A VERIFICATION OF ASSUMPTION 1

Mathematically, we can write the exact $p(\mathbf{y} \mid \tilde{\mathbf{x}})$ as

$$p(\mathbf{y} \mid \tilde{\mathbf{x}}) = \int p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x} \mid \tilde{\mathbf{x}}) d\mathbf{x}, \tag{19}$$

where from the Bayes' rule,

$$p(\mathbf{x} \mid \tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}} \mid \mathbf{x}) p(\mathbf{x})}{\int p(\tilde{\mathbf{x}} \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}}. \tag{20}$$

For NCSN/NCSNv2, recall that the likelihood $p(\tilde{\mathbf{x}} \mid \mathbf{x})$ follows a Gaussian distribution $\mathcal{N}(\tilde{\mathbf{x}}; \mathbf{x}, \beta^2 \mathbf{I})$, where β is the variance of the perturbed Gaussian noise and, by definition, $\beta \to 0$ in later steps of the reverse diffusion process in NCSN/NCSNv2. As a result, from (20), the likelihood $p(\tilde{\mathbf{x}} \mid \mathbf{x})$ will dominate the posterior $p(\mathbf{x} \mid \tilde{\mathbf{x}})$ as $\beta \to 0$ and therefore $p(\mathbf{x} \mid \tilde{\mathbf{x}}) \propto p(\tilde{\mathbf{x}} \mid \mathbf{x})$, indicating that the prior $p(\mathbf{x})$ becomes uninformative. Note that this is particularly the case when the entropy of $p(\mathbf{x})$ is high, which is usually the case for generative models capable of generating diverse images. While this assumption is crude at the beginning of the reverse diffusion process when the noise variance β is large, it is asymptotically accurate as the process goes forward. The effectiveness of this assumption is also empirically supported by various experiments in Section 4.

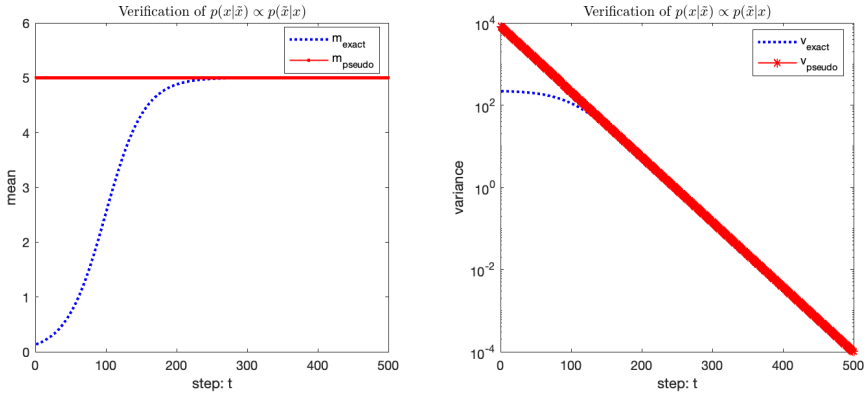


Figure 6: Comparison of the exact mean and variance of $p(x \mid \tilde{x})$ with the pseudo mean and variance under the uninformative assumption, i.e., $p(x \mid \tilde{x}) \propto p(\tilde{x} \mid x)$ in the toy scalar Gaussian example. In this plot, we set $\beta_{\max} = 90$, $\beta_{\min} = 0.01$, $T = 500$, which are the same as the setting of NCSNv2 for CelebA. The $\tilde{x}$ and the prior standard deviation σ_0 is set to be $\sigma_0 = 15$. It can be seen that the approximated values approach the exact values very quickly, verifying the effectiveness of the Assumption 1 for this toy example.

A toy example: In the following, we illustrate the assumption in a toy example where $\mathbf{x}$ reduces to a scalar random variable x and the associated prior $p(x)$ follows a Gaussian distribution, i.e., $p(x) = \mathcal{N}(x; 0, \sigma_0^2)$, where σ^2 is the prior variance. The likelihood $p(\tilde{\mathbf{x}} \mid \mathbf{x})$ in this case is simply $p(\tilde{x} \mid x) = \mathcal{N}(\tilde{x}; x, \beta^2)$. Then, from (20), after some algebra, it can be computed that the posterior distribution $p(x \mid \tilde{x})$ is

$$p(x \mid \tilde{x}) = \mathcal{N}(\tilde{x}; m_{\text{exact}}, v_{\text{exact}}) \tag{21}$$

where

$$m_{\text{exact}} = \frac{\sigma_0^2}{\sigma_0^2 + \beta^2} \tilde{x}, \; v_{\text{exact}} = \frac{\sigma_0^2 \beta^2}{\sigma_0^2 + \beta^2}. \tag{22}$$

Under the Assumption 1, i.e., $p(x \mid \tilde{x}) \propto p(\tilde{x} \mid x)$, we obtain an approximation of $p(x \mid \tilde{x})$ as follows

$$p(x \mid \tilde{x}) \simeq \tilde{p}(x \mid \tilde{x}) = \mathcal{N}(\tilde{x}; m_{\text{pseudo}}, v_{\text{pseudo}}), \tag{23}$$

where

$$m_{\text{pseudo}} = \tilde{x}, \; v_{\text{pseudo}} = \beta^2. \tag{24}$$

By comparing the exact result (22) and approximation result (24), it can be easily seen that for a fixed $\sigma_0^2 > 0$, as $\beta \to 0$, we have $m_{\text{pseudo}} \to m_{\text{post}}$ and $v_{\text{pseudo}} \to v_{\text{post}}$. In Figure 6, similar to NCSN/NCSNv2, we anneal β as $\beta_t = \beta_{\max}(\frac{\beta_{\min}}{\beta_{\max}})^{\frac{t-1}{T-1}}$ geometrically and compare $m_{\text{pseudo}}, v_{\text{pseudo}}$ with $m_{\text{exact}}, v_{\text{exact}}$ as t increase from 1 to T. It can be seen in Figure 6 that the approximated values $m_{\text{pseudo}}, v_{\text{pseudo}}$, especially the variance v_{pseudo}, approach to the exact values $m_{\text{exact}}, v_{\text{exact}}$ very quickly, verifying the effectiveness of the Assumption 1 under this toy example.

B VERIFICATION OF ASSUMPTION 2

Before verifying it in the i.i.d. Gaussian case, we first briefly add a comment on the general case when $\mathbf{A}\mathbf{A}^T$ is not an exact diagonal matrix. As demonstrated later in Appendix C, the Assumption 2 is introduced to ensure that the covariance matrix $\sigma^2\mathbf{I} + \beta_t^2\mathbf{A}\mathbf{A}^T$ of Gaussian noise is diagonal so that we can obtain a closed-form solution for the likelihood score in the quantized case, which is otherwise intractable. If

Suppose that the elements of $\mathbf{A}$ follow i.i.d. Gaussian, i.e., $A_{ij} \sim \mathcal{N}(0, \sigma^2)$. Next, we investigate the elements of the matrix $\mathbf{C} = \mathbf{A}\mathbf{A}^T = \{C_{ij}\}, i, j = 1...M$.

Regarding the diagonal elements C_{ii}, by definition, it reads as

$$C_{ii} = \sum_{n=1...N} A_{in}^2,\ i = 1...M. \tag{25}$$

As C_{ii} is the sum of square of N i.i.d. Gaussian random variables $A_{ij},\ j = 1...N$, C_{ii} follows a Gamma distribution, i.e., $\Gamma\big(\frac{N}{2}, 2\sigma^2\big)$. The mean and variance of C_{ii} can be computed as $N\sigma^2$ and $2N\sigma^4$.

Regarding the off-diagonal elements $C_{ij}, i \neq j$, by definition, it reads as

$$C_{ij} = \sum_{n=1...N} A_{in}A_{jn},\ i, j = 1...M, i \neq j. \tag{26}$$

As A_{in} and A_{jn} are independent Gaussian for $i \neq j$, it can be computed that the mean and variance of C_{ij} are 0 and $N\sigma^4$, respectively. When $\sigma^2 = 1/M$ and $M = \alpha N$, where $\alpha > 0$ is the constant measurement ratio, the variance of C_{ij} is $\frac{1}{\alpha^2 N} \to 0$ as $N \to \infty$. As a result, all the off-diagonal elements of $\mathbf{A}\mathbf{A}^T$ will tend to zero as $N \to \infty$. From another perspective, in Figure 7, we compute the ratio between the average magnitude of off-diagonal elements C_{ij} and the diagonal elements C_{ii} when $M = \alpha N$ with $\alpha = 0.5$. It can be seen that as N increases, the magnitude of the off-diagonal elements becomes negligible. As a result, when $\mathbf{A}$ is i.i.d. Gaussian, the matrix $\mathbf{A}\mathbf{A}^T$ can be well approximated as a diagonal matrix in the high-dimensional case.

C PROOF OF THEOREM 1

Proof. Let us denote $\mathbf{z} = \mathbf{A}\mathbf{x}$. For each noise scale $\beta_t > 0$, under the Assumption 1, we obtain

$$\begin{aligned} p_{\beta_t}(\mathbf{x} \mid \tilde{\mathbf{x}}) &\propto p_{\beta_t}(\tilde{\mathbf{x}} \mid \mathbf{x}) \\ &\sim \mathcal{N}(\mathbf{x}; \tilde{\mathbf{x}}, \beta_t^2\mathbf{I}) \end{aligned} \tag{27}$$

then we can equivalently write

$$\mathbf{x} = \tilde{\mathbf{x}} + \beta_t\mathbf{w}, \tag{28}$$

where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. As a result, $\mathbf{z} = \mathbf{A}\mathbf{x} = \mathbf{A}(\tilde{\mathbf{x}} + \beta_t\mathbf{w}) = \mathbf{A}\tilde{\mathbf{x}} + \beta_\mathbf{t}\mathbf{A}\mathbf{w}$. Then, from (2), we obtain

$$\mathbf{y} = \mathsf{Q}\left(\mathbf{A}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}\right), \tag{29}$$

where $\tilde{\mathbf{n}} = \mathbf{n} + \beta_t\mathbf{A}\mathbf{w}$. Since $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and are independent to each other, it can be concluded that $\tilde{\mathbf{n}}$ is also Gaussian with mean zero and covariance $\sigma^2\mathbf{I} + \beta_t^2\mathbf{A}\mathbf{A}^T$, i.e., $\tilde{\mathbf{n}} \sim \mathcal{N}(\tilde{\mathbf{n}}; \mathbf{0}, \sigma^2\mathbf{I} + \beta_t^2\mathbf{A}\mathbf{A}^T)$.

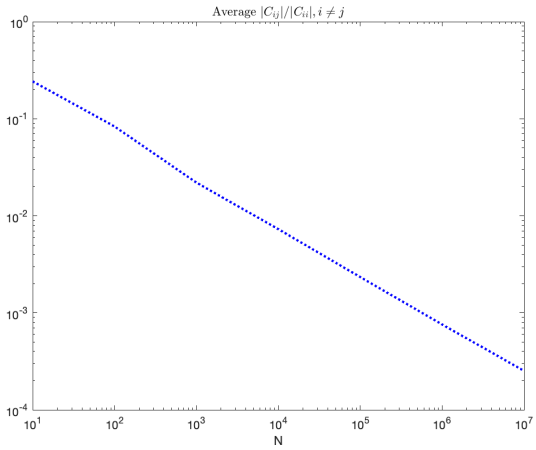


Figure 7: The ratio between the average magnitude of the diagonal elements C_{ii} and off-diagonal elements C_{ij} when $M = \alpha N$ with $\alpha = 0.5$.

Subsequently, under the Assumption 2, i.e., $\mathbf{A}\mathbf{A}^T$ is a diagonal matrix, each element $\tilde{n}_m \sim \mathcal{N}(\tilde{n}_m; 0, \sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2)$ of $\tilde{\mathbf{n}}$ will be independent to each other and thus from (38) we can obtain a closed-form solution for the likelihood distribution (we will refer it as a pseudo-likelihood due to the assumptions used) $p(\mathbf{y}|\hat{\mathbf{z}} = \mathbf{A}\tilde{\mathbf{x}})$ as follows

$$p(\mathbf{y}|\hat{\mathbf{z}} = \mathbf{A}\tilde{\mathbf{x}}) = \prod_{m=1}^{M} p\left(y_m \mid \hat{z}_m = \mathbf{a}_m^T\tilde{\mathbf{x}}\right) \tag{30}$$

where, from the definition of quantizer Q,

$$p\left(y_m \mid \hat{z}_m = \mathbf{a}_m^T\tilde{\mathbf{x}}\right) = p\left(l_{y_m} \leq \hat{z}_m + \tilde{n}_m < u_{y_m}\right) \tag{31}$$

$$= \Phi\left(\frac{-\hat{z}_m + u_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2}}\right) - \Phi\left(\frac{-\hat{z}_m + l_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2}}\right) \tag{32}$$

$$= \Phi\left(-\tilde{u}_{y_m}\right) - \Phi\left(-\tilde{l}_{y_m}\right) \tag{33}$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$ is the cumulative distribution function of the standard normal distribution and

$$\tilde{u}_{y_m} = \frac{\mathbf{a}_m^T\tilde{\mathbf{x}} - u_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2}},\ \tilde{l}_{y_m} = \frac{\mathbf{a}_m^T\tilde{\mathbf{x}} - l_{y_m}}{\sqrt{\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2}}. \tag{34}$$

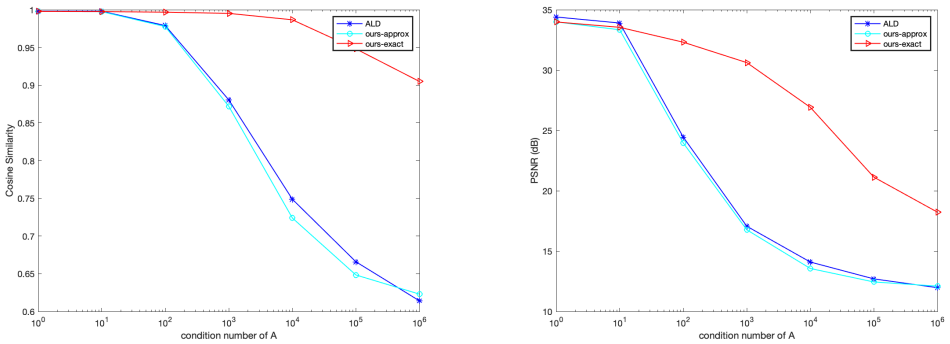
As a result, it can be calculated that the noise-perturbed pseudo-likelihood score $\nabla_{\tilde{\mathbf{x}}} \log p_{\beta_t}(\mathbf{y} \mid \tilde{\mathbf{x}})$ for the quantized measurements $\mathbf{y}$ in (2) can be computed as

$$\nabla_{\tilde{\mathbf{x}}} \log p_{\beta_t}(\mathbf{y} \mid \tilde{\mathbf{x}}) = \mathbf{A}^T \mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \tilde{\mathbf{x}}) \tag{35}$$

where $\mathbf{G}(\beta_t, \mathbf{y}, \mathbf{A}, \tilde{\mathbf{x}}) = [g_1, g_2, ..., g_M]^T \in \mathbb{R}^{M \times 1}$ with each element being

$$g_m = \frac{\exp\left(-\frac{\tilde{u}_{y_m}^2}{2}\right) - \exp\left(-\frac{\tilde{l}_{y_m}^2}{2}\right)}{\sqrt{\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2} \int_{\tilde{l}_{y_m}}^{\tilde{u}_{y_m}} \exp\left(-\frac{t^2}{2}\right) dt},\ m = 1, 2, ..., M, \tag{36}$$

which completes the proof. □



(a) Cosine Similarity (b) PSNR

Figure 8: Averaged Cosine Similarity and PSNR of reconstructed MNIST images of ALD in Jalal et al. (2021a) and ours (with formula (18) and (16), respectively) when $M = 200, \sigma = 0.1$ for the different condition number of the matrix $\mathbf{A}$ is cond$(\mathbf{A}) = 1000$. It can be seen that our method with (16) significantly outperforms ALD in Jalal et al. (2021a) at high condition number while performing similarly at low condition number. Ours with diagonal approximation (18) is about the same as ALD as expected.

D Proof of Corollary 1.2

Proof. Similarly in the proof of Theorem 1, let us denote $\mathbf{z} = \mathbf{A}\mathbf{x}$. For each noise scale $\beta_t > 0$, under the Assumption 1, we can equivalently write

$$\mathbf{x} = \tilde{\mathbf{x}} + \beta_t \mathbf{w}, \tag{37}$$

where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. As a result, $\mathbf{z} = \mathbf{A}\mathbf{x} = \mathbf{A}(\tilde{\mathbf{x}} + \beta_t \mathbf{w}) = \mathbf{A}\tilde{\mathbf{x}} + \beta_t \mathbf{A}\mathbf{w}$. Then, in the case of the linear model, from (1), we obtain

$$\mathbf{y} = \mathbf{A}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{38}$$

where $\tilde{\mathbf{n}} = \mathbf{n} + \beta_t \mathbf{A}\mathbf{w}$. Since $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and are independent to each other, it can be concluded that $\tilde{\mathbf{n}}$ is also Gaussian with mean zero and covariance $\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T$, i.e., $\tilde{\mathbf{n}} \sim \mathcal{N}(\tilde{\mathbf{n}}; \mathbf{0}, \sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T)$. Therefore, a closed-form solution for the likelihood distribution $p(\mathbf{y}|\hat{\mathbf{z}} = \mathbf{A}\tilde{\mathbf{x}})$ can be obtained as follows

$$p(\mathbf{y}|\hat{\mathbf{z}} = \mathbf{A}\tilde{\mathbf{x}}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}})^T \left(\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T\right)^{-1} (\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}})\right)}{\sqrt{(2\pi)^M \det\left(\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T\right)}}. \tag{39}$$

As a result, we can readily obtain a closed-form solution for the noise-perturbed pseudo-likelihood score $\nabla_{\tilde{\mathbf{x}}} \log p_{\beta_t}(\mathbf{y} \mid \tilde{\mathbf{x}})$ as follows

$$\nabla_{\tilde{\mathbf{x}}} \log p_{\beta_t}(\mathbf{y} \mid \tilde{\mathbf{x}}) = \mathbf{A}^T \left(\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T\right)^{-1} (\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}). \tag{40}$$

Furthermore, if $\mathbf{A}\mathbf{A}^T$ is a diagonal matrix, the inverse of matrix $\left(\sigma^2 \mathbf{I} + \beta_t^2 \mathbf{A}\mathbf{A}^T\right)$ become trivial since it is a diagonal matrix with the m-th diagonal element being $\sigma^2 + \beta_t^2 \left\|\mathbf{a}_m^T\right\|_2^2$. After some simple algebra, we can obtain the equivalent representation in (18), which completes the proof. □

E Comparison with ALD in Jalal et al. (2021a) in the linear case

As shown in Corollary 1.2, in the special case without quantization, our results in Theorem 1 can be reduced to a form similar to the ALD in Jalal et al. (2021a). However, there are several different important differences. First, our results are derived in a principled way as noise-perturbed pseudo-likelihood score and admit closed-form solutions, while the results in Jalal et al. (2021a) are obtained

(a) Truth (b) ALD (Jalal et al., 2021a) (c) Ours (approx (18)) (d) Ours (with (16))

Figure 9: Tyical recovered MNIST images of ALD in Jalal et al. (2021a) and ours (with formula (18) and (16), respectively) when $M = 200, \sigma = 0.05$ and the condition number of matrix $\mathbf{A}$ is $\text{cond}(\mathbf{A}) = 1000$. It can be seen that our method with (16) significantly outperforms ALD in Jalal et al. (2021a), which performs about the same as ours with diagonal approximation (18).

heuristically by adding an additional hyper-parameter (and thus needs fine-tuning). Second, the results in Jalal et al. (2021a) are similar to an approximate version (18) of ours, which holds only when $\mathbf{A}\mathbf{A}^T$ is a diagonal matrix. For general matrices $\mathbf{A}$, we have a closed-form approximation (16). We compare our results with ALD (Jalal et al., 2021a) and the results are shown in Figure 8 and Figure 9. It can be seen that when at low condition number of $\mathbf{A}$ when $\mathbf{A}\mathbf{A}^T$ is approximately a diagonal matrix, our with diagonal approximation (18) performs similarly as (16), both of which are similar to ALD (Jalal et al., 2021a), as expected. However, when the condition number of $\mathbf{A}$ is large so that $\mathbf{A}\mathbf{A}^T$ is far from a diagonal matrix, our method with the (16) significantly outperforms ALD in Jalal et al. (2021a).

F Detailed Experimental Settings

In training NCSNv2 for MNIST, we used a similar training setup as Song & Ermon (2020) for Cifar10 as follows.

Training: batch-size: 128 n-epochs: 500000 n-iters: 300001 snapshot-freq: 50000 snapshot-sampling: true anneal-power: 2 log-all-sigmas: false.

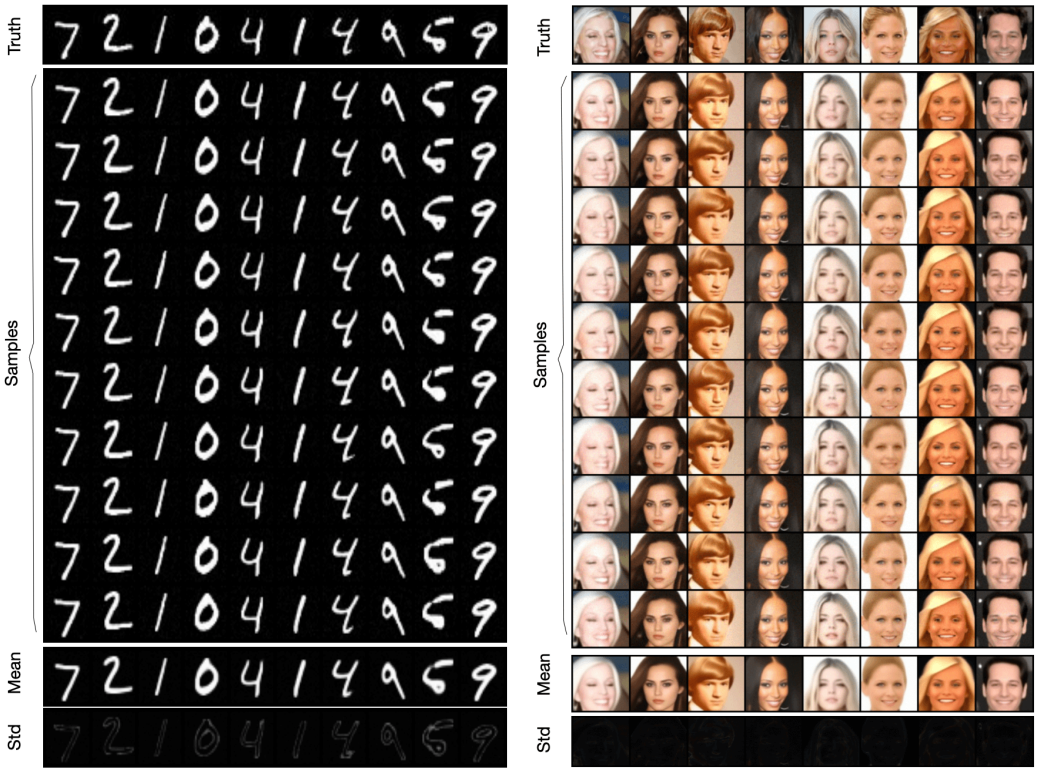
Please refer to Song & Ermon (2020) and associated open-sourced code for details of training. For Cifar-10, CelebA, and FFHQ, we directly use the pre-trained models available in this Link.

When performing posterior sampling using the QCS-SGM in 1, for simplicity, we set a constant value $\epsilon = 0.0002$ for all quantized measurements (e.g., 1-bit, 2-bit, 3-bit) for MNIST, Cifar10 and CelebA. For the high-resolution FFHQ 256×256, we set $\epsilon = 0.00005$ for 1-bit and $\epsilon = 0.00002$ for 2-bit and 3-bit case, respectively. For all linear measurements for MNIST, Cifar10, and CelebA, we set $\epsilon = 0.00002$. It is believed that some improvement can be achieved with further fine-tuning of ϵ for different scenarios. For MNIST and Cifar-10, we set $\beta_1 = 50, \beta_T = 0.01, T = 232$; for CelebA, we set $\beta_1 = 90, \beta_T = 0.01, T = 500$; for FFHQ, we set $\beta_1 = 348, \beta_T = 0.01, T = 2311$ which are the same as Song & Ermon (2020). The number of steps K in QCS-SGM for each noise scale is set to be $K = 5$ in all experiments. For more details, please refer to the submitted code.

G Multiple Samples and Uncertainty Estimates

As one kind of posterior sampling method, QCS-SGM can yield multiple samples with different random initialization so that we can easily obtain confidence intervals or uncertainty estimates of the reconstructed results. For example, typical samples as well as mean and std are shown in Figure 10 for MNIST and CelebA in the 1-bit case.

(a) MNIST, $M = 200, \sigma = 0.05$ (b) CelebA, $M = 4000, \sigma = 0.001$



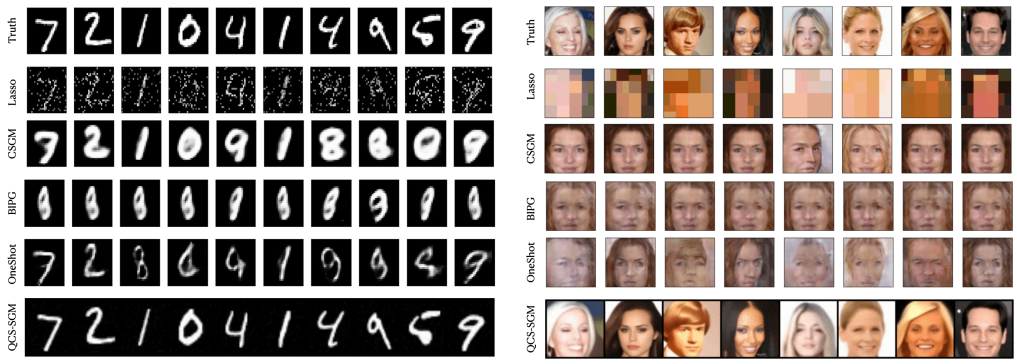
(c) MNIST, $M = 400, \sigma = 0.05$ (d) CelebA, $M = 10000, \sigma = 0.001$

Figure 10: Multiple samples of QCS-SGM on MNIST ($M = 200, 400, \sigma = 0.05$) and CelebA datasets ($M = 4000, 1000, \sigma = 0.001$) from 1-bit measurements. The mean and std of the samples are also shown.

H ADDITIONAL RESULTS

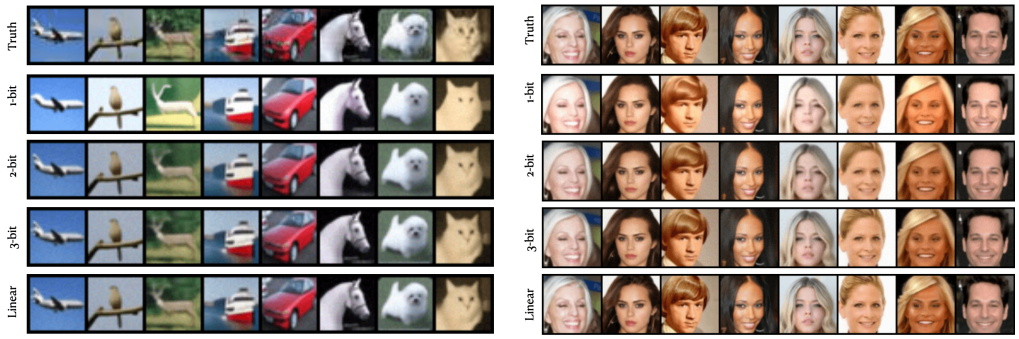
Some additional results are shown in this section.

Figure 11 and Figure 12 show results with relatively large value of M in the same setting as Figure 2 and Figure 4, respectively.



(a) MNIST, $M = 400, \sigma = 0.05$ (b) CelebA, $M = 10000, \sigma = 0.001$

Figure 11: Typical reconstructed images from 1-bit measurements on MNIST ($M = 400$) and CelebA ($M = 10000$).



(a) Cifar-10, $M = 4000$, $\sigma = 0.001$ (b) CelebA, $M = 16000$ $\sigma = 0.001$

Figure 12: Results of QCS-SGM for Cifar-10 ($M = 4000$) and CelebA ($M = 16000$) images under different quantization bits.

Figure 13 shows the quantitative results of QCS-SGM for different quantization bits.

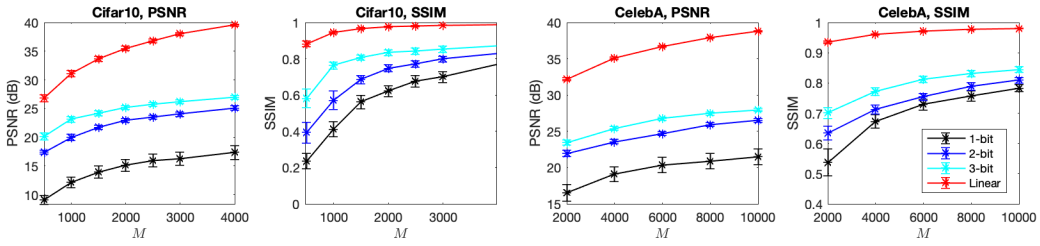
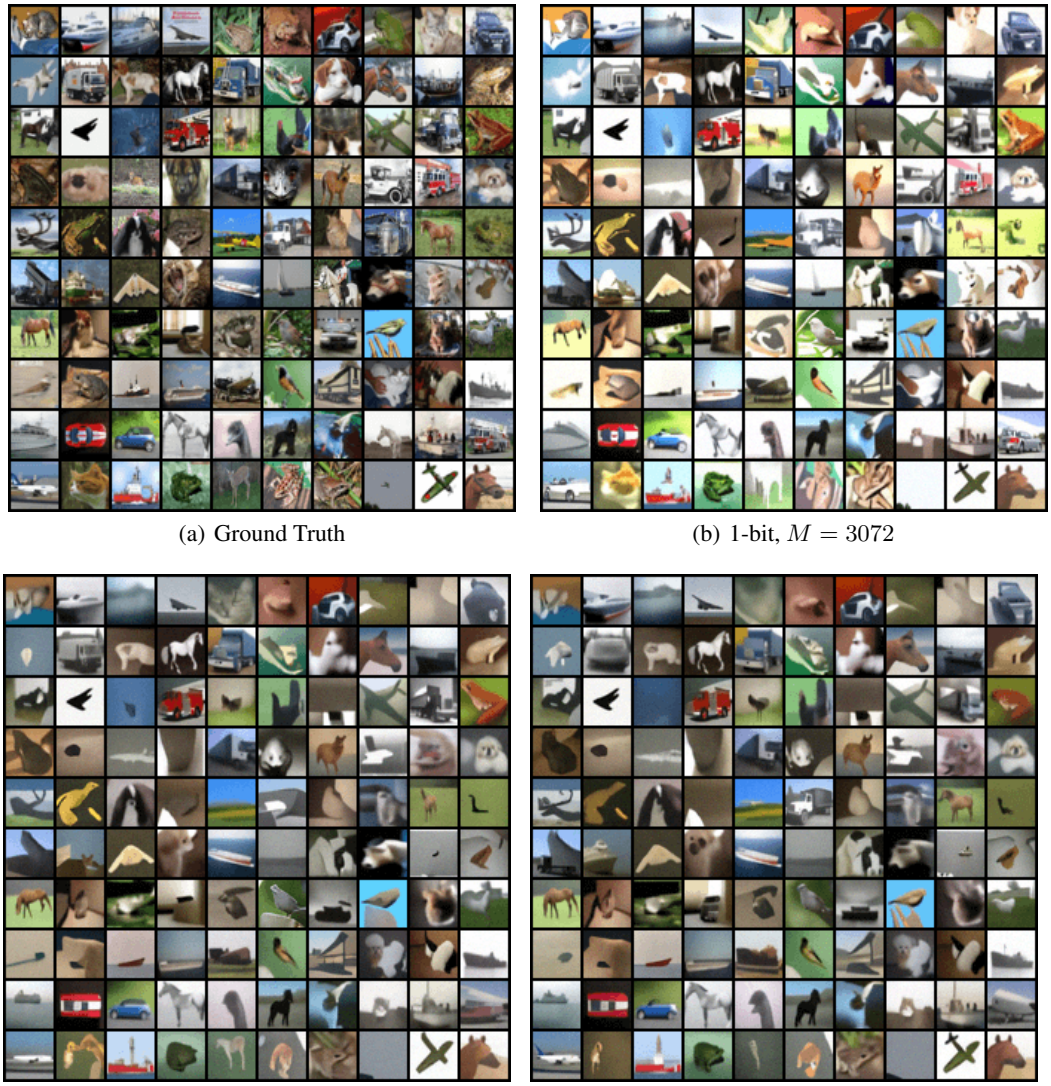


Figure 13: Quantitative results of QCS-SGM for different quantization bits, $\sigma = 0.001$.

Figure 14 and Figure 15 show the reconstructed images of QCS-SGM for Cifar10 and CelebA in the fixed budget case of $Q \times M = 3072$ and $Q \times M = 12288$, respectively.



(a) Ground Truth

(b) 1-bit, $M = 3072$

(c) 2-bit, $M = 1536$

(d) 3-bit, $M = 1024$

Figure 14: Reconstructed images of QCS-SGM for Cifar10 in the fixed budget case ($Q \times M = 3072$).

(a) Ground Truth

(b) 1-bit, $M = 12288$

(c) 2-bit, $M = 6144$

(d) 3-bit, $M = 4096$

Figure 15: Reconstructed images of QCS-SGM for CelebA in the fixed budget case ($Q \times M = 12288$) in the same setting as Table 1.

Method	Cifar10 CS2		Cifar10 CS8		CelebA CS2		CelebA CS8	
	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑
QCS-SGM (ours)	**0.9712**	**34.33**	**0.8634**	**26.20**	**0.9694**	**36.80**	**0.9156**	**31.13**
Neumann networks (Gilton et al., 2019)	-	33.83	-	25.15	-	35.12	-	28.38

Table 2: Quantitative comparison (PSNR (dB) and SSIM) of QCS-SGM with the Neumann networks in Gilton et al. (2019) for CS in the linear case, i.e., without quantization. Both Cifar10 32×32 and CelebA 64×64 are considered. It can be seen that QCS-SGM outperforms the Neumann networks (Gilton et al., 2019) in all cases. As in Gilton et al. (2019). the values reported are the median across a test set of size 256. Reconstructed images of QCS-SGM are shown in Figure 16 - 19.

Figure 16: Results of QCS-SGM on Cifar10 for CS2, i.e., $M = 1536, N = 3072$. Noiseless case.

I Comparison with Neumann Networks Gilton et al. (2019)

In the case of no quantization, we compare our method with one popular method called Neumann Networks (Gilton et al., 2019). The results for Cifar10 are shown in Table 2. It can be seen that our method outperforms Neumann Networks by a large margin. We do not show quantization comparisons for quantized CS since Neumann Networks does not support it.

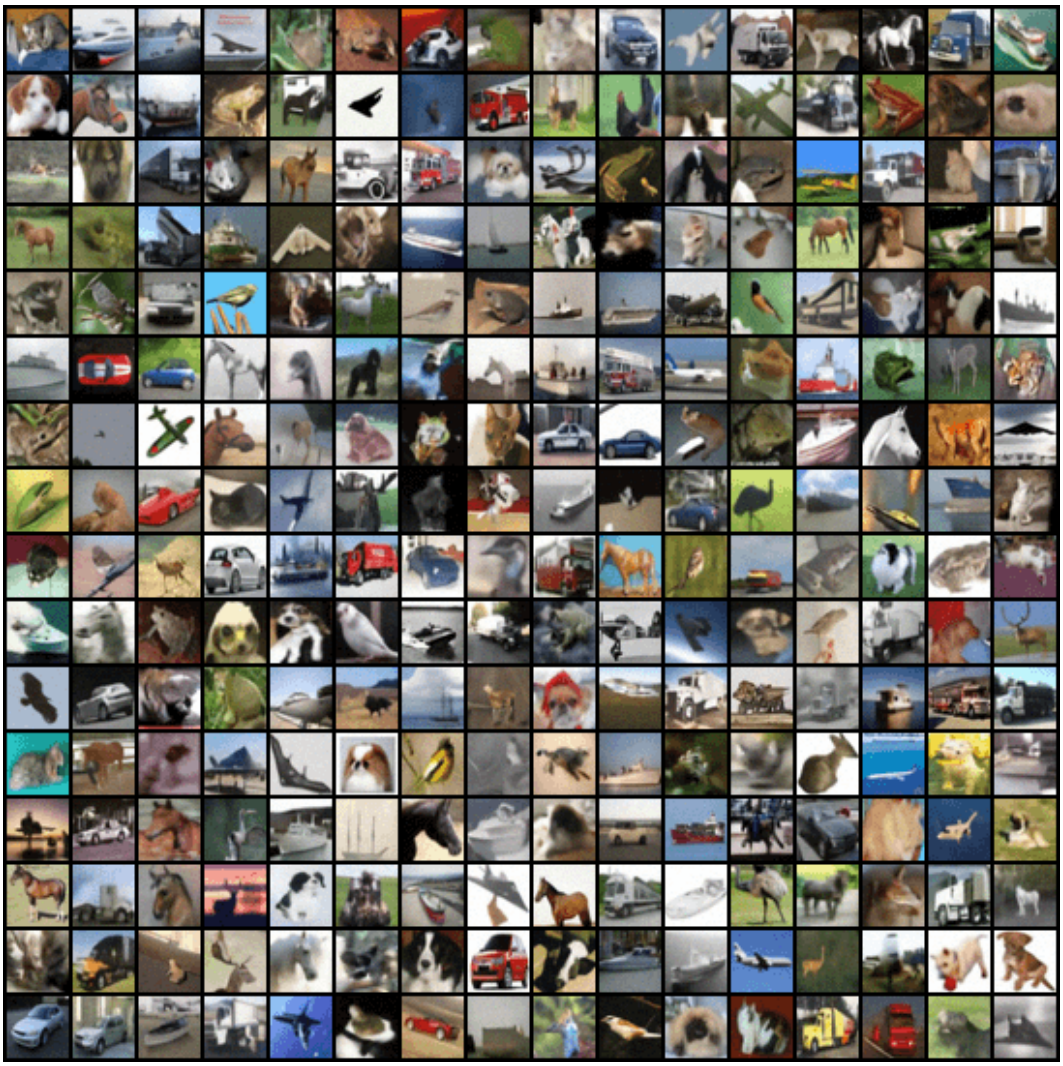
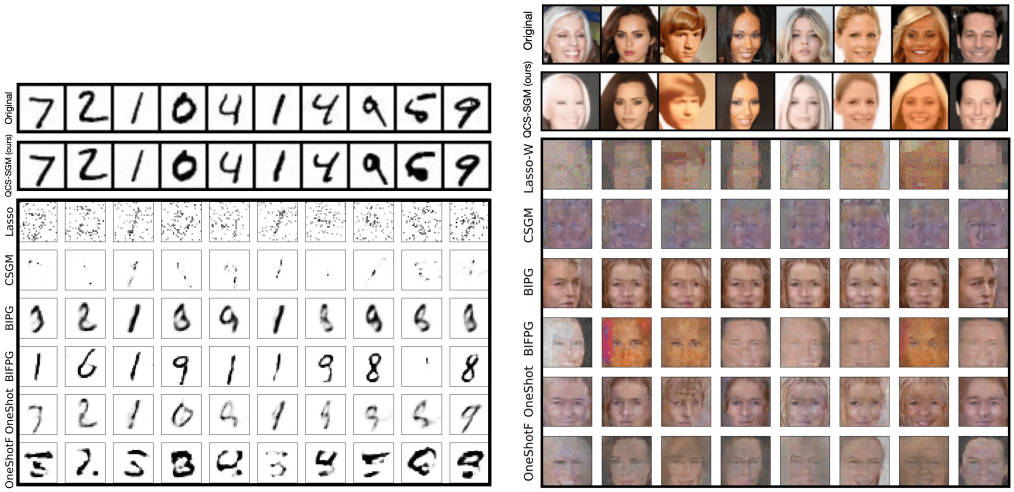


Figure 17: Results results of QCS-SGM on Cifar10 for CS8, i.e., $M = 1536, N = 3072$. Noiseless case.

Figure 18: Results of QCS-SGM on CelebA for CS2, i.e., $M = 6144, N = 12288$. Noiseless case.

Figure 19: Results of QCS-SGM on CelebA for CS8, i.e., $M = 1536, N = 12288$. Noiseless case.

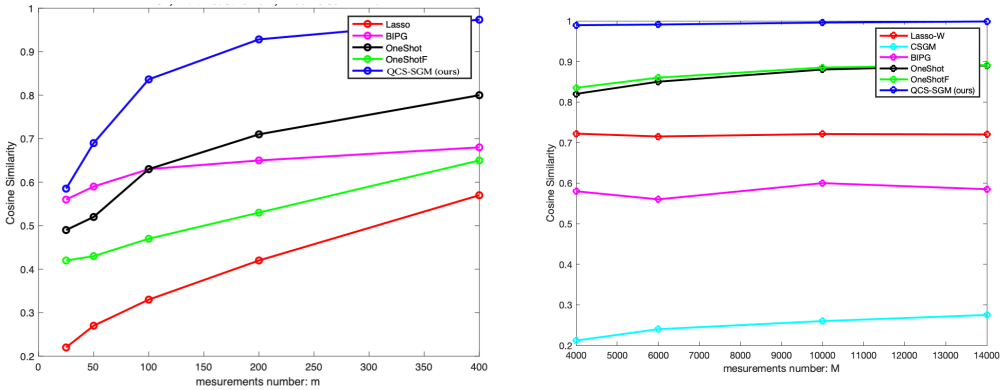


(a) MNIST, $M = 200, \sigma = 1$ (b) CelebA, $M = 4000, \sigma = 0.01$

Figure 20: Results of reconstructed images from 1-bit measurements on MNIST and CelebA images, following exactly the same setting as Liu & Liu (2022), i.e., $A_{ij} \sim \mathcal{N}(0, 1)$, and $n_i \sim \mathcal{N}(0, \sigma^2)$.

J Comparison in the exactly same setting as Liu & Liu (2022)

Note that there is a slight difference in the modeling of measurement matrix $\mathbf{A}$ and noise $\mathbf{n}$ between ours and that in Liu & Liu (2022). In Liu & Liu (2022), it is assumed that $\mathbf{A}$ is an i.i.d. Gaussian matrix where each element A_{ij} follows $A_{ij} \sim \mathcal{N}(0, 1)$ and that $\mathbf{n}$ is an i.i.d. Gaussian with variance σ^2, i.e., $n_i \sim \mathcal{N}(0, \sigma^2)$. However, in practice, the measurement matrix $\mathbf{A}$ is usually normalized so that the norm of each column equals 1. As a result, in our setting, we assume that each element of i.i.d. Gaussian matrix follows $A_{ij} \sim \mathcal{N}(0, 1/M)$. Mathematically, there is a one-to-one correspondence between the two settings, but the simulation setting is different due to the different measurement size M. As a result, for an exact comparison with results in Liu & Liu (2022), we also conducted experiments assuming exactly the same setting as Liu & Liu (2022), i.e., $A_{ij} \sim \mathcal{N}(0, 1)$, and $n_i \sim \mathcal{N}(0, \sigma^2)$. The results are shown in Figures 20, 21, 22, 23.

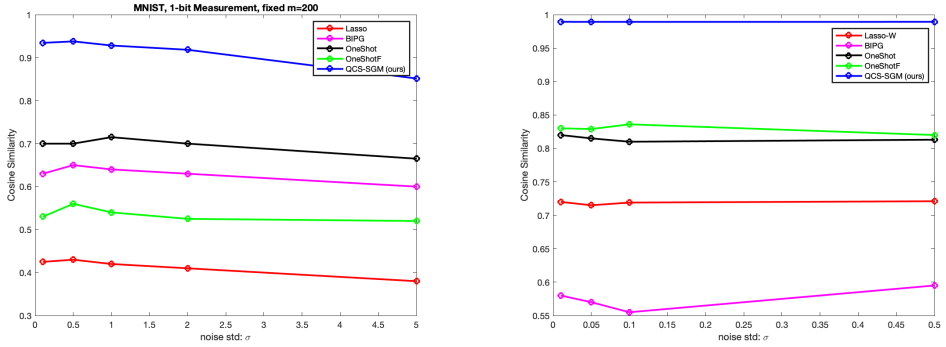


(a) MNIST, $\sigma = 1$ (b) CelebA, $\sigma = 0.01$

Figure 21: Quantitative comparisons based on cosine similarity for 1-bit MNIST and CelebA images, following exactly same setting as Liu & Liu (2022), i.e., $A_{ij} \sim \mathcal{N}(0, 1)$, and $n_i \sim \mathcal{N}(0, \sigma^2)$.

(a) MNIST, $M = 400, \sigma = 1$ (b) CelebA, $M = 10000, \sigma = 0.01$

Figure 22: Results of reconstructed images from 1-bit measurements on MNIST and CelebA images, following exactly the same setting as Liu & Liu (2022), i.e., $A_{ij} \sim \mathcal{N}(0, 1)$, and $n_i \sim \mathcal{N}(0, \sigma^2)$.



(a) MNIST, $M = 200$ (b) CelebA, $M = 4000$

Figure 23: Results of reconstructed images from 1-bit measurements on MNIST and CelebA images for different σ, following exactly the same setting as Liu & Liu (2022), i.e., $A_{ij} \sim \mathcal{N}(0, 1)$, and $n_i \sim \mathcal{N}(0, \sigma^2)$.