Supplementary Material
Non-Iterative Recovery from Nonlinear Observations using Generative Models (CVPR 2022)

Jiulong Liu and Zhaoqiang Liu

A. Introduction

This document presents the supplementary materials omitted from the main paper due to the space limit. In Appendix B, we provide the proofs of auxiliary lemmas for Theorem 1. In Appendix C, we provide the proof of Corollary 1. We present experimental results with adversarial noises in Appendix D, and we present visualization of samples generated from the pre-trained generative models in Appendix E. Numbered citations refer to the reference list in the main paper.

B. Proof of Theorem 1 (Recovery Guarantee for OneShot)

Before providing the proof, we present some useful auxiliary results.

B.1. Auxiliary Results for Proving Theorem 1

First, we have a simple tail bound for a random Gaussian variable.

Lemma 2. (Gaussian tail bounds [63, Example 2.1]) *Suppose that $X \sim \mathcal{N}(\alpha, \sigma^2)$ is a random Gaussian variable with mean α and variance σ^2. Then, for any $t > 0$,*

$$\mathbb{P}(|X-\alpha| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}. \tag{23}$$

Based on Lemma 2, we provide the proof of Lemma 1.

Proof of Lemma 1. Since $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a matrix with i.i.d. standard Gaussian entries, $\mathbf{P}\mathbf{A}^T$ is independent with $\mathbf{P}^\perp\mathbf{A}^T$. Note that the columns of $\mathbf{P}\mathbf{A}^T$ are $\langle \mathbf{a}_i, \mathbf{x}\rangle\mathbf{x}$, $y_i = f_i(\langle \mathbf{a}_i, \mathbf{x}\rangle)$ with f_i being i.i.d. realizations of f, and the columns of $\mathbf{P}^\perp\mathbf{A}^T$ are $(\mathbf{I}_n - \mathbf{x}\mathbf{x}^T)\mathbf{a}_i$. Therefore, for any fixed $\mathbf{s}$, $\langle(\mathbf{I}_n - \mathbf{x}\mathbf{x}^T)\mathbf{a}_i, \mathbf{s}\rangle$ and y_i are mutually independent. In addition,

$$\frac{1}{m}\sum_{i=1}^m y_i \left\langle \mathbf{P}^\perp \mathbf{a}_i, \mathbf{s}\right\rangle = \frac{1}{m}\sum_{i=1}^m y_i \left\langle \left(\mathbf{I}_n - \mathbf{x}\mathbf{x}^T\right)\mathbf{a}_i, \mathbf{s}\right\rangle = \frac{1}{m}\sum_{i=1}^m y_i(\langle \mathbf{a}_i, \mathbf{s}\rangle - \langle \mathbf{x}, \mathbf{s}\rangle\langle \mathbf{a}_i, \mathbf{x}\rangle) \tag{24}$$

$$= \frac{\|\mathbf{s}\|_2}{m}\sum_{i=1}^m y_i(\langle \mathbf{a}_i, \bar{\mathbf{s}}\rangle - \langle \mathbf{x}, \bar{\mathbf{s}}\rangle\langle \mathbf{a}_i, \mathbf{x}\rangle), \tag{25}$$

where $\bar{\mathbf{s}} = \mathbf{s}/\|\mathbf{s}\|_2$. Let $g_i = \langle \mathbf{a}_i, \mathbf{x}\rangle \sim \mathcal{N}(0,1)$. Then, since $\mathrm{Cov}[\langle \mathbf{a}_i, \bar{\mathbf{s}}\rangle, g_i] = \langle \mathbf{x}, \bar{\mathbf{s}}\rangle$, $\langle \mathbf{a}_i, \bar{\mathbf{s}}\rangle$ can be written as

$$\langle \mathbf{a}_i, \bar{\mathbf{s}}\rangle = \langle \mathbf{x}, \bar{\mathbf{s}}\rangle g_i + \sqrt{1 - \langle \mathbf{x}, \bar{\mathbf{s}}\rangle^2}\, t_i, \tag{26}$$

where $t_i \sim \mathcal{N}(0,1)$ is independent with g_i. Then, we obtain

$$\frac{\|\mathbf{s}\|_2}{m}\sum_{i=1}^m y_i(\langle \mathbf{a}_i, \bar{\mathbf{s}}\rangle - \langle \mathbf{x}, \bar{\mathbf{s}}\rangle\langle \mathbf{a}_i, \mathbf{x}\rangle) = \frac{\sqrt{\|\mathbf{s}\|_2^2 - \langle \mathbf{x}, \mathbf{s}\rangle^2}}{m}\sum_{i=1}^m y_i t_i, \tag{27}$$

with t_i being standard normal random variables that are independent of y_i. Conditioned on the event $\mathcal{E}$, combining (25) and (27), we have that $\frac{1}{m}\sum_{i=1}^m y_i \left\langle \mathbf{P}^\perp \mathbf{a}_i, \mathbf{s}\right\rangle$ is zero-mean Gaussian with the variance being

$$\frac{\left(\|\mathbf{s}\|_2^2 - \langle \mathbf{x}, \mathbf{s}\rangle^2\right)\sum_{i=1}^m y_i^2}{m^2}. \tag{28}$$

From Lemma 2, we have for any $u > 0$ that

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m y_i \left\langle \mathbf{P}^\perp \mathbf{a}_i, \mathbf{s}\right\rangle\right| \geq u\right) = \exp\left(-\Omega\left(\frac{mu^2}{\left(\|\mathbf{s}\|_2^2 - \langle \mathbf{x}, \mathbf{s}\rangle^2\right)\sum_{i=1}^m y_i^2/m}\right)\right) \tag{29}$$

$$= \exp\left(-\Omega\left(\frac{mu^2}{\|\mathbf{s}\|_2^2\xi^2}\right)\right), \tag{30}$$

where (30) follows from $\|\mathbf{s}\|_2^2 - \langle \mathbf{x}, \mathbf{s}\rangle^2 \le \|\mathbf{s}\|_2^2$ and $\sum_{i=1}^m y_i^2/m \le 2\xi^2$. Setting $\varepsilon = \frac{mu^2}{\|\mathbf{s}\|_2^2\xi^2}$, we have $u = \frac{\xi\|\mathbf{s}\|_2\sqrt{\varepsilon}}{\sqrt{m}}$, and that the desired inequality (13) holds. □

Next, from the Chebyshev's inequality, we have the following simple lemma, which gives (12) and an inequality that is useful in the proof of Theorem 1.

Lemma 3. *For any $t > 0$, with probability at least $1 - \frac{\rho^2}{mt^2}$ (cf. (5)),*

$$\left|\frac{1}{m}\sum_{i=1}^m y_i\langle \mathbf{a}_i, \mathbf{x}\rangle - \mu\right| < t. \tag{31}$$

Similarly, with probability at least $1 - \frac{\theta^4}{m\xi^4}$ (cf. (4) and (6)),

$$\frac{1}{m}\sum_{i=1}^m y_i^2 \le 2\xi^2. \tag{32}$$

Proof. Let $X = \frac{1}{m}\sum_{i=1}^m y_i\langle \mathbf{a}_i, \mathbf{x}\rangle - \mu$ and $X_i = y_i\langle \mathbf{a}_i, \mathbf{x}\rangle - \mu$. Then, we have $\mathbb{E}[X] = 0$ and $\mathrm{Var}[X] = \mathrm{Var}\big[\sum_{i=1}^m X_i/m\big] = \mathrm{Var}[X_1]/m = \rho^2/m$. From the Chebyshev's inequality, we obtain the desired inequality (31). Similarly, we have $\mathrm{Var}[\sum_{i=1}^m (y_i^2 - \xi^2)/m] = \mathrm{Var}[\sum_{i=1}^m y_i^2/m] = \mathrm{Var}[y_1^2]/m = \theta^4/m$. From the Chebyshev's inequality, we have

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m y_i^2 - \xi^2\right| > \xi^2\right) \le \frac{\mathrm{Var}\left[\sum_{i=1}^m (y_i^2 - \xi^2)/m\right]}{\xi^4} = \frac{\theta^4}{m\xi^4}, \tag{33}$$

which gives (32). □

Now we are ready to present the proof of Theorem 1.

B.2. Proof of Theorem 1

Since $\hat{\mathbf{x}} = \mathcal{P}_G\big(\frac{1}{m}\mathbf{A}^T\mathbf{y}\big)$ and $\mu\mathbf{x} \in \mathcal{R}(G)$, we have

$$\left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \hat{\mathbf{x}}\right\|_2 \le \left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}\right\|_2. \tag{34}$$

Taking square on both sides, we obtain

$$\left\|\left(\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}\right) + (\mu\mathbf{x} - \hat{\mathbf{x}})\right\|_2^2 \le \left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}\right\|_2^2, \tag{35}$$

which leads to

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2^2 \le 2\left\langle \frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle. \tag{36}$$

For $\mathbf{P}^\perp = \mathbf{I}_n - \mathbf{x}\mathbf{x}^T$, we have

$$\frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x} = \frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y} + \frac{1}{m}\mathbf{x}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\mathbf{x} \tag{37}$$

$$= \frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y} + \left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}. \tag{38}$$

Recall that $\mathcal{E}$ is the event that $\frac{1}{m}\sum_{i=1}^m y_i^2 \le 2\xi^2$, with ξ^2 being defined in (4). For any $u > 0$,

$$\mathbb{P}\left(\left|\left\langle \frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| > u\right) \le \mathbb{P}(\mathcal{E}^c) + \mathbb{P}\left(\left|\left\langle \frac{1}{m}\mathbf{P}^\perp\mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| > u \middle| \mathcal{E}\right). \tag{39}$$

From Lemma 3, we have $\mathbb{P}(\mathcal{E}^c) \le \frac{\theta^4}{m\xi^4}$, where θ^4 is defined in (6). Then, setting $u = C\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right)(\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 + \delta)$ with $C > 0$ being sufficiently large, from Lemma 1 and a chaining argument similar to that in [4], we have with probability $1 - e^{-\Omega\left(k\log\frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4}$ that[4]

$$\left|\left\langle \frac{1}{m}\mathbf{P}^{\perp}\mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| \le u. \tag{40}$$

Moreover, we have

$$\left|\left\langle \left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| \le \left|\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right| \cdot \|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2. \tag{41}$$

Then, from Lemma 3, for $\epsilon > 0$ and the ρ^2 defined in (5), we obtain with probability at least $1 - \frac{\rho^2}{m\epsilon^2}$ that

$$\left|\left\langle \left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| \le \epsilon\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2. \tag{42}$$

Setting $\epsilon = \xi\sqrt{(k\log\frac{Lr}{\delta})/m}$, we obtain with probability at least $1 - \frac{\rho^2}{\xi^2 k\log\frac{Lr}{\delta}}$ that

$$\left|\left\langle \left(\frac{1}{m}\mathbf{x}^T\mathbf{A}^T\mathbf{y} - \mu\right)\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| \le \xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}} \cdot \|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2. \tag{43}$$

Combining (38), (40) and (43), we obtain with probability $1 - e^{-\Omega\left(k\log\frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4} - \frac{\rho^2}{\xi^2 k\log\frac{Lr}{\delta}}$ that

$$\left|\left\langle \frac{1}{m}\mathbf{A}^T\mathbf{y} - \mu\mathbf{x}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right)(\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 + \delta). \tag{44}$$

In addition, from (36), we have

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2^2 = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right)(\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 + \delta). \tag{45}$$

Then, if

$$\delta = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right), \tag{46}$$

we obtain

$$\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right), \tag{47}$$

which completes the proof.

B.3. Proof of (40)

Based on Lemma 1 and a chaining argument similar to that in [4], we arrive at the following lemma that concerns (40).

Lemma 4. *Conditioned on $\mathcal{E}$, we have that for any $\delta > 0$ satisfying $Lr = \Omega(\delta n)$, with probability $1 - e^{-\Omega\left(k\log\frac{Lr}{\delta}\right)}$,*

$$\left|\left\langle \frac{1}{m}\mathbf{P}^{\perp}\mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \mu\mathbf{x}\right\rangle\right| = O\left(\xi\sqrt{\frac{k\log\frac{Lr}{\delta}}{m}}\right)(\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 + \delta). \tag{48}$$

[4] For completeness, the proof of (40) is presented at the end of this section, namely Appendix B.3.

Proof. For fixed $\delta > 0$ and a positive integer ℓ, let $M = M_0 \subseteq M_1 \subseteq \ldots \subseteq M_\ell$ be a chain of nets of $B_2^k(r)$ such that M_i is a $\frac{\delta_i}{L}$-net with $\delta_i = \frac{\delta}{2^i}$. There exists such a chain of nets with

$$\log |M_i| \leq k \log \frac{4Lr}{\delta_i}. \tag{49}$$

By the L-Lipschitz continuity of G, we have for any $i \in [\ell]$ that $G(M_i)$ is a δ_i-net of $\mathcal{R}(G) = G(B_2^k(r))$.

Then, we write

$$\hat{\mathbf{x}} - \mu\mathbf{x} = (\hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell) + \sum_{i=1}^{\ell}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1}) + (\hat{\mathbf{x}}_0 - \mu\mathbf{x}), \tag{50}$$

where $\hat{\mathbf{x}}_i \in G(M_i)$ for all $i \in [\ell]$, and $\|\hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\| \leq \frac{\delta}{2^\ell}$, $\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1}\|_2 \leq \frac{\delta}{2^{i-1}}$ for all $i \in [\ell]$. Therefore, the triangle inequality gives

$$\|\hat{\mathbf{x}} - \hat{\mathbf{x}}_0\|_2 < 2\delta. \tag{51}$$

Setting $\varepsilon = Ck \log \frac{Lr}{\delta}$ with $C > 0$ being a sufficiently large constant in Lemma 1, and taking the union bound over $G(M_0)$, we have that with probability $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)}$, for *all* $\mathbf{s} \in G(M_0)$,

$$\left|\left\langle \frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}, \mathbf{s} - \mu\mathbf{x}\right\rangle\right| = O\left(\frac{\xi\|\mathbf{s} - \mu\mathbf{x}\|_2\sqrt{k \log \frac{Lr}{\delta}}}{\sqrt{m}}\right), \tag{52}$$

which gives

$$\left|\left\langle \frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}}_0 - \mu\mathbf{x}\right\rangle\right| = O\left(\frac{\xi\|\hat{\mathbf{x}}_0 - \mu\mathbf{x}\|_2\sqrt{k \log \frac{Lr}{\delta}}}{\sqrt{m}}\right). \tag{53}$$

In addition, similarly to that in [4, 37], we have that if setting $\ell = \lceil \log n \rceil$, when $Lr = \Omega(\delta n)$, with probability $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)}$, it holds that

$$\sum_{i=1}^{\ell}\left|\left\langle \frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1}\right\rangle\right| = O\left(\frac{\xi\delta\sqrt{k \log \frac{Lr}{\delta}}}{\sqrt{m}}\right). \tag{54}$$

Moreover, for any $\varepsilon > 0$ used in Lemma 1, taking a union bound over $\mathbf{s} \in \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, we have with probability $1 - ne^{-\Omega(\varepsilon)}$ that

$$\left\|\frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}\right\|_\infty \leq \frac{\xi\sqrt{\varepsilon}}{\sqrt{m}}. \tag{55}$$

Then, we have

$$\left|\left\langle \frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\right\rangle\right| \leq \left\|\frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}\right\|_\infty \cdot \|\hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\|_1 \tag{56}$$

$$\leq \frac{\xi\sqrt{\varepsilon}}{\sqrt{m}} \cdot \sqrt{n}\|\hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\|_2 \tag{57}$$

$$= O\left(\frac{\xi\delta\sqrt{\varepsilon}}{\sqrt{m}}\right), \tag{58}$$

where we use $\|\hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\|_2 \leq \frac{\delta}{2^\ell}$ and the setting $\ell = \lceil \log n \rceil$ in (58). Setting $\varepsilon = Ck \log \frac{Lr}{\delta}$ with C being a sufficiently large positive constant in (58), we obtain with probability $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)}$ that

$$\left|\left\langle \frac{1}{m}\mathbf{P}^\perp \mathbf{A}^T\mathbf{y}, \hat{\mathbf{x}} - \hat{\mathbf{x}}_\ell\right\rangle\right| = O\left(\frac{\xi\delta\sqrt{k \log \frac{Lr}{\delta}}}{\sqrt{m}}\right). \tag{59}$$

Combining (50), (53), (54), and (59), we obtain the desired result. $\square$

C. Proof of Corollary 1 (Extension of Theorem 1)

Before providing the proof of Corollary 1, we present the following useful lemma.

Lemma 5. ([37, Lemma 2]) *Let $G : B_2^k(r) \to \mathbb{R}^n$ be L-Lipschitz and $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. For $\alpha > 0$ and $\delta > 0$, if $m = \Omega\left(\frac{k}{\alpha^2} \log \frac{Lr}{\delta}\right)$, then with probability $1 - e^{-\Omega(\alpha^2 m)}$, we have for all $\mathbf{x}_1, \mathbf{x}_2 \in G(B_2^k(r))$ that*

$$\frac{1}{\sqrt{m}} \|\mathbf{A}\mathbf{x}_1 - \mathbf{A}\mathbf{x}_2\|_2 \leq (1+\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|_2 + \delta. \tag{60}$$

Proof of Corollary 1. Since $\hat{\mathbf{x}} = \mathcal{P}_G\left(\frac{1}{m}\mathbf{A}^T\mathbf{y}\right)$ and $\tilde{\mathbf{x}} \in \mathcal{R}(G)$, we have

$$\left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \hat{\mathbf{x}}\right\|_2 \leq \left\|\frac{1}{m}\mathbf{A}^T\mathbf{y} - \tilde{\mathbf{x}}\right\|_2. \tag{61}$$

Then, similarly to (36), we obtain

$$\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 \leq 2\left\langle \frac{1}{m}\mathbf{A}^T\mathbf{y} - \tilde{\mathbf{x}}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle. \tag{62}$$

Let $\tilde{\mathbf{y}} = [f_1(\langle \mathbf{a}_1, \mathbf{x}\rangle), \ldots, f_m(\langle \mathbf{a}_m, \mathbf{x}\rangle)]^T \in \mathbb{R}^m$. We have

$$\left|\left\langle \frac{1}{m}\mathbf{A}^T\mathbf{y} - \tilde{\mathbf{x}}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle\right| \leq \left|\left\langle \frac{1}{m}\mathbf{A}^T(\mathbf{y} - \tilde{\mathbf{y}}), \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle\right| + \left|\left\langle \frac{1}{m}\mathbf{A}^T\tilde{\mathbf{y}} - \mu\mathbf{x}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle\right| + |\langle \mu\mathbf{x} - \tilde{\mathbf{x}}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\rangle|. \tag{63}$$

Setting $\alpha = 0.5$ in Lemma 5, we obtain that when $m = \Omega(k \log \frac{Lr}{\delta})$, with probability $1 - e^{-\Omega(m)}$,

$$\begin{aligned}
\left|\left\langle \frac{1}{m}\mathbf{A}^T(\mathbf{y} - \tilde{\mathbf{y}}), \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle\right| &= \left|\left\langle \frac{1}{\sqrt{m}}(\mathbf{y} - \tilde{\mathbf{y}}), \frac{1}{\sqrt{m}}\mathbf{A}(\hat{\mathbf{x}} - \tilde{\mathbf{x}})\right\rangle\right| && (64)\\
&\leq \left\|\frac{1}{\sqrt{m}}(\mathbf{y} - \tilde{\mathbf{y}})\right\|_2 \cdot \left\|\frac{1}{\sqrt{m}}\mathbf{A}(\hat{\mathbf{x}} - \tilde{\mathbf{x}})\right\|_2 && (65)\\
&\leq \nu \cdot O\left(\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 + \delta\right). && (66)
\end{aligned}$$

Similarly to (44), we obtain with probability $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4} - \frac{\rho^2}{\xi^2 k \log \frac{Lr}{\delta}}$ that

$$\left|\left\langle \frac{1}{m}\mathbf{A}^T\tilde{\mathbf{y}} - \mu\mathbf{x}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\right\rangle\right| = O\left(\xi\sqrt{\frac{k \log \frac{Lr}{\delta}}{m}}\right)(\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 + \delta). \tag{67}$$

From the triangle inequality, we have

$$|\langle \mu\mathbf{x} - \tilde{\mathbf{x}}, \hat{\mathbf{x}} - \tilde{\mathbf{x}}\rangle| \leq \|\mu\mathbf{x} - \tilde{\mathbf{x}}\|_2 \cdot \|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2. \tag{68}$$

Combining (62), (63), (66), (67) and (68), we obtain that when $m = \Omega\left(k \log \frac{Lr}{\delta}\right)$, with probability $1 - e^{-\Omega\left(k \log \frac{Lr}{\delta}\right)} - \frac{\theta^4}{m\xi^4} - \frac{\rho^2}{\xi^2 k \log \frac{Lr}{\delta}}$,

$$\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 \leq \left(\xi\sqrt{\frac{k \log \frac{Lr}{\delta}}{m}} + \nu + \|\mu\mathbf{x} - \tilde{\mathbf{x}}\|_2\right)(\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 + \delta). \tag{69}$$

From the triangle inequality $\|\hat{\mathbf{x}} - \mu\mathbf{x}\|_2 \leq \|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 + \|\tilde{\mathbf{x}} - \mu\mathbf{x}\|_2$, and similarly to (47), we obtain the desired result. □

D. Supplementary Experimental Results with Adversarial Noise

The supplementary numerical results for the SIM (*cf.* (2)) with adversarial noise are presented in this section, for a noisy 1-bit measurement model

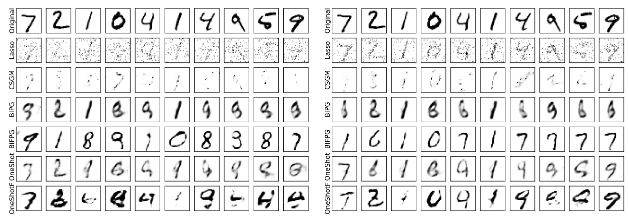
$$y_i = \mathrm{sign}(\langle \mathbf{a}_i, \mathbf{x} + e_i \rangle), \quad i \in [m], \tag{70}$$

where e_i are i.i.d. realizations of $\mathcal{N}\big(0, \sigma^2\big)$, and a noisy cubic measurement model

$$y_i = \langle \mathbf{a}_i, \mathbf{x} + \eta_i \rangle^3, \quad i \in [m], \tag{71}$$

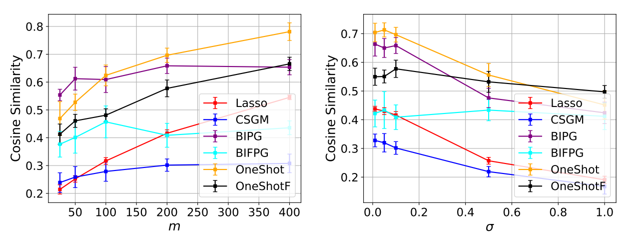
where η_i are i.i.d. realizations $\mathcal{N}\big(0, \sigma^2\big)$.

The reconstructed results from 1-bit and cubic measurements are shown in Figures 10 and 12 respectively. We can observe that `OneShot` outperforms `Lasso`, `CSGM`, `BIPG` and `BIFPG` (or `PGD` and `FPGD`) by a large margin and it also leads to superior performance over `OneShotF`. In addition, the cosine similarities plotted in Figures 11 and 13 illustrate that our method `OneShot` mostly outperforms all other competing methods.



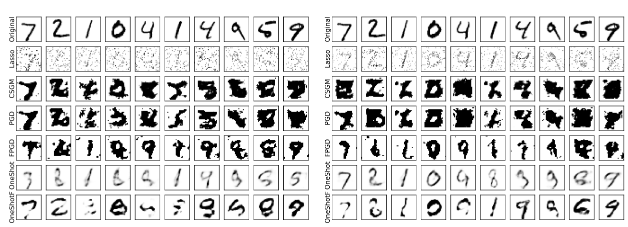
(a) $\sigma = 0.1$ and $m = 200$ (b) $\sigma = 0.01$ and $m = 400$

Figure 10. Examples of reconstructed images from adversarially corrupted 1-bit measurements on MNIST images.



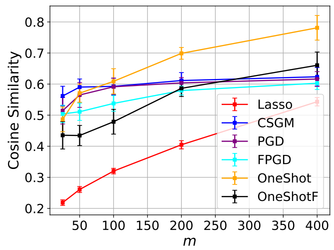
(a) Fixing $\sigma = 0.1$ and varying m (b) Fixing $m = 200$ and varying σ

Figure 11. Quantitative comparisons according to the cosine similarity for adversarially corrupted 1-bit measurements on MNIST images.

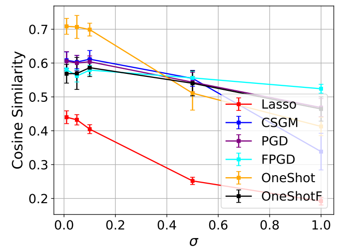


(a) $\sigma = 0.1$ and $m = 200$ (b) $\sigma = 0.01$ and $m = 400$

Figure 12. Examples of reconstructed images from adversarially corrupted cubic measurements on MNIST images.



(a) Fixing $\sigma = 0.1$, varying m



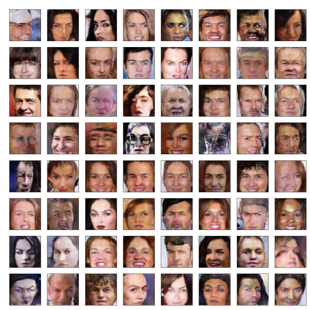
(b) Fixing $m = 200$, varying σ

Figure 13. Quantitative comparisons according to the cosine similarity for adversarially corrupted cubic measurements on MNIST images.

E. Visualization of Samples Generated from the Pre-trained Generative Models

The samples generated from the pre-trained VAE and the pre-trained DCGAN that are used for the gradient-based projection method in this paper are shown in Figure 14. Though some samples generated from the two classic generative models are not perfect and they are distinguishable from images in the datasets, our proposed method still achieves the SOTA performance. We may investigate our method using the SOTA generative models such as those in [17,27,54,55] in future works to further improve the performance.

(a) Samples generated from the pre-trained VAE



(b) Samples generated from the pre-trained DCGAN

Figure 14. Visualization of samples generated from the pre-trained generative models.