Recovery Guarantees for Posterior Sampling of One-Bit Compressed Sensing

Jing Ma[⋆] **Yujia Wu**[⋆] **Zhaoqiang Liu**[†]
University of Electronic Science and Technology of China

Abstract

We study the sample complexity of noisy one-bit compressed sensing for signals drawn from a prior distribution. By characterizing the effective distributional complexity of the prior via its approximate covering number, we prove that posterior sampling achieves accurate recovery with high probability when the number of measurements scales with the logarithm of the approximate covering number, up to a one-bit separation gap factor. This upper bound is robust to learned prior mismatch. Specifically, we show that posterior sampling with an approximate prior remains reliable, provided that the learned prior distribution is sufficiently close to the true signal distribution in Wasserstein distance. In addition, we establish a sample complexity lower bound for any reliable method of noisy one-bit compressed sensing, showing that our upper bound is nearly matched in its main prior dependent term. To approximate the ideal posterior sampling process for real world scenarios, we instantiate posterior sampling through a plug-and-play algorithm with diffusion priors. Experiments on the FFHQ and ImageNet datasets demonstrate the effectiveness of our proposed approach.

1 Introduction

Compressed sensing studies how to reliably recover high dimensional signals from a number of measurements far below the ambient dimension. Its success relies on a central premise: High dimensional signals often have low intrinsic dimension [5, 7, 35]. Classical compressed sensing theory formalizes this structure through models such as sparsity [10], unions of subspaces [26], low rankness [58], or low dimensional manifolds [5], and establishes recovery guarantees when the number of measurements is sufficiently large relative to the intrinsic complexity of the underlying signal class [10, 23, 5, 58]. However, in many modern sensing systems, the measurement process is often further constrained by low precision acquisition devices, limited storage, or restricted communication bandwidth [34, 43, 61, 60, 30]. In such scenarios, the sensor may record only the signs of linear measurements, leading to the one-bit compressed sensing model [8, 55]:

$$y = \mathrm{sign}(Ax^* + \xi) \in \{-1, +1\}^m, \tag{1}$$

where $x^* \in \mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the unknown signal, $A \in \mathbb{R}^{m\times n}$ is the measurement matrix, and $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$ denotes Gaussian noise with noise level $\sigma > 0$. For one-bit measurements, there is a loss of the magnitude information on the signal [8, 56, 42] and we focus on directional recovery and measure reconstruction error by geodesic distance on $\mathbb{S}^{n-1}$.

Classical compressed sensing often adopts a set based viewpoint, seeking recovery over a prescribed structured class such as sparse vectors [53, 52], low dimensional manifolds [32, 25], or range of generative model [7, 31]. While this perspective has been successful, it can be overly pessimistic for expressive priors since it ignores how probability mass is distributed within the signal model [35, 59]. A distributional viewpoint instead treats the signal prior itself as the object of study and asks for high

[⋆]Equal contribution. [†]Corresponding author (zqliu12@gmail.com).

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

probability recovery when the signal x^* is drawn from that prior [36, 72, 59, 2]. In this setting, it is more natural to measure the complexity of the distribution through the approximate covering number of a high probability region [35, 1], rather than through the covering number of the entire support. For linear Gaussian measurements, it is shown that the approximate covering number governs the sample complexity of posterior sampling for arbitrary data priors [35]. This motivates the central question of this paper: For signals drawn from an arbitrary data prior distribution on the unit sphere, how should the sample complexity of reliable recovery from noisy one-bit measurements be characterized in terms of the prior distribution?

1.1 Related works

Early work on one-bit compressed sensing studied sensing systems in which only the signs of linear measurements are retained [8]. Since such sign observations discard amplitude information, traditional one-bit compressed sensing focuses on directional recovery, typically under a normalization constraint on the signal [55, 34]. For sparse signals, foundational results establish recovery guarantees from noisy [39] or noiseless sign measurements using convex programs and related optimization methods [55, 56, 34, 3]. A key geometric principle underlying these results is that random hyperplanes can provide binary embeddings: As first proven in [34] and soon after in [56], the normalized Hamming distance between sign patterns concentrates around the angular distance between signals, uniformly over low complexity sets. Subsequent works further develop robust recovery guarantees under measurement noise, sign flips, and other quantization effects [34, 42]. These classical results show that one-bit measurements can preserve sufficient angular information for reliable recovery over structured signal classes, despite the severe information loss caused by quantization.

With the development of deep learning, the seminal work [7] shows that signals lying near the range of a generative model can be recovered from significantly fewer linear measurements than the ambient dimension by optimizing over the latent space of the generator. This viewpoint replaces hand crafted structures, such as sparsity or low rankness, with the range of a learned generator, and has motivated a broad line of work on compressed sensing with generative priors [46, 13, 14]. Motivated by these developments, recent studies have extended generative prior methods to one-bit compressed sensing, where recovery is typically performed by searching for a latent code whose generated signal is consistent with the observed signs [38, 57, 47]. Theoretical analyses in this direction often rely on binary stable embeddings or uniform concentration over the generator range, leading to sample complexity bounds for Lipschitz generative models under Gaussian hyperplane measurements [47, 46, 57]. More recently, pretrained diffusion generative models [33, 63, 64] have emerged as powerful deep generators, and are widely leveraged as highly flexible priors for signal recovery [18, 41, 69, 68, 75, 70, 73, 4, 76, 19, 71, 20, 9, 11, 37, 48], including subsequent extensions to one-bit compressed sensing [51, 65, 16]. However, the theoretical guarantees for these diffusion based methods remain relatively underexplored. Unlike previous set based guarantees over prescribed signal classes or range of generator [47, 45, 13, 12], we study arbitrary data prior distributions on the unit sphere and characterize the sample complexity through the approximate covering complexity [35] of high probability regions.

The work [35] is perhaps the most closely relevant work, which develops a posterior sampling framework for linear compressed sensing with arbitrary data priors. Their theory introduces approximate covering numbers as a central complexity measure for distributional compressed sensing and shows that posterior sampling is instance optimal under linear Gaussian measurements. We extend this idea to one-bit measurements. This extension is not a simple technical modification. Since one-bit observations retain only binary sign patterns, residual comparisons and likelihood separation arguments from continuous linear measurements are no longer directly applicable, and we can only distinguish spherical signals using a separation gap defined through the expected Hamming distance. Moreover, the nonlinearity of the one-bit measurement model makes the analysis of noise and model mismatch more difficult. Additive Gaussian noise, after passing through the sign function, no longer behaves as a simple continuous perturbation. When the true prior and the sampling prior are mismatched, we use a Rényi type stability [66] argument to control how this perturbation affects the posterior distribution. In addition, the analysis is carried out on the unit sphere with geodesic distance, which requires spherical versions of covering, separation, and Wasserstein coupling arguments.

1.2 Contributions

Our work extends the distributional posterior sampling framework from continuous linear measurements to the noisy one-bit setting, together with the corresponding analysis of noisy binary separation, recovery guarantees under prior mismatch, and a nearly matching information-theoretic necessary condition in the leading prior complexity term. Our main results are summarized as follows:

- In Section 3.1, we derive an upper bound on the sample complexity with Gaussian measurement matrices. The result gives the upper bound of sample complexity conditions for achieving geodesic recovery accuracy η, and is robust to mismatch between the learned prior distribution and the true signal distribution when they are sufficiently close in Wasserstein distance.
- In Section 3.2, we derive a nearly matching lower bound on the sample complexity required by any reliable recovery method, showing that the approximate covering number in the upper bound is unavoidable.
- In Section 4, we develop a specialized Plug-and-Play method with diffusion priors to approximate the posterior sampling process for noisy one-bit compressed sensing. The experimental results verify the effectiveness of this approach in image recovery tasks.

2 Preliminaries

Definition 1 (Signal and observation distances)**.** *Since signals are normalized to the unit sphere $\mathbb{S}^{n-1}$, we measure signal distance by the normalized geodesic distance $\mathrm{d_S}(x_1, x_2) := \frac{1}{\pi}\arccos\langle x_1, x_2\rangle$, where $x_1, x_2 \in \mathbb{S}^{n-1}$. For binary observations $u, v \in \{-1, +1\}^m$, we measure their discrepancy by the normalized Hamming distance $\mathrm{d_H}(u, v) := \frac{1}{m}\sum_{i=1}^m \mathbf{1}\{u_i \neq v_i\}$.*

Definition 2 ((η, δ)-approximate covering number [35])**.** *For $x_0 \in \mathbb{S}^{n-1}$ and $0 < \eta \leq 1$, define the geodesic η-cap $\mathcal{B}(x_0, \eta) := \{x \in \mathbb{S}^{n-1} : \mathrm{d_S}(x, x_0) \leq \eta\}$. For a distribution R on $\mathbb{S}^{n-1}$ and $\delta \in [0, 1)$, the (η, δ)-approximate covering number is*

$$\mathrm{Cov}_{\eta,\delta}(R) := \min\left\{k \in \mathbb{N} : R\left(\bigcup_{i=1}^k \mathcal{B}(x_i, \eta)\right) \geq 1 - \delta,\ x_i \in \mathbb{S}^{n-1}\right\}. \tag{2}$$

Definition 3 (Geodesic Wasserstein distance)**.** *Let μ and ν be probability distributions on $\mathbb{S}^{n-1}$, and let $\Pi(\mu, \nu)$ denote the set of all couplings of μ and ν. For $p \geq 1$, the geodesic Wasserstein-p distance between μ and ν is defined as*

$$\mathcal{W}_{p,\mathrm{geo}}(\mu, \nu) := \left(\inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{S}^{n-1}\times\mathbb{S}^{n-1}} \mathrm{d_S}(x, y)^p \, d\gamma(x, y)\right)^{1/p}. \tag{3}$$

The corresponding geodesic Wasserstein-∞ distance is defined as

$$\mathcal{W}_{\infty,\mathrm{geo}}(\mu, \nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \operatorname*{ess\,sup}_{(x,y)\sim\gamma} \mathrm{d_S}(x, y). \tag{4}$$

Definition 4 (One-bit separation gap)**.** *For the noisy Gaussian one-bit model with noise level σ, let $f_\sigma(w) = \frac{1}{\pi}\arccos\left(\frac{\cos(\pi w)}{\sqrt{1+\sigma^2}}\right)$, $w \in [0, 1]$. For $u \in (0, 1)$ and $s \in (1, 1/u)$, the one-bit separation gap is defined by $\Delta_\sigma(s, u) := f_\sigma(su) - f_\sigma(u)$. Since $s > 1$, we have $su > u$, by the monotonicity of f_σ, $\Delta_\sigma(s, u) \geq 0$.*

Definition 5 (Maximum row ℓ_2 norm)**.** *Let $A \in \mathbb{R}^{m\times n}$ and let $a_i^\top$ denote the i-th row of A. We define $\|A\|_{2,\infty} := \max_{i\in[m]} \|a_i\|_2$.*

Definition 6 (One-bit probit conditional likelihood)**.** *Under the one-bit probit model, the conditional likelihood of observing $y \in \{-1, +1\}^m$ given x and A is defined as $\Pr(y|x, A) := \prod_{i=1}^m \Phi\left(\frac{y_i a_i^\top x}{\sigma}\right)$, where Φ is the standard Gaussian cumulative distribution function and $\sigma > 0$ is the noise level.*

3 Main results

In this section, we study the sample complexity of reliable signal recovery by presenting both upper bound and lower bound. Specifically, the upper bound provides a sufficient measurement condition

for reliable recovery, whereas the lower bound provides a necessary measurement condition that must be satisfied by any reliable recovery method. In the upper bound analysis, we assume an i.i.d. Gaussian sensing matrix $A_{ij} \sim \mathcal{N}(0,1)$, which ensures that geodesic separation on the sphere is translated into predictable Hamming separation between one-bit sign patterns. In the lower bound analysis, we consider both deterministic and Gaussian sensing matrices.

3.1 Upper bound

In this section, we establish an upper bound on the sample complexity of posterior sampling for one-bit compressed sensing. The key idea is that, under sufficiently many independent Gaussian measurements, geodesically separated signals induce one-bit observations that can be distinguished through their Hamming disagreement.

Theorem 3.1 (One-bit posterior sampling upper bound)**.** *Let R and P be two probability distributions on the unit sphere $\mathbb{S}^{n-1}$.*[1] *Let $p \geq 1$, $0 < \eta < 1$, $0 < \delta < 1/4$, and assume $\mathcal{W}_{p,\mathrm{geo}}(R,P) \leq \rho$. Define the effective mismatch level $\varepsilon_\delta := \rho/\delta^{1/p}$, and define the parameter $\bar{\eta} := \eta + \varepsilon_\delta$. We take $c \in \left(1, \frac{1-\varepsilon_\delta}{\bar{\eta}} - 1\right)$, assume further that $2\eta + 3\varepsilon_\delta < 1$, which guarantees that $\left(1, \frac{1-\varepsilon_\delta}{\bar{\eta}} - 1\right)$ is a nonempty interval. Given (y, A), $\widehat{x} \sim P(\cdot|y,A)$ is the posterior sampling estimator based on P. Furthermore, define $r_{\varepsilon_\delta} := 2\sin\left(\frac{\pi\varepsilon_\delta}{2}\right)$, assume that the model mismatch is sufficiently small, such that $\frac{r_{\varepsilon_\delta}^2}{\sigma^2} \leq \frac{1}{6}\Delta_\sigma(c,\bar{\eta})^2$. For $\zeta \in (0,1)$, when $m \geq C\frac{\log \mathrm{Cov}_{\eta,\delta}(R) + \log 2 + \log(1/\zeta)}{\Delta_\sigma(c,\bar{\eta})^2}$ for some $C > 0$, with probability at least $1-\zeta$ over the draw of A,*

$$\Pr_{x^* \sim R, \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\bar{\eta} + \varepsilon_\delta\right) \leq 2\delta + \zeta. \tag{5}$$

In particular, if $c > 1$ is also a fixed constant and $\varepsilon_\delta = O(\eta)$, then $\bar{\eta} = \Theta(\eta)$. In this case, to ensure $\mathrm{d_S}(x^*, \widehat{x}) = O(\eta)$ with probability at least $1 - 2\delta - \zeta$ conditionally on a measurement matrix A, it suffices to take

$$m = \begin{cases} O\left(\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log(1/\zeta)\right)\eta^{-2}\right), & \sigma \lesssim \eta, \\ O\left(\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log(1/\zeta)\right)\sigma^2\eta^{-4}\right), & \sigma \gtrsim \eta. \end{cases} \tag{6}$$

Remark 1. The condition $\frac{r_{\varepsilon_\delta}^2}{\sigma^2} \leq \frac{1}{6}\Delta_\sigma(c,\bar{\eta})^2$ requires the model mismatch to be small compared with the effective noise level and the one-bit near–far separation gap. Since $r_{\varepsilon_\delta} = 2\sin\left(\frac{\pi\varepsilon_\delta}{2}\right), \varepsilon_\delta = \frac{\rho}{\delta^{1/p}}$, for small ε_δ this condition is equivalent up to constants to $\rho \lesssim \delta^{1/p}\sigma\Delta_\sigma(c,\bar{\eta})$. Thus, the learned sampling prior P must be sufficiently close to the true prior R. If the mismatch is too large, the main probability masses of R and P may be separated on the sphere, so even after Gaussian smoothing their induced one-bit observation distributions may not overlap enough. Consequently, the posterior induced by P may place little mass near the true signal $x^* \sim R$, causing the stability guarantee for posterior sampling to fail. Moreover, extensive studies have shown the generated distribution P can be close to the true distribution R under appropriate conditions [15, 44, 27, 50, 28].

Remark 2. Compared with the linear posterior sampling upper bound [35], the one-bit setting contains an additional factor $\Delta_\sigma(c,\bar{\eta})^{-2}$. This is due to the one-bit observation process, which preserves only sign information and discards the amplitude information available in linear measurements. Consequently, the geometric separation between two covering caps is not directly reflected in the observation space, but can only be detected through the difference between their induced expected Hamming disagreements. This difference is precisely $\Delta_\sigma(c,\bar{\eta}) = f_\sigma(c\bar{\eta}) - f_\sigma(\bar{\eta})$. To reliably distinguish the covering caps, the empirical Hamming distance must concentrate at the scale of $\Delta_\sigma(c,\bar{\eta})$, which requires $m\Delta_\sigma(c,\bar{\eta})^2 \gtrsim \log \mathrm{Cov}_{\eta,\delta}(R)$. Therefore, the sample complexity naturally contains the factor $\Delta_\sigma(c,\bar{\eta})^{-2}$. This term reflects the intrinsic loss of information caused by one-bit quantization and noise.

We first outline the proof of Theorem 3.1. Since one-bit observations retain only sign information, the Euclidean residuals used in linear models are no longer directly available. Therefore, we need to establish a new notion of separability in the binary observation space. Specifically, we first characterize the relationship between spherical geodesic distance and the Hamming disagreement rate of one-bit observations, and then use this relationship to construct a near–far test. We then combine

[1] R is regarded as the true signal distribution and P is regarded as learned distribution corresponding to the generative model.

approximate covering with a union bound to control the probability that the posterior sample falls into an incorrect far-away region, and finally use a Rényi type stability to handle model mismatch.

We first need to characterize how distinguishable two spherical signals are after one-bit observation. Due to the rotational invariance of Gaussian measurements, the sign disagreement probability of two signals depends only on their geodesic distance. The following lemma gives an explicit expression for this relationship.

Lemma 3.2 (Expected Hamming distance between noisy and noiseless one-bit measurements)**.** *Consider two signals $x_1, x_2 \in \mathbb{S}^{n-1}$ defined on the unit Euclidean sphere. Define the observations of these two signals as $u = \operatorname{sign}(Ax_1 + \xi)$ and $v = \operatorname{sign}(Ax_2)$. We have:*

$$\mathbb{E}[\mathrm{d_H}(u, v)] = \frac{1}{\pi} \arccos\left(\frac{\cos(\pi \mathrm{d_S}(x_1, x_2))}{\sqrt{1+\sigma^2}}\right). \tag{7}$$

The previous lemma shows that geodesic separation is transformed by the one-bit channel into a gap in expected Hamming disagreement. Hence the quantity $\Delta_\sigma(c, \eta) = f_\sigma(c\eta) - f_\sigma(\eta)$ measures the observational margin between signals inside $\mathcal{B}(x_2, \eta)$ and signals outside $\mathcal{B}(x_2, c\eta)$-cap. This margin allows us to construct the following near–far test.

Lemma 3.3 (One-bit near–far test lemma)**.** *Fix a reference point $x_0 \in \mathbb{S}^{n-1}$. Let $A_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$ independently, and $u = \operatorname{sign}(Ax + \xi)$ for any fixed $x \in \mathbb{S}^{n-1}$. Let the testing threshold be $\tau_{\eta,c,\sigma} := \frac{f_\sigma(\eta)+f_\sigma(c\eta)}{2}$, and define the test function $\phi_{x_0,\eta}(u; A) := \mathbf{1}\{\mathrm{d_H}(u, \operatorname{sign}(Ax_0)) \geq \tau_{\eta,c,\sigma}\}$. Then the following two testing error bounds hold in the annealed sense:*

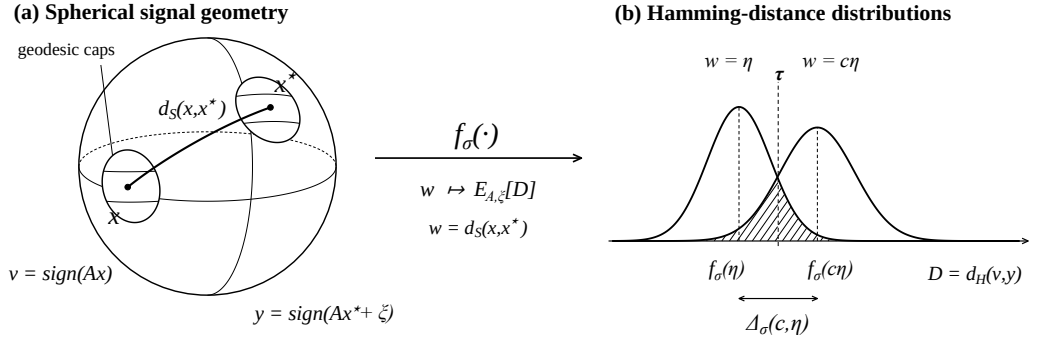
$$\sup_{x:\, \mathrm{d_S}(x,x_0)\leq\eta} \Pr_{A,\xi|x}(\phi_{x_0,\eta}(u; A) = 1) \leq \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right), \tag{8}$$

and

$$\sup_{x:\, \mathrm{d_S}(x,x_0)\geq c\eta} \Pr_{A,\xi|x}(\phi_{x_0,\eta}(u; A) = 0) \leq \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{9}$$

Here the probability is taken with respect to the joint randomness of the measurement matrix A and the noise ξ, conditional on the fixed reference point x. Hence the result is annealed, in the sense that the probability is averaged over the measurement matrix A, rather than a conditional in A bound for a fixed realization of A.

Remark 3. Lemma 3.2 and Lemma 3.3 are closely related to the distance-preservation principle underlying binary stable embeddings. When $\sigma = 0$, we have $f_0(w) = w$, so Lemma 3.2 reduces to the classical random-hyperplane identity $\mathbb{E}[\mathrm{d_H}(\operatorname{sign}(Ax_1), \operatorname{sign}(Ax_2))] = \mathrm{d_S}(x_1, x_2)$, with $\Delta_0(c, \eta) = (c-1)\eta$, while Lemma 3.3 yields concentration of the empirical Hamming distance and hence a near–far separation guarantee. Unlike standard binary stable embeddings, our posterior sampling analysis only requires such near–far separation over an approximate cover of a high-probability region of the prior, rather than uniform distance preservation over all signal pairs. For $\sigma > 0$, $f_\sigma(\cdot)$ and $\Delta_\sigma(c, \eta)$ quantify the additional distortion induced by Gaussian noise before quantization. A visual explanation of this section can be found in the diagram below, the shaded area represents the cases in which we recover from errors.



We now extend the near–far test to a general prior by covering its high probability region with finitely many η-caps. Applying the same far region control to each cap and taking a union bound over the covering centers yields the $\log \mathrm{Cov}_{\eta,\delta}$ term in the sample complexity.

Lemma 3.4 (Overall recovery upper bound for one-bit posterior sampling under a covering partition)**.** *Let Q be a probability distribution on $\mathbb{S}^{n-1}$. Assume that there exist centers $x_1, \ldots, x_N \in \mathbb{S}^{n-1}$ such that, with $\mathcal{C}_j := \{x \in \mathbb{S}^{n-1} : \mathrm{d_S}(x, x_j) \le \eta\}, G := \bigcup_{j=1}^N \mathcal{C}_j$. Fix $\alpha \in (0, \frac{1}{2})$, and assume that*

$$Q = (1-\alpha)Q_g + \alpha Q_b \qquad Q_g(G) = 1, \tag{10}$$

Here Q_g denotes a covered good component of Q, of mixture weight $1-\alpha$, contained in the high-probability region G. The distribution Q_b denotes the remaining component of Q, with mixture weight α. Let $x^* \sim Q_g$, and let $\widehat{x} \sim Q(\cdot|y, A)$ be the posterior sampling estimator based on the same distribution Q. Suppose $0 < \eta < \frac{1}{2}$, and $c \in (1, \frac{1}{\eta} - 1)$. Then we have:

$$\Pr_{x^* \sim Q_g, A, \xi, \widehat{x}} (\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\eta) \le \frac{1}{1-\alpha} N \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{11}$$

The previous lemma handles the matched prior finite cover case. To extend it to the general setting with distribution mismatch, the following lemma controls the probability loss under $\mathcal{W}_{\infty,\mathrm{geo}}$ mismatch through a Rényi type stability argument [66].

Lemma 3.5 (Rényi stability under spherical $W_{\infty,\mathrm{geo}}$ mismatch)**.** *Let R_0 and S_0 be two probability distributions defined on the unit sphere $\mathbb{S}^{n-1}$. Assume that R_0 and S_0 satisfy the spherical Wasserstein-∞ mismatch condition $\mathcal{W}_{\infty,\mathrm{geo}}(R_0, S_0) \le \epsilon$, where $0 < \epsilon \le 1$. Let D be a probability distribution on $\mathbb{S}^{n-1}$. Let $x^* \sim R_0, z^* \sim S_0$. Let Π be a coupling of R_0 and S_0 such that, for $(x^*, z^*) \sim \Pi$, $\mathrm{d_S}(x^*, z^*) \le \epsilon$ almost surely. Let $\widehat{x}$ and $\widehat{z}$ be posterior samples with respect to the model prior D, conditioned on (A, y) and (A, u). Consider the noisy one-bit measurement model*

$$y = \mathrm{sign}(Ax^* + \xi), \qquad u = \mathrm{sign}(Az^* + \xi'), \tag{12}$$

where $A_{ij} \sim \mathcal{N}(0,1)$, $\xi, \xi' \sim \mathcal{N}(0, \sigma^2 I_m)$, and A, ξ, and ξ' are mutually independent. Let $r_\epsilon = 2\sin\left(\frac{\pi\epsilon}{2}\right)$. Then, for any $d > 0$ and any parameter $0 < \lambda \le \min\left\{1, \frac{\sigma^2}{3r_\epsilon^2}\right\}$, we have

$$\Pr\left(\mathrm{d_S}(x^*, \widehat{x}) \ge d + \epsilon\right) \le \left(1 - \frac{\lambda(1+\lambda)r_\epsilon^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \Pr\left(\mathrm{d_S}(z^*, \widehat{z}) \ge d\right)^{\frac{\lambda}{1+\lambda}}. \tag{13}$$

The proof of Theorem 3.1 is obtained by combining Lemmas A.1, 3.4 and 3.5. First, Lemma A.1 reduces the general mismatched priors (R, P) to their high probability components (R', P'), losing only 2δ probability mass. Since R' and P' are coupled within geodesic distance $\varepsilon_\delta = \rho/\delta^{1/p}$, the covering radius is enlarged from η to $\bar{\eta} = \eta + \varepsilon_\delta$. Then Lemma 3.4 gives the posterior sampling recovery bound for the covered good component under the learned prior distribution P. Finally, Lemma 3.5 transfers this bound from P' to R', adding the Rényi stability factor and an extra ε_δ in the recovery radius. At this point, the error probability is controlled in expectation over the random matrix A. Hence, for any $\beta \in (0,1)$, Markov's inequality implies that the corresponding conditional error probability, given A, is at most $1/\beta$ times this expectation with probability at least $1-\beta$ over A. Combining these terms and simplifying yields Lemma A.2. Theorem 3.1 then follows by specializing the auxiliary parameters in Lemma A.2 and simplifying the resulting expression under the small-mismatch condition.

For the specific proof methods and processes, please refer to Section A in Appendix.

3.2 Lower bound

The one-bit observation model induces a stronger information bottleneck than the linear sensing model. In the Gaussian linear model, each real valued measurement may carry an amount of information that scales as a logarithmic signal-to-noise ratio. In contrast, under one-bit quantization, the observation $y \in \{-1, +1\}^m$ is a binary string, and therefore the total information transmitted by the sensing process is intrinsically limited by the entropy of $m \log 2$ nats. Consequently, the mutual information argument used for continuous compressed sensing must be replaced by a channel bound adapted to the one-bit model. This section formalizes this replacement and combines it with a spherical Fano argument to obtain a lower bound on the number of measurements. Throughout this section, logarithms are natural logarithms.

Theorem 3.6 (Information lower bound for one-bit compressed sensing)**.** *Let R be a prior distribution defined on the high dimensional unit Euclidean sphere $\mathbb{S}^{n-1}$, where $x^* \sim R$. Define the recovery accuracy $\eta \in (0, \frac{1}{3})$, suppose that $\delta < 0.1$ and that there exists a recovery algorithm capable of outputting $\widehat{x} \in \mathbb{S}^{n-1}$ based solely on the observation y and the matrix A, such that:*

$$\Pr\left(d_S(\widehat{x}, x^*) \leq \eta\right) \geq 1 - \delta. \tag{14}$$

If A is a deterministic matrix, then the number of measurements m must satisfy:

$$m \geq \frac{0.072\pi\sigma^2}{2(\log 2)\|A\|_{2,\infty}^2} \left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta) - 25\right). \tag{15}$$

If $A_{ij} \sim \mathcal{N}(0,1)$, then the number of measurements m must satisfy:

$$m \geq \frac{0.072\pi}{2(\log 2)\arcsin\left(\frac{1}{1+\sigma^2}\right)} \left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta) - 25\right). \tag{16}$$

Remark 4 (Comparison between the upper and lower bounds). We now compare the upper bound in Theorem 3.1 with the lower bound in Theorem 3.6. Then Theorem 3.1 shows that when $m \gtrsim \frac{\log \mathrm{Cov}_{\eta,\delta}(R)}{\Delta_\sigma(c,\eta)^2}$, posterior sampling achieves accurate recovery with high probability. In contrast, Theorem 3.6 shows that any estimator that achieves $O(\eta)$ accurate recovery with constant success probability must satisfy $m \gtrsim \log \mathrm{Cov}_{3\eta,4\delta}(R)$, where the constants may depend on the one-bit channel and the noise level.

Therefore, the two bounds nearly match in the leading term $\log \mathrm{Cov}_{\eta,\delta}(R)$ depending on the prior R. This shows that the approximate covering number is the correct distributional complexity measure for one-bit compressed sensing with arbitrary data priors. In particular, the complexity term appearing in the posterior sampling upper bound generally cannot be removed or substantially improved.

The remaining difference is mainly the factor $\Delta_\sigma(c,\eta)^{-2}$ in the upper bound. This factor comes from the one-bit observation process. Since the measurements only retain the signs, two signals separated on the sphere can be distinguished only through the gap between their expected Hamming distances. This gap is exactly measured by $\Delta_\sigma(c,\eta)$. If c is chosen as a fixed numerical constant, for instance $c = 2$, then in the small-noise regime, we have $\Delta_\sigma(c,\eta) \asymp \eta$. Therefore, the upper bound requires sufficiently many measurements for this Hamming-distance gap to concentrate. Thus, the upper and lower bounds are nearly matched in terms of $\log \mathrm{Cov}_{\eta,\delta}(R)$.

Remark 5. For the lower bound, we explicitly consider arbitrary fixed deterministic sensing matrices and i.i.d. Gaussian random matrices, which already cover many commonly used sensing settings. In particular, since the deterministic result holds for any fixed A, it also allows A to be optimized using knowledge of the prior distribution R, provided that A is independent of the realized signal x^*. Our current analysis does not establish a universal lower bound for more general randomized or adaptive sensing strategies. In such settings, the sensing matrix may depend on additional randomness or side information correlated with the realized signal, in which case the independence argument used in Lemma B.1 no longer applies directly and a separate information-theoretic analysis is required.

Following the method in [35], we establish an upper bound on the mutual information conveyed by the one-bit observation channel. We derive our first lemma by starting from the mutual information conveyed between the observation y and the ground truth x^*.

Lemma 3.7 (Mutual Information Bound)**.** *Assume that the measurement process is $y = \mathrm{sign}(Ax^* + \xi)$, where $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$. If A is a deterministic matrix, we have*

$$I(y; x^*) \leq \frac{2m \log 2}{\pi\sigma^2} \|A\|_{2,\infty}^2. \tag{17}$$

If $A_{ij} \sim \mathcal{N}(0,1)$, we have

$$I(y; x^* | A) \leq \frac{2m \log 2}{\pi} \arcsin\left(\frac{1}{1+\sigma^2}\right). \tag{18}$$

Next, utilizing the properties of Markov chains, we express the mutual information between x^* and $\widehat{x}$. Here, we directly refer to the conclusion in [35] and obtain that if A is a Gaussian matrix, we

have $I(x^*;\widehat{x}) \le I(y;x^*|A)$; if A is a fixed matrix, we have $I(x^*;\widehat{x}) \le I(y;x^*)$ (see Lemma B.1 for details).

We next establish a spherical Fano type inequality, which converts the high probability recovery guarantee into a lower bound on the mutual information between the signal and its estimate.

Lemma 3.8 (Variant of spherical Fano inequality)**.** *Let $(x,\widehat{x})$ be random variables with a joint distribution on the unit Euclidean sphere $\mathbb{S}^{n-1}\times\mathbb{S}^{n-1}$, where $x\sim R$, and suppose that the estimator satisfies*

$$\Pr\left(d_S(x,\widehat{x})\le\eta\right)\ge 1-\delta. \tag{19}$$

Assuming δ is sufficiently small, then for any $\tau\in(0,1-2\delta)$, the following holds:

$$(1-\delta)\tau(1-2\delta)\log \mathrm{Cov}_{3\eta,\tau+3\delta}(R)\le I(x;\widehat{x})+2(1-\delta). \tag{20}$$

Now, based on Lemma 3.8, we obtain the requirement of the mutual information on the effective geometric complexity, that is, $(1-\delta)\tau(1-2\delta)\log \mathrm{Cov}_{3\eta,\tau+3\delta}(R)\le I(x;\widehat{x})+2(1-\delta)$. On the other hand, according to Lemma 3.7 and Lemma B.1 in Appendix, we derive the upper bound of the mutual information: When A is a fixed matrix, we have $I(\widehat{x};x^*)\le\frac{2m\log 2}{\pi\sigma^2}\|A\|_{2,\infty}^2$ and when A is a Gaussian matrix, we have $I(\widehat{x};x^*|A)\le\frac{2m\log 2}{\pi}\arcsin\left(\frac{1}{1+\sigma^2}\right)$. Therefore, substituting the corresponding mutual information upper bound into the Fano inequality requirement and solving for m yields the desired lower bounds in Theorem 3.6.

For the specific proof methods and processes, please refer to Section B in Appendix.

4 Algorithm and experiments

4.1 Plug-and-play approach for one-bit measurements

While our theory establishes guarantees for ideal posterior sampling, exact inference in high dimensions is intractable. To provide a practical counterpart, we introduce PnP-OneBit[2], which approximates this sampling process using diffusion models as the learned prior. Specifically, in one-bit quantized signal recovery, given measurements y and the forward matrix A, our objective is to reconstruct x^* in Eq. (1) by sampling from the target posterior:

$$p(x|y,A)\propto p(y|x,A)p(x)=\exp\left(-\mathcal{L}(x;y)-\mathcal{P}(x)\right), \tag{21}$$

where $\mathcal{P}(x):=-\log p_{\text{data}}(x)$ is the negative log-prior function, and $\mathcal{L}(x;y):=-\log p(y|x,A)$ is the negative log-likelihood function. In one-bit signal recovery, we can exactly express $\mathcal{L}(x;y)=-\sum_{i=1}^m\log\Phi\left(\frac{y_i a_i^\top x}{\sigma}\right)$ based on the cumulative distribution function of the standard normal distribution $\Phi(\cdot)$.

While existing one-bit recovery methods like Diff-OneBit [16] typically seek a single optimal estimate via half quadratic splitting (HQS) [29], we propose PnP-OneBit for more robust signal reconstruction. Earlier works have explored sampling-based plug-and-play approaches for inverse problems. In particular, the work [20] incorporated diffusion-based generative priors into a split Gibbs sampling framework, while work [9] developed a related generative plug-and-play framework for posterior sampling. Building on these developments, we adapt the PnP-DM framework [71] and the split Gibbs sampler (SGS) [67] to the nonlinear one-bit likelihood. This alternating sampling strategy aligns with the algorithmic structure of [71], which has been shown to be provably robust for image recovery. By introducing an auxiliary variable $z\in\mathbb{R}^n$ and a coupling parameter $\varrho>0$, we construct a joint distribution

$$\Pi(x,z)\propto\exp\left(-\mathcal{L}(z;y)-\mathcal{P}(x)-\frac{1}{2\varrho^2}\|x-z\|_2^2\right). \tag{22}$$

Starting from an initialization $x^{(0)}$, SGS alternates between sampling from the following two conditional distributions at each iteration $k=1,2,\dots,K$:

$$z^{(k)}\sim\Pi(z|x^{(k-1)})\propto\exp\left(-\mathcal{L}(z;y)-\frac{1}{2\varrho^2}\|z-x^{(k-1)}\|_2^2\right), \tag{23}$$

[2]The corresponding code is available at https://github.com/Dugehude/pnp-onebit.

$$x^{(k)} \sim \Pi(x|z^{(k)}) \propto \exp\left(-\mathcal{P}(x) - \frac{1}{2\varrho^2}\|x - z^{(k)}\|_2^2\right). \tag{24}$$

Algorithm 1 PnP-OneBit

Require: iteration number K
1: Sample $x^{(0)} \sim \mathcal{N}(0, I_n)$
2: **for** $k = 0, 1, \ldots, K-1$ **do**
3: Sample $z^{(k)} \sim \Pi(z|x^{(k)})$ via Langevin Monte Carlo
4: Sample $x^{(k+1)} \sim \Pi(x|z^{(k)})$ via DDIM denoising
5: **end for**
6: **return** $x^{(K)}$

Sampling from Eq. (23) can be achieved via Langevin Monte Carlo due to the differentiability of $\mathcal{L}(z; y)$ in the one-bit setting. However, due to $\mathcal{P}(x)$, directly sampling from Eq. (24) is intractable. The decoupling between data fidelity and prior allows us to utilize a pretrained diffusion model [63, 33] as a plug-and-play prior. The forward process of a diffusion model perturbs the data distribution to a noise distribution over continuous time $t \in [0, T]$, characterized by the conditional distribution $q(x_t|x_0)$ as:

$$x_t|x_0 \sim \mathcal{N}(x_t; \alpha_t x_0, \sigma_t^2 I_n), \tag{25}$$

where $x_0 \sim p_{\text{data}}$ denotes the clean data, and α_t and σ_t dictate the noise schedule. This forward process can be rewritten as a stochastic differential equation (SDE) [64]:

$$dx_t = f(t)x_t dt + g(t)dw_t, \tag{26}$$

where $f(t)x_t$ and $g(t)$ are the drift and diffusion coefficients, and w_t is the standard Wiener process. To match the marginal distribution of Eq. (25) and Eq. (26), the drift and diffusion terms should satisfy [49]:

$$f(t) = \frac{d\log\alpha_t}{dt}, \quad g^2(t) = \frac{d\sigma_t^2}{dt} - 2f(t)\sigma_t^2. \tag{27}$$

To simulate the reverse generative process, a neural network $s_\theta(x_t, t)$ is pretrained via score matching [64] to approximate the intractable score function $s_\theta(x_t, t) \approx \nabla_{x_t} \log p_t(x_t)$ where $x_t \sim p_t$. The reverse process is then governed by SDE [64] as:

$$dx_t = \left(f(t)x_t - g^2(t)\nabla_{x_t}\log p_t(x_t)\right)dt + g(t)d\overline{w}_t, \tag{28}$$

where $\overline{w}_t$ is the inverse-time standard Wiener process. Detailed derivations of the diffusion model are provided in Section C.1 in Appendix.

We establish that the prior sampling step in Eq. (24) mathematically aligns with the diffusion posterior $p(x_0|x_t)$, when the diffusion time t matches the condition $\varrho = \frac{\sigma_t}{\alpha_t}$ and $z^{(k)} = x_t/\alpha_t$ (the corresponding derivation is deferred to Section C.2 in Appendix). Consequently, we can obtain an approximate sample $x^{(k)}$ by initializing the reverse SDE in Eq. (28) at $x_t = \alpha_t z^{(k)}$ and integrating it back to $t = 0$. By alternating the likelihood sampling in Eq. (23) and the diffusion-based prior sampling in Eq. (24) over K iterations, we ultimately obtain the reconstructed signal $x^{(K)}$, which approximates a sample from the one-bit posterior distribution. Algorithm 1 presents our algorithm in a summarized manner. The complete sampling algorithm is summarized in Section C.3 in Appendix.

4.2 Quantitative results

Our experiments are conducted on the FFHQ [40] and ImageNet [62] datasets at a spatial resolution of 256×256. Following standard practices in one-bit compressed sensing [51, 65, 16], the elements of the measurement matrix A are sampled from $\mathcal{N}(0, 1/m)$. We compare the proposed PnP-OneBit against several models, including DiffPIR [77], DPS [18], DAPS [73], QCS-SGM [51], SIM-DMIS [65], and Diff-OneBit [16]. For a fair comparison, all evaluated algorithms utilize the same pretrained diffusion models and operate on identical measurements y. For each dataset, we randomly select 100 test images. We evaluate each method five times per image and report the mean $\pm$ standard deviation for peak signal-to-noise ratio (PSNR), structural similarity index measure

Table 1: Quantitative results for one-bit recovery on the FFHQ dataset under different compression ratios n/m with additive Gaussian noise $\sigma = 0.5$. Best results are shown in **bold**.

Methods	$n/m = 16$			$n/m = 32$		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DiffPIR	11.71±1.26	0.30±.05	0.73±.06	11.48±0.63	0.24±.04	0.76±.05
DPS	14.88±1.21	0.52±.06	0.50±.07	13.81±0.78	0.50±.06	0.56±.05
DAPS	14.84±1.61	0.43±.06	0.59±.06	12.02±0.82	0.31±.06	0.61±.06
QCS-SGM	18.11±2.43	0.52±.10	0.52±.07	17.12±1.84	0.49±.09	0.59±.06
SIM-DMIS	18.73±1.94	0.52±.08	0.50±.07	17.36±1.34	0.48±.08	0.55±.08
Diff-OneBit	21.95±1.23	0.53±.08	0.47±.06	18.68±0.98	0.51±.06	0.53±.06
PnP-OneBit	**22.96**±1.86	**0.69**±.09	**0.33**±.07	**19.64**±1.23	**0.58**±.07	**0.49**±.07

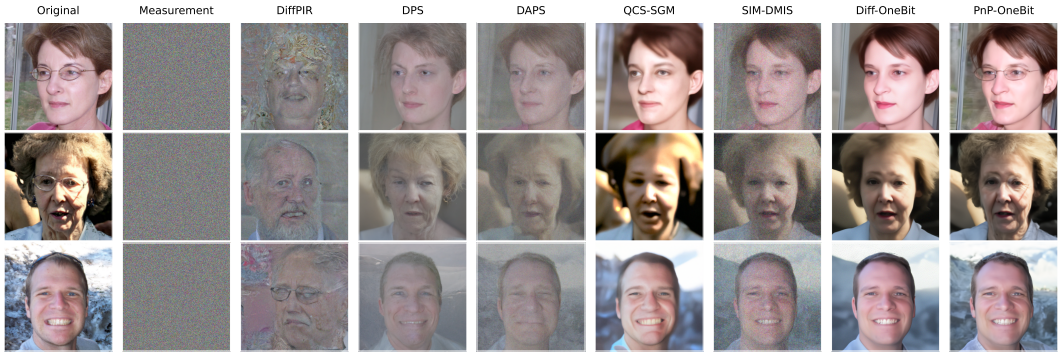


Figure 1: Qualitative results on FFHQ dataset with $\sigma = 0.5$ and $n/m = 16$.

(SSIM), and learned perceptual image patch similarity (LPIPS) [74]. Further implementation details are provided in Section D.1 in Appendix.

Table 1 summarizes the quantitative results on the FFHQ dataset under different compression ratios n/m. Additional evaluations on the ImageNet dataset, experiments under varying noise levels, and comprehensive further studies are deferred to Section D in Appendix. Across both datasets, the proposed PnP-OneBit consistently achieves the great reconstruction performance on all evaluated metrics. Notably, PnP-OneBit maintains strong performance even at severe compression ratios ($n/m = 32$).

Furthermore, visual comparisons of the reconstructed images are provided in Figure 1. While methods seeking a single point estimate, such as Diff-OneBit, may oversmooth high-frequency information under the information bottleneck of one-bit quantization, PnP-OneBit mitigates this issue. By balancing the data likelihood and the diffusion prior via posterior sampling, it effectively retains high-frequency textures.

5 Conclusion

We studied noisy one-bit compressed sensing for signals drawn from arbitrary data prior distributions on the sphere. We introduced an approximate covering number on the sphere and established a posterior sampling upper bound related to the logarithm of the approximate covering number and the recovery accuracy under noisy one-bit measurements, a robustness guarantee under learned prior mismatch, and an information theoretic lower bound for any reliable estimator. To complement these theoretical findings, we presented a one-bit plug-and-play reconstruction method and demonstrated its effectiveness on high dimensional image recovery tasks. Overall, our results show that reliable recovery under severe nonlinear quantization can be precisely characterized by the high probability geometric complexity of the signal prior. We focus on the noisy one-bit model with additive Gaussian noise before quantization. An open problem is whether similar robust posterior sampling bounds can be established when the model noise is adversarial.

Acknowledgments and Disclosure of Funding

Funding. The authors received no external funding or third-party support for this work during the 36 months preceding submission.

Competing Interests. The authors declare no financial relationships outside the submitted work during this period that could reasonably be perceived as influencing this research.

References

[1] Ben Adcock and Nick Huang. How many measurements are enough? Bayesian recovery in inverse problems with general distributions. In *NeurIPS*, 2025.

[2] Shuchin Aeron, Venkatesh Saligrama, and Manqi Zhao. Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56(10):5111–5130, 2010.

[3] Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-Gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014.

[4] Ismail Alkhouri, Shijun Liang, Cheng-Han Huang, Jimmy Dai, Qing Qu, Saiprasad Ravishankar, and Rongrong Wang. SITCOM: Step-wise triple-consistent diffusion sampling for inverse problems. In *ICML*, 2025.

[5] Richard G Baraniuk and Michael B Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009.

[6] Vidmantas Bentkus. On Hoeffding's inequalities. *Annals of Probability*, 2004.

[7] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *ICML*, 2017.

[8] Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *CISS*, 2008.

[9] Charles A. Bouman and Gregery T. Buzzard. Generative plug and play: Posterior sampling for inverse problems. In *Allerton*, 2023.

[10] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[11] Jinyuan Chang, Chenguang Duan, Yuling Jiao, Ruoxuan Li, Jerry Zhijian Yang, and Cheng Yuan. Provable diffusion posterior sampling for Bayesian inversion. *arXiv preprint arXiv:2512.08022*, 2025.

[12] Junren Chen and Zhaoqiang Liu. Efficient algorithms for non-Gaussian single index models with generative priors. In *AAAI*, 2024.

[13] Junren Chen, Jonathan Scarlett, Michael Ng, and Zhaoqiang Liu. A unified framework for uniform signal recovery in nonlinear generative compressed sensing. In *NeurIPS*, 2023.

[14] Junren Chen, Michael K. Ng, and Zhaoqiang Liu. Solving quadratic systems with full-rank matrices using sparse or generative priors. *IEEE Transactions on Signal Processing*, 73:477–492, 2025.

[15] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *ICLR*, 2023.

[16] Youming Chen and Zhaoqiang Liu. Diffusion model based signal recovery under 1-bit quantization. In *AAAI*, 2026.

[17] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. In *ICCV*, 2021.

[18] Hyungjin Chung, Jeongsol Kim, Michael T McCann, Marc L Klasky, and Jong Chul Ye. diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023.

[19] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *ICLR*, 2024.

[20] Florentin Coeurdoux, Nicolas Dobigeon, and Pierre Chainais. Plug-and-play split Gibbs sampler: Embedding deep generative priors in Bayesian inference. *IEEE Transactions on Image Processing*, 33:3496–3507, 2024.

[21] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[22] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *NeurIPS*, 2021.

[23] David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4): 1289–1306, 2006.

[24] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

[25] Armin Eftekhari and Michael B Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39 (1):67–109, 2015.

[26] Yonina C. Eldar and Moshe Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory*, 55(11):5302–5316, 2009.

[27] Xuefeng Gao, Hoang M Nguyen, and Lingjiong Zhu. Wasserstein convergence guarantees for a general class of score-based generative models. *Journal of Machine Learning Research*, 26(43): 1–54, 2025.

[28] Yihang Gao, Michael K. Ng, and Mingjie Zhou. Approximating probability distributions by using Wasserstein generative adversarial networks. *SIAM Journal on Mathematics of Data Science*, 5(4):949–976, 2023.

[29] D. Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946, 1995.

[30] Mingda Han, Huanqi Yang, Tao Ni, Di Duan, Mengzhe Ruan, Yongliang Chen, Jia Zhang, and Weitao Xu. mmSign: mmWave-based few-shot online handwritten signature verification. *ACM Transactions on Sensor Networks*, 20(4):1–31, 2024.

[31] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *NeurIPS*, 2018.

[32] Chinmay Hegde, Michael Wakin, and Richard Baraniuk. Random projections for manifold learning. In *NeurIPS*, 2007.

[33] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

[34] Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

[35] Ajil Jalal, Sushrut Karmalkar, Alexandros G Dimakis, and Eric Price. Instance-optimal compressed sensing via posterior sampling. In *ICML*, 2021.

[36] Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[37] Yuchen Jiao, Na Li, Changxiao Cai, Yuxin Chen, and Gen Li. Provable diffusion-based posterior sampling for linear inverse problems via DDIM. *arXiv preprint arXiv:2607.19333*, 2026.

[38] Geethu Joseph, Swatantra Kafle, and Pramod K Varshney. One-bit compressed sensing using generative models. In *ICASSP*, 2020.

[39] Swatantra Kafle, Thakshila Wimalajeewa, and Pramod K. Varshney. Noisy one-bit compressed sensing with side-information. *IEEE Transactions on Signal Processing*, 72:3792–3804, 2024.

[40] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

[41] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, 2022.

[42] Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Transactions on Information Theory*, 62(5):2748–2758, 2016.

[43] Jason N. Laska, Zaiwen Wen, Wotao Yin, and Richard G. Baraniuk. Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements. *IEEE Transactions on Signal Processing*, 59(11):5289–5301, 2011.

[44] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *ALT*, 2023.

[45] Jiulong Liu and Zhaoqiang Liu. Non-iterative recovery from nonlinear observations using generative models. In *CVPR*, 2022.

[46] Zhaoqiang Liu and Jonathan Scarlett. Information-theoretic lower bounds for compressive sensing with generative models. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 292–303, 2020.

[47] Zhaoqiang Liu, Selwyn Gomes, Avtansh Tiwari, and Jonathan Scarlett. Sample complexity bounds for 1-bit compressive sensing and binary stable embeddings with generative priors. In *ICML*, 2020.

[48] Zhaoqiang Liu, Tongyao Pang, Ruibing Wang, and Yang Zheng. A posterior-dynamics framework for imaging inverse problems with pretrained diffusion priors. *arXiv preprint arXiv:2608.15144*, 2026.

[49] Cheng Lu, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *NeurIPS*, 2022.

[50] Yulong Lu and Jianfeng Lu. A universal approximation theorem of deep neural networks for expressing probability distributions. In *NeurIPS*, 2020.

[51] Xiangming Meng and Yoshiyuki Kabashima. Quantized compressed sensing with score-based generative models. In *ICLR*, 2023.

[52] Deanna Needell and Joel A Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[53] Deanna Needell and Roman Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3): 317–334, 2009.

[54] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[55] Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1): 482–494, 2013.

[56] Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013.

[57] Shuang Qiu, Xiaohan Wei, and Zhuoran Yang. Robust one-bit recovery via ReLU generative networks: Near-optimal statistical rate and global landscape analysis. In *ICML*, 2020.

[58] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[59] Galen Reeves and Michael Gastpar. The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing. *IEEE Transactions on Information Theory*, 58(5):3065–3092, 2012.

[60] Mengzhe Ruan, Yunhe Li, Weizhou Zhang, Linqi Song, and Weitao Xu. Optimal power control for over-the-air federated learning with gradient compression. In *ICPADS*, 2024.

[61] Mengzhe Ruan, Guangfeng Yan, Yuanzhang Xiao, Linqi Song, and Weitao Xu. Adaptive top-k in SGD for communication-efficient distributed learning in multi-robot collaboration. *IEEE Journal of Selected Topics in Signal Processing*, 18(3):487–501, 2024.

[62] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[63] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

[64] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021.

[65] Anqi Tang, Youming Chen, Shuchen Xue, and Zhaoqiang Liu. Learning single index models with diffusion priors. In *ICML*, 2025.

[66] Tim van Erven and Peter Harremos. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[67] Maxime Vono, Nicolas Dobigeon, and Pierre Chainais. Split-and-augmented Gibbs sampler—application to large-scale inference problems. *IEEE Transactions on Signal Processing*, 67(6):1648–1661, 2019.

[68] Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. DMPlug: A plug-in method for solving inverse problems with diffusion models. In *NeurIPS*, 2024.

[69] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR*, 2023.

[70] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *NeurIPS*, 2024.

[71] Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *NeurIPS*, 2024.

[72] Guoshen Yu and Guillermo Sapiro. Statistical compressed sensing of Gaussian mixture models. *IEEE Transactions on Signal Processing*, 59(12):5842–5858, 2011.

[73] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *CVPR*, 2025.

[74] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

[75] Yang Zheng, Wen Li, and Zhaoqiang Liu. Integrating intermediate layer optimization and projected gradient descent for solving inverse problems with diffusion models. In *ICML*, 2025.

[76] Yang Zheng, Wen Li, and Zhaoqiang Liu. Image restoration via diffusion models with dynamic resolution. In *ICML*, 2026.

[77] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *CVPRW*, 2023.

A Upper bound proofs

Proof roadmap. The proof consists of four steps. First, using the rotational invariance of Gaussian measurements, we show that the expected Hamming distance between one-bit observations is determined only by the geodesic distance between the underlying signals. Second, based on this relation, we establish a one-bit near–far test to quantify the separability between nearby and faraway signals in the one-bit observation space, and prove that its testing error decays exponentially with the number of measurements m. Third, we construct an approximate geodesic cover for the high probability mass of the signal distribution, and apply the above test together with a union bound over all covering centers, thereby obtaining the posterior sampling recovery guarantee in the matched-prior setting. Fourth, to handle model mismatch between the true distribution and the learned distribution, we use a geodesic Wasserstein truncation argument to decompose the two distributions into high probability regular components and small residual components, the regular components are close in $W_{\infty,\mathrm{geo}}$ and can be controlled by a discrete distribution whose support size is at most e^k. Finally, combining this with a Rényi stability lemma transfers the matched-prior recovery guarantee to the mismatched prior setting, while the residual components contribute only an additional failure probability of 2δ.

A.1 Proof of Lemma 3.2

Define the pre-quantization continuous latent variables for the i-th measurement as $z_{1,i} = a_i^T x_1 + \xi_i$ and $z_{2,i} = a_i^T x_2$. According to Sheppard's formula [54], for jointly zero-mean Gaussian variables $z_{1,i}, z_{2,i}$, defining ρ as the correlation coefficient between them, we have:

$$\Pr\left(\mathrm{sign}(z_{1,i}) \neq \mathrm{sign}(z_{2,i})\right) = \frac{1}{\pi}\arccos(\rho). \tag{29}$$

Next, we consider calculating the correlation coefficient ρ between $z_{1,i}, z_{2,i}$.

First, we calculate the variances of $z_{1,i}, z_{2,i}$. Since both $a_i^T x_1$ and ξ_i can be regarded as zero-mean Gaussian distributions and are mutually independent, by the superposition property of Gaussian distributions, we obtain:

$$\begin{aligned}
\mathrm{Var}(z_{1,i}) &= \mathrm{Var}(a_i^T x_1) + \mathrm{Var}(\xi_i), && (30)\\
&= \|x_1\|^2 + \sigma^2, && (31)\\
&= 1 + \sigma^2. && (32)
\end{aligned}$$

Similarly, since the second observation is noiseless, we obtain:

$$\mathrm{Var}(z_{2,i}) = \|x_2\|_2^2 = 1. \tag{33}$$

Next, we compute the covariance between the pre-quantization variables $z_{1,i}, z_{2,i}$. Since the noise ξ_i is independent of a_i and has mean zero:

$$\begin{aligned}
\mathrm{Cov}(z_{1,i}, z_{2,i}) &= \mathbb{E}[(a_i^T x_1 + \xi_i)(a_i^T x_2)], && (34)\\
&= \mathbb{E}[x_1^T a_i a_i^T x_2] + \mathbb{E}[\xi_i a_i^T x_2], && (35)\\
&= \mathbb{E}[x_1^T a_i a_i^T x_2], && (36)\\
&= x_1^T x_2. && (37)
\end{aligned}$$

Utilizing the variance Eq. (30) and the covariance Eq. (34), we get:

$$\rho = \frac{x_1^T x_2}{\sqrt{1+\sigma^2}}. \tag{38}$$

Therefore, the normalized Hamming distance between the observation vectors u, v is ultimately expressed as:

$$\mathbb{E}[\mathrm{d_H}(u,v)] = \frac{1}{m}\sum_{i=1}^{m} \Pr\left(\mathrm{sign}(z_{1,i}) \neq \mathrm{sign}(z_{2,i})\right), \tag{39}$$

$$= \frac{1}{\pi}\arccos\left(\frac{\langle x_1, x_2\rangle}{\sqrt{1+\sigma^2}}\right), \tag{40}$$

$$= \frac{1}{\pi}\arccos\left(\frac{\cos(\pi \mathrm{d_S}(x_1,x_2))}{\sqrt{1+\sigma^2}}\right). \tag{41}$$

A.2 Proof of Lemma 3.3

Fix an arbitrary $x \in \mathbb{S}^{n-1}$. When the true signal is x, write

$$u = \mathrm{sign}(Ax + \xi). \tag{42}$$

For each $i = 1, \ldots, m$, define

$$U_i := \mathbf{1}\left\{\mathrm{sign}(a_i^\top x + \xi_i) \neq \mathrm{sign}(a_i^\top x_0)\right\}. \tag{43}$$

Then

$$\mathrm{d_H}\left(u, \mathrm{sign}(Ax_0)\right) = \frac{1}{m}\sum_{i=1}^{m} U_i. \tag{44}$$

Under the joint randomness of (A, ξ) conditional on x, (a_i, ξ_i) are independent and identically distributed. Therefore, $U_1, \ldots, U_m$ are independent and identically distributed Bernoulli random variables.

By Lemma 3.2, we have

$$\mathbb{E}_{a_i,\xi_i|x}[U_i] = f_\sigma(\mathrm{d_S}(x, x_0)). \tag{45}$$

Note that $f_\sigma(\cdot)$ is nondecreasing on $[0,1]$.

We first consider the near region. If $\mathrm{d_S}(x, x_0) \leq \eta$, then

$$\mathbb{E}_{a_i,\xi_i|x}[U_i] \leq f_\sigma(\eta). \tag{46}$$

Thus, by Hoeffding's inequality [6],

$$\Pr_{A,\xi|x}\left(\phi_{x_0,\eta}(u;A) = 1\right) = \Pr_{A,\xi|x}\left(\mathrm{d_H}\left(u, \mathrm{sign}(Ax_0)\right) \geq \tau_{\eta,c,\sigma}\right), \tag{47}$$

$$\leq \exp\left(-2m\left(\tau_{\eta,c,\sigma} - f_\sigma(\eta)\right)^2\right), \tag{48}$$

$$= \exp\left(-\frac{m}{2}\left(f_\sigma(c\eta) - f_\sigma(\eta)\right)^2\right), \tag{49}$$

$$= \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{50}$$

Taking the supremum over all x satisfying $\mathrm{d_S}(x, x_0) \leq \eta$ yields the first claim.

We next consider the far region. If $\mathrm{d_S}(x, x_0) \geq c\eta$, then

$$\mathbb{E}_{a_i,\xi_i|x}[U_i] \geq f_\sigma(c\eta). \tag{51}$$

By Hoeffding's inequality [6],

$$\Pr_{A,\xi|x}\left(\phi_{x_0,\eta}(u;A) = 0\right) = \Pr_{A,\xi|x}\left(\mathrm{d_H}\left(u, \mathrm{sign}(Ax_0)\right) \leq \tau_{\eta,c,\sigma}\right), \tag{52}$$

$$\leq \exp\left(-2m\left(f_\sigma(c\eta) - \tau_{\eta,c,\sigma}\right)^2\right), \tag{53}$$

$$= \exp\left(-\frac{m}{2}\left(f_\sigma(c\eta) - f_\sigma(\eta)\right)^2\right), \tag{54}$$

$$= \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{55}$$

Taking the supremum over all x satisfying $\mathrm{d_S}(x, x_0) \geq c\eta$ yields the second claim.

A.3 Proof of Lemma 3.4

The proof proceeds in three steps. In the first step, we convert a possibly overlapping covering by spherical caps into a pairwise disjoint measurable partition. In the second step, for each partition block S_j, we reduce the error event that the posterior sample falls into the far region to a binary posterior testing problem between the near region and the far region. In the third step, we sum over all j, and use a counting argument to improve the final coefficient from the crude $N+1$ to N.

Through $x_1, \ldots, x_N \in \mathbb{S}^{n-1}$, we define the far region

$$\mathcal{F}_j(c) := \left\{ x \in \mathbb{S}^{n-1} : \mathrm{d_S}(x, x_j) \geq c\eta \right\}. \tag{56}$$

Since these spherical caps $\mathcal{C}_j$ for $j = 1, \ldots, N$ cover the whole mass of Q_g, namely

$$Q_g \left[\bigcup_{j=1}^{N} \mathcal{C}_j \right] = 1. \tag{57}$$

Denote the covered region by

$$G := \bigcup_{j=1}^{N} \mathcal{C}_j. \tag{58}$$

Since the spherical caps $\mathcal{C}_1, \ldots, \mathcal{C}_N$ may overlap, we define the following disjoint partition. Let

$$S_1 := \mathcal{C}_1, \tag{59}$$

and for $j = 2, \ldots, N$, let

$$S_j := \mathcal{C}_j \setminus \bigcup_{i=1}^{j-1} \mathcal{C}_i. \tag{60}$$

By construction, $S_1, \ldots, S_N$ are pairwise disjoint, and satisfy

$$S_j \subseteq \mathcal{C}_j, \qquad G = \bigsqcup_{j=1}^{N} S_j. \tag{61}$$

Fix an arbitrary $j \in \{1, \ldots, N\}$. If $x^* \in S_j$, then since $S_j \subseteq \mathcal{C}_j$, we have

$$\mathrm{d_S}(x^*, x_j) \leq \eta. \tag{62}$$

On the other hand, if the recovery error satisfies $\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta$, then by the triangle inequality we obtain

$$\mathrm{d_S}(x_j, \widehat{x}) \geq \mathrm{d_S}(x^*, \widehat{x}) - \mathrm{d_S}(x^*, x_j) \geq (c+1)\eta - \eta = c\eta. \tag{63}$$

Therefore, we can conclude that when $\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta$, we have $\widehat{x} \in \mathcal{F}_j(c)$. Thus we obtain the event inclusion

$$\left\{ x^* \in S_j,\ \mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta \right\} \subseteq \left\{ x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c) \right\}. \tag{64}$$

Since $G = \bigsqcup_{j=1}^{N} S_j$, we have

$$\left\{ \mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta \right\} \subseteq \left\{ x^* \notin G \right\} \cup \bigcup_{j=1}^{N} \left\{ x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c) \right\}. \tag{65}$$

Using the covering assumption $Q_g(G) = 1$, we obtain

$$\Pr\left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta \right) \leq \Pr\left(x^* \notin G \right) + \sum_{j=1}^{N} \Pr\left(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c) \right), \tag{66}$$

$$\leq \sum_{j=1}^{N} \Pr\left(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c) \right). \tag{67}$$

It remains to bound each term $\Pr(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c))$.

Fix j. If $Q_g(S_j) = 0$ or $Q(\mathcal{F}_j(c)) = 0$, then

$$\Pr(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c)) = 0, \tag{68}$$

the claim is immediate. Hence in what follows we only consider the case $Q_g(S_j) > 0, Q(\mathcal{F}_j(c)) > 0$.

Define

$$\pi_j^{\mathrm{in}} := Q(S_j), \qquad \pi_j^{\mathrm{far}} := Q(\mathcal{F}_j(c)). \tag{69}$$

Since $S_j \subseteq \mathcal{C}_j$ and $c > 1$, every $x \in S_j$ satisfies $\mathrm{d_S}(x, x_j) \le \eta < c\eta$. Therefore

$$S_j \cap \mathcal{F}_j(c) = \varnothing. \tag{70}$$

Define the middle region

$$M_j(c) := \mathbb{S}^{n-1} \setminus (S_j \cup \mathcal{F}_j(c)), \tag{71}$$

and let

$$\pi_j^{\mathrm{mid}} := Q(M_j(c)). \tag{72}$$

Then

$$\pi_j^{\mathrm{in}} + \pi_j^{\mathrm{far}} + \pi_j^{\mathrm{mid}} = 1. \tag{73}$$

Let Q_j^{in}, $Q_{j,c}^{\mathrm{far}}$, and Q_j^{mid} denote the normalized restrictions of Q to S_j, $\mathcal{F}_j(c)$, and $M_j(c)$, respectively. That is, for any Borel set $B \subseteq \mathbb{S}^{n-1}$, define

$$Q_j^{\mathrm{in}}(B) := \frac{Q(B \cap S_j)}{\pi_j^{\mathrm{in}}}, \tag{74}$$

and

$$Q_{j,c}^{\mathrm{far}}(B) := \frac{Q(B \cap \mathcal{F}_j(c))}{\pi_j^{\mathrm{far}}}. \tag{75}$$

When $\pi_j^{\mathrm{mid}} > 0$, further define

$$Q_j^{\mathrm{mid}}(B) := \frac{Q(B \cap M_j(c))}{\pi_j^{\mathrm{mid}}}. \tag{76}$$

For fixed A, write

$$p_{A,x}(y) := \Pr(y|A, x). \tag{77}$$

Accordingly, define the marginal likelihoods induced by the near region, the far region, and the middle region:

$$h_j^{\mathrm{in}}(y; A) := \int p_{A,x}(y)\, dQ_j^{\mathrm{in}}(x), \tag{78}$$

$$h_j^{\mathrm{far}}(y; A) := \int p_{A,x}(y)\, dQ_{j,c}^{\mathrm{far}}(x), \tag{79}$$

and, when $\pi_j^{\mathrm{mid}} > 0$,

$$h_j^{\mathrm{mid}}(y; A) := \int p_{A,x}(y)\, dQ_j^{\mathrm{mid}}(x). \tag{80}$$

Let $Q_{g,j}^{\mathrm{in}}$ denote the normalized restriction of Q_g to S_j, and let $h_{g,j}^{\mathrm{in}}(y; A) := \int p_{A,x}(y)\, dQ_{g,j}^{\mathrm{in}}(x)$.

Given y and A, the posterior probability assigned to the far region $\mathcal{F}_j(c)$ is

$$Q(\mathcal{F}_j(c)|y, A) = \frac{\pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A)}{\pi_j^{\mathrm{in}} h_j^{\mathrm{in}}(y; A) + \pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A) + \pi_j^{\mathrm{mid}} h_j^{\mathrm{mid}}(y; A)}. \tag{81}$$

Since the middle term satisfies

$$\pi_j^{\mathrm{mid}} h_j^{\mathrm{mid}}(y; A) \ge 0, \tag{82}$$

Dropping this nonnegative term from the denominator can only increase the ratio. Therefore

$$Q(\mathcal{F}_j(c)|y, A) \le \frac{\pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A)}{\pi_j^{\mathrm{in}} h_j^{\mathrm{in}}(y; A) + \pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A)}. \tag{83}$$

Define the far-region posterior probability under the reduced two-group model by

$$\Gamma_j^{\mathrm{far}}(y, A) := \frac{\pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A)}{\pi_j^{\mathrm{in}} h_j^{\mathrm{in}}(y; A) + \pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A)}. \tag{84}$$

Thus

$$\Pr\left(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c)\right) = \mathbb{E}_A \sum_{y \in \{-1,+1\}^m} Q_g(S_j) h_{g,j}^{\mathrm{in}}(y; A) Q(\mathcal{F}_j(c)|y, A), \tag{85}$$

$$\leq Q_g(S_j) \mathbb{E}_A \sum_{y \in \{-1,+1\}^m} h_{g,j}^{\mathrm{in}}(y; A) \Gamma_j^{\mathrm{far}}(y, A). \tag{86}$$

Now let $\phi : \{-1, +1\}^m \times \mathbb{R}^{m \times n} \to [0, 1]$ be any measurable test function. Since $Q = (1 - \alpha)Q_g + \alpha Q_b$, we have

$$\pi_j^{\mathrm{in}} h_j^{\mathrm{in}}(y; A) = \int_{S_j} p_{A,x}(y)\, dQ(x) \geq (1 - \alpha) \int_{S_j} p_{A,x}(y)\, dQ_g(x) = (1 - \alpha) Q_g(S_j) h_{g,j}^{\mathrm{in}}(y; A). \tag{87}$$

For any fixed (y, A), we have

$$Q_g(S_j) h_{g,j}^{\mathrm{in}}(y; A) \Gamma_j^{\mathrm{far}}(y, A) \leq \frac{1}{1 - \alpha} \left(\pi_j^{\mathrm{in}} h_j^{\mathrm{in}}(y; A) \phi(y; A) + \pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A) \left(1 - \phi(y; A)\right) \right). \tag{88}$$

This is because on the region selected by ϕ, we use the trivial upper bound

$$\Gamma_j^{\mathrm{far}}(y, A) \leq 1; \tag{89}$$

whereas on the complementary region, we use

$$Q_g(S_j) h_{g,j}^{\mathrm{in}}(y; A) \Gamma_j^{\mathrm{far}}(y, A) \leq \frac{1}{1 - \alpha} \pi_j^{\mathrm{far}} h_j^{\mathrm{far}}(y; A). \tag{90}$$

Therefore

$$\Pr\left(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c)\right) \leq \frac{1}{1 - \alpha} \left(\pi_j^{\mathrm{in}} \alpha_j(\phi) + \pi_j^{\mathrm{far}} \beta_j(\phi) \right), \tag{91}$$

where

$$\alpha_j(\phi) := \mathbb{E}_A \sum_{y \in \{-1,+1\}^m} h_j^{\mathrm{in}}(y; A) \phi(y; A), \tag{92}$$

and

$$\beta_j(\phi) := \mathbb{E}_A \sum_{y \in \{-1,+1\}^m} h_j^{\mathrm{far}}(y; A) \left(1 - \phi(y; A)\right). \tag{93}$$

Next, choose the near-far Hamming test centered at x_j. Define

$$H_j(y; A) := \mathrm{d_H}\left(y, \mathrm{sign}(A x_j)\right), \tag{94}$$

and let

$$\phi_j(y; A) := \mathbf{1}\left\{ H_j(y; A) \geq \frac{f_\sigma(\eta) + f_\sigma(c\eta)}{2} \right\}. \tag{95}$$

Since Q_j^{in} is supported on $S_j \subseteq \mathcal{C}_j$, for any $x \in \mathrm{supp}(Q_j^{\mathrm{in}})$, we have

$$\mathrm{d_S}(x, x_j) \leq \eta. \tag{96}$$

By Lemma 3.3, we obtain

$$\alpha_j(\phi_j) \leq \exp\left(-\frac{m}{2} \Delta_\sigma(c, \eta)^2 \right). \tag{97}$$

On the other hand, $Q_{j,c}^{\mathrm{far}}$ is supported on $\mathcal{F}_j(c)$, so for any $x \in \mathrm{supp}(Q_{j,c}^{\mathrm{far}})$, we have

$$\mathrm{d_S}(x, x_j) \geq c\eta. \tag{98}$$

Again by Lemma 3.3, we obtain

$$\beta_j(\phi_j) \le \exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{99}$$

Therefore

$$\Pr\left(x^* \in S_j,\ \widehat{x} \in \mathcal{F}_j(c)\right) \le \frac{1}{1-\alpha}\left(\pi_j^{\mathrm{in}} + \pi_j^{\mathrm{far}}\right)\exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right) \tag{100}$$

$$= \frac{1}{1-\alpha}\left(Q(S_j) + Q(\mathcal{F}_j(c))\right)\exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{101}$$

Substituting the result of (100) into the event decomposition in (66) gives

$$\Pr_{x^*,A,\xi,\widehat{x}}\left(\mathrm{d_S}(x^*,\widehat{x}) \ge (c+1)\eta\right) \le \frac{1}{1-\alpha}\sum_{j=1}^{N}\left(Q(S_j) + Q(\mathcal{F}_j(c))\right)\exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{102}$$

Since $S_1,\dots,S_N$ form a disjoint partition of G, we have

$$\sum_{j=1}^{N} Q(S_j) = Q(G). \tag{103}$$

Thus

$$\Pr_{x^*,A,\xi,\widehat{x}}\left(\mathrm{d_S}(x^*,\widehat{x}) \ge (c+1)\eta\right) \le \frac{1}{1-\alpha}\left(Q(G) + \sum_{j=1}^{N} Q(\mathcal{F}_j(c))\right)\exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{104}$$

We next prove that $Q(G) + \sum_{j=1}^{N} Q(\mathcal{F}_j(c)) \le N$. Observe that

$$Q(G) + \sum_{j=1}^{N} Q(\mathcal{F}_j(c)) = \int_{\mathbb{S}^{n-1}}\left(\mathbf{1}_G(x) + \sum_{j=1}^{N}\mathbf{1}_{\mathcal{F}_j(c)}(x)\right)dQ(x). \tag{105}$$

For any $x \in G$, there exists at least one $k \in \{1,\dots,N\}$ such that $x \in \mathcal{C}_k$. Therefore

$$\mathrm{d_S}(x,x_k) \le \eta < c\eta, \tag{106}$$

where the last step follows from $c > 1$. Hence

$$x \notin \mathcal{F}_k(c). \tag{107}$$

Therefore, when $x \in G$, the point x belongs to at most $N-1$ far regions $\mathcal{F}_1(c),\dots,\mathcal{F}_N(c)$. Thus

$$\mathbf{1}_G(x) + \sum_{j=1}^{N}\mathbf{1}_{\mathcal{F}_j(c)}(x) \le 1 + (N-1) = N. \tag{108}$$

When $x \notin G$, it is clear that

$$\mathbf{1}_G(x) + \sum_{j=1}^{N}\mathbf{1}_{\mathcal{F}_j(c)}(x) = \sum_{j=1}^{N}\mathbf{1}_{\mathcal{F}_j(c)}(x) \le N. \tag{109}$$

Therefore, for any $x \in \mathbb{S}^{n-1}$, we have

$$\mathbf{1}_G(x) + \sum_{j=1}^{N}\mathbf{1}_{\mathcal{F}_j(c)}(x) \le N. \tag{110}$$

Integrating the above inequality with respect to Q, we obtain

$$Q(G) + \sum_{j=1}^{N} Q(\mathcal{F}_j(c)) \le N. \tag{111}$$

Therefore

$$\Pr_{x^*,A,\xi,\widehat{x}}\left(\mathrm{d_S}(x^*,\widehat{x}) \ge (c+1)\eta\right) \le \frac{N}{1-\alpha}\exp\left(-\frac{m}{2}\Delta_\sigma(c,\eta)^2\right). \tag{112}$$

The proof is complete.

Lemma A.1. *(Adopted from [35, Lemma A.1]) Let R, P be arbitrary distributions on $\mathbb{S}^{n-1}$. Let $p \geq 1$ and $\eta, \rho, \delta > 0$ be parameters. If $\mathcal{W}_{p,geo}(R, P) \leq \rho$ and $\min\{\log \mathrm{Cov}_{\eta,\delta}(P), \log \mathrm{Cov}_{\eta,\delta}(R)\} \leq k$, then there exist distributions R', R'', P', P'' and a finite discrete distribution Q with a support size $|supp(Q)| \leq e^k$, satisfying the following conditions:*

1. $\min\{\mathcal{W}_{\infty,geo}(P', Q), \mathcal{W}_{\infty,geo}(R', Q)\} \leq \eta$.
2. $\mathcal{W}_{\infty,geo}(R', P') \leq \frac{\rho}{\delta^{1/p}}$.
3. $P = (1-2\delta)P' + (2\delta)P''$ *and* $R = (1-2\delta)R' + (2\delta)R''$.

A.4 Proof of Lemma 3.5

The central idea of the proof is to relate the true signal x^* to the auxiliary signal z^* through an optimal coupling associated with $\mathcal{W}_{\infty,\mathrm{geo}}(R_0, S_0)$. Since x^* and z^* differ by at most ϵ in spherical geodesic distance, if the posterior-generated signal $\widehat{x}$ is at distance at least $d + \epsilon$ from the true signal x^*, then it is at distance at least d from the auxiliary signal z^*. Therefore, the error event under the true model can first be reduced to a tail event centered at the auxiliary signal z^*. We then use a Rényi/Hölder change of measure to convert the posterior sampling error under the true observation y into the corresponding posterior sampling error under the auxiliary observation u.

First, by the choice of the coupling Π, we have

$$\mathrm{d_S}(x^*, z^*) \leq \epsilon \qquad \Pi\text{-almost surely.} \tag{113}$$

Therefore, under this coupling, the spherical triangle inequality implies that if

$$\mathrm{d_S}(x^*, \widehat{x}) \geq d + \epsilon, \tag{114}$$

then

$$\mathrm{d_S}(z^*, \widehat{x}) \geq \mathrm{d_S}(x^*, \widehat{x}) - \mathrm{d_S}(x^*, z^*) \geq d. \tag{115}$$

Thus, we have the event inclusion

$$\{\mathrm{d_S}(x^*, \widehat{x}) \geq d + \epsilon\} \subseteq \{\mathrm{d_S}(z^*, \widehat{x}) \geq d\}. \tag{116}$$

Consequently,

$$\Pr\left(\mathrm{d_S}(x^*, \widehat{x}) \geq d + \epsilon\right) \leq \Pr\left(\mathrm{d_S}(z^*, \widehat{x}) \geq d\right). \tag{117}$$

Next, fix a realization $z \in \mathbb{S}^{n-1}$ of the auxiliary signal and define the tail set centered at z by

$$E_z := \left\{w \in \mathbb{S}^{n-1} : \mathrm{d_S}(z, w) \geq d\right\}. \tag{118}$$

Let $K_A(s, \cdot)$ denote the posterior kernel induced by the prior D and the one-bit likelihood, that is, for any measurable set $B \subseteq \mathbb{S}^{n-1}$,

$$K_A(s, B) := \Pr_{w \sim D}(w \in B \,|\, \mathrm{sign}(Aw + \xi) = s,\ A). \tag{119}$$

More explicitly, if $\mu_{w,A}$ denotes the distribution of the one-bit observation $\mathrm{sign}(Aw + \xi)$ conditional on w and A, then

$$K_A(s, B) = \frac{\int_B \mu_{w,A}(s)\, dD(w)}{\int_{\mathbb{S}^{n-1}} \mu_{v,A}(s)\, dD(v)}. \tag{120}$$

It follows that $\widehat{x}, \widehat{z}$ satisfy

$$\widehat{x} \sim K_A(y, \cdot), \qquad \widehat{z} \sim K_A(u, \cdot). \tag{121}$$

After conditioning on the random variables $x^* = x$ and $z^* = z$, the set E_z is fixed, while the remaining randomness comes from the measurement matrix A, the noise ξ, and the posterior sample $\widehat{x}$ induced by the observation y. Let $\mathsf{M}_x^{(m)}$ denote the distribution of the joint random variable $(A, y) = (A, \mathrm{sign}(Ax + \xi))$ when the true signal is fixed at x. Then

$$\Pr\left(\mathrm{d_S}(z, \widehat{x}) \geq d | x^* = x, z^* = z\right) = \int K_A(y, E_z)\, d\mathsf{M}_x^{(m)}(A, y). \tag{122}$$

This is because, conditional on (A, y), the posterior-generated signal satisfies $\widehat{x} \sim K_A(y, \cdot)$, and hence

$$\Pr\left(\widehat{x} \in E_z | A, y\right) = K_A(y, E_z). \tag{123}$$

Similarly, let $\mathsf{M}_z^{(m)}$ denote the distribution of the joint random variable $(A, u) = (A, \operatorname{sign}(Az + \xi'))$ when the auxiliary signal is fixed at z.

We now perform a change of measure for fixed $x, z \in \mathbb{S}^{n-1}$. Since the Gaussian noise assigns positive probability to every one-bit pattern, we have

$$\mathsf{M}_x^{(m)} \ll \mathsf{M}_z^{(m)}. \tag{124}$$

Denote the corresponding likelihood ratio by

$$L_{x,z}(A, s) := \frac{d\mathsf{M}_x^{(m)}}{d\mathsf{M}_z^{(m)}}(A, s), \qquad s \in \{-1, +1\}^m. \tag{125}$$

Thus, by Eq. (122), we obtain

$$\int K_A(y, E_z)\, d\mathsf{M}_x^{(m)}(A, y) = \int K_A(s, E_z) L_{x,z}(A, s)\, d\mathsf{M}_z^{(m)}(A, s). \tag{126}$$

Let

$$q := 1 + \lambda, \qquad \theta := \frac{q-1}{q} = \frac{\lambda}{1+\lambda}. \tag{127}$$

Applying Hölder's inequality to the preceding display gives

$$\int K_A(s, E_z)\, d\mathsf{M}_x^{(m)}(A, s) \le \left(\int L_{x,z}(A, s)^q\, d\mathsf{M}_z^{(m)}(A, s) \right)^{1/q} \left(\int K_A(s, E_z)^{q/(q-1)}\, d\mathsf{M}_z^{(m)}(A, s) \right)^{(q-1)/q}. \tag{128}$$

Since$0 \le K_A(s, E_z) \le 1$ and $q/(q-1) > 1$, we have

$$K_A(s, E_z)^{q/(q-1)} \le K_A(s, E_z). \tag{129}$$

Therefore,

$$\int K_A(s, E_z)\, d\mathsf{M}_x^{(m)}(A, s) \le \left(\int L_{x,z}(A, s)^q\, d\mathsf{M}_z^{(m)}(A, s) \right)^{1/q} p_z^{\theta}, \tag{130}$$

where

$$p_z := \int K_A(s, E_z)\, d\mathsf{M}_z^{(m)}(A, s). \tag{131}$$

Here, p_z denotes the probability that the posterior-generated signal $\widehat{z}$ falls into the tail set E_z when the auxiliary signal is fixed at z and the observation is $u = \operatorname{sign}(Az + \xi')$.

We next control the Rényi-type moment in Eq. (130). To this end, introduce the unquantized joint Gaussian observation laws. Let $\widetilde{\mathsf{M}}_x^{(m)}$ denote the distribution of the joint random variable $(A, V_x) = (A, Ax+\xi)$, and let $\widetilde{\mathsf{M}}_z^{(m)}$ denote the distribution of the joint random variable $(A, V_z) = (A, Az+\xi')$. Since the one-bit observation is obtained through the mapping

$$(A, v) \mapsto (A, \operatorname{sign}(v)), \tag{132}$$

By the data processing inequality for classical Rényi divergence of order q>1, applied to the measurable map $(A, v) \to (A, sign(v))$,we have

$$D_q\left(\mathsf{M}_x^{(m)} \| \mathsf{M}_z^{(m)}\right) \le D_q\left(\widetilde{\mathsf{M}}_x^{(m)} \| \widetilde{\mathsf{M}}_z^{(m)}\right). \tag{133}$$

Equivalently,

$$\int L_{x,z}(A, s)^q\, d\mathsf{M}_z^{(m)}(A, s) \le \int \left(\frac{d\widetilde{\mathsf{M}}_x^{(m)}}{d\widetilde{\mathsf{M}}_z^{(m)}}(A, v) \right)^q d\widetilde{\mathsf{M}}_z^{(m)}(A, v). \tag{134}$$

We now compute the right-hand side. Since A has the same marginal distribution under both laws, and conditional on A,

$$V_x | A \sim \mathcal{N}(Ax, \sigma^2 I_m), \qquad V_z | A \sim \mathcal{N}(Az, \sigma^2 I_m), \tag{135}$$

the Rényi moment formula for Gaussian distributions with the same covariance gives

$$\int \left(\frac{d\widetilde{\mathsf{M}}_x^{(m)}}{d\widetilde{\mathsf{M}}_z^{(m)}}(A,v) \right)^q d\widetilde{\mathsf{M}}_z^{(m)}(A,v) = \mathbb{E}_A \exp\left(\frac{q(q-1)}{2\sigma^2} \|A(x-z)\|_2^2 \right). \tag{136}$$

Let

$$h := x - z. \tag{137}$$

If $h \neq 0$, since the rows of A are independent standard Gaussian vectors, we have

$$\frac{\|Ah\|_2^2}{\|h\|_2^2} \sim \chi_m^2. \tag{138}$$

Hence, as long as $\frac{q(q-1)\|h\|_2^2}{\sigma^2} < 1$, we have

$$\mathbb{E}_A \exp\left(\frac{q(q-1)}{2\sigma^2} \|Ah\|_2^2 \right) = \left(1 - \frac{q(q-1)\|h\|_2^2}{\sigma^2} \right)^{-m/2}. \tag{139}$$

If $h = 0$, then the above exponential moment equals 1, and the formula can be understood by continuous extension.

We now control the Euclidean distance by the spherical mismatch radius. Since $x, z \in \mathbb{S}^{n-1}$, we have

$$\|x - z\|_2 = 2 \sin\left(\frac{\pi \mathrm{d_S}(x,z)}{2} \right). \tag{140}$$

Moreover, under the coupling Π, we have $\mathrm{d_S}(x,z) \leq \epsilon$, and since $0 < \epsilon \leq 1$,

$$\|x - z\|_2 \leq 2 \sin\left(\frac{\pi \epsilon}{2} \right) = r_\epsilon. \tag{141}$$

Combining this with $q = 1 + \lambda$ and the condition$\lambda(1+\lambda) r_\epsilon^2 < \sigma^2$,we obtain

$$\frac{q(q-1)\|x-z\|_2^2}{\sigma^2} \leq \frac{\lambda(1+\lambda) r_\epsilon^2}{\sigma^2} < 1. \tag{142}$$

Therefore,

$$\left(\int L_{x,z}(A,s)^q \, d\mathsf{M}_z^{(m)}(A,s) \right)^{1/q} \leq \left(1 - \frac{\lambda(1+\lambda) r_\epsilon^2}{\sigma^2} \right)^{-\frac{m}{2(1+\lambda)}}. \tag{143}$$

Substituting this into Eq. (130), we obtain, for any fixed signal pair x, z satisfying $\mathrm{d_S}(x,z) \leq \epsilon$,

$$\int K_A(s, E_z) \, d\mathsf{M}_x^{(m)}(A,s) \leq \left(1 - \frac{\lambda(1+\lambda) r_\epsilon^2}{\sigma^2} \right)^{-\frac{m}{2(1+\lambda)}} p_z^{\frac{\lambda}{1+\lambda}}. \tag{144}$$

Finally, we take expectation with respect to the coupling $(x^*, z^*) \sim \Pi$. By Eq. (117) and Eq. (144), we get

$$\Pr\left(\mathrm{d_S}(x^*, \widehat{x}) \geq d + \epsilon \right) \leq \left(1 - \frac{\lambda(1+\lambda) r_\epsilon^2}{\sigma^2} \right)^{-\frac{m}{2(1+\lambda)}} \mathbb{E}_{z^*} \left[p_{z^*}^{\frac{\lambda}{1+\lambda}} \right]. \tag{145}$$

Since

$$0 < \frac{\lambda}{1+\lambda} < 1, \tag{146}$$

the function $t \mapsto t^{\lambda/(1+\lambda)}$ is concave. Therefore, by Jensen's inequality,

$$\mathbb{E}_{z^*} \left[p_{z^*}^{\frac{\lambda}{1+\lambda}} \right] \leq \left(\mathbb{E}_{z^*}[p_{z^*}] \right)^{\frac{\lambda}{1+\lambda}}. \tag{147}$$

By the definition of p_z,

$$\mathbb{E}_{z^*}[p_{z^*}] = \Pr\left(\mathrm{d_S}(z^*, \widehat{z}) \geq d \right), \tag{148}$$

where

$$u = \mathrm{sign}(Az^* + \xi'), \qquad \widehat{z} \sim K_A(u, \cdot). \tag{149}$$

Hence,

$$\Pr\left(\mathrm{d_S}(x^*, \widehat{x}) \geq d + \epsilon \right) \leq \left(1 - \frac{\lambda(1+\lambda) r_\epsilon^2}{\sigma^2} \right)^{-\frac{m}{2(1+\lambda)}} \Pr\left(\mathrm{d_S}(z^*, \widehat{z}) \geq d \right)^{\frac{\lambda}{1+\lambda}}. \tag{150}$$

This completes the proof.

A.5 Proof of Theorem 3.1

Based on the above lemmas, we derive the following lemma, which is closely related to Theorem 3.1.

Lemma A.2. *Let R and P be two probability distributions on the unit sphere $\mathbb{S}^{n-1}$. Let $p \geq 1$, $0 < \eta < 1$, $0 < \delta < 1/4$, and assume that $\mathcal{W}_{p,\mathrm{geo}}(R, P) \leq \rho$.*

Define the effective mismatch level $\varepsilon_\delta := \rho/\delta^{1/p}$, and define the parameter $\bar{\eta} := \eta + \varepsilon_\delta$. We take $c \in \left(1, \frac{1-\varepsilon_\delta}{\bar{\eta}} - 1\right)$, assume further that $2\eta + 3\varepsilon_\delta < 1$, which guarantees that $\left(1, \frac{1-\varepsilon_\delta}{\bar{\eta}} - 1\right)$ is a nonempty interval. Given (y, A), $\widehat{x} \sim P(\cdot|y, A)$ is the posterior sampling estimator based on P. Furthermore, define $r_{\varepsilon_\delta} := 2\sin\left(\frac{\pi\varepsilon_\delta}{2}\right)$, for any $0 < \lambda \leq \min\left\{1, \frac{\sigma^2}{3r_{\varepsilon_\delta}^2}\right\}$ (if $\rho = 0$,we take the $\lambda = 1$), and any $\beta \in (0, 1)$, with probability at least $1 - \beta$ over the draw of A, we have

$$\begin{aligned}&\Pr_{x^* \sim R, \xi, \widehat{x}|A}\left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\bar{\eta} + \varepsilon_\delta\right) \\ &\quad \leq 2\delta + \frac{1}{\beta}\left(1 - \frac{\lambda(1+\lambda)r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right)\right).\end{aligned} \tag{151}$$

In particular, when $R = P$, we have $\rho = 0$. Taking $\lambda = 1$, we obtain

$$\Pr_{x^* \sim R, \xi, \widehat{x}|A}\left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\eta\right) \leq 2\delta + \frac{1}{\beta}\exp\left(\frac{1}{2}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \eta)^2\right)\right). \tag{152}$$

Proof. By assumption, $\mathcal{W}_{p,\mathrm{geo}}(R, P) \leq \rho$, and we also have

$$\min\{\log \mathrm{Cov}_{\eta,\delta}(P), \log \mathrm{Cov}_{\eta,\delta}(R)\} \leq \log \mathrm{Cov}_{\eta,\delta}(R). \tag{153}$$

By $Lemma\ A.1$, there exist probability distributions R', R'', P', P'', and a finite discrete distribution Q, such that

$$|\operatorname{supp}(Q)| \leq \mathrm{Cov}_{\eta,\delta}(R), \tag{154}$$

and

$$\mathcal{W}_{\infty,\mathrm{geo}}(R', P') \leq \frac{\rho}{\delta^{1/p}} = \varepsilon_\delta. \tag{155}$$

Moreover, R and P can be decomposed as

$$R = (1-2\delta)R' + 2\delta R'', \qquad P = (1-2\delta)P' + 2\delta P''. \tag{156}$$

In addition, the lemma also guarantees that

$$\min\{\mathcal{W}_{\infty,\mathrm{geo}}(P', Q), \mathcal{W}_{\infty,\mathrm{geo}}(R', Q)\} \leq \eta. \tag{157}$$

This step converts the global W_p mismatch into a W_∞ mismatch on the good parts. The residual components R'' and P'' may be arbitrary distributions, but their weights in the original distributions are at most 2δ.

We next prove that P' can be fully covered by at most $\mathrm{Cov}_{\eta,\delta}(R)$ spherical caps of radius $\bar{\eta}$.

By Eq. (157), there are two cases.

The first case is $\mathcal{W}_{\infty,\mathrm{geo}}(P', Q) \leq \eta$.

By the definition of the W_∞ distance, there exists a coupling $\gamma \in \Pi(P', Q)$ such that

$$\mathrm{d_S}(x, q) \leq \eta \qquad \gamma\text{-a.s.} \tag{158}$$

Since Q is a finite discrete distribution and the second marginal of γ is Q, we have

$$\gamma\left(\{(x, q) : q \in \operatorname{supp}(Q),\ \mathrm{d_S}(x, q) \leq \eta\}\right) = 1. \tag{159}$$

Therefore,

$$P'\left(\bigcup_{q \in \operatorname{supp}(Q)} \mathcal{B}(q, \eta)\right) = 1. \tag{160}$$

By the definition of the $(\eta, 0)$-covering number, it follows that

$$\mathrm{Cov}_{\eta,0}(P') \le |\operatorname{supp}(Q)| \le \mathrm{Cov}_{\eta,\delta}(R). \tag{161}$$

Moreover, since $\bar{\eta} \ge \eta$, we obtain

$$\mathrm{Cov}_{\bar{\eta},0}(P') \le \mathrm{Cov}_{\eta,0}(P') \le \mathrm{Cov}_{\eta,\delta}(R). \tag{162}$$

The second case is $\mathcal{W}_{\infty,\mathrm{geo}}(R', Q) \le \eta$.

Combining this with Eq. (155) and using the triangle inequality for the W_∞ distance, we have

$$\mathcal{W}_{\infty,\mathrm{geo}}(P', Q) \le \mathcal{W}_{\infty,\mathrm{geo}}(P', R') + \mathcal{W}_{\infty,\mathrm{geo}}(R', Q) \le \varepsilon_\delta + \eta = \bar{\eta}. \tag{163}$$

Thus, there exists a coupling $\tilde{\gamma} \in \Pi(P', Q)$ such that

$$\mathrm{d_S}(x, q) \le \bar{\eta} \qquad \tilde{\gamma}\text{-a.s.} \tag{164}$$

Similarly, we obtain

$$P'\left(\bigcup_{q \in \operatorname{supp}(Q)} \mathcal{B}(q, \bar{\eta})\right) = 1. \tag{165}$$

Hence,

$$\mathrm{Cov}_{\bar{\eta},0}(P') \le |\operatorname{supp}(Q)| \le \mathrm{Cov}_{\eta,\delta}(R). \tag{166}$$

Combining the two cases, we have

$$\mathrm{Cov}_{\bar{\eta},0}(P') \le \mathrm{Cov}_{\eta,\delta}(R). \tag{167}$$

Moreover, defining

$$G_P := \bigcup_{q \in \operatorname{supp}(Q)} \mathcal{B}(q, \bar{\eta}), \tag{168}$$

the previous argument shows that

$$P'(G_P) = 1. \tag{169}$$

Since $P = (1 - 2\delta)P' + 2\delta P''$, it follows that

$$P(G_P) \ge 1 - 2\delta. \tag{170}$$

Consequently,

$$\mathrm{Cov}_{\bar{\eta},2\delta}(P) \le \mathrm{Cov}_{\eta,\delta}(R). \tag{171}$$

Consider the following matched good-component experiment:

$$z^* \sim P', \qquad u = \operatorname{sign}(Az^* + \xi'), \tag{172}$$

where $\xi' \sim \mathcal{N}(0, \sigma^2 I_m)$, and ξ' is independent of all other random variables. Let $\widehat{z} \sim P(\cdot|u, A)$ be the posterior sampling estimator based on the full learned prior distribution$P$.

By Eq. (167), P' can be fully covered by at most $\mathrm{Cov}_{\eta,\delta}(R)$ spherical caps of radius $\bar{\eta}$. Moreover, by Eq. (156),

$$P = (1 - 2\delta)P' + 2\delta P''.$$

Therefore, applying $Lemma\ 3.4$ with full prior P, good component P', bad component P'', mixture parameter $\alpha = 2\delta$, covering radius $\bar{\eta}$, we get $N = |\operatorname{supp}(Q)| \le \mathrm{Cov}_{\eta,\delta}(R)$.Then we obtain

$$\begin{aligned}
\Pr_{z^* \sim P', A, \xi', \widehat{z}}(\mathrm{d_S}(z^*, \widehat{z}) \ge (c+1)\bar{\eta}) &\le \frac{1}{1 - 2\delta} \mathrm{Cov}_{\bar{\eta},2\delta}(P) \exp\Big(-\frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\Big), \\
&\le \frac{1}{1 - 2\delta} \exp\Big(\log \mathrm{Cov}_{\eta,\delta}(R) - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\Big), \\
&= \exp\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log \frac{1}{1 - 2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right).
\end{aligned} \tag{173}$$

By Eq. (155), we have

$$\mathcal{W}_{\infty,\mathrm{geo}}(R', P') \le \varepsilon_\delta. \tag{174}$$

Therefore, there exists a coupling $\Pi \in \Pi(R', P')$ such that, if $(x^*, z^*) \sim \Pi$, then almost surely

$$\mathrm{d_S}(x^*, z^*) \le \varepsilon_\delta. \tag{175}$$

Now consider two one-bit observation models:

$$y = \mathrm{sign}(Ax^* + \xi), \qquad u = \mathrm{sign}(Az^* + \xi'). \tag{176}$$

Here $x^* \sim R'$ and $z^* \sim P'$. The posterior samples are both taken with respect to the full learned prior distribution$P$, namely

$$\widehat{x} \sim P(\cdot|y, A), \qquad \widehat{z} \sim P(\cdot|u, A). \tag{177}$$

Applying Lemma 3.5 with

$$R_0 = R', \qquad S_0 = P', \qquad D = P, \qquad \epsilon = \varepsilon_\delta, \qquad d = (c+1)\bar{\eta},$$

we obtain

$$\begin{aligned}
&\Pr_{x^* \sim R', A, \xi, \widehat{x}} \left(\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right) \\
&\quad \le \left(1 - \frac{\lambda(1+\lambda) r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \Pr_{z^* \sim P', A, \xi', \widehat{z}} \left(\mathrm{d_S}(z^*, \widehat{z}) \ge (c+1)\bar{\eta}\right)^{\frac{\lambda}{1+\lambda}}.
\end{aligned} \tag{178}$$

Substituting Eq. (173) into Eq. (178), we get

$$\begin{aligned}
&\Pr_{x^* \sim R', A, \xi, \widehat{x}} \left(\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right) \\
&\quad \le \left(1 - \frac{\lambda(1+\lambda) r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\!\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right)\right).
\end{aligned} \tag{179}$$

For each realization of A, define

$$G(A) := \Pr_{x^* \sim R', \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right). \tag{180}$$

Then Eq. (179) gives

$$\mathbb{E}_A\left[G(A)\right] \le \left(1 - \frac{\lambda(1+\lambda) r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\!\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right)\right). \tag{181}$$

By Markov's inequality, for every $\beta \in (0, 1)$, we obtain that, with probability at least $1 - \beta$ over A,

$$\begin{aligned}
&\Pr_{x^* \sim R', \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right) \\
&\quad \le \frac{1}{\beta}\left(1 - \frac{\lambda(1+\lambda) r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\!\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right)\right).
\end{aligned} \tag{182}$$

By the decomposition Eq. (156), we have

$$R = (1 - 2\delta)R' + 2\delta R''. \tag{183}$$

Therefore, for every fixed realization of A, the probability that the true signal x^* comes from the residual component R'' is 2δ. For any event $\mathcal{E}$, we have

$$\Pr_{x^* \sim R, \xi, \widehat{x}|A}(\mathcal{E}) \le 2\delta + \Pr_{x^* \sim R', \xi, \widehat{x}|A}(\mathcal{E}). \tag{184}$$

Let

$$\mathcal{E} = \left\{\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right\}. \tag{185}$$

Combining this with Eq. (182), we obtain that, with probability at least $1 - \beta$ over A,

$$\begin{aligned}
&\Pr_{x^* \sim R, \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \ge (c+1)\bar{\eta} + \varepsilon_\delta\right) \\
&\quad \le 2\delta + \frac{1}{\beta}\left(1 - \frac{\lambda(1+\lambda) r_{\varepsilon_\delta}^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\!\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta_\sigma(c, \bar{\eta})^2\right)\right).
\end{aligned} \tag{186}$$

This completes the proof. □

Proof of Theorem 3.1. Let $\Delta := \Delta_\sigma(c, \bar{\eta}), r := r_{\varepsilon_\delta}$. By the definition of the one-bit separation gap, we have $\Delta_\sigma(c, \bar{\eta}) = f_\sigma(c\bar{\eta}) - f_\sigma(\bar{\eta})$. Since $c > 1$, and f_σ is monotonically increasing on the corresponding interval, it follows that $\Delta > 0$. Moreover, since the range of f_σ is contained in $[0, 1]$, we have

$$0 < \Delta \leq 1.$$

Set

$$\beta := \exp\left(-\frac{m}{48}\Delta^2\right). \tag{187}$$

By Lemma A.2, for any parameter $0 < \lambda \leq \min\left\{1, \frac{\sigma^2}{3r^2}\right\}$, with probability at least $1 - \beta$ over the draw of A, we have

$$\begin{aligned}
&\Pr_{x^* \sim R, \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\bar{\eta} + \varepsilon_\delta\right), \\
&\quad \leq 2\delta + \frac{1}{\beta}\left(1 - \frac{\lambda(1+\lambda)r^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log\frac{1}{1-2\delta} - \frac{m}{2}\Delta^2\right)\right), \\
&\quad \leq 2\delta + \frac{1}{\beta}\left(1 - \frac{\lambda(1+\lambda)r^2}{\sigma^2}\right)^{-\frac{m}{2(1+\lambda)}} \exp\left(\frac{\lambda}{1+\lambda}\left(\log \mathrm{Cov}_{\eta,\delta}(R) + \log 2 - \frac{m}{2}\Delta^2\right)\right),
\end{aligned} \tag{188}$$

where the last step uses $\delta < 1/4$, and hence $\log\frac{1}{1-2\delta} \leq \log 2$.

We now take $\lambda = 1$. First, we verify that this choice is admissible. By the small model-mismatch assumption in the corollary, we have

$$\frac{r^2}{\sigma^2} \leq \frac{1}{6}\Delta^2 \leq \frac{1}{6}. \tag{189}$$

Thus

$$1 \leq \min\left\{1, \frac{\sigma^2}{3r^2}\right\}. \tag{190}$$

Hence $\lambda = 1$ lies in the range allowed by the main theorem. Substituting $\lambda = 1$ into (188) yields

$$\begin{aligned}
&\Pr_{x^* \sim R, \xi, \widehat{x}|A} \left(\mathrm{d_S}(x^*, \widehat{x}) \geq (c+1)\bar{\eta} + \varepsilon_\delta\right) \\
&\quad \leq 2\delta + \frac{1}{\beta}\left(1 - \frac{2r^2}{\sigma^2}\right)^{-m/4} \exp\left(\frac{1}{2}(\log \mathrm{Cov}_{\eta,\delta}(R) + \log 2) - \frac{m}{4}\Delta^2\right).
\end{aligned} \tag{191}$$

Next, we control the multiplicative factor induced by the model mismatch. Let $u := \frac{2r^2}{\sigma^2}$. By the assumption,

$$0 \leq u \leq \frac{\Delta^2}{3} \leq \frac{1}{3}. \tag{192}$$

Using the elementary inequality

$$-\log(1-u) \leq \frac{u}{1-u} \leq \frac{3}{2}u, \qquad 0 \leq u \leq \frac{1}{3}, \tag{193}$$

we obtain

$$-\log\left(1 - \frac{2r^2}{\sigma^2}\right) \leq 3\frac{r^2}{\sigma^2}. \tag{194}$$

Therefore,

$$\left(1 - \frac{2r^2}{\sigma^2}\right)^{-m/4} = \exp\left(-\frac{m}{4}\log\left(1 - \frac{2r^2}{\sigma^2}\right)\right) \leq \exp\left(\frac{3m}{4}\frac{r^2}{\sigma^2}\right) \leq \exp\left(\frac{m}{8}\Delta^2\right), \tag{195}$$

where the last step uses $\frac{r^2}{\sigma^2} \leq \frac{1}{6}\Delta^2$.

Substituting (195) into (191), we get

$$\Pr_{x^*\sim R,\xi,\widehat{x}|A}\left(\mathrm{d_S}(x^*,\widehat{x})\geq (c+1)\bar{\eta}+\varepsilon_\delta\right) \tag{196}$$

$$\leq 2\delta+\frac{1}{\beta}\exp\left(\frac{1}{2}(\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2)-\frac{m}{4}\Delta^2+\frac{m}{8}\Delta^2\right) \tag{197}$$

$$=2\delta+\frac{1}{\beta}\exp\left(\frac{1}{2}(\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2)-\frac{m}{8}\Delta^2\right). \tag{198}$$

If, furthermore, $m\geq \frac{8(\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2)}{\Delta^2}$, then

$$\frac{1}{2}(\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2)-\frac{m}{8}\Delta^2\leq -\frac{m}{16}\Delta^2. \tag{199}$$

Together with the choice $\beta=\exp\left(-\frac{m}{48}\Delta^2\right)$, this gives

$$\frac{1}{\beta}\exp\left(\frac{1}{2}(\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2)-\frac{m}{8}\Delta^2\right)\leq \exp\left(\frac{m}{48}\Delta^2-\frac{m}{16}\Delta^2\right)=\exp\left(-\frac{m}{24}\Delta^2\right). \tag{200}$$

It follows from Eq. (196) that, with probability at least $1-\exp\left(-\frac{m}{48}\Delta_\sigma(c,\bar{\eta})^2\right)$ over the draw of A,

$$\Pr_{x^*\sim R,\xi,\widehat{x}|A}\left(\mathrm{d_S}(x^*,\widehat{x})\geq (c+1)\bar{\eta}+\varepsilon_\delta\right)\leq 2\delta+\exp\left(-\frac{m}{24}\Delta_\sigma(c,\bar{\eta})^2\right). \tag{201}$$

This proves the first probability upper bound.

We next prove the order-wise sample complexity statement. To ensure that the second term on the right-hand side and the exceptional probability over A are no larger than the prescribed $\zeta\in(0,1)$, it suffices that

$$\exp\left(-\frac{m}{24}\Delta_\sigma(c,\bar{\eta})^2\right)\leq\zeta. \tag{202}$$

Equivalently, it suffices that $m\geq 24\frac{\log(1/\zeta)}{\Delta_\sigma(c,\bar{\eta})^2}$. On the other hand, to use the exponential bound proved above, we also need $m\geq 8\frac{\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2}{\Delta_\sigma(c,\bar{\eta})^2}$. Therefore, it suffices that

$$m\geq C\frac{\log \mathrm{Cov}_{\eta,\delta}(R)+\log(1/\zeta)+\log 2}{\Delta_\sigma(c,\bar{\eta})^2} \tag{203}$$

where $C>0$ is an absolute constant. For example, one may take $C=48$. Hence, with probability at least $1-\zeta$ over the draw of A,

$$\Pr_{x^*\sim R,\xi,\widehat{x}|A}\left(\mathrm{d_S}(x^*,\widehat{x})\geq (c+1)\bar{\eta}+\varepsilon_\delta\right)\leq 2\delta+\zeta. \tag{204}$$

Finally, we analyze the order of $\Delta_\sigma(c,\bar{\eta})$ under different noise levels. From

$$f_\sigma(d)=\frac{1}{\pi}\arccos\left(\frac{\cos(\pi d)}{\sqrt{1+\sigma^2}}\right) \tag{205}$$

direct differentiation gives

$$f'_\sigma(d)=\frac{\sin(\pi d)}{\sqrt{\sigma^2+\sin^2(\pi d)}}. \tag{206}$$

Thus

$$\Delta_\sigma(c,\bar{\eta})=\int_{\bar{\eta}}^{c\bar{\eta}}\frac{\sin(\pi t)}{\sqrt{\sigma^2+\sin^2(\pi t)}}\,dt. \tag{207}$$

Since $c>1$ is fixed and $\bar{\eta}=\Theta(\eta)$, the integral is over an interval of length $\Theta(\eta)$, and $t=\Theta(\eta)$ throughout the interval. Therefore,

$$\sin(\pi t)\asymp t\asymp\eta.$$

If $\sigma\lesssim\eta$, then

$$\frac{\sin(\pi t)}{\sqrt{\sigma^2+\sin^2(\pi t)}}\asymp 1, \tag{208}$$

and hence

$$\Delta_\sigma(c,\bar{\eta}) \asymp \eta. \tag{209}$$

If $\sigma \gtrsim \eta$, then

$$\frac{\sin(\pi t)}{\sqrt{\sigma^2+\sin^2(\pi t)}} \asymp \frac{\eta}{\sigma}, \tag{210}$$

and hence

$$\Delta_\sigma(c,\bar{\eta}) \asymp \frac{\eta^2}{\sigma}. \tag{211}$$

Substituting these estimates into $m \geq C\frac{\log \mathrm{Cov}_{\eta,\delta}(R)+\log 2+\log(1/\zeta)}{\Delta_\sigma(c,\bar{\eta})^2}$ gives

$$m = \begin{cases} O\big((\log \mathrm{Cov}_{\eta,\delta}(R)+\log(1/\zeta))\,\eta^{-2}\big), & \sigma \lesssim \eta, \\ O\big((\log \mathrm{Cov}_{\eta,\delta}(R)+\log(1/\zeta))\,\sigma^2\eta^{-4}\big), & \sigma \gtrsim \eta. \end{cases} \tag{212}$$

This completes the proof. □

B Lower bound proofs

B.1 Proof of Lemma 3.7

We prove the two claims separately, according to whether the measurement matrix is random Gaussian or deterministic.

Case 1: When $A_{ij} \sim \mathcal{N}(0,1)$, we prove an upper bound on the conditional mutual information $I(y;x^*|A)$. By definition,

$$I(y;x^*|A) = \mathbb{E}_{A_0}\big[I(y;x^*|A=A_0)\big]. \tag{213}$$

Therefore, in the inner mutual information, the measurement matrix A is fixed; the expectation over A is only taken at the very end in the outer layer.

For a fixed realization of A, the observations $y_1,\dots,y_m$ are conditionally independent given x^*. Thus,

$$\begin{aligned} I(y;x^*|A) &= H(y|A) - H(y|x^*,A) && (214)\\ &\leq \sum_{i=1}^m H(y_i|A) - \sum_{i=1}^m H(y_i|x^*,A) && (215)\\ &= \sum_{i=1}^m I(y_i;x^*|A). && (216) \end{aligned}$$

For each i, conditioned on A and x^*, we have

$$\Pr\left(y_i=1|x^*,A\right) = \Phi\left(\frac{a_i^\top x^*}{\sigma}\right). \tag{217}$$

Denote

$$p_i(A,x^*) := \Phi\left(\frac{a_i^\top x^*}{\sigma}\right). \tag{218}$$

Since $y_i \in \{-1,+1\}$, for a fixed A, we have

$$I(y_i;x^*|A) = H(y_i|A) - H(y_i|x^*,A) \leq \log 2 - \mathbb{E}_{x^*}\big[h(p_i(A,x^*))|A\big], \tag{219}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function.

Using the basic inequality

$$\log 2 - h(p) \leq 4\log 2\left(p-\frac{1}{2}\right)^2, \qquad p\in[0,1], \tag{220}$$

we obtain

$$I(y_i; x^*|A) \leq 4 \log 2 \, \mathbb{E}_{x^*} \left[\left(\Phi \left(\frac{a_i^\top x^*}{\sigma} \right) - \frac{1}{2} \right)^2 \middle| A \right]. \tag{221}$$

Therefore, taking the outer expectation of the above inequality with respect to A and summing over $i = 1, \ldots, m$, we obtain

$$I(y; x^*|A) \leq 4 \log 2 \sum_{i=1}^{m} \mathbb{E}_{A,x^*} \left[\left(\Phi \left(\frac{a_i^\top x^*}{\sigma} \right) - \frac{1}{2} \right)^2 \right]. \tag{222}$$

Now, the expectation is the joint expectation over both A and x^*. Since A is independent of x^*, and $\|x^*\|_2 = 1$, for any fixed x^*, we have $a_i^\top x^* \sim \mathcal{N}(0, 1)$.Therefore, for $Z \sim \mathcal{N}(0, 1)$, we have

$$\mathbb{E}_{A,x^*} \left[\left(\Phi \left(\frac{a_i^\top x^*}{\sigma} \right) - \frac{1}{2} \right)^2 \right] = \mathbb{E}_Z \left[\left(\Phi \left(\frac{Z}{\sigma} \right) - \frac{1}{2} \right)^2 \right]. \tag{223}$$

By Claim 1, taking $c = 1/\sigma$, we obtain

$$\mathbb{E}_Z \left[\left(\Phi \left(\frac{Z}{\sigma} \right) - \frac{1}{2} \right)^2 \right] = \frac{1}{2\pi} \arcsin \left(\frac{1}{1 + \sigma^2} \right). \tag{224}$$

Combining the above results, we obtain

$$I(y; x^*|A) \leq 4 \log 2 \sum_{i=1}^{m} \frac{1}{2\pi} \arcsin \left(\frac{1}{1 + \sigma^2} \right). \tag{225}$$

Thus,

$$I(y; x^*|A) \leq \frac{2m \log 2}{\pi} \arcsin \left(\frac{1}{1 + \sigma^2} \right). \tag{226}$$

Case 2: When A is a deterministic matrix. Since, conditional on x^*, the observations $y_1, \ldots, y_m$ are independent, we have

$$\begin{aligned} I(y; x^*) &= H(y) - H(y|x^*) && (227) \\ &\leq \sum_{i=1}^{m} H(y_i) - \sum_{i=1}^{m} H(y_i|x^*) && (228) \\ &= \sum_{i=1}^{m} I(y_i; x^*). && (229) \end{aligned}$$

For each i, we have

$$\Pr(y_i = 1|x^*) = \Phi \left(\frac{a_i^\top x^*}{\sigma} \right). \tag{230}$$

Therefore, using the Eq. (222), we obtain

$$I(y_i; x^*) \leq 4 \log 2 \, \mathbb{E}_{x^*} \left[\left(\Phi \left(\frac{a_i^\top x^*}{\sigma} \right) - \frac{1}{2} \right)^2 \right]. \tag{231}$$

According to the Lagrange mean value theorem, we have

$$\begin{aligned} |\Phi(t) - \Phi(0)| &= |\Phi'(\xi) \cdot (t - 0)|, && (232) \\ &\leq \frac{|t|}{\sqrt{2\pi}}. && (233) \end{aligned}$$

Taking $t = \frac{a_i^\top x^*}{\sigma}$,we obtain

$$\left(\Phi \left(\frac{a_i^\top x^*}{\sigma} \right) - \frac{1}{2} \right)^2 \leq \frac{(a_i^\top x^*)^2}{2\pi\sigma^2}. \tag{234}$$

Therefore,

$$I(y_i; x^*) \leq \frac{2\log 2}{\pi\sigma^2}\mathbb{E}_{x^*}\left[(a_i^\top x^*)^2\right]. \tag{235}$$

Summing over $i = 1, \ldots, m$, we get

$$I(y; x^*) \leq \frac{2\log 2}{\pi\sigma^2}\sum_{i=1}^{m}\mathbb{E}_{x^*}\left[(a_i^\top x^*)^2\right], \tag{236}$$

$$\leq \frac{2\log 2}{\pi\sigma^2}\sum_{i=1}^{m}\|a_i\|_2^2, \tag{237}$$

$$\leq \frac{2m\log 2}{\pi\sigma^2}\|A\|_{2,\infty}^2. \tag{238}$$

Claim 1. *For a standard normal random variable $Z \sim \mathcal{N}(0,1)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, we have:*

$$\mathbb{E}_Z\left[\left(\Phi(cZ) - \frac{1}{2}\right)^2\right] = \frac{1}{2\pi}\arcsin\left(\frac{c^2}{1+c^2}\right). \tag{239}$$

Proof. We introduce two new standard normal random variables $U_1, U_2 \sim \mathcal{N}(0,1)$ that are mutually independent and independent of Z. Then, conditioned on Z:

$$\Phi(cZ) = \Pr\left(U_1 \leq cZ | Z\right) = \Pr\left(U_1 - cZ \leq 0 | Z\right), \tag{240}$$

$$\Phi(cZ)^2 = \Pr\left(U_1 \leq cZ, U_2 \leq cZ | Z\right) = \Pr\left(U_1 - cZ \leq 0, U_2 - cZ \leq 0 | Z\right). \tag{241}$$

Define the random variables: $W_1 = U_1 - cZ$, $W_2 = U_2 - cZ$. Since $U_1, U_2, Z \sim \mathcal{N}(0,1)$, W_1 and W_2 follow a joint zero-mean bivariate normal distribution. By the additivity of Gaussian distributions, we have

$$\mathrm{Var}(W_1) = \mathrm{Var}(U_1) + c^2\mathrm{Var}(Z) = 1 + c^2. \tag{242}$$

$$\mathrm{Cov}(W_1, W_2) = \mathbb{E}[(U_1 - cZ)(U_2 - cZ)] = c^2\mathbb{E}[Z^2] = c^2. \tag{243}$$

Therefore, the correlation coefficient between W_1 and W_2 is:

$$\rho = \frac{\mathrm{Cov}(W_1, W_2)}{\sqrt{\mathrm{Var}(W_1)\mathrm{Var}(W_2)}}, \tag{244}$$

$$= \frac{c^2}{1+c^2}. \tag{245}$$

For joint bivariate normal variables with mean 0 and correlation coefficient ρ, the probability that both are simultaneously less than or equal to 0 is strictly given by the orthant probability formula for the standard bivariate normal distribution [54]:

$$\Pr\left(W_1 \leq 0, W_2 \leq 0\right) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho). \tag{246}$$

Combining the above equations, we have:

$$\mathbb{E}_Z\left[\left(\Phi(cZ) - \frac{1}{2}\right)^2\right] = \mathbb{E}_Z[\Phi(cZ)^2] - \mathbb{E}_Z[\Phi(cZ)] + \frac{1}{4}, \tag{247}$$

$$= \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho) - \frac{1}{2} + \frac{1}{4}, \tag{248}$$

$$= \frac{1}{2\pi}\arcsin(\frac{c^2}{1+c^2}). \tag{249}$$

□

Lemma B.1 (Lemma 4.3 in [35])**.** *Consider the setting of Theorem 3.6. If the measurement matrix A is deterministic, we have:*

$$I(x^*; \widehat{x}) \leq I(y; x^*). \tag{250}$$

If A is a Gaussian matrix, then:

$$I(x^*; \widehat{x}) \leq I(y; x^*|A). \tag{251}$$

Lemma B.2 (Adapted from [35]). *Let Q be a distribution defined on a finite discrete set S containing N points. Assume there exists a sufficiently large integer N_2 such that for each u in the support, $Q(u) = j\alpha$, where $j \in \mathbb{N}$, and the quantization step is $\alpha := 1/N_2$. Let $(x, \widehat{x})$ be jointly distributed, where $x \sim Q$, and $\widehat{x}$ lies in an arbitrary countable set. If the recovery condition is satisfied:*

$$\Pr\left(d(x, \widehat{x}) \leq \epsilon\right) \geq 1 - \delta. \tag{252}$$

Then we have:

$$\tau(1-\delta) \log \mathrm{Cov}_{2\epsilon, \tau+\delta}(Q) \leq I(x; \widehat{x}) + 2. \tag{253}$$

B.2 Proof of Lemma 3.8

Let $E = 1_{\mathrm{d_S}(x,\widehat{x}) \leq \epsilon}$ denote the indicator variable.

Let $\epsilon = \eta$. Since the continuous unit sphere $\mathbb{S}^{n-1}$ is a compact manifold, we construct a finite spherical γ-net $\mathcal{N}_\gamma$ with resolution γ ($\gamma > 0$) on it. We map the continuous signal x and the recovered result $\widehat{x}$ to their nearest discrete points $\bar{x}$ and $\widehat{\bar{x}}$ on $\mathcal{N}_\gamma$. By bounding, we obtain the following discretization error:

$$\begin{aligned} d_S(\bar{x}, \widehat{\bar{x}}) &\leq d_S(\bar{x}, x) + d_S(x, \widehat{x}) + \mathrm{d_S}(\widehat{x}, \widehat{\bar{x}}), & (254)\\ &\leq \gamma + \epsilon + \gamma, & (255)\\ &= \epsilon + 2\gamma. & (256) \end{aligned}$$

Therefore, the high-probability recovery event is transformed into $\Pr\left(\mathrm{d_S}(\bar{x}, \widehat{\bar{x}}) \leq \epsilon + 2\gamma\right) \geq 1 - \delta$.

Since $\bar{x}$ and $\widehat{\bar{x}}$ are deterministic functions of x and $\widehat{x}$ respectively, according to Lemma B.1, the data processing does not increase the mutual information, so we have:

$$I(\bar{x}; \widehat{\bar{x}}) \leq I(x; \widehat{x}). \tag{257}$$

Let $\bar{x}$ follow the discrete distribution $\bar{R}$. Since $\mathcal{N}_\gamma$ is a finite set, the support of $\bar{R}$ is naturally bounded. To satisfy the probability quantization requirement of Lemma B.2, we introduce a minimum quantization step size α, and quantize the probability values in $\bar{R}$ downwards to integer multiples of α to construct a finite discrete distribution Q'. The total probability loss in this step does not exceed $\alpha|\mathcal{N}_\gamma|$. We choose α sufficiently small such that $\alpha|\mathcal{N}_\gamma| \leq \delta$.

Given the concavity of mutual information with respect to the marginal distribution under a fixed conditional distribution [21], the adjusted mutual information satisfies:

$$I(\bar{x}; \widehat{\bar{x}}) \geq (1-\delta) I_{Q'}(\bar{x}; \widehat{\bar{x}}). \tag{258}$$

Meanwhile, under the distribution Q', the lower bound of the successful recovery probability decreases to $1 - \delta - \delta = 1 - 2\delta$.

Now Q' is a finite distribution that strictly satisfies the probability quantization. Applying Lemma B.2 in the metric space, we have:

$$\tau(1-2\delta) \log \mathrm{Cov}_{2(\epsilon+2\gamma), \tau+2\delta}(Q') \leq I_{Q'}(\bar{x}; \widehat{\bar{x}}) + 2. \tag{259}$$

Substituting Eq. (258) into the above equation yields:

$$\tau(1-2\delta) \log \mathrm{Cov}_{2\epsilon+4\gamma, \tau+2\delta}(Q') \leq \frac{I(x; \widehat{x})}{1-\delta} + 2. \tag{260}$$

Furthermore, we need to generalize the approximate covering bound established on the discrete distribution Q' to the true continuous prior distribution R. Considering that Q' truncates at most a probability measure of δ relative to R, this means that any subset capable of covering a measure of $1 - (\tau + 2\delta)$ in Q' will necessarily cover a measure of at least $1 - (\tau + 3\delta)$ in R. At the same time, to accommodate the geometric error introduced by spatial discretization, we relax the bound of the covering radius by γ to $2\epsilon + 5\gamma$. At this point, if we choose the grid resolution $\gamma = \epsilon/5$, the new covering radius corresponds exactly to $3\epsilon = 3\eta$. Substituting the above conclusions and multiplying both sides of the inequality by $(1-\delta)$, we can prove:

$$(1-\delta)\tau(1-2\delta) \log \mathrm{Cov}_{3\eta, \tau+3\delta}(R) \leq I(x; \widehat{x}) + 2(1-\delta). \tag{261}$$

B.3 Proof of Theorem 3.6

Let $N(R, \gamma)$ denote the minimal set of 3η-geodesic balls covering at least $1-\gamma$ probability mass of R, such that its cardinality is $\mathrm{Cov}_{3\eta,\gamma}(R)$. Let B be the ball with the smallest marginal probability in $N(R, 10\delta)$. Define the subset $S = N(R, 10\delta) \setminus B$, which strictly covers less than $1-10\delta$ of the probability mass. We decompose the distribution R into a mixture distribution $R = (1-c)R' + cR''$, where R' and R'' are the conditional distributions restricted to S and its complement S^c, respectively. By the construction of S, we have $c > 10\delta$. Since R'' contributes at least 10δ of the mass to R, any algorithm that succeeds with probability $\geq 1-\delta$ over the entire distribution R must succeed with probability ≥ 0.9 on R''.

We now apply the variant of the spherical Fano inequality (Lemma 3.8) to the distribution R''. Substituting the free parameter $\tau = 0.1$ and the failure probability $\delta_{R''} = 0.1$ into Lemma 3.8 yields:

$$(1-0.1)(0.1)(1-0.2)\log \mathrm{Cov}_{3\eta,0.1+3(0.1)}(R'') \leq I(x^*;\widehat{x}) + 2(1-0.1). \tag{262}$$

Simplifying this inequality yields:

$$I(x^*;\widehat{x}) \geq 0.072 \log \mathrm{Cov}_{3\eta,0.4}(R'') - 1.8. \tag{263}$$

Next, we strictly relate the covering number of R'' back to the original distribution R, without loose relaxation. Since R'' contains at least 10δ of the mass of R, the set $N(R'', 0.4)$ which covers $\frac{3}{5}$ of the mass in R'', essentially covers at least 6δ of the mass of R. Meanwhile, $N(R, 10\delta)$ covers $1-10\delta$ of the mass of R. The union of these two sets must cover at least $1-10\delta+6\delta = 1-4\delta$ of the mass of R. Therefore, we establish the cardinality bound:

$$|N(R'', 0.4)| + |N(R, 10\delta)| \geq |N(R, 4\delta)| \implies |N(R'', 0.4)| \geq |N(R, 4\delta)| - |N(R, 10\delta)|. \tag{264}$$

Furthermore, we use a counting argument to strictly relate $|N(R, 4\delta)|$ and $|N(R, 10\delta)|$. Assuming the balls in $N(R, 4\delta)$ are sorted in descending order of their marginal probabilities, the bottom proportion of $\frac{10\delta-4\delta}{1-4\delta} = \frac{6\delta}{1-4\delta}$ balls can contain at most $10\delta - 4\delta = 6\delta$ of the mass (relative to the total of $1-4\delta$). Alternatively, observing that the top proportion of $\frac{1-10\delta}{1-4\delta}$ balls in $N(R, 4\delta)$ must contain at least $1-10\delta$ of the mass, we strictly obtain:

$$\frac{1-10\delta}{1-4\delta}|N(R, 4\delta)| \geq |N(R, 10\delta)|. \tag{265}$$

Substituting Eq. (265) into Eq. (264), we obtain:

$$\begin{aligned} |N(R'', 0.4)| &\geq |N(R, 4\delta)| - \frac{1-10\delta}{1-4\delta}|N(R, 4\delta)|, && (266)\\ &= \left(\frac{1-4\delta-(1-10\delta)}{1-4\delta}\right)|N(R, 4\delta)|, && (267)\\ &\geq 6\delta |N(R, 4\delta)|. && (268) \end{aligned}$$

Taking the logarithm of both sides and using the definition $|N(R, \gamma)| = \mathrm{Cov}_{3\eta,\gamma}(R)$, we obtain the tight geometric relation:

$$\log \mathrm{Cov}_{3\eta,0.4}(R'') \geq \log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta). \tag{269}$$

Substituting the tight relation Eq. (269) into Eq. (263) yields:

$$I(x^*;\widehat{x}) \geq 0.072\,(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta)) - 1.8. \tag{270}$$

Now, we establish the upper bound on mutual information. According to the Markov chain $x^* \to y \to \widehat{x}$ and the data processing inequality (Lemma B.1), for a Gaussian matrix A, we have $I(x^*;\widehat{x}) \leq I(y; x^*|A)$ (for a deterministic matrix A, we have $I(x^*;\widehat{x}) \leq I(y; x^*)$). Utilizing the bound on the one-bit quantized channel capacity from Lemma 3.7, we obtain:

Case 1: When $A_{ij} \sim \mathcal{N}(0,1)$, we have

$$\frac{2m\log 2}{\pi} \arcsin\left(\frac{1}{1+\sigma^2}\right) \geq 0.072\,(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta)) - 1.8. \tag{271}$$

Dividing both sides by $\frac{2\log 2}{\pi}\arcsin\left(\frac{1}{1+\sigma^2}\right)$, we conclude:

$$m \geq \frac{0.072\pi}{2(\log 2)\arcsin\left(\frac{1}{1+\sigma^2}\right)}\left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta) - 25\right). \tag{272}$$

Case 2: When A is a deterministic matrix, we have:

$$\frac{2m\log 2}{\pi\sigma^2}\|A\|_{2,\infty}^2 \geq 0.072\left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta)\right) - 1.8. \tag{273}$$

Dividing both sides by $\frac{2\log 2}{\pi\sigma^2}\|A\|_{2,\infty}^2$, we conclude:

$$m \geq \frac{0.072\pi\sigma^2}{2(\log 2)\|A\|_{2,\infty}^2}\left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log(6\delta) - 25\right). \tag{274}$$

This completes the proof.

C Detailed PnP-OneBit

C.1 Preliminary for diffusion

By employing a continuous-time stochastic process, diffusion models construct a generative paradigm that systematically degrades a complex data distribution into a manageable prior. Suppose an uncorrupted data point $x_0 \in \mathbb{R}^n$ is sampled from the target distribution p_{data}. The forward corruption mechanism over a continuous timeline $t \in [0, T]$ is defined by the following stochastic differential equation (SDE):

$$dx_t = f(t)x_t dt + g(t)dw_t, \tag{275}$$

where $x_t \in \mathbb{R}^n$ is the state at time t, $w_t \in \mathbb{R}^n$ represents the standard Wiener process, $f(t)$ serves as the drift factor, and $g(t)$ denotes the diffusion coefficient. For any given t, the transition probability from the original sample x_0 to the perturbed state x_t is characterized in [33] as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \alpha_t x_0, \sigma_t^2 I_n), \tag{276}$$

where the terms α_t and σ_t denote differentiable schedule functions for scale and noise, with α_t monotonically decreasing and σ_t monotonically increasing as t progresses. To guarantee that the SDE presented in Eq. (275) accurately produces the marginal distribution $q(x_t|x_0)$, the drift and diffusion terms are tied to the schedule functions as follows [49]:

$$f(t) = \frac{d\log\alpha_t}{dt}, \quad g^2(t) = \frac{d\sigma_t^2}{dt} - 2\sigma_t^2\frac{d\log\alpha_t}{dt}. \tag{277}$$

At the final time T, the boundary condition dictates that the distribution closely mirrors a standard normal distribution, where $x_T \sim \mathcal{N}(\mathbf{0}, I_n)$.

To reverse this diffusion process and reconstruct the target data, a parameterized neural network $\epsilon_\theta(x_t, t)$ is optimized to estimate the additive noise $\epsilon \sim \mathcal{N}(0, I_n)$. The training procedure minimizes the subsequent loss function:

$$\mathbb{E}_{t\sim\mathcal{U}(0,T), x_0\sim p_{\text{data}}, \epsilon\sim\mathcal{N}(0,I_d)}\left[\|\epsilon_\theta(\alpha_t x_0 + \sigma_t\epsilon, t) - \epsilon\|_2^2\right]. \tag{278}$$

According to Song et al. [64], the generative process running backward from $t = T$ to $t = 0$ corresponds to the following reverse-time SDE:

$$dx_t = \left(f(t)x_t - g^2(t)\nabla_{x_t}\log p_t(x_t)\right)dt + g(t)d\overline{w}_t, \tag{279}$$

where $\overline{w}_t$ indicates a backward-in-time Wiener process, and p_t represents the marginal density of x_t. The required score function $\nabla_{x_t}\log p_t(x_t)$ is empirically estimated using the trained noise predictor via $\nabla_{x_t}\log p_t(x_t) \approx -\epsilon_\theta(x_t, t)/\sigma_t$. Additionally, an equivalent probability flow ordinary differential equation (ODE) exists, which strictly maintains identical marginal distributions to the SDE [64]:

$$dx_t = \left(f(t)x_t - \frac{1}{2}g^2(t)\nabla_{x_t}\log p_t(x_t)\right)dt. \tag{280}$$

For practical implementation, Eq. (280) is evaluated via numerical integration across a predefined discrete schedule $\{t_i\}_{i=0}^N$, constrained by $t_i > t_{i+1}$, starting at $t_0 = T$ and ending at $t_N = 0$. At a specific step t_i, the expected clean data can be approximated using Tweedie's formula [24] as follows:

$$\hat{x}_{0|t_i} = \frac{x_{t_i} + \sigma_{t_i}^2 \nabla_{x_{t_i}} \log p_{t_i}(x_{t_i})}{\alpha_{t_i}} \approx \frac{x_{t_i} - \sigma_{t_i} \epsilon_\theta(x_{t_i}, t_i)}{\alpha_{t_i}}. \tag{281}$$

Finally, the update step from t_i to t_{i+1} operates through the DDIM [63] integration scheme with sampling noise level η_{DDIM}:

$$x_{t_{i+1}} = \alpha_{t_{i+1}} \hat{x}_{0|t_i} + \sqrt{1 - \eta_{\text{DDIM}}} \sigma_{t_{i+1}} \epsilon_\theta(x_{t_i}, t_i) + \eta_{\text{DDIM}} \xi_i,\ \xi_i \sim \mathcal{N}(0, I_n). \tag{282}$$

C.2 Align between likelihood and prior

Recall from Section 4 that the prior sampling step requires drawing $x^{(k)}$ from the following conditional distribution:

$$x^{(k)} \sim \Pi(x|z^{(k)}) \propto \exp\left(-\mathcal{P}(x) - \frac{1}{2\varrho^2}\|x - z^{(k)}\|_2^2\right), \tag{283}$$

where $\mathcal{P}(x) = -\log p(x)$.

According to the diffusion formulation detailed in Section C.1, the posterior distribution of the clean data x_0 given the noised data x_t can be expressed via Bayes' theorem as:

$$\begin{aligned} p(x_0|x_t) &\propto p(x_t|x_0)p(x_0) \\ &\propto \exp\left(\log p(x_0) + \log p(x_t|x_0)\right) \\ &\propto \exp\left(-\mathcal{P}(x_0) - \frac{1}{2\sigma_t^2}\|x_t - \alpha_t x_0\|_2^2\right) \\ &\propto \exp\left(-\mathcal{P}(x_0) - \frac{\alpha_t^2}{2\sigma_t^2}\|x_0 - x_t/\alpha_t\|_2^2\right). \end{aligned} \tag{284}$$

By comparing Eq. (284) with Eq. (283), it becomes evident that the two distributions are mathematically equivalent under the substitutions $x = x_0$, $\varrho = \sigma_t/\alpha_t$, and $z^{(k)} = x_t/\alpha_t$. Consequently, drawing a sample from Eq. (283) is structurally identical to sampling x_0 from $p(x_0|x_t)$. While the exact posterior $p(x_0|x_t)$ is intractable in practice, a pretrained diffusion model provides an effective approximation of this reverse transition. Therefore, we can reliably employ the diffusion model to perform the sampling step for Eq. (283) by setting the initial state as $x_t = \alpha_t z^{(k)}$ and matching the noise schedule such that $\varrho = \sigma_t/\alpha_t$.

C.3 Detailed algorithm

In this section, we first introduce the formulation of the measurement likelihood $\Pr(y|x, A)$ and the gradient $\nabla_x \mathcal{L}(x; y)$, followed by the detailed algorithmic workflow of our PnP-OneBit.

Given the one-bit observation model $A^T = [a_1^T, \ldots, a_m^T]$, the probability of acquiring a single measurement y_i can be derived as:

$$\Pr(y_i|x, A) = \Pr\left(y_i(a_i^T x + \sigma\zeta_i) \geq 0\right) = \Pr\left(\zeta_i \leq \frac{y_i a_i^T x}{\sigma}\right) = \Phi\left(\frac{y_i a_i^T x}{\sigma}\right), \tag{285}$$

where $\zeta_i \sim \mathcal{N}(0, 1)$ is the standard Gaussian noise, and $\Phi(\cdot)$ represents the cumulative distribution function of the standard Gaussian distribution. Assuming conditional independence among the measurements, the joint likelihood is given by:

$$\Pr(y|x, A) = \prod_{i=1}^m \Phi\left(\frac{y_i a_i^T x}{\sigma}\right). \tag{286}$$

Consequently, the negative log-likelihood $\mathcal{L}(x; y)$ can be rewritten as:

$$\mathcal{L}(x; y) = -\log \Pr(y|x, A) = -\sum_{i=1}^m \log \Phi\left(\frac{y_i a_i^T x}{\sigma}\right). \tag{287}$$

Algorithm 2 PnP-OneBit

Require: Pretrained noise predictor model ϵ_θ, iteration schedule $\{\varrho_k\}_{k=0}^{K-1}$, diffusion schedule $\{t_i\}_{t=0}^{T}$, learning rate κ, MCMC step J, DDIM noise level η, measurement y, forward matrix A

Sample $x^{(0)} \sim \mathcal{N}(0, I_n)$
for $k = 0, 1, \ldots, K-1$ **do**
  $z_0 \leftarrow x^{(k)}$
  for $j = 0, 1, \ldots, J$ **do**
    $g_j \leftarrow \nabla_{z_j} \mathcal{L}(z_j, y)$
    Sample $\xi_j \sim \mathcal{N}(0, I_n)$
    $z_{j+1} \leftarrow z_j - \kappa g_j - \frac{\kappa}{\varrho_k^2}(z_j - x^{(k)}) + \sqrt{2\kappa}\xi_j$
  end for
  $z^{(k)} \leftarrow z_J$
  Solve $\varrho_k = \sigma_{t^\star}/\alpha_{t^\star}$ to get $t^\star$
  $l \leftarrow \min\{i | t^\star > t_i\}$ ▷ seek l satisfying $t_{l-1} \geq t^\star > t_l$
  $x_{t^\star} \leftarrow \alpha_{t_\star} z^{(k)}$
  $\hat{\epsilon}^\star = \epsilon_\theta(x^\star, t^\star)$
  $\hat{x}_{0|t^\star} = \frac{x_{t^\star} - \sigma_{t^\star}\,\epsilon_\theta(x_{t^\star}, t^\star)}{\alpha_{t^\star}}$
  Sample $\xi^\star \sim \mathcal{N}(0, I_n)$
  $x_{t_l} = \alpha_{t_l}\hat{x}_{0|t^\star} + \sqrt{1 - \eta^2_{\text{DDIM}}}\,\sigma_{t_l}\,\hat{\epsilon}^\star + \eta_{\text{DDIM}}\,\sigma_{t_l}\xi^\star$ ▷ DDIM sample to t_l from $t^\star$
  for $i = l, l+1, \ldots, T-1$ **do**
    $\hat{\epsilon}_i = \epsilon_\theta(x_{t_i}, t_i)$
    $\hat{x}_{0|t_i} = \frac{x_{t_i} - \sigma_{t_i}\,\epsilon_\theta(x_{t_i}, t_i)}{\alpha_{t_i}}$
    Sample $\xi_i \sim \mathcal{N}(0, I_n)$
    $x_{t_{i+1}} = \alpha_{t_{i+1}}\hat{x}_{0|t_i} + \sqrt{1 - \eta^2_{\text{DDIM}}}\,\sigma_{t_{i+1}}\,\hat{\epsilon}_i + \eta_{\text{DDIM}}\,\sigma_{t_{i+1}}\,\xi_i$
    ▷ DDIM sample to t_{i+1} from t_i
  end for
  $x^{(k+1)} \leftarrow x_T$
end for
return $x^{(K)}$

Taking the derivative with respect to the signal x yields the analytical gradient:

$$\nabla_x \mathcal{L}(x; y) = -A^T \left(\frac{y}{\sigma} \odot \frac{\phi(r)}{\Phi(r)} \right), \tag{288}$$

where $r = \frac{y \odot (Ax)}{\sigma}$ and the operator $\odot$ denotes the element-wise Hadamard product. By utilizing the explicit gradient in Eq. (288), we can apply Markov chain Monte Carlo (MCMC) methods in PnP-DM [70] to sample from the following likelihood-based conditional distribution:

$$z^{(k)} \sim \Pi(z | x^{(k-1)}) \propto \exp\left(-\mathcal{L}(z; y) - \frac{1}{2\varrho^2} \| z - x^{(k-1)} \|_2^2 \right). \tag{289}$$

The complete procedure of our proposed approach is summarized in Algorithm 2.

D Experimental results

D.1 Implementation details

Our experiments are conducted on the FFHQ [40] and ImageNet [62] datasets. FFHQ is licensed under Creative Commons BY-NC-SA 4.0, and ImageNet is used under its standard research license. We employ the pre-trained diffusion models provided by Dhariwal & Nichol [22] and Choi et al. [17], both of which are released under the MIT License. We maintain their original noise schedules α_t and σ_t. Following the settings in PnP-DM, the coupling parameter ϱ_k is updated via $\varrho_k = \max\{\nu^k \varrho_0, \varrho_{\min}\}$, where $\varrho_0 = 10$, $\nu = 0.9$, and $\varrho_{\min} = 0.3$. The hyperparameters for the iterative process are set as follows: MCMC steps $J = 100$, learning rate $\kappa = 0.01$, diffusion timesteps $T = 100$, DDIM noise level $\eta = 0.5$, and outer iterations $K = 100$. For the baseline methods, including DiffPIR [77],

DPS [18], DAPS [73], QCS-SGM [51], SIM-DMIS [65], and Diff-OneBit [16], we use the default parameter settings provided in their respective publications. All experiments were conducted on a single Nvidia RTX 4090 (24GB).

D.2 Quantitative results

We evaluate the robustness of the algorithms on the FFHQ dataset under a higher noise level of $\sigma = 1.0$ at compression ratios $n/m = 16$ and $n/m = 32$. As detailed in Table 2, the proposed algorithm consistently maintains the highest performance under increased noise corruption. Furthermore, Table 3 presents the evaluation results on the ImageNet dataset. These results confirm that PnP-OneBit effectively recovers one-bit signals not only for domain-specific distributions such as human faces but also for diverse, complex natural scenes. In addition, compared to Diff-OneBit [16], our method exhibits a slightly larger standard deviation. This numerical variance originates from the fundamental difference in the problem formulation. Diff-OneBit employs deterministic optimization techniques to find a single point estimate, which mathematically suppresses variance but can lead to overly smoothed textures. In contrast, PnP-OneBit functions strictly as a posterior sampler. The stochastic transitions used to explore the target distribution inherently introduce algorithmic randomness. However, this distributional modeling approach effectively preserves fine high-frequency details, which contributes to the overall improvement in reconstruction quality.

Table 2: Quantitative results of one bit recovery on FFHQ dataset with different m and the additive Gaussian noise of $\sigma = 1.0$. The best results are **bold**.

Methods	$n/m = 16$			$n/m = 32$		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DiffPIR	11.50±1.13	0.24±.05	0.78±.06	11.20±0.94	0.18±.05	0.80±.06
DPS	14.53±1.23	0.51±.06	0.52±.07	13.68±0.78	0.49±.06	0.57±.05
DAPS	14.50±1.48	0.40±.05	0.61±.07	13.57±0.98	0.48±.06	0.56±.07
QCS-SGM	18.43±2.47	0.52±.10	0.54±.07	17.00±1.76	0.49±.09	0.61±.06
SIM-DMIS	18.32±1.79	0.51±.08	0.53±.06	16.89±1.17	0.47±.08	0.59±.08
Diff-OneBit	21.15±1.17	0.40±.06	0.49±.06	18.32±0.92	0.48±.06	0.54±.06
PnP-OneBit	**22.22**±1.73	**0.58**±.09	**0.35**±.07	**19.37**±1.18	**0.57**±.07	**0.49**±.07

Table 3: Quantitative results of one bit recovery on ImageNet dataset with different m and the additive Gaussian noise of $\sigma = 0.5$. The best results are **bold**.

Methods	$n/m = 16$			$n/m = 32$		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DiffPIR	10.73±0.69	0.09±.03	0.77±.07	9.80±0.54	0.06±.02	0.80±.07
DPS	14.38±1.36	0.38±.05	0.69±.08	13.14±0.76	0.35±.04	0.76±.08
DAPS	14.17±1.51	0.24±.04	0.67±.07	12.27±0.77	0.33±.05	0.78±.08
QCS-SGM	16.28±1.92	0.39±.08	0.66±.08	14.13±1.64	0.37±.08	0.72±.09
SIM-DMIS	17.03±1.23	0.41±.06	0.60±.07	15.42±0.98	0.40±.06	0.68±.08
Diff-OneBit	18.27±0.87	0.43±.05	0.53±.05	16.46±0.61	0.42±.05	0.67±.07
PnP-OneBit	**19.77**±1.32	**0.53**±.07	**0.47**±.06	**18.03**±0.94	**0.46**±.05	**0.58**±.06

Furthermore, we provide additional visual comparisons to illustrate the reconstruction performance. Figure 2 displays more reconstruction examples on the FFHQ dataset under the compression ratio $n/m = 16$ and noise level $\sigma = 0.5$. Compared to the evaluated baseline methods, PnP-OneBit recovers facial details with fewer visual artifacts. In Addition, Figure 3 compares PnP-OneBit with Diff-OneBit on the ImageNet dataset across varying noise levels ($\sigma = 0.5$ and $\sigma = 1.0$) at $n/m = 16$.

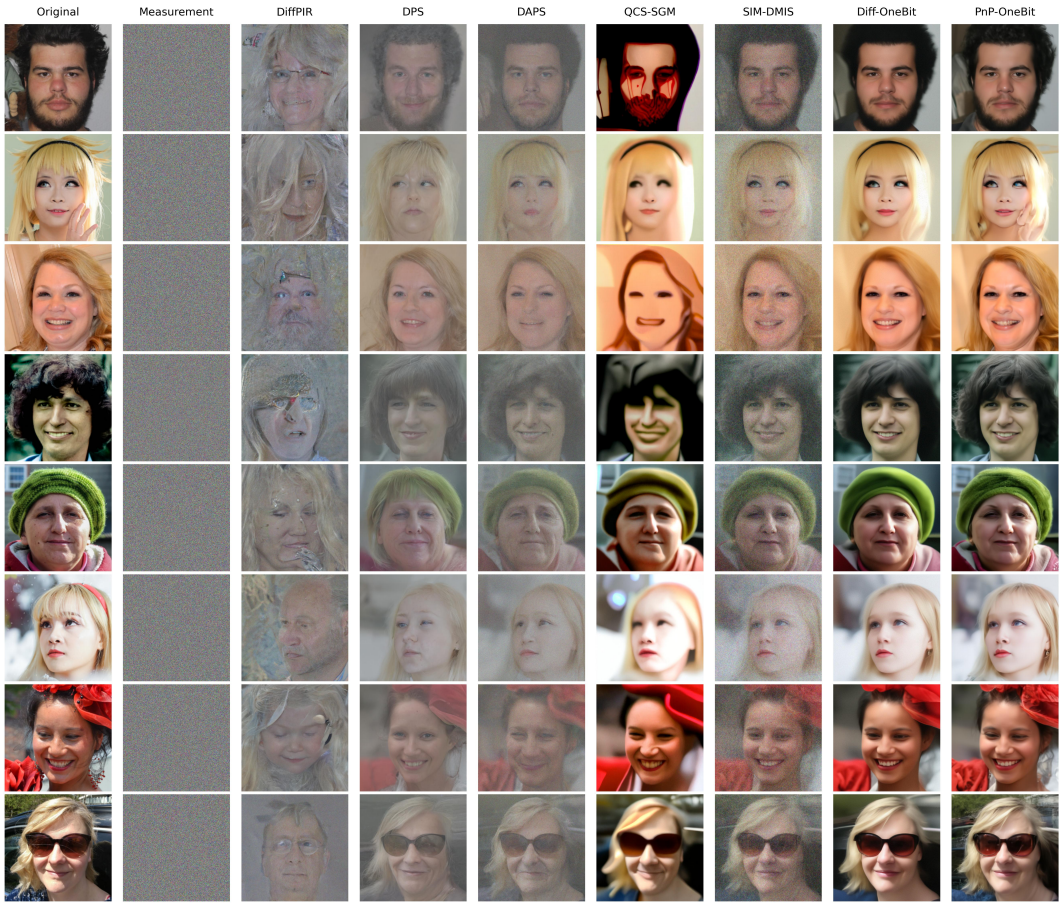


Figure 2: Qualitative comparison of different methods on the FFHQ dataset with compression ratio $n/m = 16$ and additive Gaussian noise $\sigma = 0.5$.

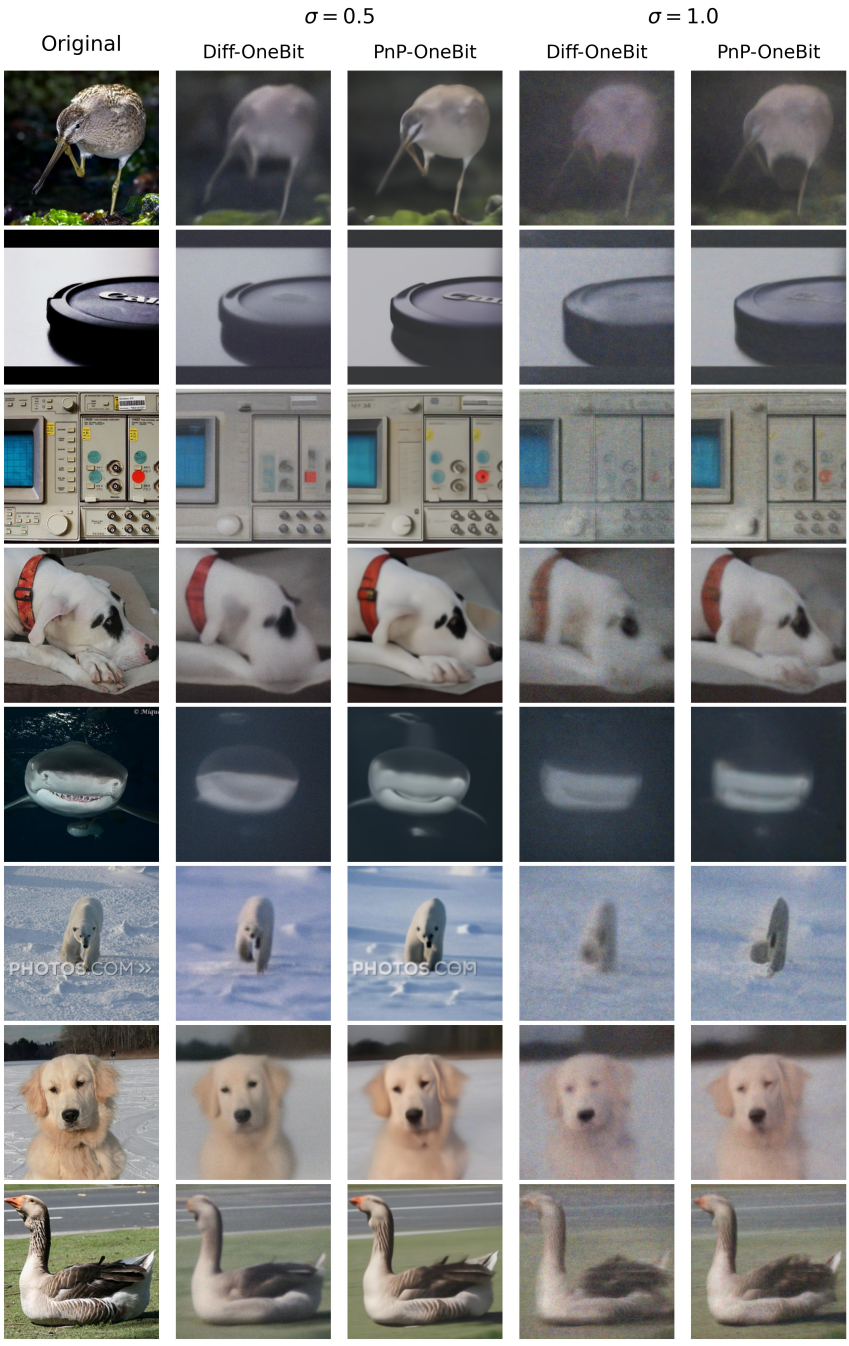


Figure 3: Qualitative comparison between Diff-OneBit and PnP-OneBit on the ImageNet dataset with compression ratio $n/m = 16$ under different noise levels ($\sigma = 0.5$ and $\sigma = 1.0$).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims presented in our abstract and introduction, including both the theoretical recovery guarantees for posterior sampling and the empirical superiority of our proposed algorithm, are fully supported by the mathematical proofs and experimental results detailed throughout the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We acknowledge the limitations of our current theoretical analysis, and provide a discussion on these constraints and potential directions for future improvements in Conclusion section.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theoretical claims are presented with their necessary assumptions clearly stated. We provide intuitive proof sketches in the main manuscript and include complete, rigorous formal proofs for every theorem and lemma in Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide comprehensive implementation details in the appendix, including all open-source datasets and pre-trained models, a detailed algorithmic pseudo-code, and the complete set of hyperparameters used. These resources are sufficient for readers to independently replicate our algorithm and experimental findings.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: All datasets used in this work are open-source with access provided in the appendix, while the source code will be released upon acceptance. We have included comprehensive pseudo-code and implementation details in the supplemental material to ensure reproducibility.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We fully disclose all training and evaluation configurations, including hyperparameter choices and hardware specifications, across the main manuscript and supplemental material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: All main experimental results include standard deviations across multiple independent trials to demonstrate statistical significance, as detailed in the quantitative comparison tables.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the necessary hardware specifications required to reproduce our experiments, including the specific GPU model and available memory, in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: All research conducted in this paper, including data usage and methodology, adheres strictly to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This work focuses on the mathematical and algorithmic foundations of 1-bit inverse problems. As a piece of foundational signal recovery research not tied to a specific deployment, we believe it does not have an immediate societal impact beyond the academic community.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This research focuses on the technical aspects of signal restoration and does not involve releasing high-risk generative models or newly scraped datasets with risks of misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All assets, including datasets and pre-trained models, are properly cited. Their specific licenses (e.g., CC BY-NC-SA 4.0 and MIT) are detailed in Appendix.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: This paper does not release new datasets, software, or model weights with the current submission. The proposed methodology is a mathematical and algorithmic contribution fully documented for reproducibility in the appendix.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This research is purely algorithmic and does not involve any crowdsourcing or research conducted with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: As this study does not involve human participants, private human data, or clinical trials, institutional review board (IRB) approval was not required.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were used only for professional editing, language polishing, and LaTeX formatting purposes and did not influence the core methodology, scientific rigor, or original contributions of this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.