Supplementary Material

Misspecified Phase Retrieval with Generative Priors (NeurIPS 2022)

Zhaoqiang Liu, Xinshao Wang, and Jiulong Liu

A Auxiliary Results

In this section, we first provide some useful auxiliary results that are general, and then some that are specific to our setup.

A.1 General Auxiliary Results

First, we state the following standard definitions for a sub-Gaussian random variable and the associated sub-Gaussian norm.

Definition 2. *A random variable X is said to be sub-Gaussian if there exists a positive constant C such that $\left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \le C\sqrt{p}$ for all $p \ge 1$. The sub-Gaussian norm of a sub-Gaussian random variable X is defined as $\|X\|_{\psi_2} := \sup_{p\ge 1} p^{-1/2}\left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}$.*

Recall that the definitions of a sub-exponential random variable and the associated sub-exponential norm have been provided in Definition 1 in the main document. The following lemma states that the product of two sub-Gaussian random variables is sub-exponential.

Lemma 1. ([87, Lemma 2.7.7]) *Let X and Y be sub-Gaussian random variables (not necessarily independent). Then XY is sub-exponential, and satisfies*

$$\|XY\|_{\psi_1} \le \|X\|_{\psi_2} \cdot \|Y\|_{\psi_2}. \tag{26}$$

We consider sub-Weibull random variables that generalize sub-Gaussian and sub-exponential random variables.

Definition 3. *For any $\alpha > 0$, a random variable X is said to be sub-Weibull of order α if it has a bounded ψ_α-norm, where the ψ_α-norm of X is defined as*

$$\|X\|_{\psi_\alpha} := \inf\left\{K \in (0,\infty)\,:\, \mathbb{E}\left[\exp\left(|X|^\alpha/K^\alpha\right)\right] \le 2\right\}. \tag{27}$$

In particular, when $\alpha = 2$ or 1, sub-Weibull random variables reduce to sub-Gaussian or sub-exponential random variables respectively. The smaller the α is, the heavier tail a sub-Weibull random variable has. Moreover, it follows readily from Definition 3 that X is sub-exponential if and only if $|X|^{1/\alpha}$ is sub-Weibull of order α. We have the following concentration inequality for the sum of independent sub-Weibull random variables.

Lemma 2. ([33, Theorem 3.1]) *Suppose that $X_1, X_2, \dots, X_N$ are independent sub-Weibull random variables that are of order α, and $K = \max_i \|X_i\|_{\psi_\alpha}$. Then, there exists a positive constant C_α only depending on α such that for any $\mathbf{b} = [b_1, b_2, \dots, b_N]^T \in \mathbb{R}^N$ and $u > 2$, with probability at least $1 - e^{-u}$, it holds that*

$$\left|\sum_{i=1}^N b_i X_i - \mathbb{E}\left[\sum_{i=1}^N b_i X_i\right]\right| \le C_\alpha K\left(\|\mathbf{b}\|_2 \cdot \sqrt{u} + \|\mathbf{b}\|_\infty \cdot u^{1/\alpha}\right). \tag{28}$$

In addition, we have the following lemma concerning the Two-sided Set-Restricted Eigenvalue Condition (TS-REC).

Lemma 3. ([51, Lemma 2]) *Let $G\,:\,B^k(r) \to \mathbb{R}^n$ be L-Lipschitz continuous and $\mathbf{a}_1, \dots, \mathbf{a}_m$ be i.i.d. realizations of $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. For $\varepsilon \in (0,1)$ and $\delta > 0$, if $m = \Omega\left(\frac{k}{\varepsilon^2}\log\frac{Lr}{\delta}\right)$, then with probability $1 - e^{-\Omega(\varepsilon^2 m)}$, the following holds for all $\mathbf{x}_1, \mathbf{x}_2 \in G(B^k(r))$:*

$$(1-\varepsilon)\|\mathbf{x}_1 - \mathbf{x}_2\|_2 - \delta \le \frac{1}{\sqrt{m}} \cdot \sqrt{\sum_{i=1}^m \left(\mathbf{a}_i^T(\mathbf{x}_1 - \mathbf{x}_2)\right)^2} \le (1+\varepsilon)\|\mathbf{x}_1 - \mathbf{x}_2\|_2 + \delta. \tag{29}$$

A.2 Auxiliary Results for Our Setup

From Chebyshev's inequality and the definition of a sub-exponential random variable (*cf.* Definition 1), as well as Lemma 2, we obtain the following lemma. Here and in subsequent results where it is clear from the context, for simplicity of presentation, we think of $\mathbf{a}_i$ and y_i as random variables, instead of realizations of corresponding random variables.

Lemma 4. *When $m \ge \log^3 m$, the event*

$$
\begin{aligned}
\mathcal{E}: \quad & \max_{i\in[m]} |y_i| \le 5K_y \cdot \log m, \quad \frac{1}{m}\sum_{i=1}^m y_i^2 \le 8K_y^2, \quad \frac{1}{m}\sum_{i=1}^m y_i^2(\mathbf{a}_i^T\mathbf{x})^2 \le 32\sqrt{3}K_y^2, \\
& \left|\frac{1}{m}\sum_{i=1}^m (\mathbf{a}_i^T\mathbf{x})^2 - 1\right| \le C\sqrt{\frac{\log m}{m}}, \quad \left|\frac{1}{m}\sum_{i=1}^m y_i - M_y\right| \le CK_y \cdot \sqrt{\frac{\log m}{m}}, \\
& \left|\frac{1}{m}\sum_{i=1}^m y_i(\mathbf{a}_i^T\mathbf{x})^2 - (\nu + M_y)\right| \le CK_y \cdot \sqrt{\frac{\log m}{m}},
\end{aligned}
\tag{30}
$$

occurs with probability $1 - O(1/m)$, where $\nu := \mathrm{Cov}\left[y, (\mathbf{a}^T\mathbf{x})^2\right]$ (cf. (7))*, $M_y := \mathbb{E}[y]$, $K_y := \|y\|_{\psi_1}$* (cf. *Section 2.2), and C is an absolute constant.*

Proof. Since y_i is assumed to be sub-exponential with the sub-exponential norm being K_y, from the definition of a sub-exponential random variable, we obtain for any $i \in [m]$ and $u > 0$ that

$$\mathbb{P}(|y_i| > u) \le \exp(1 - u/K_y). \tag{31}$$

Then, setting $u = 5K_y \cdot \log m$ and taking a union bound over $i \in [m]$, we obtain with probability at least $1 - \frac{e}{m^4}$ that

$$\max_{i\in[m]} |y_i| \le 5K_y \cdot \log m. \tag{32}$$

Note that from

$$K_y = \|y\|_{\psi_1} = \sup_{p\ge 1} p^{-1} \left(\mathbb{E}\left[|y|^p\right]\right)^{1/p}, \tag{33}$$

we obtain

$$|M_y| \le \mathbb{E}[|y|] \le K_y, \quad \mathbb{E}[y^2] \le (2K_y)^2, \quad \mathbb{E}[y^4] \le (4K_y)^4, \quad \mathbb{E}[y^8] \le (8K_y)^8. \tag{34}$$

In addition, from Chebyshev's inequality, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m y_i^2 - \mathbb{E}[y^2]\right| \ge \epsilon\right) \le \frac{\mathrm{Var}[y^2]}{m\epsilon^2}. \tag{35}$$

Setting $\epsilon = 4K_y^2$, we obtain

$$\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m y_i^2 \ge 8K_y^2\right) \le \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m y_i^2 \ge \mathbb{E}[y^2] + 4K_y^2\right) \le \frac{\mathrm{Var}[y^2]}{16mK_y^4} \le \frac{\mathbb{E}[y^4]}{16mK_y^4} \le \frac{16}{m}, \tag{36}$$

where we use (34) in the first and last inequalities. Moreover, from

$$\mathbb{E}\left[y_i^2(\mathbf{a}_i^T\mathbf{x})^2\right] \le (\mathbb{E}[y_i^4])^{1/2}(\mathbb{E}[(\mathbf{a}_i^T\mathbf{x})^4])^{1/2} = (\mathbb{E}[y_i^4])^{1/2}(\mathbb{E}[g^4])^{1/2} \le (4K_y)^2 \cdot \sqrt{3} = 16\sqrt{3}K_y^2, \tag{37}$$

where $g \sim \mathcal{N}(0,1)$ represents a standard normal random variable, similarly to (36), we obtain

$$
\begin{aligned}
\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m y_i^2(\mathbf{a}_i^T\mathbf{x})^2 \ge 32\sqrt{3}K_y^2\right) &\le \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m y_i^2(\mathbf{a}_i^T\mathbf{x})^2 \ge \mathbb{E}\left[y^2(\mathbf{a}^T\mathbf{x})^2\right] + 16\sqrt{3}K_y^2\right) && (38)\\
&\le \frac{\mathrm{Var}\left[y^2(\mathbf{a}^T\mathbf{x})^2\right]}{768mK_y^4} \le \frac{\mathbb{E}[y^4(\mathbf{a}^T\mathbf{x})^4]}{768mK_y^4} && (39)\\
&\le \frac{(\mathbb{E}[y^8])^{1/2} \cdot (\mathbb{E}[g^8])^{1/2}}{768mK_y^4} && (40)\\
&\le \frac{(8K_y)^4 \cdot \sqrt{105}}{768mK_y^4} = \frac{16\sqrt{105}}{3m}, && (41)
\end{aligned}
$$

where $g \sim \mathcal{N}(0,1)$ and (41) follows from (34) . Furthermore, since $\mathbb{E}[y(\mathbf{a}^T\mathbf{x})^2] = \mathrm{Cov}[y, (\mathbf{a}^T\mathbf{x})^2] + \mathbb{E}[y] \cdot \mathbb{E}[(\mathbf{a}^T\mathbf{x})^2] = \nu + M_y$ and $y(\mathbf{a}^T\mathbf{x})^2$ is sub-Weibull of order $\alpha = 1/2$ with the corresponding constant $C_\alpha \le CK_y$,[5] from Lemma 2, we obtain that for any $u > 2$, with probability at least $1 - e^{-u}$,

$$\left| \frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i^T\mathbf{x})^2 - (\nu + M_y) \right| \le C' K_y \left(\sqrt{\frac{u}{m}} + \frac{u^2}{m} \right). \tag{42}$$

Setting $u = \log m$, we obtain that when $m \ge \log^3 m$, with probability at least $1 - 1/m$,

$$\left| \frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i^T\mathbf{x})^2 - (\nu + M_y) \right| \le 2C' K_y \cdot \sqrt{\frac{\log m}{m}}. \tag{43}$$

Since y is sub-exponential with $\|y\|_{\psi_1} = K_y$ and $(\mathbf{a}^T\mathbf{x})^2$ is sub-exponential with the sub-exponential norm being upper bounded by C, similarly to (43), we have with probability at least $1 - 2/m$ that

$$\left| \frac{1}{m} \sum_{i=1}^m y_i - M_y \right| \le CK_y \cdot \sqrt{\frac{\log m}{m}}, \tag{44}$$

$$\left| \frac{1}{m} \sum_{i=1}^m (\mathbf{a}_i^T\mathbf{x})^2 - 1 \right| \le C\sqrt{\frac{\log m}{m}}. \tag{45}$$

□

B Proof of Theorem 1 (Guarantees for the First Step of Algorithm 1)

Before proving the theorem, we provide some additional auxiliary results.

B.1 Useful Lemmas for Theorem 1

Recall that $\mathbf{V}$ is defined in (10) and ν is defined in (7). First, we present the following useful lemma.

Lemma 5. *Let $\mathbf{E} = \mathbf{V} - \nu\mathbf{x}\mathbf{x}^T$. For any $u > 2$ satisfying $m = \Omega\big(u \cdot \log m\big)$, conditioned on the event $\mathcal{E}$ (*cf. (30)*), we have for any fixed $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$, with probability $1 - O\big(e^{-u}\big)$ that*

$$\left|\mathbf{s}_1^T\mathbf{E}\mathbf{s}_2\right| = O\left(K_y\sqrt{\frac{u \cdot \log m}{m}}\right) \cdot \|\mathbf{s}_1\|_2 \cdot \|\mathbf{s}_2\|_2. \tag{46}$$

Proof. First, it is easy to calculate that $\mathbb{E}[\mathbf{V}] = \nu\mathbf{x}\mathbf{x}^T$ (see, e.g., [47, Lemma 8]) and thus $\mathbb{E}[\mathbf{E}] = \mathbf{0}$. Without loss of generality, we assume that $\|\mathbf{s}_1\|_2 = \|\mathbf{s}_2\|_2 = 1$, and we also assume that $\mathbf{s}_1 \ne \mathbf{x}$ and $\mathbf{s}_2 \ne \mathbf{x}$.[6] From the definition of $\mathbf{V}$ in (10), we have

$$\mathbf{s}_1^T\mathbf{E}\mathbf{s}_2 = \mathbf{s}_1^T(\mathbf{V} - \nu\mathbf{x}\mathbf{x}^T)\mathbf{s}_2 = \frac{1}{m}\sum_{i=1}^m y_i \left((\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2) - \mathbf{s}_1^T\mathbf{s}_2\right) - \nu(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x}). \tag{47}$$

We focus on dealing with the term

$$I := \frac{1}{m}\sum_{i=1}^m y_i (\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2). \tag{48}$$

We decompose $\mathbf{s}_1$ as

$$\mathbf{s}_1 = (\mathbf{s}_1^T\mathbf{x})\mathbf{x} + \sqrt{1 - (\mathbf{s}_1^T\mathbf{x})^2} \cdot \mathbf{t}_1, \tag{49}$$

---

[5] Since y is sub-exponential with $\|y\|_{\psi_1} = K_y$, from (27), we obtain that $\sqrt{|y|}$ is sub-Gaussian with the sub-Gaussian norm being $\sqrt{K_y}$. From Lemma 1, $\sqrt{|y|} \cdot (\mathbf{a}^T\mathbf{x})$ is sub-exponential with the sub-exponential being upper bounded by $\sqrt{CK_y}$, where C is an absolute constant. Again from (27), we obtain $|y_i| \cdot (\mathbf{a}^T\mathbf{x})^2$ is sub-Weibull of order $\alpha = 1/2$ with the corresponding constant $C_\alpha \le CK_y$.

[6] We will see from the proof that the case that $\mathbf{s}_1 = \mathbf{x}$ or $\mathbf{s}_2 = \mathbf{x}$ is easier to handle.

where $\|\mathbf{t}_1\|_2 = 1$ and $\mathbf{t}_1^T\mathbf{x} = 0$. Similarly, letting $w_{12} = \sqrt{1 - (\mathbf{s}_2^T\mathbf{x})^2 - (\mathbf{s}_2^T\mathbf{t}_1)^2}$, $\mathbf{s}_2$ can be written as

$$\mathbf{s}_2 = (\mathbf{s}_2^T\mathbf{x})\mathbf{x} + (\mathbf{s}_2^T\mathbf{t}_1)\mathbf{t}_1 + w_{12}\mathbf{t}_2, \tag{50}$$

where $\|\mathbf{t}_2\|_2 = 1$ and $\mathbf{t}_2^T\mathbf{x} = \mathbf{t}_2^T\mathbf{t}_1 = 0$. Note that from (49) and (50), we obtain

$$\mathbf{s}_1^T\mathbf{s}_2 = (\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x}) + \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{t}_1). \tag{51}$$

Let $g_i = \mathbf{a}_i^T\mathbf{x}$, $h_{i,1} = \mathbf{a}_i^T\mathbf{t}_1$, and $h_{i,2} = \mathbf{a}_i^T\mathbf{t}_2$; the three are independent standard normal random variables. From (49) and (50), I in (48) can be written as

$$\begin{aligned}
I &= \frac{1}{m}\sum_{i=1}^m y_i(\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2) \\
&= \frac{1}{m}\sum_{i=1}^m y_i\left((\mathbf{s}_1^T\mathbf{x})g_i + \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot h_{i,1}\right)\left((\mathbf{s}_2^T\mathbf{x})g_i + (\mathbf{s}_2^T\mathbf{t}_1)h_{i,1} + w_{12}h_{i,2}\right) \qquad (52)\\
&= (\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i^2 + \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{t}_1)\cdot\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}^2 \\
&\quad + \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot w_{12}\cdot\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2} \\
&\quad + \left((\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{t}_1) + \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{x})\right)\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i h_{i,1} + (\mathbf{s}_1^T\mathbf{x})w_{12}\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i h_{i,2}. \qquad (53)
\end{aligned}$$

In the following, we deal with the five terms in (53) separately.

- The first term $\underline{(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i^2}$: From Lemma 4, we have conditioned on the event $\mathcal{E}$ that

$$\left|\frac{1}{m}\sum_{i=1}^m y_i g_i^2 - (\nu + M_y)\right| < CK_y\sqrt{\frac{\log m}{m}}, \tag{54}$$

  which gives

$$|(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})|\cdot\left|\frac{1}{m}\sum_{i=1}^m y_i g_i^2 - (\nu + M_y)\right| < |(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})|\cdot CK_y\sqrt{\frac{\log m}{m}} \le CK_y\sqrt{\frac{\log m}{m}}. \tag{55}$$

- The second term $\underline{\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{t}_1)\cdot\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}^2}$: Since y_i are independent of $h_{i,1}$, $\mathbb{E}[h_{i,1}^2] = 1$, and $h_{i,1}^2$ are sub-exponential with the sub-exponential norm being upper bounded by an absolute constant C, from Lemma 2, we obtain that for any $u > 2$, with probability at least $1 - e^{-u}$,

$$\left|\frac{1}{m}\sum_{i=1}^m y_i(h_{i,1}^2 - 1)\right| \le C\left(\frac{\sqrt{u}\cdot\sqrt{\sum_{i=1}^m y_i^2/m}}{\sqrt{m}} + \frac{u\cdot\max_{i\in[m]}|y_i|}{m}\right). \tag{56}$$

  From Lemma 4, we obtain that when $m = \Omega(u\cdot\log m)$, conditioned on the event $\mathcal{E}$,

$$\left|\frac{1}{m}\sum_{i=1}^m y_i(h_{i,1}^2 - 1)\right| \le CK_y\cdot\sqrt{\frac{u\cdot\log m}{m}}. \tag{57}$$

  Then,

$$\begin{aligned}
\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot|\mathbf{s}_2^T\mathbf{t}_1|\cdot\left|\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}^2 - M_y\right| &\le \left|\frac{1}{m}\sum_{i=1}^m y_i(h_{i,1}^2 - 1)\right| + \left|\frac{1}{m}\sum_{i=1}^m y_i - M_y\right| \qquad (58)\\
&\le \frac{CK_y\cdot\sqrt{u\cdot\log m}}{\sqrt{m}}. \qquad (59)
\end{aligned}$$

- The third term $\underline{\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot w_{12}\cdot\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2}}$: We have

$$\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2}\right] = \mathbb{E}[y_1]\mathbb{E}[h_{1,1}]\mathbb{E}[h_{1,2}] = 0. \tag{60}$$

Since $\|h_{i,1}h_{i,2}\|_{\psi_1} \le \|h_{i,1}\|_{\psi_2}\|h_{i,2}\|_{\psi_2} = C$, from Lemma 2, we have that for fixed y_i and any $u \in (2,m)$, with probability at least $1-e^{-u}$,

$$\left|\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2}\right| \le C\left(\frac{\sqrt{u}\cdot\sqrt{\sum_{i=1}^m y_i^2/m}}{\sqrt{m}} + \frac{u\cdot\max_i|y_i|}{m}\right). \tag{61}$$

When $m = \Omega(u\cdot\log m)$, conditioned on the event $\mathcal{E}$, we have with probability at least $1-e^{-u}$ that

$$\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot w_{12}\cdot\left|\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2}\right| \le \left|\frac{1}{m}\sum_{i=1}^m y_i h_{i,1}h_{i,2}\right| \le CK_y\sqrt{\frac{u\cdot\log m}{m}}. \tag{62}$$

- The fourth to fifth terms in (53) can be controlled in a same way. For example, for the fourth term $\underline{((\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{t}_1)+\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{x}))\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i h_{i,1}}$: We have for fixed g_i and y_i that

$$\frac{1}{m}\sum_{i=1}^m y_i g_i h_{i,1} \sim \mathcal{N}\left(0, \sum_{i=1}^m y_i^2 g_i^2/m^2\right). \tag{63}$$

Then, from the standard Gaussian concentration [89, Example 2.1], we obtain that for any $u>0$, with probability at least $1-2e^{-u}$,

$$\left|\left((\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{t}_1)+\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{x})\right)\cdot\frac{1}{m}\sum_{i=1}^m y_i g_i h_{i,1}\right| \le \left|\frac{2}{m}\sum_{i=1}^m y_i g_i h_{i,1}\right| \tag{64}$$

$$\le \sqrt{\frac{8u\cdot\sum_{i=1}^m y_i^2 g_i^2/m}{m}} = O\left(K_y\sqrt{\frac{u}{m}}\right), \tag{65}$$

where the last inequality follows from Lemma 4. We have a similar result for the fifth term.

Combining (53) with (55), (59), (62) and (65), we obtain that when $m = \Omega\big(u\cdot\log m\big)$ and conditioned on the event $\mathcal{E}$, with probability $1-O\big(e^{-u}\big)$,

$$\left|\frac{1}{m}\sum_{i=1}^m y_i(\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2) - (\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})\nu - (\mathbf{s}_1^T\mathbf{s}_2)M_y\right|$$

$$= \left|\frac{1}{m}\sum_{i=1}^m y_i(\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2) - (\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})(\nu+M_y) - \sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{t}_1)M_y\right| \tag{66}$$

$$\le \frac{CK_y\cdot\sqrt{u\cdot\log m}}{\sqrt{m}}, \tag{67}$$

where (66) follows from 51. Then, from (47), we have

$$\left|\mathbf{s}_1^T\mathbf{E}\mathbf{s}_2\right| = \left|\frac{1}{m}\sum_{i=1}^m \big(y_i(\mathbf{a}_i^T\mathbf{s}_1)(\mathbf{a}_i^T\mathbf{s}_2) - (\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})\nu - (\mathbf{s}_1^T\mathbf{s}_2)M_y\big) + \frac{(\mathbf{s}_1^T\mathbf{s}_2)}{m}\sum_{i=1}^m(M_y-y_i)\right| \tag{68}$$

$$\le \frac{CK_y\cdot\sqrt{u\cdot\log m}}{\sqrt{m}}, \tag{69}$$

where (69) follows from (67) and the definite of the event $\mathcal{E}$ in (30). For general $\mathbf{s}_1$ and $\mathbf{s}_2$ (beyond unit vectors), when considering K_y as a fixed positive constant, we obtain (46) as desired. □

Based on Lemma 5, we obtain the following lemma.

Lemma 6. *Let $\mathbf{E} = \mathbf{V} - \nu\mathbf{x}\mathbf{x}^T$. For any pair of finite sets S_1, S_2 in $\mathbb{R}^n$ satisfying $m = \Omega((\log(|S_1| \cdot |S_2|)) \cdot (\log m))$, we have that with probability $1 - e^{-\Omega(\log(|S_1|\cdot|S_2|))} - O(1/m)$, for all $\mathbf{s}_1 \in S_1$ and $\mathbf{s}_2 \in S_2$, it holds that*

$$\left|\mathbf{s}_1^T\mathbf{E}\mathbf{s}_2\right| \le CK_y\sqrt{\frac{(\log(|S_1|\cdot|S_2|))\cdot(\log m)}{m}} \cdot \|\mathbf{s}_1\|_2 \cdot \|\mathbf{s}_2\|_2. \tag{70}$$

In addition, we have that $\|\mathbf{E}\|_{2\to 2} = O((K_y \cdot n\log m)/m)$ with probability $1 - O(1/m)$.

Proof. Note that from Lemma 4, the event $\mathcal{E}$ occurs with probability $1 - O(1/m)$. Setting $u = \log(|S_1| \cdot |S_2|)$ in Lemma 5, and taking a union bound over all $\mathbf{s}_1 \in S_1$ and $\mathbf{s}_2 \in S_2$, we obtain (70). In addition, according to [86, Lemma 5.4], we have

$$\|\mathbf{E}\|_{2\to 2} = \left\|\mathbf{V} - \nu\mathbf{x}\mathbf{x}^T\right\|_{2\to 2} = \sup_{\mathbf{r}\in\mathcal{S}^{n-1}} \left|\mathbf{r}^T(\mathbf{V} - \nu\mathbf{x}\mathbf{x}^T)\mathbf{r}\right| \le 2\sup_{\mathbf{r}\in\mathcal{C}_{1/4}} \left|\mathbf{r}^T(\mathbf{V} - \nu\mathbf{x}\mathbf{x}^T)\mathbf{r}\right|, \tag{71}$$

where $\mathcal{C}_{1/4}$ is a $(1/4)$-net of the unit sphere $\mathcal{S}^{n-1}$. In addition, according to [86, Lemma 5.2], we have $\left|\mathcal{C}_{1/4}\right| \le 9^n$. Similarly to Lemma 5,[7] we obtain that for any $\mathbf{r} \in \mathcal{S}^{n-1}$ and any $u > 2$ satisfying $u = \Omega(m)$, with probability at least $1 - e^{-u}$,

$$\left|\mathbf{r}^T\mathbf{E}\mathbf{r}\right| \le CK_y \cdot \frac{u \cdot \log m}{m}. \tag{72}$$

Setting $u = Cn$ in (72) and taking a union bound over all $\mathbf{r} \in \mathcal{C}_{1/4}$, we obtain that with probability $1 - e^{-\Omega(n)} - O(1/m)$, $\|\mathbf{E}\|_{2\to 2} = O((K_y \cdot n\log m)/m)$. $\square$

In addition, we have the following lemma according to [50, Theorem 2].[8]

Lemma 7. (Adapted from [50, Theorem 2]) *Suppose that the data matrix $\mathbf{V} \in \mathbb{R}^{n\times n}$ can be written as $\mathbf{V} = \bar{\mathbf{V}} + \mathbf{E}$ with $\bar{\mathbf{V}}$ being a positive definite* rank-one *matrix and $\mathbf{E}$ satisfying the following conditions: 1) For any two finite sets S_1, S_2 in $\mathbb{R}^n$ satisfying $m = \Omega((\log(|S_1|\cdot|S_2|)) \cdot (\log m))$, we have for all $\mathbf{s}_1 \in S_1$ and $\mathbf{s}_2 \in S_2$ that $\left|\mathbf{s}_1^T\mathbf{E}\mathbf{s}_2\right| \le CK_y\sqrt{\frac{\log(|S_1|\cdot|S_2|)\cdot(\log m)}{m}} \cdot \|\mathbf{s}_1\|_2 \cdot \|\mathbf{s}_2\|_2$; 2) $\|\mathbf{E}\|_{2\to 2} = O((K_y \cdot n\log m)/m)$. Then, if there exists $t_0 \in \mathbb{N}$ such that $\mathbf{x}^T\mathbf{w}^{(t_0)} \ge c_0$ with c_0 being a sufficiently small positive constant and $m = \Omega((k\log(nLr)) \cdot (\log m))$ with a large enough implied constant, we have that after one projected power iteration in the first step of Algorithm 1 (beyond t_0),*

$$\|\mathbf{w}^{(t)} - \mathbf{x}\|_2 \le \frac{CK_y}{c_0}\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}, \tag{73}$$

i.e., this equation holds for all $t > t_0$.

B.2 Proof of Theorem 1

Combining the results of Lemmas 6 and 7, we obtain the desired result of Theorem 1.

C Proof of Theorem 2 (Guarantees for the Second Step of Algorithm 1)

Before presenting the proof of the theorem, we provide some useful lemmas.

C.1 Useful Lemmas for Theorem 2

Recall that $\bar{y}$, $\hat{\nu}^{(t)}$ are defined in (12) and (13) respectively and $M_y := \mathbb{E}[y]$ and $K_y := \|y\|_{\psi_1}$ (*cf.* Section 2.2). First, we have the following lemma.

Lemma 8. *Conditioned on the event $\mathcal{E}$ (*cf. *(30)), when $m = \Omega((k\log(nLr)) \cdot (\log m))$, for any $t \in \{0, 1, \dots, T_2 - 1\}$, we have with probability $1 - e^{-\Omega(k\log(nLr))}$ that*

$$\left|\hat{\nu}^{(t)} - \nu\right| \le CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + \left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2^2 \cdot \nu. \tag{74}$$

[7] More precisely, only (57) (and thus (59)) and (62) need to be modified accordingly.

[8] We consider the $\bar{\gamma} = 0$ case therein.

Proof. For any $\delta \in (0,1)$, let M be a (δ/L)-net of $B^k(r)$. According to [86, Lemma 5.2], there exists such a net with the cardinality satisfies

$$\log|M| \leq k \log \frac{4Lr}{\delta}. \tag{75}$$

Due to the L-Lipschitz continuity of G, $G(M)$ is a δ-net of $G(B^k(r))$. Then, since $\mathbf{x}^{(t)} \in \mathrm{Range}(G) = G(B^k(r))$, it can be written as

$$\mathbf{x}^{(t)} = \mathbf{s}^{(t)} + \mathbf{e}^{(t)}, \tag{76}$$

where $\mathbf{s}^{(t)} \in G(M)$ and $\|\mathbf{e}^{(t)}\|_2 \leq \delta$. We obtain

$$\left|\hat{\nu}^{(t)} - \nu\right| = \left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{x}^{(t)}\right)^2 - \nu\right| \tag{77}$$

$$= \left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \left(\mathbf{s}^{(t)} + \mathbf{e}^{(t)}\right)\right)^2 - \nu\right| \tag{78}$$

$$\leq \left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{s}^{(t)}\right)^2 - \nu\right| + \left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{e}^{(t)}\right)^2\right|$$
$$+ \left|\frac{2}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{s}^{(t)}\right) \cdot \left(\mathbf{a}_i^T \mathbf{e}^{(t)}\right)\right|. \tag{79}$$

From Lemma 2 and by taking a union bound over $[m]$, we obtain with probability $1 - me^{-\Omega(n)}$ that

$$\max_{i \in [m]} \|\mathbf{a}_i\|_2 \leq \sqrt{2n}. \tag{80}$$

In addition, we have

$$\frac{1}{m}\sum_{i=1}^m |y_i - \bar{y}| \leq \sqrt{\frac{\sum_{i=1}^m (y_i - \bar{y})^2}{m}} \tag{81}$$

$$= \sqrt{\frac{\sum_{i=1}^m y_i^2 - m\bar{y}^2}{m}} \tag{82}$$

$$\leq \sqrt{\frac{\sum_{i=1}^m y_i^2}{m}}. \tag{83}$$

Then, from the definition of the event $\mathcal{E}$ in Lemma 4, we obtain

$$\frac{1}{m}\sum_{i=1}^m |y_i - \bar{y}| \leq \sqrt{\frac{\sum_{i=1}^m y_i^2}{m}} \leq 2\sqrt{2}K_y. \tag{84}$$

Therefore, since $\|\mathbf{e}^{(t)}\|_2 \leq \delta$, conditioned on the event in (80) and event $\mathcal{E}$, we have

$$\left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{e}^{(t)}\right)^2\right| \leq \frac{1}{m}\sum_{i=1}^m |y_i - \bar{y}| \cdot 2n\delta^2 \tag{85}$$

$$\leq 2\sqrt{2}K_y \cdot 2n\delta^2. \tag{86}$$

In addition, since $\mathbf{s}^{(t)} \in G(M) \subseteq \mathcal{S}^{n-1}$, similarly to (86), we obtain

$$\left|\frac{2}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T \mathbf{s}^{(t)}\right) \cdot \left(\mathbf{a}_i^T \mathbf{e}^{(t)}\right)\right| \leq 4\sqrt{2}K_y \cdot 2n\delta. \tag{87}$$

Then, it remains to control the first term in (79), namely $\left|\frac{1}{m}\sum_{i=1}^m (y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{s}^{(t)}\right)^2-\nu\right|$. For any fixed $\mathbf{s}\in\mathcal{S}^{n-1}$, we obtain

$$\left|\frac{1}{m}\sum_{i=1}^m (y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{s}\right)^2-\nu\right| \le \left|\frac{1}{m}\sum_{i=1}^m (y_i-M_y)\cdot\left(\mathbf{a}_i^T\mathbf{s}\right)^2-\nu\right| + \left|\frac{1}{m}\sum_{i=1}^m (M_y-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{s}\right)^2\right| \tag{88}$$

$$\le \left|\frac{1}{m}\sum_{i=1}^m (y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{s})^2-\left(\mathbf{x}^T\mathbf{s}\right)^2\nu\right| + \left|1-\left(\mathbf{x}^T\mathbf{s}\right)^2\right|\cdot\nu + \left|\frac{|M_y-\bar{y}|}{m}\sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{s}\right)^2\right| \tag{89}$$

$$\le \left|\frac{1}{m}\sum_{i=m}^m (y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{s})^2-\left(\mathbf{x}^T\mathbf{s}\right)^2\nu\right| + \frac{|M_y-\bar{y}|}{m}\sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{s}\right)^2 + \|\mathbf{x}-\mathbf{s}\|_2^2\cdot\nu, \tag{90}$$

where we use $\left|1-\left(\mathbf{x}^T\mathbf{s}\right)^2\right| = \left(1+\mathbf{x}^T\mathbf{s}\right)\cdot\left(1-\mathbf{x}^T\mathbf{s}\right)\le 2\left(1-\mathbf{x}^T\mathbf{s}\right) = \|\mathbf{x}-\mathbf{s}\|_2^2$ in (90). Since $\mathbf{s}$ is a unit vector, from Lemma 2, we obtain with probability $1-e^{-\Omega(m)}$ that

$$\frac{1}{m}\sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{s}\right)^2 \le 2. \tag{91}$$

In addition, conditioned on the event $\mathcal{E}$, we have

$$|\bar{y}-M_y| \le CK_y\sqrt{\frac{\log m}{m}}. \tag{92}$$

Then, we only need to focus on the first term of (90). We write $\mathbf{s}\in\mathcal{S}^{n-1}$ as

$$\mathbf{s} = (\mathbf{s}^T\mathbf{x})\mathbf{x} + \sqrt{1-(\mathbf{s}^T\mathbf{x})^2}\cdot\mathbf{t}, \tag{93}$$

where $\mathbf{x}^T\mathbf{t}=0$ and $\|\mathbf{t}\|_2=1$. Hence,

$$\left(\mathbf{a}_i^T\mathbf{s}\right)^2 = \left(\mathbf{x}^T\mathbf{s}\right)^2\left(\mathbf{a}_i^T\mathbf{x}\right)^2 + \left(1-\left(\mathbf{x}^T\mathbf{s}\right)^2\right)\cdot\left(\mathbf{a}_i^T\mathbf{t}\right)^2 + 2\left(\mathbf{x}^T\mathbf{s}\right)\cdot\sqrt{1-\left(\mathbf{x}^T\mathbf{s}\right)^2}\cdot\left(\mathbf{a}_i^T\mathbf{x}\right)\cdot\left(\mathbf{a}_i^T\mathbf{t}\right). \tag{94}$$

Then, the first term of (90) can be upper bounded as

$$\begin{aligned}&\left|\frac{1}{m}\sum_{i=1}^m (y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{s})^2-\left(\mathbf{x}^T\mathbf{s}\right)^2\nu\right| \\ &\le \frac{\left(\mathbf{x}^T\mathbf{s}\right)^2}{m}\cdot\left|\sum_{i=1}^m (y_i-M_y)(\mathbf{a}_i^T\mathbf{x})^2-\nu\right| + \frac{1-\left(\mathbf{x}^T\mathbf{s}\right)^2}{m}\cdot\left|\sum_{i=1}^m (y_i-M_y)(\mathbf{a}_i^T\mathbf{t})^2\right| \\ &\quad + \frac{2\left|\mathbf{x}^T\mathbf{s}\right|\cdot\sqrt{1-\left(\mathbf{x}^T\mathbf{s}\right)^2}}{m}\cdot\left|\sum_{i=1}^m (y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{x})\cdot(\mathbf{a}_i^T\mathbf{t})\right|.\end{aligned} \tag{95}$$

Conditioned on the event $\mathcal{E}$, the first term in (95) can be upper bounded by

$$\frac{\left(\mathbf{x}^T\mathbf{s}\right)^2}{m}\left|\sum_{i=m+1}^{2m} (y_i-M_y)(\mathbf{a}_i^T\mathbf{x})^2-\nu\right| \le CK_y\sqrt{\frac{\log m}{m}}. \tag{96}$$

In addition, since $y=f(\mathbf{a}^T\mathbf{x})$ is independent of $\mathbf{a}^T\mathbf{t}$, from Lemma 2, for any $u>2$ and fixed y_i, we have with probability at least $1-e^{-u}$ that

$$\frac{1-\left(\mathbf{x}^T\mathbf{s}\right)^2}{m}\cdot\left|\sum_{i=1}^m (y_i-M_y)(\mathbf{a}_i^T\mathbf{t})^2\right| \le \frac{1}{m}\left|\sum_{i=1}^m (y_i-M_y)((\mathbf{a}_i^T\mathbf{t})^2-1)\right| + \frac{1}{m}\left|\sum_{i=1}^m (y_i-M_y)\right| \tag{97}$$

$$\le C\left(\frac{\sqrt{u}\cdot\sqrt{\sum_{i=1}^m (y_i-M_y)^2/m}}{\sqrt{m}} + \frac{u\cdot\max_i|y_i-M_y|}{m}\right) + \frac{1}{m}\left|\sum_{i=1}^m (y_i-M_y)\right|. \tag{98}$$

Conditioned on the event $\mathcal{E}$ and using the inequality $|M_y| \leq K_y$ (*cf.* (34)), we obtain that when $m = \Omega(u \cdot \log m)$, with probability at least $1 - e^{-u}$, it holds that

$$\frac{1 - \left(\mathbf{x}^T\mathbf{s}\right)^2}{m} \cdot \left|\sum_{i=1}^m (y_i - M_y)(\mathbf{a}_i^T\mathbf{t})^2\right| \leq CK_y \cdot \sqrt{\frac{u \cdot \log m}{m}}. \tag{99}$$

Moreover, similarly to (65), we obtain that conditioned on the event $\mathcal{E}$, with probability at least $1 - 2e^{-u}$, the third term in (95) is upper bounded as

$$\frac{2\left(\mathbf{x}^T\mathbf{s}\right) \cdot \sqrt{1 - \left(\mathbf{x}^T\mathbf{s}\right)^2}}{m} \cdot \left|\sum_{i=1}^m (y_i - M_y) \cdot (\mathbf{a}_i^T\mathbf{x}) \cdot (\mathbf{a}_i^T\mathbf{t})\right| = O\left(K_y \cdot \sqrt{\frac{u}{m}}\right). \tag{100}$$

Combining (90), (91), (92), (95), (96), (99) and (100), we obtain that conditioned on the event $\mathcal{E}$, for any fixed $\mathbf{s} \in \mathcal{S}^{n-1}$ and $u > 2$, when $m = \Omega(u \cdot \log m)$, with probability $1 - e^{-\Omega(u)}$,

$$\left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T\mathbf{s}\right)^2 - \nu\right| \leq CK_y \cdot \sqrt{\frac{u \cdot \log m}{m}} + \|\mathbf{x} - \mathbf{s}\|_2^2 \cdot \nu. \tag{101}$$

Taking a union bound for all $\mathbf{s} \in G(M)$ and setting $u = k \log \frac{Lr}{\delta}$, we obtain that when conditioned on the event $\mathcal{E}$ and $m = \Omega\big(\big(k \log \frac{Lr}{\delta}\big) \cdot (\log m)\big)$, with probability $1 - e^{-\Omega(k \log \frac{Lr}{\delta})}$, for *all* $\mathbf{s} \in G(M)$,

$$\left|\frac{1}{m}\sum_{i=1}^m (y_i - \bar{y}) \cdot \left(\mathbf{a}_i^T\mathbf{s}\right)^2 - \nu\right| \leq CK_y \cdot \sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}} + \|\mathbf{x} - \mathbf{s}\|_2^2 \cdot \nu, \tag{102}$$

and this gives an upper bound for the first term of (79) by substituting $\mathbf{s}^{(t)}$ for $\mathbf{s}$. Combining (79), (86), (87) and (102), we obtain that when $m = \Omega\big(\big(k \log \frac{Lr}{\delta}\big) \cdot (\log m)\big)$, with probability $1 - e^{-\Omega(k \log \frac{Lr}{\delta})}$,

$$\begin{aligned}
\left|\hat{\nu}^{(t)} - \nu\right| &\leq 4\sqrt{2}K_y n\delta(2 + \delta) + CK_y \cdot \sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}} + \|\mathbf{x} - \mathbf{s}^{(t)}\|_2^2 \cdot \nu & (103)\\
&\leq CK_y\left(n\delta + \sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}}\right) + \left(\left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2 + \delta\right)^2 \cdot \nu, & (104)
\end{aligned}$$

where (104) follows from (76). Setting $\delta = \frac{1}{n\sqrt{m}}$ and using $n = \Omega(m)$, we obtain

$$\begin{aligned}
&CK_y\left(n\delta + \sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}}\right) + \left(\left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2 + \delta\right)^2 \cdot \nu \\
&\leq CK_y\sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}} + \left(\left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2^2 + 2\delta \cdot \left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2 + \delta^2\right) \cdot \nu & (105)\\
&\leq CK_y\sqrt{\frac{\left(k \log \frac{Lr}{\delta}\right) \cdot (\log m)}{m}} + \left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2^2 \cdot \nu, & (106)
\end{aligned}$$

where (106) follows from $\left\|\mathbf{x} - \mathbf{x}^{(t)}\right\|_2 \leq 2$ and $\nu > 0$ is a fixed constant (recall that the value of C may differ from line to line). $\square$

Next, we present the following useful lemma.

Lemma 9. *For any $u > 2$ satisfying $m = \Omega(u \cdot \log m)$, conditioned on the event $\mathcal{E}$ (*cf.* (30)), we have that for any $\mathbf{s}_1 \in \mathcal{S}^{n-1}$ and $\mathbf{s}_2 \in \mathbb{R}^n$, with probability $1 - e^{-\Omega(u)}$,*

$$\left|\frac{1}{m}\sum_{i=1}^m \left((y_i - \bar{y}) \cdot (\mathbf{a}_i^T\mathbf{s}_1) - \nu(\mathbf{a}_i^T\mathbf{x})\right) \cdot (\mathbf{a}_i^T\mathbf{s}_2)\right| \leq \left(\frac{CK_y\sqrt{u \cdot \log m}}{\sqrt{m}} + \|\mathbf{s}_1 - \mathbf{x}\|_2^2 \cdot \nu\right) \cdot \|\mathbf{s}_2\|_2. \tag{107}$$

Proof. Without loss of generality, we assume that $\|\mathbf{s}_2\|_2 = 1$. We have

$$\frac{1}{m}\sum_{i=1}^m \big((y_i-\bar{y})\cdot(\mathbf{a}_i^T\mathbf{s}_1)-\nu(\mathbf{a}_i^T\mathbf{x})\big)\cdot(\mathbf{a}_i^T\mathbf{s}_2)$$
$$=\frac{1}{m}\sum_{i=1}^m \big((y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{s}_1)-\nu(\mathbf{a}_i^T\mathbf{x})\big)\cdot(\mathbf{a}_i^T\mathbf{s}_2)+\frac{1}{m}\sum_{i=1}^m(M_y-\bar{y})\cdot(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2). \quad (108)$$

In the second term of (108), we observe that from Lemma 1, $(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2)$ are i.i.d. sub-exponential random variables with mean $\mathbf{s}_1^T\mathbf{s}_2$ and the sub-exponential norm being upper bounded by C. From Lemma 2, we have with probability $1-e^{-u}$ that

$$\left|\frac{1}{m}\sum_{i=1}^m(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2)-(\mathbf{s}_1^T\mathbf{s}_2)\right|\le C\sqrt{\frac{u}{m}}. \quad (109)$$

Then, we obtain

$$\left|\frac{1}{m}\sum_{i=1}^m(M_y-\bar{y})\cdot(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2)\right|\le|M_y-\bar{y}|\cdot\left|\frac{1}{m}\sum_{i=1}^m(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2)\right| \quad (110)$$
$$\le CK_y\sqrt{\frac{\log m}{m}}\cdot\left|\frac{1}{m}\sum_{i=1}^m(\mathbf{a}_i^T\mathbf{s}_1)\cdot(\mathbf{a}_i^T\mathbf{s}_2)\right| \quad (111)$$
$$\le CK_y\sqrt{\frac{\log m}{m}}\cdot\left(1+C\sqrt{\frac{u}{m}}\right) \quad (112)$$
$$\le C'K_y\sqrt{\frac{\log m}{m}}, \quad (113)$$

where (111) follows from the definition of the event $\mathcal{E}$ in Lemma 4, (112) follows from (109) and $|\mathbf{s}_1^T\mathbf{s}_2|\le\|\mathbf{s}_1\|_2\cdot\|\mathbf{s}_2\|_2=1$, and (113) follows from the condition $m=\Omega\big(u\cdot\log m\big)$. Then, it remains to control the first term of (108). Similarly to that in the proof of Lemma 5, we decompose $\mathbf{s}_1$ as (*cf.* (49))

$$\mathbf{s}_1=(\mathbf{s}_1^T\mathbf{x})\mathbf{x}+\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot\mathbf{t}_1, \quad (114)$$

where $\|\mathbf{t}_1\|_2=1$ and $\mathbf{t}_1^T\mathbf{x}=0$. Similarly, letting $w_{12}=\sqrt{1-(\mathbf{s}_2^T\mathbf{x})^2-(\mathbf{s}_2^T\mathbf{t}_1)^2}$, $\mathbf{s}_2$ can be written as (*cf.* (50))

$$\mathbf{s}_2=(\mathbf{s}_2^T\mathbf{x})\mathbf{x}+(\mathbf{s}_2^T\mathbf{t}_1)\mathbf{t}_1+w_{12}\mathbf{t}_2, \quad (115)$$

where $\|\mathbf{t}_2\|_2=1$ and $\mathbf{t}_2^T\mathbf{x}=\mathbf{t}_2^T\mathbf{t}_1=0$. Let $g_i=\mathbf{a}_i^T\mathbf{x}\sim\mathcal{N}(0,1)$, $h_{i,1}=\mathbf{a}_i^T\mathbf{t}_1\sim\mathcal{N}(0,1)$, and $h_{i,2}=\mathbf{a}_i^T\mathbf{t}_2\sim\mathcal{N}(0,1)$; the three are independent. Therefore, we obtain

$$\frac{1}{m}\sum_{i=1}^m\big((y_i-M_y)\cdot(\mathbf{a}_i^T\mathbf{s}_1)-\nu(\mathbf{a}_i^T\mathbf{x})\big)\cdot(\mathbf{a}_i^T\mathbf{s}_2)=\frac{1}{m}\sum_{i=1}^m\big((y_i-M_y)\cdot(\mathbf{s}_1^T\mathbf{x})-\nu\big)\cdot(\mathbf{s}_2^T\mathbf{x})g_i^2$$
$$+\frac{1}{m}\sum_{i=1}^m\left((y_i-M_y)\cdot\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{t}_1)\right)h_{i,1}^2$$
$$+\frac{1}{m}\sum_{i=1}^m\left((y_i-M_y)\cdot\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\right)\cdot w_{12}h_{i,1}h_{i,2}$$
$$+\frac{1}{m}\sum_{i=1}^m\left((y_i-M_y)\cdot\sqrt{1-(\mathbf{s}_1^T\mathbf{x})^2}\cdot(\mathbf{s}_2^T\mathbf{x})+(y_i-M_y)\cdot(\mathbf{s}_1^T\mathbf{x})\cdot(\mathbf{s}_2^T\mathbf{t}_1)-\nu(\mathbf{s}_2^T\mathbf{t}_1)\right)\cdot g_ih_{i,1}$$
$$+\frac{1}{m}\sum_{i=1}^m\big((y_i-M_y)\cdot(\mathbf{s}_1^T\mathbf{x})-\nu\big)\cdot w_{12}g_ih_{i,2}. \quad (116)$$

The equality (116) is similar to (53), with the major difference being that y_i is replaced by y_i-M_y, which has zero mean. In the following, we focus on bounding the first term in (116), and other terms can be similarly bounded as those in the proof of Lemma 5. In particular, for the first term in (116), we have

$$\frac{1}{m}\sum_{i=1}^{m}\big((y_i - M_y)\cdot(\mathbf{s}_1^T\mathbf{x}) - \nu\big)\cdot(\mathbf{s}_2^T\mathbf{x})g_i^2$$
$$= \frac{(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})}{m}\sum_{i=1}^{m}(y_i - M_y - \nu)g_i^2 + \frac{1}{m}\sum_{i=1}^{m}\nu(\mathbf{s}_2^T\mathbf{x})((\mathbf{s}_1^T\mathbf{x})-1)g_i^2 \tag{117}$$
$$= \frac{(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})}{m}\sum_{i=1}^{m}(y_ig_i^2 - M_y - \nu) + \frac{(\mathbf{s}_1^T\mathbf{x})(\mathbf{s}_2^T\mathbf{x})}{m}\sum_{i=1}^{m}(\nu + M_y)(1-g_i^2)$$
$$+ \frac{1}{m}\sum_{i=1}^{m}\nu(\mathbf{s}_2^T\mathbf{x})((\mathbf{s}_1^T\mathbf{x})-1)g_i^2. \tag{118}$$

By Lemma 4 and similarly to (109), as well as using $|M_y| \le K_y$ (*cf.* (34)) and $\nu < CK_y$,[9] we obtain

$$\left|\frac{1}{m}\sum_{i=1}^{m}\big((y_i - M_y)\cdot(\mathbf{s}_1^T\mathbf{x}) - \nu\big)\cdot(\mathbf{s}_2^T\mathbf{x})g_i^2\right| \le CK_y\sqrt{\frac{\log m}{m}} + C'\|\mathbf{s}_1 - \mathbf{x}\|_2^2\cdot\nu, \tag{119}$$

where $C' > 0$ can be choose to be slightly larger than $\frac{1}{2}$. For the last four terms in (116), similarly to (67), we obtain that when $m = \Omega\big(u\cdot\log m\big)$, with probability $1 - O\big(e^{-u}\big)$, the sum of the absolute value of these four terms can be upper bounded by

$$\frac{CK_y\sqrt{u\cdot\log m}}{\sqrt{m}}. \tag{120}$$

Combining (108), (113), (116), (119), and (120), we obtain the desired result. $\square$

C.2 Proof of Theorem 2

Since $\mathbf{x}^{(t+1)} = \mathcal{P}_G(\tilde{\mathbf{x}}^{(t+1)})$ and $\mathbf{x} \in \mathrm{Range}(G)$, we obtain

$$\left\|\tilde{\mathbf{x}}^{(t+1)} - \mathbf{x}^{(t+1)}\right\|_2 \le \left\|\tilde{\mathbf{x}}^{(t+1)} - \mathbf{x}\right\|_2, \tag{121}$$

or equivalently,

$$\left\|\left(\tilde{\mathbf{x}}^{(t+1)} - \mathbf{x}\right) + \left(\mathbf{x} - \mathbf{x}^{(t+1)}\right)\right\|_2^2 \le \left\|\tilde{\mathbf{x}}^{(t+1)} - \mathbf{x}\right\|_2^2, \tag{122}$$

which gives

$$\|\mathbf{x}^{(t+1)} - \mathbf{x}\|_2^2 \le 2\langle\tilde{\mathbf{x}}^{(t+1)} - \mathbf{x}, \mathbf{x}^{(t+1)} - \mathbf{x}\rangle \tag{123}$$
$$= 2\left\langle \mathbf{x}^{(t)} - \frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\hat{\nu}^{(t)}\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right) - \tilde{\mathbf{y}}_i^{(t)}\right)\mathbf{a}_i - \mathbf{x}, \mathbf{x}^{(t+1)} - \mathbf{x}\right\rangle \tag{124}$$
$$= 2\left\langle \mathbf{x}^{(t)} - \frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\nu\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right) - \tilde{\mathbf{y}}_i^{(t)}\right)\mathbf{a}_i - \mathbf{x}, \mathbf{x}^{(t+1)} - \mathbf{x}\right\rangle$$
$$+ 2\left\langle \frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\nu - \hat{\nu}^{(t)}\right)\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)\mathbf{a}_i, \mathbf{x}^{(t+1)} - \mathbf{x}\right\rangle \tag{125}$$
$$= 2\left\langle \frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\tilde{\mathbf{y}}_i^{(t)} - \nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i, \mathbf{x}^{(t+1)} - \mathbf{x}\right\rangle$$
$$+ 2\left\langle \left(\mathbf{I}_n - \frac{\zeta\nu}{m}\cdot\sum_{i=1}^{m}\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{x}^{(t)} - \mathbf{x}\right), \mathbf{x}^{(t+1)} - \mathbf{x}\right\rangle$$

[9]Recall that we will assume that $\nu > 0$ (*cf.* (7)). We have $\nu = \mathbb{E}[y(\mathbf{a}^T\mathbf{x})^2] - M_y \le (\mathbb{E}[y^2])^{1/2}\cdot(\mathbb{E}[(\mathbf{a}^T\mathbf{x})^4])^{1/2} + |M_y|$. Note that $\mathbf{a}^T\mathbf{x} \sim \mathcal{N}(0,1)$. From (34), we have $|M_y| \le K_y$ and $(\mathbb{E}[y^2])^{1/2} \le 2K_y$. Then, we obtain $\nu \le (2\sqrt{3}+1)K_y$.

$$+2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\nu-\hat{\nu}^{(t)}\right)\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)\mathbf{a}_i,\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle \tag{126}$$

$$=2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle$$
$$+2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^{m}\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{x}^{(t)}-\mathbf{x}\right),\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle$$
$$+2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left(\nu-\hat{\nu}^{(t)}\right)\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)\mathbf{a}_i,\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle, \tag{127}$$

where (124) follows from the setting of $\tilde{\mathbf{x}}^{(t+1)}$ in (15), (125) follows from the decomposition $\hat{\nu}^{(t+1)}=\nu+\left(\hat{\nu}^{(t)}-\nu\right)$, (126) follows from the fact that the first term in (125) can be written as the sum of the first two terms in (126), and (127) follows from the setting of $\tilde{y}_i^{(t)}$ in (14). For any $\delta\in(0,1)$, let M be a (δ/L)-net of $B^k(r)$. Then, similarly to that in the proof of Lemma 8, we have $\log|M|\le k\log\frac{4Lr}{\delta}$ (*cf.* (75)) and $G(M)$ is a δ-net of $G(B^k(r))$. For any $t\in\{0,\dots,T_2\}$, we write

$$\mathbf{x}^{(t)}=\mathbf{s}^{(t)}+\mathbf{e}^{(t)}, \tag{128}$$

where $\mathbf{s}^{(t)}\in G(M)$ and $\|\mathbf{e}^{(t)}\|_2\le\delta$. Next, we provide upper bounds for the three terms in (127) separately. Throughout the following, we assume the occurrence of the event $\mathcal{E}$ (*cf.* (30)) and the relevant probabilities are all conditioned accordingly.

The first term of (127): From (128), we have

$$2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle$$
$$=2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle$$
$$+2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{e}^{(t+1)}\right\rangle \tag{129}$$
$$=2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{s}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle$$
$$+2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}(y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{e}^{(t)}\right)\mathbf{a}_i,\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle$$
$$+2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}\left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i,\mathbf{e}^{(t+1)}\right\rangle. \tag{130}$$

Recall that in (80), we obtain with probability $1-me^{-\Omega(n)}$ that $\max_{i\in[m]}\|\mathbf{a}_i\|_2\le\sqrt{2n}$. In addition, according to (84), we have $\frac{1}{m}\sum_{i=1}^m|y_i-\bar{y}|\le2\sqrt{2}K_y$. Then, for the second term in (130), since $\left\|\mathbf{e}^{(t)}\right\|_2\le\delta$, similarly to (86), we obtain

$$2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^{m}(y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{e}^{(t)}\right)\mathbf{a}_i,\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle\le2\zeta\cdot2n\delta\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2\cdot\frac{1}{m}\sum_{i=1}^{m}|y_i-\bar{y}| \tag{131}$$
$$\le8\sqrt{2}\zeta K_yn\delta\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2 \tag{132}$$
$$\le8\sqrt{2}\zeta K_yn\delta\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right). \tag{133}$$

Similarly, since both $\mathbf{x}$ and $\mathbf{x}^{(t)}$ are unit vectors and $\left\|\mathbf{e}^{(t+1)}\right\|_2 \le \delta$, for the third term in (130),

$$
\begin{aligned}
&2\left\langle \frac{\zeta}{m}\cdot\sum_{i=1}^m \left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i, \mathbf{e}^{(t+1)}\right\rangle \\
&\le 2\zeta\cdot 2n\delta\cdot\frac{1}{m}\sum_{i=1}^m |y_i-\bar{y}| + 2\zeta\cdot 2n\delta\cdot\nu \qquad (134)\\
&\le 4\zeta n\delta\cdot(2\sqrt{2}K_y+\nu). \qquad (135)
\end{aligned}
$$

It remains to control the first term of (130). In order to do this, we make use of Lemma 9 with taking a union bound over all $(\mathbf{s}_1,\mathbf{s}_2)\in G(M)\times(G(M)-\mathbf{x})$ and setting $u = k\log\frac{Lr}{\delta}$, and we obtain that when $m=\Omega\big(\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)\big)$, with probability $1-e^{-\Omega(k\log\frac{Lr}{\delta})}$,

$$
\begin{aligned}
&\left|2\left\langle \frac{\zeta}{m}\cdot\sum_{i=1}^m \left((y_i-\bar{y})\cdot\left(\mathbf{a}_i^T\mathbf{s}^{(t)}\right)-\nu(\mathbf{a}_i^T\mathbf{x})\right)\mathbf{a}_i, \mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle\right| \\
&\le 2\zeta\cdot\left(CK_y\sqrt{\frac{\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)}{m}}+\nu\left\|\mathbf{s}^{(t)}-\mathbf{x}\right\|_2^2\right)\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2 \qquad (136)\\
&\le 2\zeta\cdot\left(CK_y\sqrt{\frac{\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)}{m}}+\nu\left(\left\|\mathbf{x}^{(t)}-\mathbf{x}\right\|_2+\delta\right)^2\right)\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right). \qquad (137)
\end{aligned}
$$

<u>The second term of (127)</u>: By (128), we have

$$
\begin{aligned}
&2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{x}^{(t)}-\mathbf{x}\right),\mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle\\
&=2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{x}^{(t)}-\mathbf{x}\right),\mathbf{e}^{(t+1)}\right\rangle\\
&\quad+2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\mathbf{e}^{(t)},\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle\\
&\quad+2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{s}^{(t)}-\mathbf{x}\right),\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle. \qquad (138)
\end{aligned}
$$

Since from (80), we have with probability $1-me^{-\Omega(n)}$ that $\max_{i\in[m]}\|\mathbf{a}_i\|_2\le\sqrt{2n}$. Similarly to (86), we obtain

$$
\left|2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{x}^{(t)}-\mathbf{x}\right),\mathbf{e}^{(t+1)}\right\rangle\right|\le 2\left(1+\frac{2n\zeta\nu}{m}\right)\cdot\delta\cdot\left\|\mathbf{x}^{(t)}-\mathbf{x}\right\|_2, \qquad (139)
$$

and

$$
\begin{aligned}
2\left\langle\left(\mathbf{I}_n-\frac{\zeta\nu}{m}\cdot\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\right)\mathbf{e}^{(t)},\mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle&\le 2\left(1+\frac{2n\zeta\nu}{m}\right)\cdot\delta\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2 \qquad (140)\\
&\le 2\left(1+\frac{2n\zeta\nu}{m}\right)\cdot\delta\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right). \qquad (141)
\end{aligned}
$$

Fix any pair of $\mathbf{s}_1,\mathbf{s}_2\in\mathbb{R}^n$. From Lemma 2 and similarly to (109), we obtain that for any $u>2$, with probability at least $1-e^{-u}$,

$$
\zeta\nu(\mathbf{s}_1^T\mathbf{s}_2)-C\zeta\nu\sqrt{\frac{u}{m}}\cdot\|\mathbf{s}_1\|_2\cdot\|\mathbf{s}_2\|_2\le\frac{\zeta\nu}{m}\left\langle\sum_{i=1}^m\mathbf{a}_i\mathbf{a}_i^T\mathbf{s}_1,\mathbf{s}_2\right\rangle\le\zeta\nu(\mathbf{s}_1^T\mathbf{s}_2)+C\zeta\nu\sqrt{\frac{u}{m}}\cdot\|\mathbf{s}_1\|_2\cdot\|\mathbf{s}_2\|_2. \qquad (142)
$$

Then, we obtain

$$\left|\left\langle\left(\mathbf{I}_n - \frac{\zeta\nu}{m}\sum_{i=1}^m \mathbf{a}_i\mathbf{a}_i^T\right)\mathbf{s}_1, \mathbf{s}_2\right\rangle - (1-\zeta\nu)(\mathbf{s}_1^T\mathbf{s}_2)\right| \le C\zeta\nu\sqrt{\frac{u}{m}}\cdot\|\mathbf{s}_1\|_2\cdot\|\mathbf{s}_2\|_2. \tag{143}$$

Taking a union bound for all pairs $(\mathbf{s}_1, \mathbf{s}_2) \in (G(M)-\mathbf{x})\times(G(M)-\mathbf{x})$ and setting $u = \frac{m}{C^2\log m}$, as well as considering ζ, ν as fixed positive constants, we obtain that when $m = \Omega\big(\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)\big)$, with probability $1 - e^{-\Omega(k\log\frac{Lr}{\delta})}$,

$$\begin{aligned}
&\left|2\left\langle\left(\mathbf{I}_n - \frac{\zeta\nu}{m}\cdot\sum_{i=1}^m \mathbf{a}_i\mathbf{a}_i^T\right)\left(\mathbf{s}^{(t)}-\mathbf{x}\right), \mathbf{s}^{(t+1)}-\mathbf{x}\right\rangle\right| \\
&\le 2\max\{1-\zeta\nu(1-1/\sqrt{\log m}), \zeta\nu(1+1/\sqrt{\log m})-1\}\cdot\left\|\mathbf{s}^{(t)}-\mathbf{x}\right\|_2\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2 && (144)\\
&\le (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\cdot\left\|\mathbf{s}^{(t)}-\mathbf{x}\right\|_2\cdot\left\|\mathbf{s}^{(t+1)}-\mathbf{x}\right\|_2 && (145)\\
&\le (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\cdot\left(\left\|\mathbf{x}^{(t)}-\mathbf{x}\right\|_2+\delta\right)\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right). && (146)
\end{aligned}$$

The third term of (127): By the TS-REC in Lemma 3, we obtain that for any $\delta \in (0,1)$, when $m = \Omega(k\log\frac{Lr}{\delta})$, with probability $1 - e^{-\Omega(m)}$,

$$\frac{1}{\sqrt{m}}\cdot\sqrt{\sum_{i=1}^m\left(\mathbf{a}_i^T\left(\mathbf{x}^{(t+1)}-\mathbf{x}\right)\right)^2} \le \left((1+c)\|\mathbf{x}^{(t+1)}-\mathbf{x}\|_2+\delta\right), \tag{147}$$

where c is a sufficiently small positive constant. By (128) and similarly to the derivation of the TS-REC, we have that when $m = \Omega(k\log\frac{Lr}{\delta})$, with probability $1 - e^{-\Omega(m)}$,

$$\frac{1}{\sqrt{m}}\cdot\sqrt{\sum_{i=1}^m\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)^2} \le (1+c+\delta). \tag{148}$$

Then, from the Cauchy-Schwarz inequality, the third term in (127) can be upper bounded as

$$\begin{aligned}
&\left|2\left\langle\frac{\zeta}{m}\cdot\sum_{i=1}^m\left(\nu-\hat{\nu}^{(t)}\right)\cdot\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)\mathbf{a}_i, \mathbf{x}^{(t+1)}-\mathbf{x}\right\rangle\right| \\
&\le \frac{2\zeta\cdot\left|\nu-\hat{\nu}^{(t)}\right|}{\sqrt{m}}\cdot\sqrt{\sum_{i=1}^m\left(\mathbf{a}_i^T\mathbf{x}^{(t)}\right)^2}\cdot\frac{1}{\sqrt{m}}\cdot\sqrt{\sum_{i=1}^m\left(\mathbf{a}_i^T\left(\mathbf{x}^{(t+1)}-\mathbf{x}\right)\right)^2} && (149)\\
&\le 2\zeta(1+c+\delta)\cdot\left|\nu-\hat{\nu}^{(t)}\right|\cdot\left((1+c)\|\mathbf{x}^{(t+1)}-\mathbf{x}\|_2+\delta\right). && (150)
\end{aligned}$$

Combining terms: Combining (127), (130), (133), (135), (137), (138), (139), (141), (146), and (150), we obtain that when $m = \Omega\big(\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)\big)$, with probability $1 - e^{-\Omega(k\log\frac{Lr}{\delta})}$,

$$\begin{aligned}
&\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2^2 \le Cn\zeta K_y\delta \\
&\quad + 2\zeta\left(CK_y\sqrt{\frac{\big(k\log\frac{Lr}{\delta}\big)\cdot(\log m)}{m}} + \nu\left(\left\|\mathbf{x}^{(t)}-\mathbf{x}\right\|_2+\delta\right)^2\right)\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right) \\
&\quad + (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\cdot\left(\left\|\mathbf{x}^{(t)}-\mathbf{x}\right\|_2+\delta\right)\cdot\left(\left\|\mathbf{x}^{(t+1)}-\mathbf{x}\right\|_2+\delta\right) \\
&\quad + 2\zeta(1+c+\delta)\cdot\left|\nu-\hat{\nu}^{(t)}\right|\cdot\left((1+c)\|\mathbf{x}^{(t+1)}-\mathbf{x}\|_2+\delta\right). && (151)
\end{aligned}$$

Considering ζ as a positive constant and setting $\delta = \frac{K_y}{mn}$, as well as using $n = \Omega(m)$, similarly to (106), we obtain

$$\begin{aligned}\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2^2 &\leq \frac{CK_y^2}{m} + \left(\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 + \frac{K_y}{mn}\right) \\ &\times \left(CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + 2\zeta\nu\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2^2 + (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\cdot\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2\right) \\ &+ 2\zeta\left(1 + c + \frac{K_y}{mn}\right)\cdot\left|\nu - \hat{\nu}^{(t)}\right|\cdot\left((1+c)\|\mathbf{x}^{(t+1)} - \mathbf{x}\|_2 + \frac{K_y}{mn}\right). \qquad (152)\end{aligned}$$

From Lemma 8, we have

$$\left|\nu - \hat{\nu}^{(t)}\right| \leq CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + \left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2^2\cdot\nu. \qquad (153)$$

Then, since $c > 0$ is sufficiently small, we obtain

$$\begin{aligned}&2\zeta\left(1 + c + \frac{K_y}{mn}\right)\cdot\left|\nu - \hat{\nu}^{(t)}\right|\cdot\left((1+c)\|\mathbf{x}^{(t+1)} - \mathbf{x}\|_2 + \delta\right) \\ &\leq \left(CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + 3\zeta\nu\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2^2\right)\cdot\left(\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 + \frac{K_y}{mn}\right). \qquad (154)\end{aligned}$$

Combining (152) and (154), we obtain

$$\begin{aligned}\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2^2 &\leq \frac{C}{m} + \left(\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 + \frac{K_y}{mn}\right) \\ &\times \left(CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + 5\zeta\nu\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2^2 + (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\cdot\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2\right). \qquad (155)\end{aligned}$$

Therefore, from the quadratic formula, we obtain

$$\begin{aligned}\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 &\leq CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} \\ &+ \left(5\zeta\nu\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2 + (2\cdot|1-\zeta\nu| + 1/\sqrt{\log m})\right)\cdot\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2. \qquad (156)\end{aligned}$$

Then, since $\frac{1}{\sqrt{\log m}} = o(1)$, if (18) holds for $t = 0$, i.e., $5\zeta\nu\cdot\left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2 + 2\cdot|1-\zeta\nu| + \beta_1 = 1 - \beta_2$, we obtain from (156) that

$$\left\|\mathbf{x}^{(1)} - \mathbf{x}\right\|_2 \leq CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + (1-\beta_2)\cdot\left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2. \qquad (157)$$

When

$$\left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2 > \frac{CK_y}{\beta_2}\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}, \qquad (158)$$

from (157), we obtain

$$\left\|\mathbf{x}^{(1)} - \mathbf{x}\right\|_2 < \left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2. \qquad (159)$$

This in turn leads to $5\zeta\nu\cdot\left\|\mathbf{x}^{(1)} - \mathbf{x}\right\|_2 + 2\cdot|1-\zeta\nu| + \beta_1 < 1 - \beta_2$, and thus from (156), we obtain

$$\left\|\mathbf{x}^{(2)} - \mathbf{x}\right\|_2 \leq CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + (1-\beta_2)\cdot\left\|\mathbf{x}^{(1)} - \mathbf{x}\right\|_2. \qquad (160)$$

Therefore, if letting $T_0 \in \mathbb{N}$ be the smallest integer such that the inequality

$$\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2 > \frac{CK_y}{\beta_2}\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}, \qquad (161)$$

is violated, by induction, we obtain that the sequence $\left\{\|\mathbf{x}^{(t)} - \mathbf{x}\|_2\right\}_{t\in[0,T_0]}$ is monotonically decreasing, and for all $t \le T_0$, we have

$$\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 \le CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + (1-\beta_2)\cdot\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2, \tag{162}$$

which gives

$$\left\|\mathbf{x}^{(t+1)} - \mathbf{x}\right\|_2 \le \left(1 + (1-\beta_2) + \ldots + (1-\beta_2)^t\right)\cdot CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} \tag{163}$$

$$+ (1-\beta_2)^{t+1}\left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2 \tag{164}$$

$$< \frac{CK_y}{\beta_2}\cdot\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + (1-\beta_2)^{t+1}\cdot\left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|_2. \tag{165}$$

In addition, since $T_0 \in \mathbb{N}$ is the smallest integer such that (161) is violated, we have

$$\left\|\mathbf{x}^{(T_0)} - \mathbf{x}\right\|_2 \le \frac{CK_y}{\beta_2}\cdot\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}. \tag{166}$$

Then, by the linear convergence rate in (165) and $\|\mathbf{x}^{(0)} - \mathbf{x}\|_2 \le 2$, we obtain $T_0 = O\big(\log\big(\frac{m}{(k\log(nLr))\cdot(\log m)}\big)\big)$. Moreover,

$$\left\|\mathbf{x}^{(T_0+1)} - \mathbf{x}\right\|_2 \le CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + (1-\beta_2)\cdot\left\|\mathbf{x}^{(T_0)} - \mathbf{x}\right\|_2 \tag{167}$$

$$\le CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} + \frac{1-\beta_2}{\beta_2}\cdot CK_y\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}} \tag{168}$$

$$= \frac{CK_y}{\beta_2}\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}. \tag{169}$$

Then, by induction, we obtain for all $t \ge T_0$ that

$$\left\|\mathbf{x}^{(t)} - \mathbf{x}\right\|_2 \le \frac{CK_y}{\beta_2}\sqrt{\frac{(k\log(nLr))\cdot(\log m)}{m}}, \tag{170}$$

which completes the proof.

D Additional Numerical Results for the MNIST Dataset

In this section, we present some additional numerical results for the MNIST dataset. We first present a figure to illustrate how quickly does Step 2 of Algorithm 1 converge in Appendix D.1, then we present how the reconstruction error of Algorithm 1 varies with respect to $1/\sqrt{m}$ in Appendix D.2. In Appendix D.3, we empirically illustrate the effect of the scale factor $\hat{\nu}^{(t)}$ used in (15). In Appendix D.4, we compare with the method proposed in [80]. A simple numerical comparison with the approximate message passing (AMP) algorithm proposed in [3] is provided in Appendix D.5.

D.1 The Convergence Rate of Step 2 of Algorithm 1

For the noisy magnitude-only measurement model (22), we fix $m = 400$ and $\sigma = 0.1$ to see how the reconstruction error decays with respect to the number of iterations of Step 2 of Algorithm 1. The results are illustrated in Figure 5. From Figure 5, we observe that the logarithm of the reconstruction error decays almost linearly during the first 20 iterations.

D.2 The Reconstruction Error against $1/\sqrt{m}$

According to Theorem 2, for a fixed generative model, the final reconstruction error of our Algorithm 1 should scale as $O\big(1/\sqrt{m}\big)$ (ignoring the $\log m$ term). This is numerically verified in Figure 6, for which we consider the noisy magnitude-only measurement model (22) with $\sigma = 0$ or $\sigma = 0.1$.

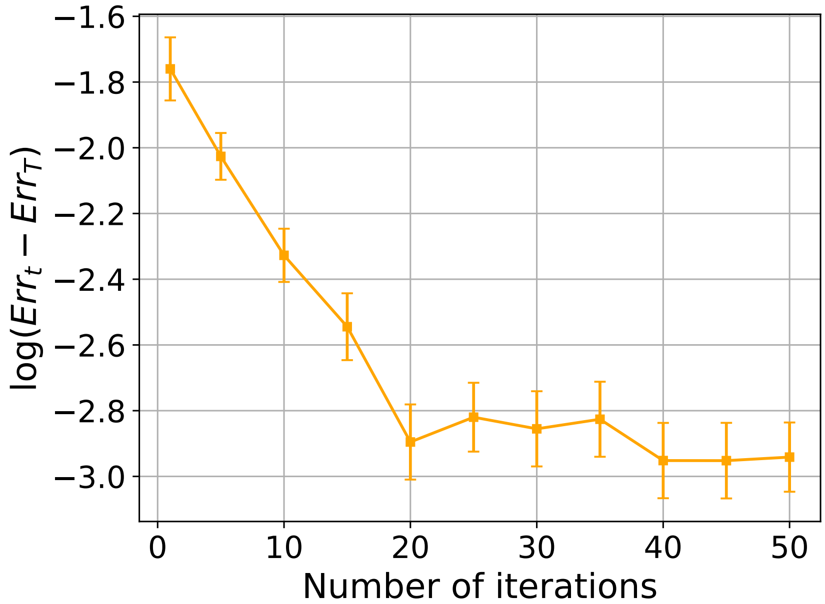
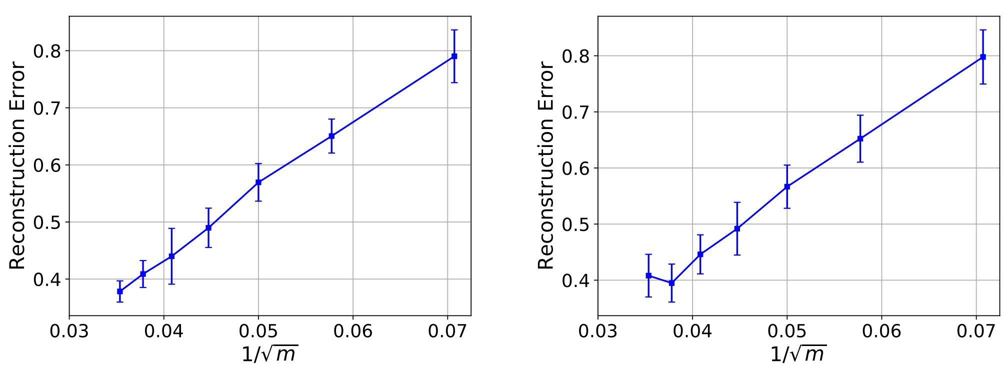


Figure 5: A plot of the log reconstruction error against the number of iterations.



(a) (22) with $\sigma = 0$

(b) (22) with $\sigma = 0.1$

Figure 6: Plots of reconstruction error of MPRG against $1/\sqrt{m}$ for the measurement model (22).

D.3 The Effect of the Scale Factor $\hat{\nu}^{(t)}$ used in (15)

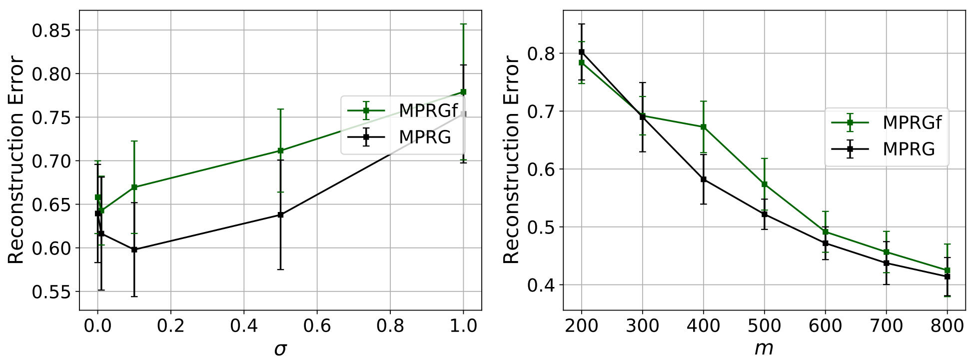
In Algorithm 1, the scale factor $\hat{\nu}^{(t)}$ is calculated according to (13) and is used in (15). We remark that this scale factor plays an important role. To illustrate this, we compare Algorithm 1 (recall that it is denoted by `MPRG`) with the case that using a fixed $\hat{\nu}^{(0)}$ in (15) (i.e., it is not varying with respect to t) during the iterations of the second step of Algorithm 1, which is denoted by `MPRGf`. For the measurement models (24) and (25), the numerical results are presented in Figures 7 and 8. From these figures, we observe that by using a varying scale factor $\hat{\nu}^{(t)}$, we obtain better reconstructed images and smaller reconstruction errors.

D.4 Comparison with the Method proposed in [80]

In this subsection, we compare with the method proposed in [80], which we denote by `prGAN`. We focus on the noiseless measurement model $y_i = |\mathbf{a}_i^T \mathbf{x}|$ for $i \in [m]$, and the numerical results are presented in Figures 9 and 10. We observe from these figures that the three methods `prGAN`, `APPGD`,

(a) (24) with $m = 300$ and $\sigma = 0.01$ (b) (25) with $m = 400$ and $\sigma = 0.5$

Figure 7: Examples of reconstructed MNIST images of MPRG and MPRGf.



(a) (24) with fixed $m = 300$ (b) (25) with fixed $\sigma = 0.5$

Figure 8: Quantitative comparisons of the performance of MPRG and MPRGf.

and MPRG lead to competitive reconstruction error. However, prGAN is not as stable as APPGD and MPRG, and it may result in reconstructed images that are not desired.

D.5 Comparison with the AMP Algorithm proposed in [3]

We follow the setting in [3] to use a ReLU neural network generative model with no offsets and zero-mean random Gaussian weights. The architecture of this neural network generative model G is the same as the decoder of the VAE model used for the MNIST dataset, i.e., the latent dimension $k = 20$, there are two hidden layers with 500 neurons, and the output dimension is 784. We randomly

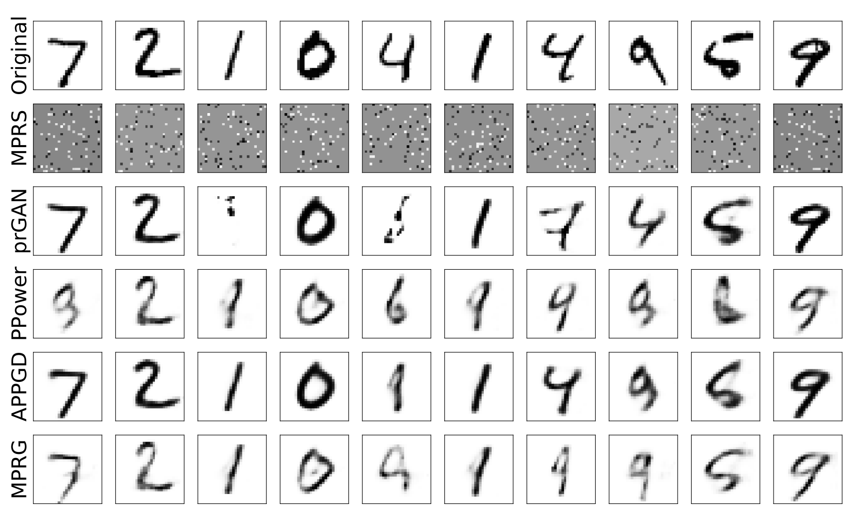


Figure 9: Examples of reconstructed MNIST images of prGAN.

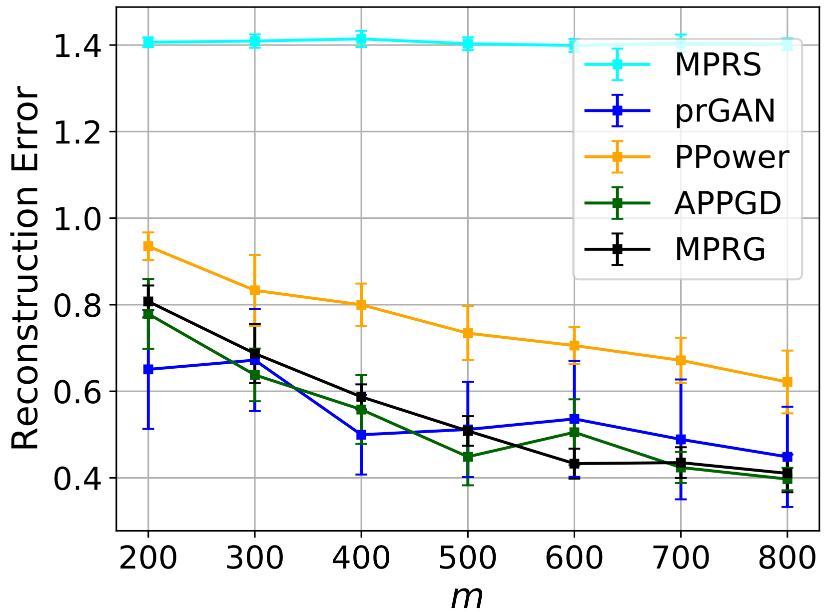
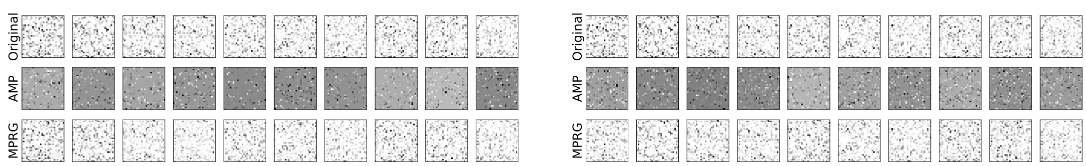


Figure 10: Quantitative comparisons to the performance of `prGAN`.



(a) $m = 200$ (b) $m = 300$

Figure 11: Examples of reconstructed signals of `MPRG` and `AMP`.

choose 10 latent vectors $\mathbf{z} \in \mathbb{R}^k$ and use the corresponding $G(\mathbf{z})$ as the signal. Under this ReLU neural network generative model and the noiseless measurement model $y_i = |\mathbf{a}_i^T\mathbf{x}|$, we compare our Algorithm 1 with the AMP algorithm used in [3]. The reconstruction error is averaged over the 10 signals and 10 random restarts. The results are presented in Figures 11 and 12. We can observe from these figures that `MPRG` leads to better reconstruction performance compared to `AMP`.

E Empirical Results for the CelebA Dataset

The CelebA dataset contains more than $200,000$ face images of celebrities. Each input image is cropped into a 64×64 RGB image and thus $n = 64 \times 64 \times 3 = 12288$. The generative model G is set to be (the normalized version of) the Deep Convolutional Generative Adversarial Networks (DCGAN) model pre-trained by the authors of [6], with the latent dimension being $k = 100$. The projection operator $\mathcal{P}_G(\cdot)$ is approximated by the Adam optimizer with a learning rate of 0.1 and 100 steps. To reduce the impact of local minima, we present the best reconstructed images among 6 random restarts. Other involved parameters are set to be the same as those for the MNIST dataset.

For the CelebA dataset, we do not compare with the sparsity-based method `MPRS` since the face images are clearly not sparse in the natural domain and we have observed from the results for the MNIST dataset that `MPRS` leads to poor reconstructions. Recall that we have performed numerical experiments for several measurement models (*cf.* (22), (23), (24) and (25)) for the MNIST dataset. In this section, we only present some proof-of-concept experimental results for the following measurement model:

$$y_i = |\mathbf{a}_i^T\mathbf{x} + \eta_i| + 5\tanh(|\mathbf{a}_i^T\mathbf{x}|), \quad i = 1, 2, \ldots m, \tag{171}$$

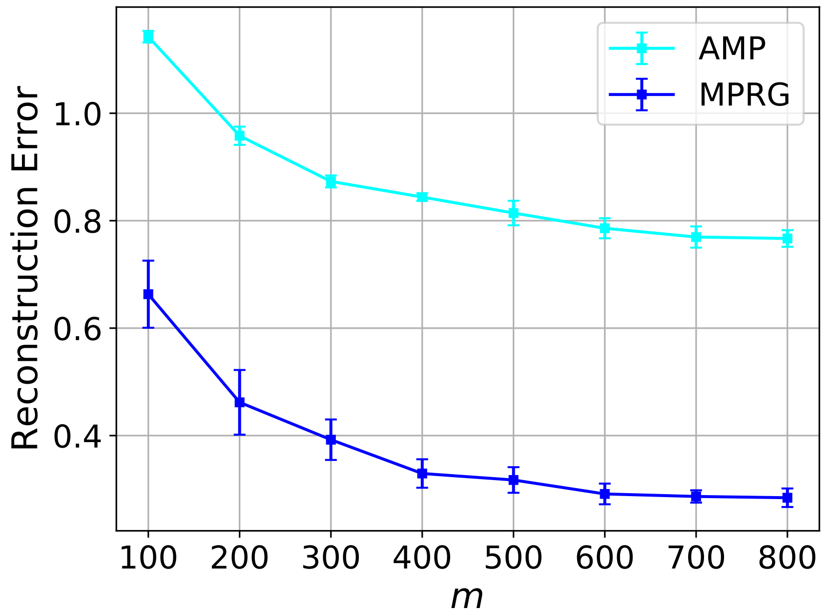


Figure 12: Quantitative comparisons to the performance of `AMP` under a ReLU neural network generative model.

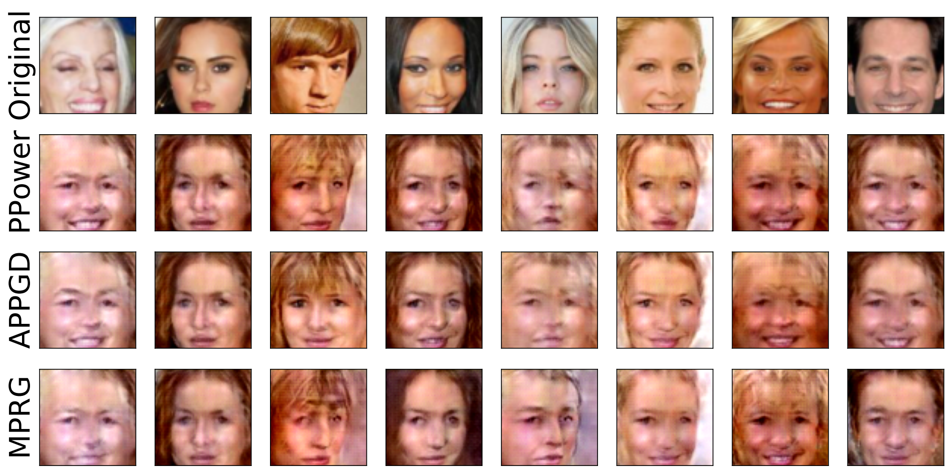


Figure 13: Examples of reconstructed images of the CelebA dataset for the measurement model (171) with $m = 6000$ and $\sigma = 0.4$.

where η_i are i.i.d. realizations of an $\mathcal{N}(0, \sigma^2)$ random variable. The model in (171) can be thought of as a misspecified version of the measurement model $y_i = |\mathbf{a}_i^T \mathbf{x} + \eta_i|$. The reconstructed images are presented in Figure 13, from which we observe that our method `MPRG` leads to the best reconstructed images compared to those of `PPower` and `APPGD`. Quantitative comparisons according to the reconstruction error are provided in Figure 14. From this figure, we observe that when $m > 4000$, `MPRG` gives smallest reconstruction error. We expect the advantage of our method to be more significant as the level of misspecification increases, i.e., the multiplier of the $\tanh$ term in (171) becomes larger.

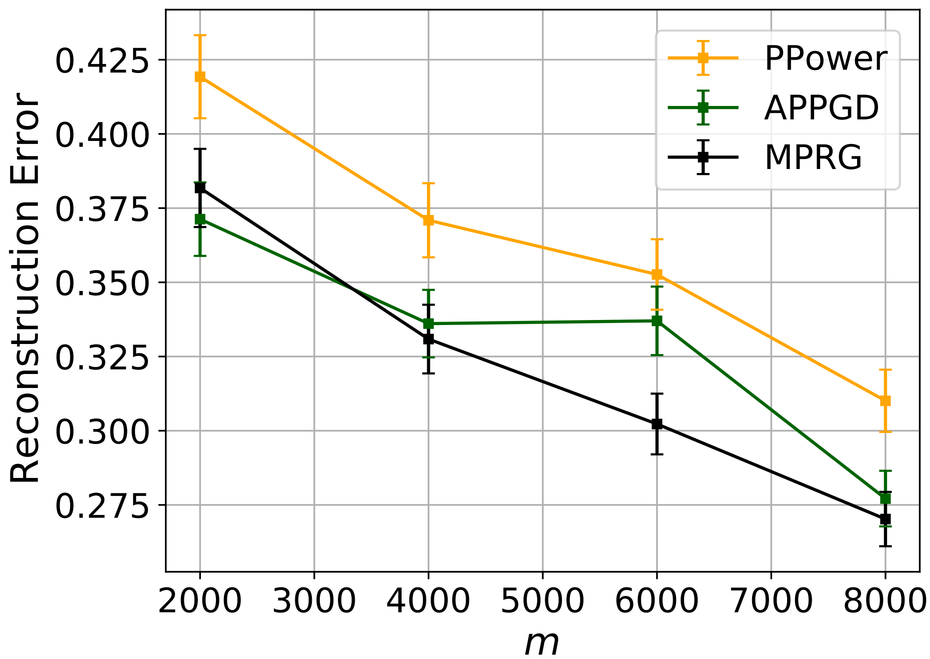


Figure 14: Quantitative comparisons of the performance for the CelebA dataset and measurement model (171) with $\sigma = 0.4$ and varying m.