have constants C_3 and c_3 such that with probability $> 1 - e^{-c_3 k}$.

$$\textbf{III} \leq C_3||Z^{(R)}||_2 \quad (55)$$

Combining equations (55,53,49) with lemma 8 and taking $c^{(C)} = C_1 + C_2$, $c^{(R)} = C_3$ with $c = min(c_1, c_2, c_3)$ concludes our proof. □

7.5. Simulations for Large values of n

We varied n between 10 and 1000, with results averaged over 20 different matrices of rank 3 at each point, and try to reconstruct them from $k = 20$. We see that preference is insensitive to n. if we take $A^{(R)}, A^{(C)} \overset{i.i.d.}{\sim} N(0,1)$ instead of $N(0, \frac{1}{n})$ we will get results as in (3)

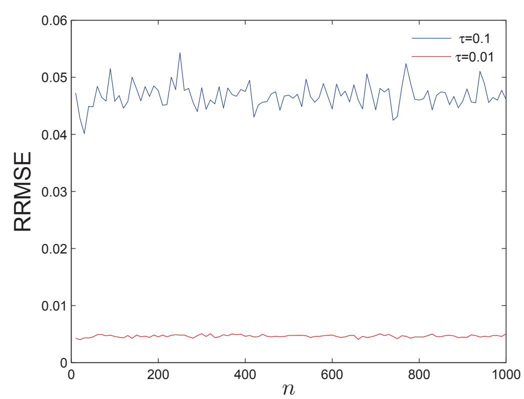

Figure 5. Reconstruction error for $n \times n$ matrix where n is varied between 10 and 1000, $k = 20$ and $r = 3$ and two different noise levels: $\tau = 0.1$ (blue) and $\tau = 0.01$ (red). Each point is an average over 20 matrices.

.

Now we take $n, k, r \to \infty$ while the ratios $\frac{n}{k} = 5$ and $\frac{k}{r} = 4$ are constant, and look at the relative error for different noise level. Again, the relative error converges rapidly to constant, independent of n, k, r .

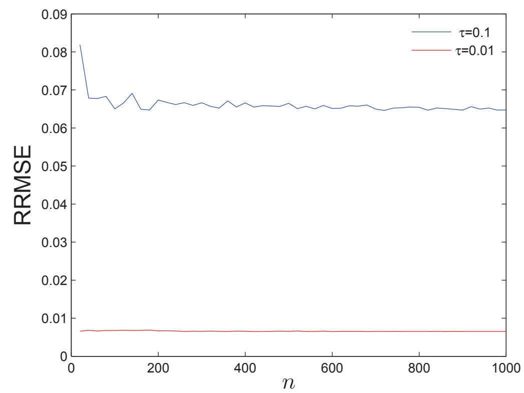

Figure 6. We reconstruct matrix X for i between 1 and 50 and $n = 20i$, $k = 4i$ and $r = i$. and for different noise level $\tau = 0.1$ (blue) and $\tau = 0.01$ (red).