Supplementary Material for 'Matrix Recovery using Deep Generative Priors with Low-Rank Deviations'

Pengbin Yu*, Jianjun Wang*, Chen Xu †
*School of mathematics and Statistics, Southwest University, Chongqing, 400715, China
†Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

A. Proof of Lemma 1

In order to take the noise into account, we define a ε-tube set U of an ensemble $\mathcal{A}$ as,

$$U_{\mathcal{A}}(\varepsilon) = \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2} : \|\mathcal{A}(\mathbf{X})\|_2 \leq \varepsilon.\}$$

In particular, $U_{\mathcal{A}}(0)$ represents the nullspace of $\mathcal{A}$. For convenience, we defined a difference function $D : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^{n_1 \times n_2}$ such that $D(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{G}(\mathbf{z}_1) - \mathbf{G}(\mathbf{z}_2), \mathbf{z}_1, \mathbf{z}_2 \in Dom(G)$. Therefore, we can define difference set $\mathcal{LR}_{r,D}$ as,

$$\begin{aligned}
&\mathcal{LR}_{r,D} = \mathcal{LR}_r(D) \\
&= \{\mathbf{X} : rank(\mathbf{X} - D(\mathbf{z}_1, \mathbf{z}_2)) \leq r, \mathbf{z}_1, \mathbf{z}_2 \in Dom(\mathbf{G}))\} \\
&= \{\mathbf{X} : rank(\mathbf{X} - \mathbf{G}(\mathbf{z}_1) + \mathbf{G}(\mathbf{z}_2)) \leq r, \mathbf{z}_1, \mathbf{z}_2 \in Dom(\mathbf{G})\}.
\end{aligned} \tag{A.1}$$

Further, we can similarly define the least possible nulcear norm for the matrices in $\mathcal{LR}_{r,D}$:

$$\phi_{r,D}(\mathbf{X}) = \inf_{\hat{\mathbf{X}} \in \mathcal{LR}_{r,D}} \|\mathbf{X} - \hat{\mathbf{X}}\|_*,$$

Now, we generalize and provide analogous theoretical proofs of Dhar et al. [1] and Bora et al.[2]. We state Lemma 4 and Lemma 5. Lemma 4 gives the sufficient condition for Nuclear Decoder to upper bound the reconstruction error. The lemma also expresses the following ideal: to make the sensing successful, the two points in $\mathcal{LR}_{r,G}$ should not be very close under $\mathcal{A}$, which is equivalent to that any point in the nullspace of $\mathcal{A}$ should not be very close to the point in $\mathcal{LR}_{2r,D}$. Since bounded noise is added in this paper, we need to get such a result on the ε-tube, not just the nullspace.

Lemma 3. *Let* $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{n_1 \times n_2}$*, then* $\phi_{2r,D}(\mathbf{X}_1 - \mathbf{X}_2) \leq \phi_{r,G}(\mathbf{X}_1) + \phi_{r,G}(\mathbf{X}_2)$.

Proof. Set $\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2 \in \mathcal{LR}_{r,G}$, then we can write them as $\hat{\mathbf{X}}_1 = \tilde{\mathbf{X}}_1 + \mathbf{G}(\mathbf{z}_1), \hat{\mathbf{X}}_2 = \tilde{\mathbf{X}}_2 + \mathbf{G}(\mathbf{z}_2)$, where $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \in \mathcal{LR}_r$. Since $\tilde{\mathbf{X}}_1 - \tilde{\mathbf{X}}_2 \in \mathcal{LR}_{2r}$, then

$$\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_2 = \tilde{\mathbf{X}}_1 - \tilde{\mathbf{X}}_2 + (\mathbf{G}(\mathbf{z}_1) - \mathbf{G}(\mathbf{z}_2)) \in \mathcal{LR}_{2r,D}.$$

Hence

$$\begin{aligned}
&\phi_{2r,D}(\mathbf{X}_1 - \mathbf{X}_2) \\
&= \inf_{\hat{\mathbf{X}} \in \mathcal{LR}_{2r,D}} \|\mathbf{X}_1 - \mathbf{X}_2 - \hat{\mathbf{X}}\|_* \\
&\leq \inf_{\hat{\mathbf{X}}_1 \in \mathcal{LR}_{r,G}, \hat{\mathbf{X}}_2 \in \mathcal{LR}_{r,G}} \|\mathbf{X}_1 - \mathbf{X}_2 - (\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_2)\|_* \\
&\leq \inf_{\hat{\mathbf{X}}_1 \in \mathcal{LR}_{r,G}} \|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|_* + \inf_{\hat{\mathbf{X}}_2 \in \mathcal{LR}_{r,G}} \|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|_* \\
&= \phi_{r,G}(\mathbf{X}_1) + \phi_{r,G}(\mathbf{X}_2).
\end{aligned} \tag{A.2}$$

■

Lemma 4. *Given a measurement ensemble $\mathcal{A}:\mathbb{R}^{n_1\times n_2}\rightarrow\mathbb{R}^m$, measurement noise ζ with $\|\zeta\|_2\leq\varepsilon$, and a generative function $\mathbf{G}:\mathbb{R}^k\rightarrow\mathbb{R}^{n_1\times n_2}$. Then there exists a decoder $\Delta:\mathbb{R}^m\rightarrow\mathbb{R}^{n_1\times n_2}$ satisfies the following Frobenius-Nuclear mixed norm approximation guarantee:*

$$\|\mathbf{X}-\Delta(\mathcal{A}\mathbf{X}+\zeta)\|_F\leq C_1r^{-t}\phi_{r,G}(\mathbf{X})+C_2\varepsilon+\tau,\tag{A.3}$$

where $\mathbf{X}\in\mathbb{R}^{n_1\times n_2}$, $C_1,C_2,\tau,t\geq 0,r>0$ are constants.

The sufficient condition for the guarantee is as follows,

$$\|\mathbf{Z}\|_F\leq\frac{C_1}{2}r^{-t}\phi_{2r,D}(\mathbf{Z})+C_2\varepsilon+\tau,\quad\forall\mathbf{Z}\in U_{\mathcal{A}}(2\varepsilon).\tag{A.4}$$

Proof. The Nuclear Decoder $\Delta:\mathbb{R}^m\rightarrow\mathbb{R}^{n_1\times n_2}$ is as follows,

$$\Delta(y)=\arg\min_{\hat{\mathbf{X}}:\|\mathcal{A}\hat{\mathbf{X}}-y\|_2\leq\varepsilon}\phi_{r,G}(\mathbf{X}).$$

Then we proof nuclear decoder satisfies (A.3). According to the definition of Nuclear Decoder, we can get:

$$\|\mathcal{A}(\mathbf{X}-\Delta(\mathcal{A}\mathbf{X}+\zeta))\|_2\leq\|\mathcal{A}\mathbf{X}+\zeta-\mathcal{A}(\Delta(\mathcal{A}\mathbf{X}+\zeta))\|_2+\varepsilon\leq 2\varepsilon.$$

This implies $\mathbf{X}-\Delta(\mathcal{A}\mathbf{X}+\zeta)\in U_{\mathcal{A}}(2\varepsilon)$. From (A.4), we have

$$\begin{aligned}&\|\mathbf{X}-\Delta(\mathcal{A}\mathbf{X}+\zeta)\|_F-\tau-C_2\varepsilon\\&\leq\frac{C_1}{2}r^{-t}\phi_{2r,D}(X-\Delta(\mathcal{A}\mathbf{X}+\zeta))\\&\leq\frac{C_1}{2}r^{-t}\Big[\phi_{r,G}(\mathbf{X})+\phi_{r,G}(\Delta(\mathcal{A}\mathbf{X}+\zeta))\Big]\\&\leq C_1r^{-t}\phi_{r,G}(\mathbf{X}).\end{aligned}\tag{A.5}$$

The second inequality is form lemma 3. The last inequality holds because it uses the definition of the decoder, which minimizes $\phi_{r,G}(\mathbf{X})$. ■

The next lemma shows that if $\mathcal{A}$ meets M-S-REC and M-RIP then the sufficient condition of lemma 4 can be satisfied.

Lemma 5. *Given a function $\mathbf{G}:\mathbb{R}^k\rightarrow\mathbb{R}^{n_1\times n_2}$ and a measurement ensemble $\mathcal{A}:\mathbb{R}^{n_1\times n_2}\rightarrow\mathbb{R}^m$. If $\mathcal{A}$ satisfies M-S-REC($\mathcal{LR}_{\frac{(a+b)r}{2},G},1-\delta,\tau$) and M-RIP($br,\delta$) for integer $a,b,r>0$, then for any matrix $\mathbf{Z}\in U_{\mathcal{A}}(\varepsilon)$, we have*

$$\|\mathbf{Z}\|_F\leq(br)^{-\frac{1}{2}}(C_1+1)\phi_{ar,D}(\mathbf{Z})+C_2\varepsilon+\tau',\tag{A.6}$$

where $C_1=(1-\delta)^{-1}\sqrt{(1+\delta)}$, $C_2=(1-\delta)^{-1}$, $\tau'=(1-\delta)^{-1}\tau$.

Proof. For any $\mathbf{Z}\in U_{\mathcal{A}}(\varepsilon)$, we choose the images of $\mathcal{A}$: $\mathbf{G}(\mathbf{z_1}),\mathbf{G}(\mathbf{z_2})$ whose left and right singular matrices are as same as those of $\mathbf{Z}$. In other words, $\mathbf{G}(\mathbf{z_1}),\mathbf{G}(\mathbf{z_2})$ are from intersection of the space spanned by the left and right singular matrices of $\mathbf{Z}$ and the image space of $\mathbf{G}$, denoting as $\hat{\mathbf{G}}$ (for convenience, we substitute $\mathbf{G}_1$ and $\mathbf{G}_2$ for $\mathbf{G}(\mathbf{z_1})$ and $\mathbf{G}(\mathbf{z_2})$ respectively).

For any matrix $\mathbf{M}$, by singular value decomposition, we can partition $\mathbf{M}$ into a sum of matrices $\mathbf{M}_0,\mathbf{M}_1,\mathbf{M}_2\cdots$ whose rank are at most $ar,br,br,\cdots$, respectively. Seriously, the rank of final matrix maybe less than br. Let $\mathbf{M}=\mathbf{U}diag(\sigma)\mathbf{V}'$ be the singular value decomposition of $\mathbf{M}$. For each $i\geq 1$, define the index set $I_0=\{1,\ldots,ar\}$, $I_i=\{br(i-1)+1,\cdots,bri\}$, and let $\mathbf{M}_0=\mathbf{U}diag(\sigma_{I_0})\mathbf{V}'$, $\mathbf{M}_i=\mathbf{U}diag(\sigma_{I_i})\mathbf{V}'$.

Then let $\mathbf{W}\in\mathcal{LR}_{ar}$ be the minimizer of $\|\mathbf{Z}-\mathbf{G}_1+\mathbf{G}_2-\mathbf{W}\|_*$. By the definition of the partition above and singular value decomposition, W is exactly $(\mathbf{Z}-\mathbf{G}_1+\mathbf{G}_2)_{I_0}$.

Given a set of indices I for a $n_1 \times n_2$ dimensional matrix, we use I^c to denote the set of indices not in I. Let $\bar{\mathbf{Z}} = \mathbf{Z} - \mathbf{G}_1 + \mathbf{G}_2$ and $\mathbf{W}$ corresponding to the rank ar truncated SVD of $\bar{\mathbf{Z}}$. Let $I_{01} = I_0 \cup I_1$. Likewise, $\bar{\mathbf{Z}}_{I_1}, \bar{\mathbf{Z}}_{I_2}, \cdots, \bar{\mathbf{Z}}_{I_s}$ are all matrices of rank at most br due to the number of non-zero singular values of $diag(\sigma_{I_i})$.

It's not hard to image that we can do some decomposition based on the index sets. Firstly, $\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c} = \mathbf{Z}_{I_{01}} - (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)$, where $\mathbf{Z}_{I_{01}}, (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}} \in \mathcal{LR}_{(a+b)r}$. Secondly, we can write $\mathbf{Z}_{I_{01}} - (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}} \in \mathcal{LR}_{(a+b)r}$ as $\mathbf{V}_1 - \mathbf{V}_2$, where $\mathbf{V}_1, \mathbf{V}_2 \in \mathcal{LR}_{\frac{(a+b)r}{2}}$, which implies that

$$\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c} = (\mathbf{G}_1 + \mathbf{V}_1) - (\mathbf{G}_2 + \mathbf{V}_2),$$

where $\mathbf{G}_1 + \mathbf{V}_1, \mathbf{G}_2 + \mathbf{V}_2 \in \mathcal{LR}_{\frac{(a+b)r}{2}, G}$.

Since $\mathcal{A}$ satisfies M-S-REC($\mathcal{LR}_{\frac{(a+b)r}{2}, G}, 1-\delta, \tau$), we have:

$$\begin{aligned}
&\left\|\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\right\|_F \\
&= \|(\mathbf{G}_1 + \mathbf{V}_1) - (\mathbf{G}_2 + \mathbf{V}_2)\|_F \\
&\leq (1-\delta)^{-1} \left\|\mathcal{A}\left[(\mathbf{G}_1 + \mathbf{V}_1) - (\mathbf{G}_2 + \mathbf{V}_2)\right]\right\| + (1-\delta)^{-1}\tau \\
&\leq (1-\delta)^{-1} \|\mathcal{A}\left[\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\right]\|_F + (1-\delta)^{-1}\tau.
\end{aligned} \tag{A.7}$$

Based on $\mathbf{Z} \in U_{\mathcal{A}}(\varepsilon)$, then we can write $\mathcal{A}(\mathbf{Z}) = \mathcal{A}(\mathbf{Z}_{I_{01}} + \mathbf{Z}_{I_2} + \cdots + \mathbf{Z}_{I_s}) = \gamma$, where $\|\gamma\|_2 \leq \varepsilon$, which follows $\mathcal{A}(\mathbf{Z}_{I_{01}}) = -\mathcal{A}(\mathbf{Z}_{I_2} + \cdots + \mathbf{Z}_{I_s})$.

Hence,

$$\begin{aligned}
&\left\|\mathcal{A}\big[\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\big]\right\|_2 \\
&= \left\|\mathcal{A}\big[(\mathbf{Z} - \mathbf{G}_1 + \mathbf{G}_2)_{I_2} + \cdots + (\mathbf{Z} - \mathbf{G}_1 + \mathbf{G}_2)_{I_s}\big] - \gamma\right\|_2 \\
&\leq \left\|\mathcal{A}(\bar{\mathbf{Z}}_{I_2}) + \cdots + \mathcal{A}(\bar{\mathbf{Z}}_{I_s}) - \gamma\right\|_2 \\
&\leq \sum_{j=2}^{s} \left\|\mathcal{A}(\bar{\mathbf{Z}}_{I_j})\right\|_2 + \|\gamma\|_2 \quad \text{(using triangle inequality)} \\
&\leq \sqrt{(1+\delta)} \sum_{j=2}^{s} \left\|\bar{\mathbf{Z}}_{I_j}\right\|_F + \varepsilon. \quad \text{(using M-RIP)}
\end{aligned} \tag{A.8}$$

Combining (A.7) and (A.8) we can get,

$$\begin{aligned}
&\left\|\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\right\|_F \\
&\leq (1-\delta)^{-1}\sqrt{(1+\delta)} \sum_{j=2}^{s} \left\|\bar{\mathbf{Z}}_{I_j}\right\|_F + (1-\delta)^{-1}\varepsilon + (1-\delta)^{-1}\tau.
\end{aligned}$$

By decomposing and applying the triangle inequality, we have:

$$\begin{aligned}
\|\mathbf{Z}\|_F &= \left\|\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c} + \mathbf{Z}_{I_{01}^c} - (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\right\|_F \\
&\leq \left\|\mathbf{Z}_{I_{01}} + (\mathbf{G}_1 - \mathbf{G}_2)_{I_{01}^c}\right\|_F + \left\|\bar{\mathbf{Z}}_{I_{01}^c}\right\|_F \\
&\leq \left[(1-\delta)^{-1}\sqrt{(1+\delta)} + 1\right] \sum_{j=2}^{s} \|\bar{\mathbf{Z}}_{I_j}\|_F + C_2\varepsilon + \tau'
\end{aligned} \tag{A.9}$$

For any $j \geq 2$, $i_1 \in I_j$ and $i_2 \in I_{j-1}$, we have $\sigma_{i_1}(\bar{\mathbf{Z}}) \leq (br)^{-1}\|\bar{\mathbf{Z}}_{I_{j-1}}\|_*$. Thus,

$$\left\|\bar{\mathbf{Z}}_{I_j}\right\|_F = \left(\sum_{i=br(j-1)}^{brj} \sigma_i^2(\bar{\mathbf{Z}})\right)^{\frac{1}{2}} \leq (br)^{-\frac{1}{2}} \|\bar{\mathbf{Z}}_{I_{j-1}}\|_*.$$

Substituting the result we obtained above in Eq.(A.9), we get,

$$\begin{aligned}\|\mathbf{Z}\|_F - \tau' - C_2\varepsilon &\le (br)^{-\frac{1}{2}}\left[(1-\delta)^{-1}\sqrt{(1+\delta)}+1\right]\sum_{j=1}^{s}\left\|\bar{\mathbf{Z}}_{I_j}\right\|_* \\ &= (br)^{-\frac{1}{2}}(C_1+1)\left\|\bar{\mathbf{Z}}_{I_0^c}\right\|_* \\ &\le (br)^{-\frac{1}{2}}(C_1+1)\phi_{ar,D}(\mathbf{Z}).\end{aligned}$$

The last equality is because the following fact:

$$\begin{aligned}\phi_{ar,D}(\mathbf{Z}) &= \inf_{\hat{\mathbf{Z}}\in\mathcal{LR}_{ar,D}}\left\|\mathbf{Z}-\hat{\mathbf{Z}}\right\|_* \\ &= \inf_{\tilde{\mathbf{z}}_1,\tilde{\mathbf{z}}_2\in Dom(\mathbf{G}),\mathbf{W}\in\mathcal{LR}_{ar}}\|\mathbf{Z}-\mathbf{G}(\tilde{\mathbf{z}}_1)+\mathbf{G}(\tilde{\mathbf{z}}_2)-\mathbf{W}\|_* \\ &= \inf_{\mathbf{z}_1,\mathbf{z}_2\in Dom(\hat{\mathbf{G}}),\mathbf{W}\in\mathcal{LR}_{ar}}\|\mathbf{Z}-\mathbf{G}(\mathbf{z}_1)+\mathbf{G}(\mathbf{z}_2)-\mathbf{W}\|_* \\ &= \inf_{\mathbf{W}\in\mathcal{LR}_{ar}}\left\|\bar{\mathbf{Z}}-\mathbf{W}\right\|_* = \left\|\bar{\mathbf{Z}}_{I_0^c}\right\|_*.\end{aligned}$$

The last second equality holds because the nuclear norm can be minimized in space spanned by the left and right singular matrices of $\mathbf{Z}$. ■

Combining Lemma 4 and Lemma 5, Lemma 1 can be derived directly when a=2 and b=1.

Proof of Lemma 2

Definition 1. *A random variable X is said to be subgamma(σ,B) if for any t>0, we have*

$$\mathbb{P}\Big\{|X-\mathbb{E}X|\ge t\Big\}\le 2\max\Big(e^{-\frac{t^2}{2\sigma^2}},e^{-\frac{Bt}{2}}\Big).$$

Lemma 6. *[3] For $0\le\delta\le1$ and $\mathcal{A}:\mathbb{R}^{n_1\times n_2}\to\mathbb{R}^m$ is a Gaussian random measurement ensemble with independent, identically, distributed (i.i.d.) Gaussian entries $\mathbf{A}_{ij}\sim N(0,\frac{1}{m})$. Then as long as $m=O(nr)$, $\mathcal{A}$ satisfies M-RIP(r,δ) with probability at least $1-\exp^{-O(m)}$, where $n=\max\{n_1,n_2\}$.*

The following lemma shows that the distance between a object and it's approximation on a net is sufficiently close in terms of the Frobenius distance give rise to imaging measurements that are also close. The Proof is an analogue of Bora et al.[2] for vector.

Lemma 7. *Let $\mathbf{G}:(\mathbb{R}^k,\|\cdot\|_{l_2})\to(\mathbb{R}^{n_1\times n_2},\|\cdot\|_F)$ be a L-Lipschitz function, where $\|\cdot\|_{l_2}$ and $\|\cdot\|_F$ denote the vector Euclidean norm (L_2 norm) and matrix Frobenius norm respectively. Let $B^k(a)$ be a L_2-ball in $\mathbb{R}^k$ with radius a, $\mathbf{S}=\mathbf{G}(B^k(a))$, and M be a $\frac{\delta}{L}$-net on $B^k(a)$ such that $|M|\le k\log\left(\frac{4La}{\tau}\right)$. Let $\mathcal{A}$ be a Gaussian random measurement ensemble $\mathbb{R}^{n_1\times n_2}\to\mathbb{R}^m$ with i.i.d. Gaussian entries from $N(0,\frac{1}{m})$. If*

$$m=O\Big(k\log\frac{La}{\tau}\Big),$$

then for any $\mathbf{X}\in\mathbf{S}$, if $\mathbf{X}'=\arg\min_{\hat{\mathbf{X}}\in\mathbf{G}(M)}\|\mathbf{X}-\hat{\mathbf{X}}\|_F$, we have $\|\mathcal{A}(\mathbf{X}-\mathbf{X}')\|=O(\tau)$ with probability at least $1-e^{-O(m)}$.

Proof. Observe that $\frac{\|\mathcal{A}X\|^2}{\|X\|_F^2}$is subgamma($\frac{1}{\sqrt{m}},\frac{1}{m}$) (As mentioned above, viewing $\mathcal{A}$ as a $m\times n_1n_2$ matrix, $\|\mathbf{X}\|_F^2=\|\mathrm{vec}(\mathbf{X})\|^2$, the result is just as same as the vector vision). Thus, for any $f\ge0$, we have

$$t\ge 2+\frac{4}{m}\log\frac{2}{f}\ge\max\big(\sqrt{\frac{2}{m}\log\frac{2}{f}},\frac{2}{m}\log\frac{2}{f}\big)$$

is sufficient to ensure that

$$\mathbb{P}\Big\{ \|\mathcal{A}X\| \geq (1+t)\|X\|_F \Big\} \leq f.$$

Now let $M = M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots \subseteq M_l$ be a chain of epsilon nets of $B^k(a)$ such that M_i is a $\frac{\tau_i}{L}$-net and $\tau_i = \frac{\tau_0}{2^i}$, with $\tau_0 = \tau$.

By the proposition of the epsilon nets on a L_2-ball, there exists nets such that

$$\log|M_i| \leq k \log\big(\frac{4La}{\tau_i}\big) \leq ik + k\log\big(\frac{4La}{\tau_0}\big).$$

Let $\mathbf{N}_i = \mathbf{G}(M_i)$. Then due to Lipschitzness of $\mathbf{G}$, for any $\mathbf{z}_1, \mathbf{z}_2 \in B^k(a)$, we have

$$\|\mathbf{G}(\mathbf{z}_1) - \mathbf{G}(\mathbf{z}_2)\|_F \leq L\|\mathbf{z}_1 - \mathbf{z}_2\| \leq L \cdot \frac{\tau_i}{L} = \tau_i,$$

which yields $\mathbf{N}_i$ is a τ_i-net of $\mathbf{S} = \mathbf{G}(B^k(a))$ and $\mathbf{N}_i$ forms a chain of epsilon nets, with $|\mathbf{N}_i| = |M_i|$.

For $i \in \{0, 1, 2, \cdots, l-1\}$, let $\mathbf{D}_i = \big\{\mathbf{X}_{i+1} - \mathbf{X}_i | \mathbf{X}_{i+1} \in \mathbf{N}_{i+1}, \mathbf{X}_i \in \mathbf{N}_i\big\}$.

Hence,

$$\begin{aligned}\log|\mathbf{D}_i| &\leq \log|\mathbf{N}_{i+1}| + \log|\mathbf{N}_i| \\ &\leq 2ik + 2k\log\big(\frac{4La}{\tau_0}\big) \\ &\leq 3ik + 2k\log\big(\frac{4La}{\tau_0}\big).\end{aligned} \tag{A.10}$$

We assume that,

$$m = 3k\log\big(\frac{4La}{\tau_0}\big), \log(f_i) = -(m + 4ik),$$

and

$$\begin{aligned} t_i = 2 + \frac{4}{m}\log\frac{2}{f_i} &= 2 + \frac{4}{m}\log 2 + 4 + \frac{16ik}{m} \\ &= O(1) + \frac{16ik}{m}. \end{aligned}$$

By choice of f_i and t_i, we have $\forall i \in [r-1], \forall \mathbf{D} \in \mathbf{D}_i$,

$$\mathbb{P}\big\{\|\mathcal{A}(\mathbf{D})\| \leq (1+t_i)\|\mathbf{D}\|_F\big\} \leq f_i.$$

Thus by union bound we obtain

$$\mathbb{P}\big\{\|\mathcal{A}(\mathbf{D})\| \leq (1+t_i)\|\mathbf{D}\|_F, \forall i, \forall \mathbf{D} \in \mathbf{D}_i\big\} \geq 1 - \sum_{i=0}^{l-1}|\mathbf{D}_i| f_i.$$

Then,

$$\begin{aligned}\log(|\mathbf{D}_i| f_i) &= \log(|\mathbf{D}_i|) + \log(f_i) \\ &\leq -k\log\big(\frac{4La}{\tau_0}\big) - ik = -\frac{m}{3} - ik.\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{i=0}^{l-1}|\mathbf{D}_i| f_i &\leq e^{-m/3}\sum_{i=0}^{l-1} e^{-ik} \\ &\leq e^{-m/3}\big(\frac{1}{1-e^{-1}}\big) \leq 2e^{-m/3}.\end{aligned}$$

Consider that for any $\mathbf{X} \in \mathbf{S}$, $\mathbf{X}$ has presentation about

$$\mathbf{X} = \mathbf{X}_0 + (\mathbf{X}_1 - \mathbf{X}_0) + (\mathbf{X}_2 - \mathbf{X}_1) + \cdots + (\mathbf{X}_l - \mathbf{X}_{l-1}) + \mathbf{X}^f,$$
$$\mathbf{X} - \mathbf{X}_0 = \sum_{i=0}^{l-1} (\mathbf{X}_{i+1} - \mathbf{X}_i) + \mathbf{X}^f,$$

where $\mathbf{X}_i \in \mathbf{N}_i$ and $\mathbf{X}^f = \mathbf{X} - \mathbf{X}_l$.

Since each $\mathbf{X}_{i+1} - \mathbf{X}_i \in \mathbf{D}_i$ with probability at least $1 - 2e^{-\frac{m}{3}}$, then

$$\begin{aligned}\sum_{i=0}^{l-1} \|\mathcal{A}(\mathbf{X}_{i+1} - \mathbf{X}_i)\| &\le \sum_{i=0}^{l-1} (1 + t_i)\|\mathbf{X}_{i+1} - \mathbf{X}_i\|_F \\ &\le \sum_{i=0}^{l-1} (1 + t_i)\tau_i \\ &= \sum_{i=0}^{l-1} \big(O(1) + \frac{16ik}{m}\frac{\tau_0}{2^i}\big) \\ &= O(\tau_0) + \tau_0 \frac{16k}{m} \sum_{i=0}^{l-1} (\frac{i}{2^i}) = O(\tau_0).\end{aligned}$$

Now, $\|\mathbf{X}^f\|_F = \|\mathbf{X} - \mathbf{X}_l\|_F \le \frac{\tau_0}{2^l}$ and $\|\mathbf{X}_{i+1} - \mathbf{X}_i\|_F \le \tau_i$ due to properties of epsilon-nets. We know that $\|\mathbf{A}\| \le 2 + \sqrt{\frac{n_1 n_2}{m}}$ holds with probability at least $1 - 2e^{-\frac{m}{2}}$ [4]. By setting $l = \log(n_1 n_2)$, then $\|\mathcal{A}\|\|\mathbf{X}^f\|_F \le \left(2 + \sqrt{\frac{n_1 n_2}{m}}\right)\frac{\tau_0}{2^l} = O(\tau_0)$ with probability at least $1 - 2e^{-m/2}$.

Combining these results, and noting that it's possible to choose $\mathbf{X}' = \mathbf{X}_0$, it follows with probability at least $1 - e^{-O(m)}$ that,

$$\begin{aligned}\|\mathcal{A}(\mathbf{X} - \mathbf{X}')\| &= \|\mathcal{A}(\mathbf{X} - \mathbf{X}_0)\| \\ &\le \sum_{i=0}^{l-1} \|\mathcal{A}(\mathbf{X}_{i+1} - \mathbf{X}_i)\| + \|\mathcal{A}\mathbf{X}^f\| \\ &= O(\tau_0) + \|\mathbf{A}\text{vec}(\mathbf{X}^f)\| \\ &= O(\tau_0) + \|\mathcal{A}\|\|\text{vec}(\mathbf{X}^f)\|_2 \\ &= O(\tau_0) + \|\mathcal{A}\|\|\mathbf{X}^f\|_F = O(\tau).\end{aligned}$$

Hence the proof is completed. ■

Lemma 2. ***(restated)*** *Let $\mathbf{G} : (B^k(a), \|\cdot\|_{l_2}) \to (\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_F)$ be an L-Lipschitz function, where $B^k(a) = \{z | z \in \mathbb{R}^k, \|z\|_2 \le a\}$ is the L_2 ball in $\mathbb{R}^k$. For $\delta \in (0, 1)$, if*

$$m = O\left(\frac{1}{\delta^2}\big(k \log(\frac{La}{\tau}) + (n_1 + n_2 + 1) r \log(\frac{1}{\tau})\big)\right),$$

then a Gaussian random measurement ensemble $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ with i.i.d. gaussian entries from $N(0, \frac{1}{m})$ satisfies the M-S-REC($\mathcal{LR}_{1.5r,G}, 1 - \delta, \tau$) and M-RIP(r,δ) with at least $1 - e^{-O(\delta^2 m)}$ probability.

Proof. Similar with the proof of lemma 7, we start with construct a $\frac{\tau}{L}$-net M on $B^k(a)$ such that $\log|M| \le k \log\left(\frac{4La}{\tau}\right)$, then $\mathbf{G}(M)$ is a τ-net of $\mathbf{G}(B^k(a))$.

We want to proof such $\mathcal{A}$ satisfies M-S-REC($\mathcal{LR}_{1.5r,G}, 1 - \delta, \tau$) condition, i.e. $\forall \mathbf{X}_1, \mathbf{X}_2 \in \mathcal{LR}_{1.5r,G}$ we have $(1 - \delta)\|\mathbf{X}_1 - \mathbf{X}_2\|_F \le \|\mathcal{A}(\mathbf{X}_1 - \mathbf{X}_2)\| + O(\tau)$.By the definition of $\mathcal{LR}_{1.5r,G}$ we can write $\mathbf{X}_1 = \mathbf{G}(\mathbf{z}_1) + \mathbf{W}_1$, $\mathbf{X}_2 = \mathbf{G}(\mathbf{z}_2) + \mathbf{W}_2$, where $\mathbf{W}_1, \mathbf{W}_2 \in \mathcal{LR}_{1.5r}$. At this time, $\mathbf{W} := \mathbf{W}_1 - \mathbf{W}_2 \in \mathcal{LR}_{3r}$. Then we just need to show that

$$
\begin{aligned}
&(1-\delta)\|\mathbf{X}_1-\mathbf{X}_2\|_F\\
&=(1-\delta)\|\mathbf{W}_1-\mathbf{W}_2+\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)\|_F\\
&=(1-\delta)\|\mathbf{W}+\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)\|_F\\
&\leq\|\mathcal{A}\big(\mathbf{W}+\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)\big)\|+O(\tau)\\
&=\|\mathcal{A}\big(\mathbf{W}_1-\mathbf{W}_2+\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)\big)\|+O(\tau).
\end{aligned}
$$

(We can use singular value decomposition to split W into W_1 and W_2, then the proof can be simplified.)

In order to control the low-rank matrix $\mathbf{W}$, according to the covering number of low-rank matrix ([3]), we construct a τ-net of $\bar{\mathbf{W}}_{3r}:=\{\mathbf{W}\in\mathbb{R}^{n_1\times n_2}: rank(\mathbf{W})\leq 3r, \|\mathbf{W}\|_F=1\}\subset\mathcal{LR}_{3r}$, noted as $\mathbf{N}$, whose covering number satisfy $\log|\mathbf{N}|\leq(n_1+n_2+1)3r\log\left(\frac{9}{\tau}\right)$. Then we just need to consider $\bar{\mathbf{W}}_{3r}$, because any low-rank matrix whose Frobenius norm not equal to 1 can be transformed by matrix in $\bar{\mathbf{W}}_{3r}$ and the covering number differ from each other by at most an absolute constant factor.

For any $\mathbf{z}_1,\mathbf{z}_2\in B^k(a)$, $\mathbf{W}\in\bar{\mathbf{W}}_{3r}$, there are $\mathbf{z}_1',\mathbf{z}_2'\in M,\mathbf{W}'\in\mathbf{N}$ such that, $\|\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_1')\|_F\leq\tau$,$\|\mathbf{G}(\mathbf{z}_2)-\mathbf{G}(\mathbf{z}_2')\|_F\leq\tau$,$\|\mathbf{W}-\mathbf{W}'\|_F\leq\tau$. It follows similarly that

$$
\begin{aligned}
&\|\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)+\mathbf{W}\|_F\\
&\leq\|\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_1')\|_F+\|\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\|_F\\
&+\|\mathbf{G}(\mathbf{z}_2)-\mathbf{G}(\mathbf{z}_2')\|_F+\|\mathbf{W}-\mathbf{W}'\|_F\\
&\leq 3\tau+\|\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\|_F.
\end{aligned}
$$

By triangle inequality,

$$
\begin{aligned}
&\|\mathcal{A}\mathbf{G}(\mathbf{z}_1')-\mathcal{A}\mathbf{G}(\mathbf{z}_2')+\mathcal{A}\mathbf{W}'\|\\
&\leq\|\mathcal{A}\mathbf{G}(\mathbf{z}_1')-\mathcal{A}\mathbf{G}(\mathbf{z}_1)\|+\|\mathcal{A}\mathbf{G}(\mathbf{z}_1)-\mathcal{A}\mathbf{G}(\mathbf{z}_2)+\mathcal{A}\mathbf{W}\|\\
&+\|\mathcal{A}\mathbf{G}(\mathbf{z}_2)-\mathcal{A}\mathbf{G}(\mathbf{z}_2')\|+\|\mathcal{A}\mathbf{W}'-\mathcal{A}\mathbf{W}\|\\
&\leq O(\tau)+\|\mathcal{A}\mathbf{G}(\mathbf{z}_1)-\mathcal{A}\mathbf{G}(\mathbf{z}_2)+\mathcal{A}\mathbf{W}\|.
\end{aligned}
$$

The last inequality holds by Lemma 7, which shows that $\|\mathcal{A}\mathbf{G}(\mathbf{z}_1)-\mathcal{A}\mathbf{G}(\mathbf{z}_1')\|=O(\tau)$, $\|\mathcal{A}\mathbf{G}(\mathbf{z}_2)-\mathcal{A}\mathbf{G}(\mathbf{z}_2')\|=O(\tau)$ with probability $1-e^{-O(m)}$. And we can proof equality $\|\mathcal{A}\mathbf{W}'-\mathcal{A}\mathbf{W}\|=O(\tau)$ holds as long as $m=O\Big((n_1+n_2+1)r\log\frac{1}{\tau}\Big)$, which just need to do some small modifications on the proof of Lemma 7.

By the Johnson-Lindenstrauss Lemma, we have,

$$
(1-\delta)\|\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\|_F\leq\|\mathcal{A}\big(\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\big)\|.
$$

holds at least $1-e^{-O(\delta^2 m)}$ probability as long as

$$
\begin{aligned}
m&=O\Big(\frac{1}{\delta^2}\big(2\log|M|+\log|\mathbf{N}|\big)\Big)\\
&=O\Big(\frac{1}{\delta^2}\big(k\log(\frac{La}{\tau})+(n_1+n_2+1)r\log(\frac{1}{\tau})\big)\Big).
\end{aligned}
$$

Then we imply that,

$$\begin{aligned}
&(1-\delta)\|\mathbf{X}_1-\mathbf{X}_2\|_F \\
&=(1-\delta)\|\mathbf{W}+\mathbf{G}(\mathbf{z}_1)-\mathbf{G}(\mathbf{z}_2)\|_F \\
&\leq(1-\delta)3\tau+(1-\delta)\|\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\|_F \\
&\leq\|\mathcal{A}\big(\mathbf{G}(\mathbf{z}_1')-\mathbf{G}(\mathbf{z}_2')+\mathbf{W}'\big)\|+O(\tau) \\
&\leq\|\mathcal{A}\mathbf{G}(\mathbf{z}_1)-\mathcal{A}\mathbf{G}(\mathbf{z}_2)+\mathcal{A}\mathbf{W}\|+O(\tau).
\end{aligned}$$

Hence, such $\mathcal{A}$ satisfies M-S-REC($\mathcal{LR}_{1.5r,G}, 1-\delta, \tau$) condition. Combining with Lemma 6, we have if $\mathcal{A}$ satisfies M-RIP(r,δ) then,

$$m=O\Big(\frac{1}{\delta^2}\big(k\log(\frac{La}{\tau})+(n_1+n_2+1)r\log(\frac{1}{\tau})\big)\Big).$$

The proof is completed. ■

Finally, we note that Theorem 1 follows directly from the statements of Lemma 1 and Lemma 2.

References

[1] M. Dhar, A. Grover, and S. Ermon, "Modeling sparse deviations for compressed sensing using generative models," in *International Conference on Machine Learning*, pp. 1214–1223, PMLR, 2018.

[2] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*, pp. 537–546, PMLR, 2017.

[3] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[4] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

B. More Results

Experimental setup

All the experiments are implemented on the platform of VMware Workstation 12 Pro, Linux operating system and Python (3.7.4) with an Intel(R) Xeon(R) Gold-5122 3.6GHz CPU and 64GB memory.

1) MNIST, F-MNIST and Omniglot.

The size of each images in MNIST, F-MNIST and Omniglot is 28×28, and each pixel value is 0 or 1. We train VAE according to the Low-Rank-Gen model to restore the original image. Through experiments, we pick the learning rate is 0.1 and $\lambda=0.2, 0.045$ in Sparse-Gen and Low-Rank-Gen for MNIST respectively and $\lambda=0.01, 0.02$ in Sparse-Gen and Low-Rank-Gen for F-MNIST respectively. Omniglot dataset is used to test the transfer performance of our algorithm.

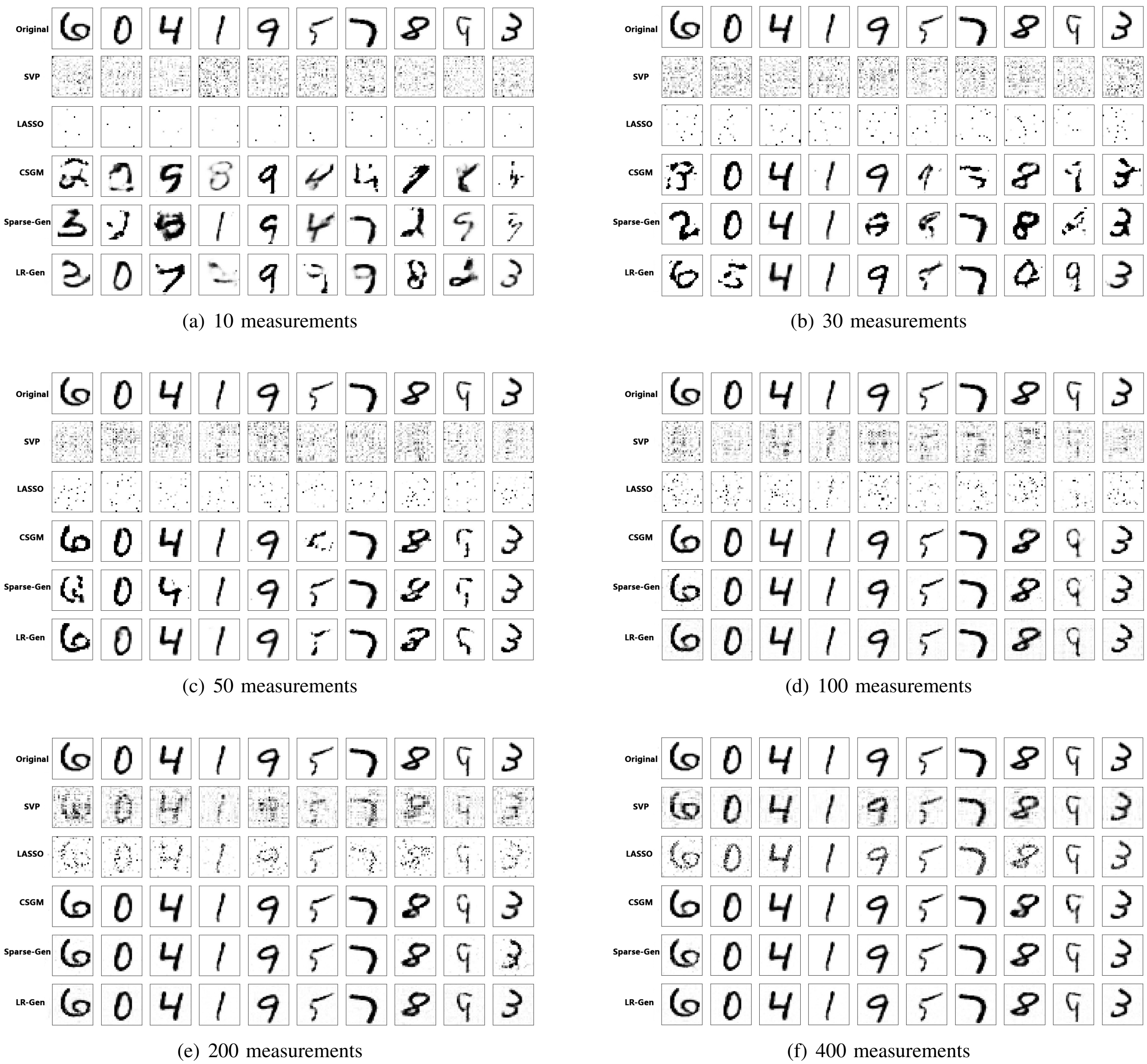
2) CIFAR-10.

The CIFAR-10 dataset consists of 60000 color images in 10 classes, with 6000 images per class. The size of each images in CIFAR-10 is $32\times 32\times 3$. We train DCGAN model as its priors [5]. Through experiments, we set the learning rate is 0.2 and $\lambda=0.3, 0.4$ in Sparse-Gen and Low-Rank-Gen.

3) CelebA.

Crop the face images to a size of 64×64 RGB so that the size of each image input is $64\times 64\times 3$. For this dataset, we also consider to train a DCGAN model as generator in Low-Rank-Gen model. In practice, we try and pick the Adam learning rate is 0.2 and $\lambda=0.03, 0.07$ in Sparse-Gen and Low-Rank-Gen respectively.

Mnist



(a) 10 measurements (b) 30 measurements

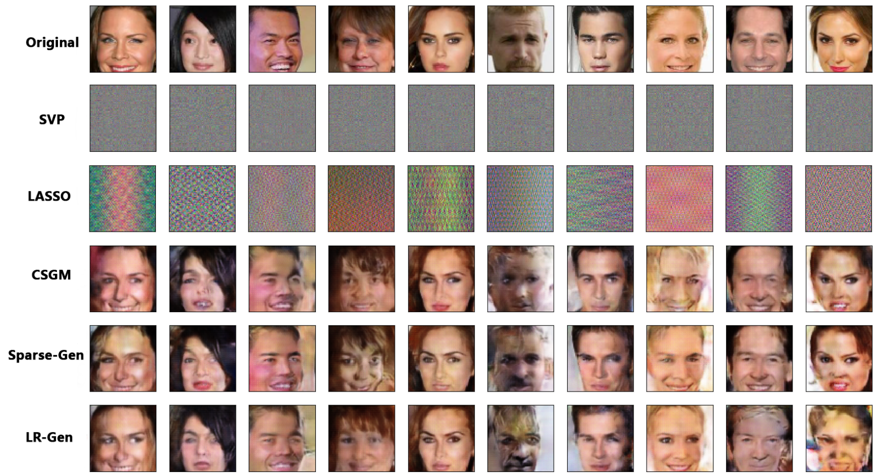
(c) 50 measurements (d) 100 measurements

(e) 200 measurements (f) 400 measurements

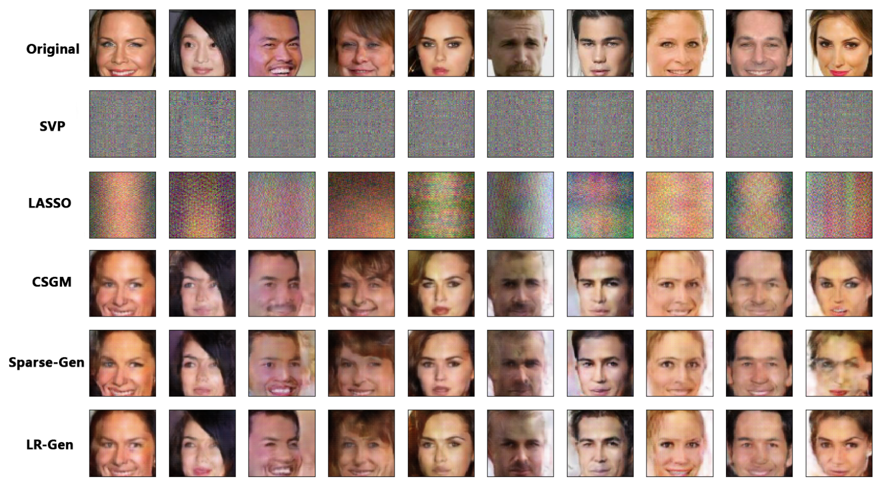
Fig. 1: MNIST dataset reconstruction results with $m = 10, 30, 50, 100, 200, 400$. Top to bottom rows are recovered images by different methods : original image, SVP, LASSO, CSGM, Sparse-Gen and Low-Rank-Gen.

CelebA

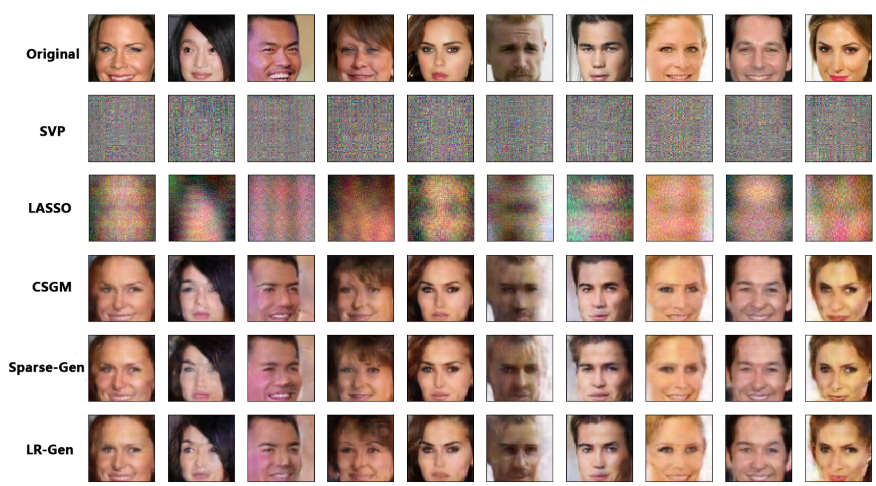
x



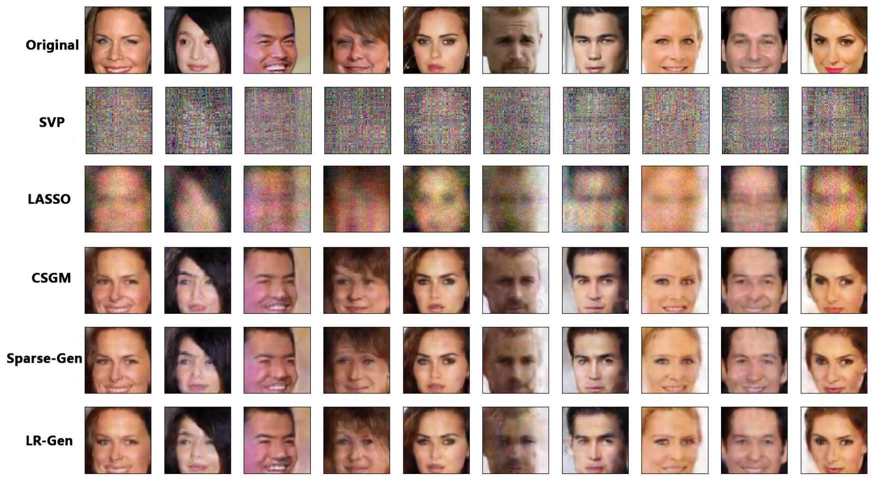
(a) 100 measurements



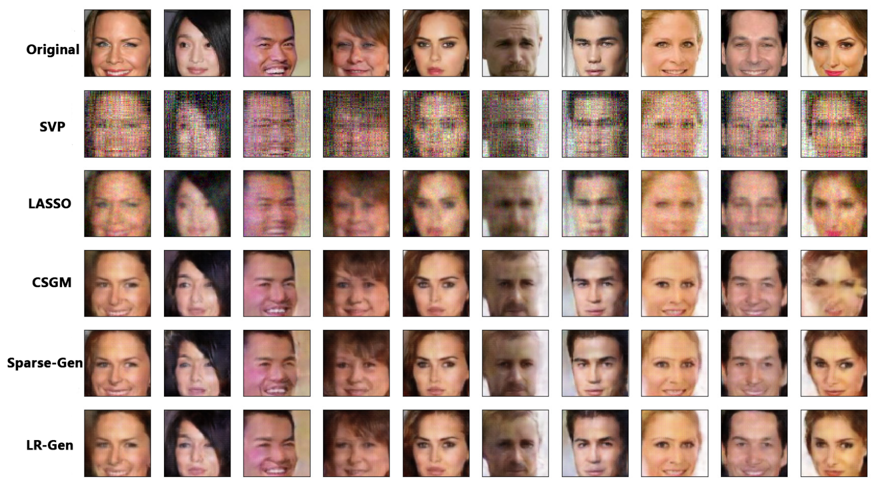
(b) 250 measurements



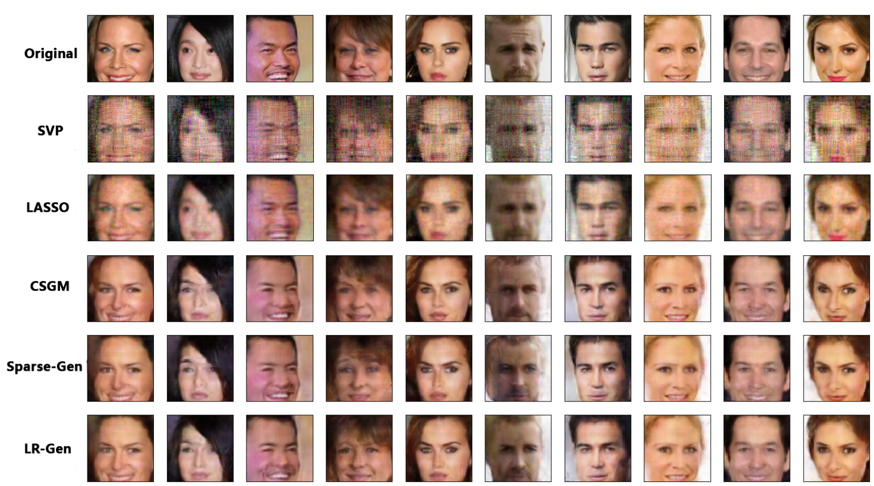
(c) 500 measurements



(d) 1000 measurements



(e) 3000 measurements



(f) 5000 measurements

Fig. 2: CelebA dataset reconstruction results with $m = 100, 250, 500, 1000, 3000, 5000$. Top to bottom rows are recoverd images by different methods : original image, SVP, LASSO, CSGM, Sparse-Gen and Low-Rank-Gen.

Under Training Situation (CIFAR-10)

We test the recovery effect in the case of under training (fewer training samples and fewer training iterations). We only use the car images in CIFAR-10 dataset. The number of traing data is 500, and 100 for testing. And we just train 1000 iterations for DCGAN (stop before stabilization). The result is shown in Fig 3.

(a) 100 measurements

(b) 300 measurements

(c) 500 measurements

(d) 1000 measurements

(e) 1500 measurements

(f) 2000 measurements

Fig. 3: Reconstruction results with measurements $m = 100, 300, 500, 1000, 1500, 2000$ on CIFAR-10(car). Top to bottom rows are recoverd images by different methods : original image, SVP, LASSO with DCT basis, CSGM, Sparse-Gen and Low-Rank-Gen.

As is depicted in Fig.3, when the model is under training, the generated model can only provide a fuzzy prior information (as can be observed from the CSGM method). Our method adopts nuclear norm regularization which retains the advantage of low rank matrix recovery, so it can further recover the data based on the fuzzy information provided by the under training model (although more samples are used). Our algorithm has higher recovery accuracy and reconstruction quality.

Transfer Recovery

The quantitative experimental results of transfer recovery are displayed in Table I. The table presents the PSNR value and the MSE recovery error (per pixel) and their standard deviation under different measurements. As shown in the table, the results of the proposed LowRank-Gen perform satisfactorily in this transfer recovery task and outperform other competitive methods under almost all measurements. The advantages are evident in the MNIST (source) to FMNIST (target) mission. Broader recovery tasks may enjoy our approach through transfer recovery.

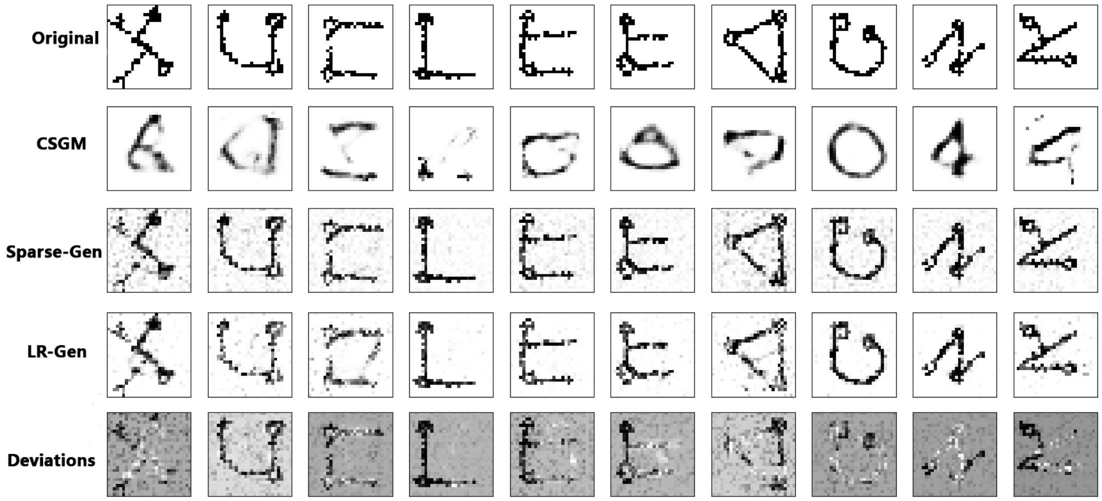
TABLE I: MSE recovery error (per pixel), PSNR and their standard deviation of transfer recovery with varying measurements.

			MNIST to Omniglot			
Algorithms	**CSGM** [?]		**Sparse-Gen** [?]		**LowRank-Gen (ours)**	
Measurements	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)
50	0.123(±0.0066)	9.119(±0.236)	0.12(±0.0065)	9.205(±0.238)	0.111(±0.0076)	9.561(±0.304)
75	0.1011(±0.0044)	9.958(±0.191)	0.094(±0.0034)	10.256(±0.16)	0.078(±0.0032)	11.059(±0.181)
100	0.0901(±0.0037)	10.457(±0.184)	0.0672(±0.003)	11.715(±0.218)	0.058(±0.0017)	12.404(±0.132)
200	0.0862(±0.0033)	10.766(±0.168)	0.043(±0.0033)	14.013(±0.341)	0.041(±0.0026)	14.294(±0.291)
250	0.0831(±0.0039)	10.807(±0.205)	0.027(±0.0012)	15.701(±0.201)	0.022(±0.0019)	16.674(±0.381)
300	0.0803(±0.0022)	10.955(±0.121)	0.013(±0.0019)	18.769(±0.684)	0.009(±0.0009)	20.405(±0.469)
350	0.0804(±0.0027)	10.952(±0.15)	0.0081(±0.0009)	20.769(±0.466)	0.007(±0.0002)	21.751(±0.183)
400	0.0784(±0.0019)	11.06(±0.108)	0.006(±0.0005)	22.585(±0.406)	0.003(±0.0004)	25.788(±0.741)
			MNIST to FMNIST			
Algorithms	**CSGM**		**Sparse-Gen**		**LowRank-Gen**	
Measurements	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)
50	0.126(±0.0062)	8.989(±0.217)	0.125(±0.0069)	9.033(±0.25)	0.101(±0.0070)	9.967(±0.297)
75	0.1098(±0.0056)	9.601(±0.224)	0.111(±0.0073)	9.595(±0.289)	0.078(±0.0052)	11.086(±0.297)
100	0.1071(±0.0067)	9.731(±0.271)	0.099(±0.0085)	10.076(±0.373)	0.067(±0.0055)	11.721(±0.348)
200	0.0888(±0.0067)	10.527(±0.348)	0.063(±0.0085)	12.022(±0.312)	0.026(±0.0055)	15.856(±0.176)
250	0.0839(±0.0068)	10.771(±0.261)	0.049(±0.0045)	13.132(±0.242)	0.017(±0.0010)	17.775(±0.361)
300	0.0793(±0.0051)	11.012(±0.261)	0.043(±0.0027)	13.686(±0.331)	0.012(±0.0013)	19.369(±0.409)
350	0.0767(±0.0048)	11.161(±0.292)	0.033(±0.0033)	14.852(±0.363)	0.007(±0.0011)	21.482(±0.332)
400	0.0763(±0.0037)	11.181(±0.211)	0.027(±0.0030)	15.745(±0.467)	0.005(±0.0005)	23.173(±0.459)
			FMNIST to MNIST			
Algorithms	**CSGM**		**Sparse-Gen**		**LowRank-Gen**	
Measurements	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)	MSE(±SD)	PSNR(±SD)
50	0.074(±0.0055)	11.29(±0.319)	0.0866(±0.0073)	10.636(±0.368)	0.0787(±0.009)	11.063(±0.504)
75	0.064(±0.0035)	11.902(±0.226)	0.0682(±0.0042)	11.665(±0.268)	0.0567(±0.0019)	12.460(±0.149)
100	0.056(±0.0020)	12.45(±0.156)	0.0576(±0.0027)	12.397(±0.206)	0.0481(±0.0030)	13.185(±0.276)
200	0.049(±0.0025)	13.09(±0.222)	0.0295(±0.0013)	15.302(±0.199)	0.0191(±0.0011)	17.190(±0.252)
250	0.047(±0.0007)	13.22(±0.0694)	0.017(±0.0022)	17.544(±0.539)	0.0106(±0.0009)	19.728(±0.384)
300	0.046(±0.0020)	13.30(±0.1821)	0.0087(±0.0012)	20.595(±0.607)	0.00568(±0.0005)	22.464(±0.381)
350	0.045(±0.0007)	13.39(±0.0731)	0.0029(±0.0005)	25.313(±0.806)	0.00229(±0.0003)	26.403(±0.348)
400	0.045(±0.0008)	13.40(±0.0789)	0.0011(±0.0001)	29.254(±0.597)	0.00067(±0.00015)	31.699(±0.431)

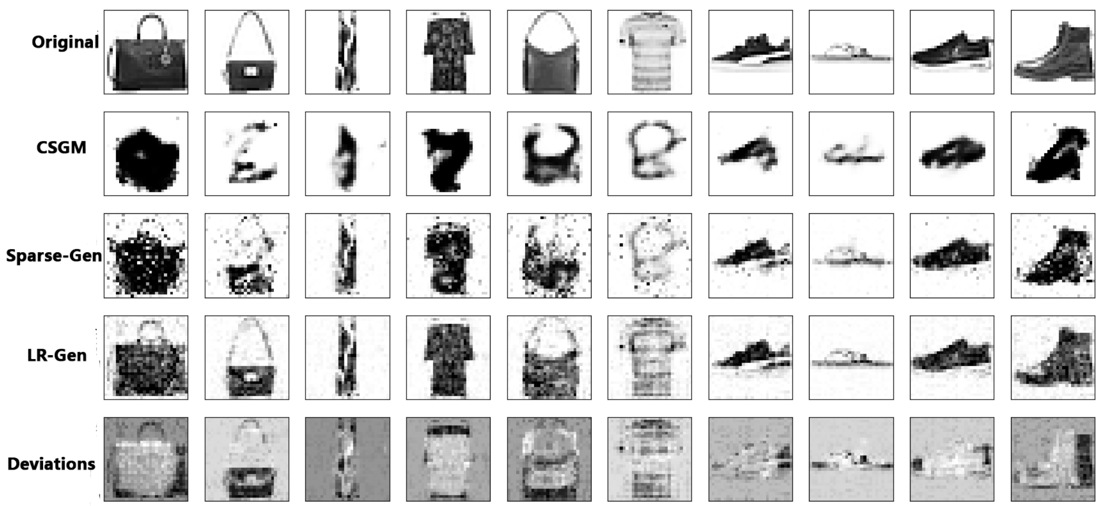
Noise Tolerance

In order to show the noise resistance of our algorithm more intuitively, we add noise with different standard deviations to the measurement, and select an image as the display. Due to the noise tolerance of SVP algorithm is poor, we do not consider this method in this experiment and we add LASSO with wavelet based method.

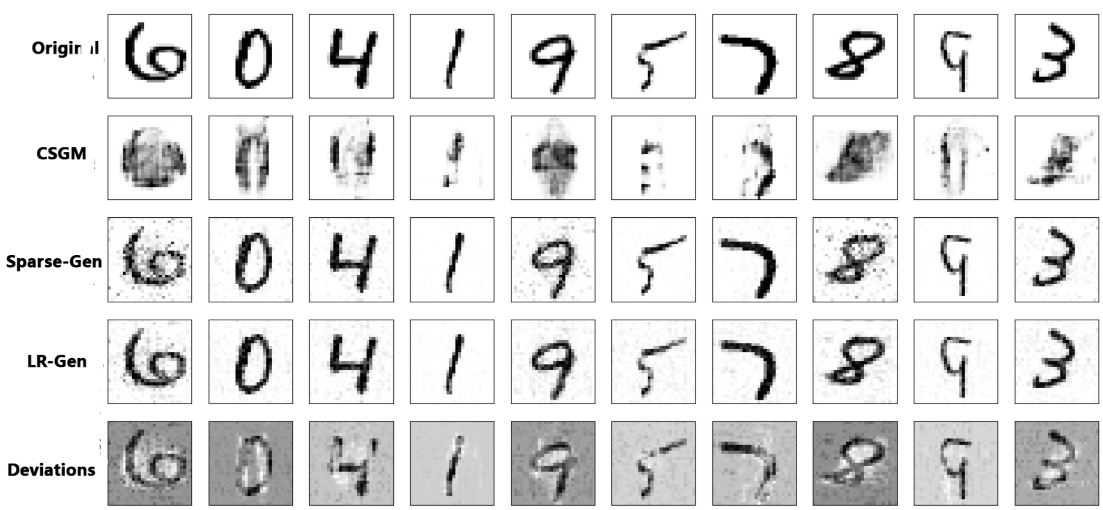
Subtle differences are hard to detect. We can clearly see the noise resistance ability of our algorithm from the PSNR and MSE graph in Fig.6.



(a) Transfer recovery of MNIST (source) to Omniglot (target)



(b) Transfer recovery of MNIST (source) to FMNIST (target)



(c) Transfer recovery of FMNIST (source) to MNIST (target)

Fig. 4: Reconstruction results of transfer recovery with measurement $m = 300$.Top to bottom rows are recovered images by different methods: original image, CSGM, Sparse-Gen, Low-Rank-Gen and deviations term of Low-Rank-Gen.

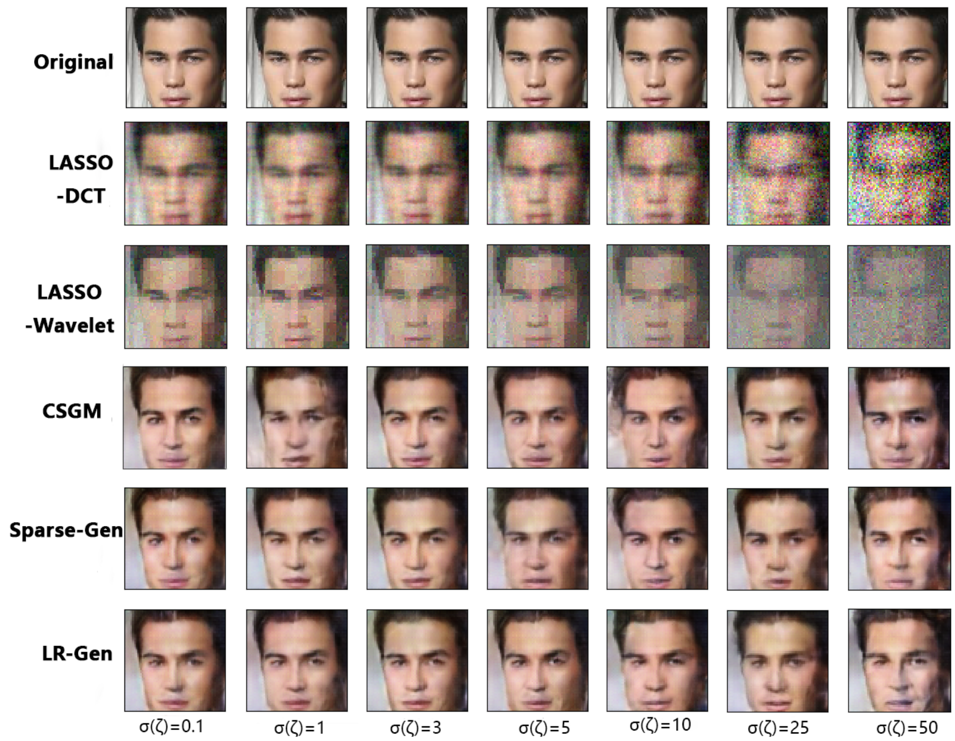


Fig. 5: Reconstruction results with $m = 2500$ and different standard deviation of noises on CelebA. Top to bottom rows are recoverd images by different methods : original image, LASSO with DCT basis, LASSO with wavelet basis, CSGM, Sparse-Gen and Low-Rank-Gen.

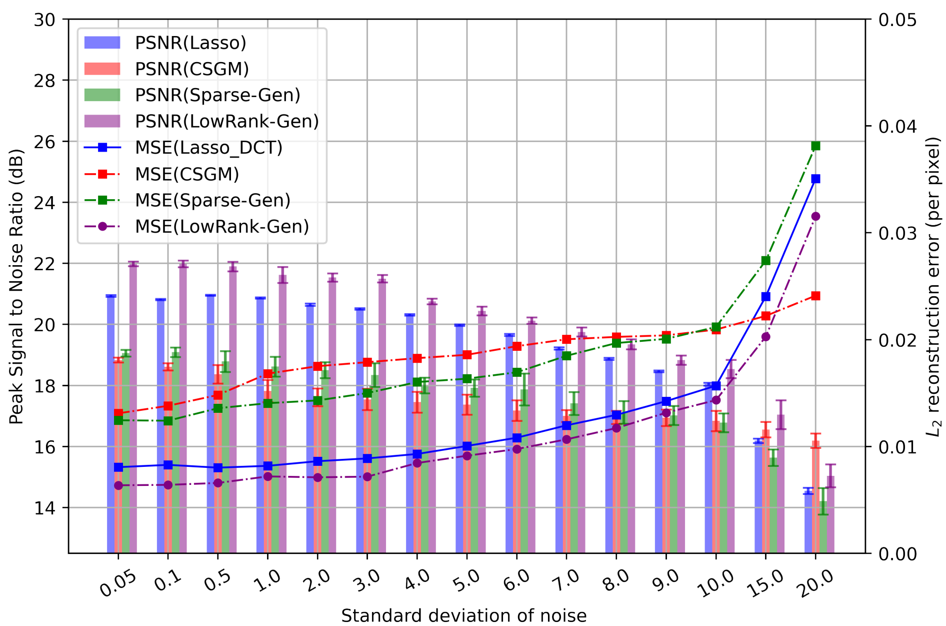


Fig. 6: PSNR and MSE results of reconstruction images in Fig.5.