Improving Noise Efficiency in Privacy-preserving Dataset Distillation

Supplementary Material

A. Theoretical Analysis

Lemma 1. *Let $\mathcal{V} = \{\boldsymbol{v}_j\}_{j=1}^{L}$ be a dataset of L samples $\boldsymbol{v}_i \in \mathbb{R}^D$. Let $\boldsymbol{P} \in \mathbb{R}^{D\times d}$ be a matrix with orthonormal rows (i.e., $\boldsymbol{P}\boldsymbol{P}^\top = \mathbb{I}_d$). Suppose that adding Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_1^2\mathbb{I}_D)$ to the sample mean ensures (ϵ, δ)-differential privacy:*

$$\hat{\mu}_{orig} = \frac{1}{L}\sum_{j=1}^{L}\boldsymbol{v}_j + \boldsymbol{\eta}.$$

Then, adding Gaussian noise $\boldsymbol{\eta}' \sim \mathcal{N}(0, \sigma_2^2\mathbb{I}_d)$ to the projected sample mean:

$$\hat{\boldsymbol{v}}_{back} = \boldsymbol{P}\left(\frac{1}{L}\sum_{j=1}^{L}\boldsymbol{P}^\top\boldsymbol{v}_j + \boldsymbol{\eta}'\right),$$

ensures (ϵ, δ)-differential privacy, provided that

$$\frac{\sigma_1}{\sigma_2} = \frac{\max_j \|\boldsymbol{v}_j\|_2}{\max_j \|\boldsymbol{P}^\top\boldsymbol{v}_j\|_2}.$$

Proof. We start by recalling that the Gaussian mechanism provides (ϵ, δ)-differential privacy when noise drawn from $\mathcal{N}(0, \sigma^2\mathbb{I})$ is added to a function $\mathcal{M}$, where the noise scale σ is proportional to the function's ℓ_2-sensitivity $\Delta_{\mathcal{M}}$.

The sensitivity of the sample mean function $\mathcal{M}_{\text{orig}}(\mathcal{V}) = \frac{1}{L}\sum_{i=1}^{L}\boldsymbol{v}_i$ is given by

$$\Delta_{\text{orig}} = \max_{\mathcal{V},\mathcal{V}'}\left\|\mathcal{M}_{\text{orig}}(\mathcal{V}) - \mathcal{M}_{\text{orig}}(\mathcal{V}')\right\|_2,$$

where $\mathcal{V}$ and $\mathcal{V}'$ differ in at most one element. The maximum change occurs when one sample is replaced, yielding

$$\Delta_{\text{orig}} = \frac{1}{L}\max_i \|\boldsymbol{v}_i\|_2.$$

Similarly, for the projected mean function $\mathcal{M}_{\text{proj}}(\mathcal{V}) = \frac{1}{L}\sum_{i=1}^{L}\boldsymbol{P}^\top\boldsymbol{v}_i$, the sensitivity is

$$\Delta_{\text{proj}} = \frac{1}{L}\max_i \left\|\boldsymbol{P}^\top\boldsymbol{v}_i\right\|_2.$$

The Gaussian mechanism requires the noise scale σ to be proportional to the sensitivity. Therefore, the ratio of the noise scales should match the ratio of sensitivities:

$$\frac{\sigma_1}{\sigma_2} = \frac{\Delta_{\text{orig}}}{\Delta_{\text{proj}}} = \frac{\max_i \|\boldsymbol{v}_i\|_2}{\max_i \left\|\boldsymbol{P}^\top\boldsymbol{v}_i\right\|_2}.$$

□

Theorem 1. *Under the same budget of differential privacy $(\epsilon, \delta$, the difference of MSE with and without projection $\boldsymbol{P}$ in estimating the true mean $\boldsymbol{\mu}$ can be decomposed into the three terms:*

$$\begin{aligned} MSE_{orig} - MSE_{back} = {} & \underbrace{\frac{1}{L}Tr\left((I - \boldsymbol{P}\boldsymbol{P}^\top)\Sigma_r\right)}_{\text{Projection Residual}} \\ & \underbrace{+\, \sigma_{proj}^2\left(\frac{\|\boldsymbol{v}_j\|_2^2}{\|\boldsymbol{P}^\top\boldsymbol{v}_j\|_2^2}D - d\right)}_{\text{Dimensional Reduction Effect}} \\ & \underbrace{-\, \|(I - \boldsymbol{P}\boldsymbol{P}^\top)\boldsymbol{\mu}\|_2^2 + \frac{1}{L}Tr\left((I - \boldsymbol{P}\boldsymbol{P}^\top)\Sigma_p\right)}_{\text{Projection Error}}. \end{aligned}$$

Proof. We analyze the MSE in both the original and projected spaces to establish the theorem.

First, consider the noisy mean in the original space:

$$\hat{\boldsymbol{\mu}}_{\text{orig}} = \boldsymbol{\mu} + \boldsymbol{\omega} + \boldsymbol{\eta}_{\text{orig}},$$

where $\boldsymbol{\omega} = \frac{1}{L}\sum_j(\boldsymbol{p}_j + \boldsymbol{r}_j)$ represents the sampling deviation from the true mean due to finite sample size and inherent data variability.

The MSE in the original space is then:

$$\text{MSE}_{\text{orig}} = \mathbb{E}\left[\|\hat{\boldsymbol{\mu}}_{\text{orig}} - \boldsymbol{\mu}\|_2^2\right] = \mathbb{E}\left[\|\boldsymbol{\omega} + \boldsymbol{\eta}_{\text{orig}}\|_2^2\right].$$

Expanding the squared norm, we obtain:

$$\text{MSE}_{\text{orig}} = \mathbb{E}\left[\|\boldsymbol{\omega}\|_2^2\right] + \mathbb{E}\left[\|\boldsymbol{\eta}_{\text{orig}}\|_2^2\right] + 2\mathbb{E}\left[\boldsymbol{\omega}^\top\boldsymbol{\eta}_{\text{orig}}\right].$$

Since $\boldsymbol{\omega}$ and $\boldsymbol{\eta}_{\text{orig}}$ are independent and both have zero mean, the cross term vanishes:

$$\mathbb{E}\left[\boldsymbol{\omega}^\top\boldsymbol{\eta}_{\text{orig}}\right] = 0.$$

Thus, the MSE in the original space simplifies to:

$$\text{MSE}_{\text{orig}} = \mathbb{E}\left[\|\boldsymbol{\omega}\|_2^2\right] + \mathbb{E}\left[\|\boldsymbol{\eta}_{\text{orig}}\|_2^2\right].$$

Next, consider the noisy mean in the projected space:

$$\hat{\boldsymbol{\mu}}_{\text{proj}} = \boldsymbol{P}^\top(\boldsymbol{\mu} + \boldsymbol{\omega}) + \boldsymbol{\eta}_{\text{proj}},$$

and the reconstructed noisy mean in the original space:

$$\hat{\boldsymbol{\mu}}_{\text{back}} = \boldsymbol{P}\hat{\boldsymbol{\mu}}_{\text{proj}} = \boldsymbol{P}\boldsymbol{P}^\top(\boldsymbol{\mu} + \boldsymbol{\omega}) + \boldsymbol{P}\boldsymbol{\eta}_{\text{proj}}.$$

We introduce an error term to account for the recover error from PCA transformation. Specifically, define:

$$\xi_{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{P}^\top\boldsymbol{\mu} - \boldsymbol{\mu},$$

which quantifies the deviation of the true mean $\boldsymbol{\mu}$ from its projection onto the subspace spanned by $\boldsymbol{P}$. If $\boldsymbol{P}$ perfectly captures the mean, then $\xi_{\boldsymbol{P}} = 0$. Otherwise, $\xi_{\boldsymbol{P}}$ represents the component of $\boldsymbol{\mu}$ orthogonal to the subspace spanned by $\boldsymbol{P}$.

Substituting this into the expression for $\hat{\boldsymbol{\mu}}_{\text{back}}$, we obtain:

$$\hat{\boldsymbol{\mu}}_{\text{back}} = \boldsymbol{\mu} + \boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega} + \boldsymbol{P}\boldsymbol{\eta}_{\text{proj}} + \xi_{\boldsymbol{P}}.$$

The MSE in the projected and reconstructed space is therefore:

$$\begin{aligned}
\text{MSE}_{\text{back}} &= \mathbb{E}\left[\|\hat{\boldsymbol{\mu}}_{\text{back}} - \boldsymbol{\mu}\|_2^2\right] \\
&= \mathbb{E}\left[\left\|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega} + \boldsymbol{P}\boldsymbol{\eta}_{\text{proj}} + \xi_{\boldsymbol{P}}\right\|_2^2\right] \\
&= \mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega}\|_2^2\right] + \mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}}\|_2^2\right] + \mathbb{E}\left[\|\xi_{\boldsymbol{P}}\|_2^2\right] \\
&+ 2\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega})^\top (\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}})\right] + 2\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega})^\top \xi_{\boldsymbol{P}}\right] \\
&+ 2\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}})^\top \xi_{\boldsymbol{P}}\right].
\end{aligned}$$

Given that $\boldsymbol{\omega}$, $\boldsymbol{\eta}_{\text{proj}}$, and $\xi_{\boldsymbol{P}}$ are all zero-mean and mutually independent, the cross terms vanish:

$$\begin{cases}
\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega})^\top (\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}})\right] = 0, \\
\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega})^\top \xi_{\boldsymbol{P}}\right] = 0, \\
\mathbb{E}\left[(\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}})^\top \xi_{\boldsymbol{P}}\right] = 0.
\end{cases}$$

Thus, the MSE in the projected and reconstructed space simplifies to:

$$\text{MSE}_{\text{back}} = \mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega}\|_2^2\right] + \mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}}\|_2^2\right] + \mathbb{E}\left[\|\xi_{\boldsymbol{P}}\|_2^2\right].$$

To evaluate these expectations, we consider the properties of covariance matrices. The covariance of $\boldsymbol{\omega}$ is:

$$\text{Cov}\left(\boldsymbol{\omega}\right) = \frac{1}{L}\left(\Sigma_p + \Sigma_r\right).$$

Thus, the first term becomes:

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\omega}\|_2^2\right] &= \text{Tr}\left(\boldsymbol{P}\boldsymbol{P}^\top \text{Cov}\left(\boldsymbol{\omega}\right)\right) \\
&= \frac{1}{L}\text{Tr}\left(\boldsymbol{P}\boldsymbol{P}^\top (\Sigma_p + \Sigma_r)\right).
\end{aligned}$$

For the second term, since $\boldsymbol{\eta}_{\text{proj}} \sim \mathcal{N}(0, \sigma_{\text{proj}}^2 I_d)$, we have:

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{P}\boldsymbol{\eta}_{\text{proj}}\|_2^2\right] &= \text{Tr}\left(\boldsymbol{P}^\top \boldsymbol{P}\mathbb{E}\left[\boldsymbol{\eta}_{\text{proj}}\boldsymbol{\eta}_{\text{proj}}^\top\right]\right) \\
&= \text{Tr}\left(\boldsymbol{P}^\top \boldsymbol{P}\sigma_{\text{proj}}^2 I_d\right) \\
&= \sigma_{\text{proj}}^2 \text{Tr}\left(\boldsymbol{P}^\top \boldsymbol{P}\right) \\
&= \sigma_{\text{proj}}^2 d.
\end{aligned}$$

The third term, $\mathbb{E}\left[\|\xi_{\boldsymbol{P}}\|_2^2\right]$, quantifies the error between the mean estimated in the subspace and its projection back to the original space compared to the true mean:

$$\mathbb{E}\left[\|\xi_{\boldsymbol{P}}\|_2^2\right] = \|\xi_{\boldsymbol{P}}\|_2^2 = \|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\mu} - \boldsymbol{\mu}\|_2^2.$$

Therefore, the MSE in the projected and reconstructed space is:

$$\text{MSE}_{\text{back}} = \frac{1}{L}\text{Tr}\left(\boldsymbol{P}\boldsymbol{P}^\top (\Sigma_p + \Sigma_r)\right) + \sigma_{\text{proj}}^2 d + \|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\mu} - \boldsymbol{\mu}\|_2^2.$$

Comparing this with the MSE in the original space:

$$\text{MSE}_{\text{orig}} = \frac{1}{L}\text{Tr}\left(\Sigma_p + \Sigma_r\right) + \sigma_{\text{orig}}^2 D,$$

we define the difference Δ as:

$$\begin{aligned}
\Delta &= \text{MSE}_{\text{orig}} - \text{MSE}_{\text{back}} \\
&= \frac{1}{L}\text{Tr}\left(\Sigma_p + \Sigma_r\right) + \sigma_{\text{orig}}^2 D \\
&- \left(\frac{1}{L}\text{Tr}\left(\boldsymbol{P}\boldsymbol{P}^\top (\Sigma_p + \Sigma_r)\right) + \sigma_{\text{proj}}^2 d + \|\boldsymbol{P}\boldsymbol{P}^\top \boldsymbol{\mu} - \boldsymbol{\mu}\|_2^2\right).
\end{aligned}$$

Simplifying the trace terms, we observe that:

$$\begin{aligned}
&\text{Tr}\left(\Sigma_p + \Sigma_r\right) - \text{Tr}\left(\boldsymbol{P}\boldsymbol{P}^\top (\Sigma_p + \Sigma_r)\right) \\
&\qquad = \text{Tr}\left((I - \boldsymbol{P}\boldsymbol{P}^\top)(\Sigma_p + \Sigma_r)\right).
\end{aligned}$$

According to Lemma 1, we have:

$$\sigma_{\text{orig}}^2 D - \sigma_{\text{proj}}^2 d = \sigma_{\text{proj}}^2 \left(\frac{\max_j \|\boldsymbol{v}_j\|_2^2}{\max_j \|\boldsymbol{P}^\top \boldsymbol{v}_j\|_2^2} D - d\right).$$

Substituting above into the expression for $\boldsymbol{\omega}$, we obtain:

$$\begin{aligned}
\Delta = &\underbrace{\frac{1}{L}\text{Tr}\left((I - \boldsymbol{P}\boldsymbol{P}^\top)\Sigma_r\right)}_{\text{Projection Residual}} \\
&+ \underbrace{\sigma_{\text{proj}}^2 \left(\frac{\max_j \|\boldsymbol{v}_j\|_2^2}{\max_j \|\boldsymbol{P}^\top \boldsymbol{v}_j\|_2^2} D - d\right)}_{\text{Dimensional Reduction Effect}} \\
&\underbrace{- \|(I - \boldsymbol{P}\boldsymbol{P}^\top)\boldsymbol{\mu}\|_2^2 + \frac{1}{L}\text{Tr}\left((I - \boldsymbol{P}\boldsymbol{P}^\top)\Sigma_p\right)}_{\text{Projection Error}}.
\end{aligned}$$

□

Theorem 2. *The process of distilling the private dataset $\mathcal{D}$ with an (ϵ_1, δ_1)-DP mechanism, supported by SER with an auxiliary dataset $\mathcal{D}_{aux}$ satisfying $(\epsilon_2, \delta 2)$-DP to $\mathcal{D}$, achieves $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP to $\mathcal{D}$.*

Proof. To prove Theorem 2, we utilize fundamental properties of differential privacy, specifically the *Basic Composition Theorem* and the *Post-Processing Theorem.*

Lemma 2 (Basic Composition Theorem [10])**.** *If a randomized mechanism $\mathcal{M}_1$ satisfies (ϵ_1, δ_1)-DP and another randomized mechanism $\mathcal{M}_2$ satisfies (ϵ_2, δ_2)-DP, then the sequential composition of these mechanisms, defined as $\mathcal{M} = \mathcal{M}_2 \circ \mathcal{M}_1$, satisfies $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP.*

Lemma 3 (Post-Processing Theorem [10]). *Any data-independent transformation of the output of a differentially private mechanism does not degrade its privacy guarantees. Formally, if $\mathcal{M}$ satisfies (ϵ, δ)-DP, then for any deterministic or randomized function f, the mechanism $f \circ \mathcal{M}$ also satisfies (ϵ, δ)-DP.*

We define the two mechanisms involved in the process as follows.

Let $\mathcal{M}_1$ represent the mechanism responsible for SER. The input to $\mathcal{M}_1$ is the private dataset $\mathcal{D}$, and its output is the auxiliary dataset $\mathcal{D}_{\text{aux}}$. By assumption, $\mathcal{M}_1$ satisfies (ϵ_1, δ_1)-differential privacy with respect to $\mathcal{D}$.

Let $\mathcal{M}_2$ represent the mechanism responsible for the distillation process. The inputs to $\mathcal{M}_2$ are the private dataset $\mathcal{D}$ and the auxiliary dataset $\mathcal{D}_{\text{aux}}$, and its output is the distilled dataset $\mathcal{Z}$. By assumption, $\mathcal{M}_2$ satisfies (ϵ_2, δ_2)-differential privacy with respect to $\mathcal{D}$.

It is important to note that $\mathcal{M}_2$ utilizes $\mathcal{D}_{\text{aux}}$, which is already the output of $\mathcal{M}_1$. However, since $\mathcal{M}_1$ ensures that $\mathcal{D}_{\text{aux}}$ is (ϵ_1, δ_1)-DP with respect to $\mathcal{D}$, any further processing of $\mathcal{D}_{\text{aux}}$ by $\mathcal{M}_2$ is considered post-processing of a DP-protected output.

Applying Lemma 3, the usage of $\mathcal{D}_{\text{aux}}$ by $\mathcal{M}_2$ does not introduce any additional privacy loss beyond what is already accounted for by $\mathcal{M}_1$. Therefore, $\mathcal{M}_2$ maintains its (ϵ_2, δ_2)-DP guarantee with respect to $\mathcal{D}$ independently of $\mathcal{D}_{\text{aux}}$.

Since $\mathcal{M}_1$ and $\mathcal{M}_2$ are applied sequentially, we apply Lemma 3. The cumulative privacy loss incurred by applying both mechanisms in sequence is the sum of their individual privacy parameters.

Formally, the overall mechanism $\mathcal{M}$, defined as:

$$\mathcal{M} = \mathcal{M}_2 \circ \mathcal{M}_1$$

satisfies:

$$\mathcal{M} \text{ satisfies } (\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)\text{-DP}.$$

By sequentially applying $\mathcal{M}_1$ and $\mathcal{M}_2$, and leveraging both the Basic Composition and Post-Processing Theorems, we conclude that the combined process satisfies $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-DP with respect to the private dataset $\mathcal{D}$. □

B. Additional Quantitative Analysis

B.1. Effect of Privacy-Budget Split

Table 3 shows how allocating the total budget (ϵ=1.0) between auxiliary data generation (ϵ_1) and DP-based optimation (ϵ_2) affects downstream accuracy. We observe that allocating ϵ_1:ϵ_2= 0.8:0.2 offers a good trade-off.

(ϵ_1, ϵ_2)	**(0.9, 0.1)**		**(0.8, 0.2)**		**(0.7, 0.3)**		**(0.6, 0.4)**		**(0.5, 0.5)**	
IPC	10	50	10	50	10	50	10	50	10	50
MNIST	96.3	96.5	96.4	96.7	95.9	96.1	94.9	95.2	93.2	94.5
FashionMNIST	80.2	82.9	80.1	83.1	79.7	82.4	78.8	80.8	76.2	79.4

Table 3. Accuracy (%) under different privacy-budget splits ϵ_1+ϵ_2=1.0, fixing $\delta_1 = \delta_2 = 5 \times 10^6$. Results show that allocating ϵ_1:ϵ_2= 0.8:0.2 offers a overall good trade-off.

B.2. SER Performance Across Varying Noise Levels & Subspace Dimensions

Figure 5 details how the mean squared error (MSE) of mean estimation evolves on MNIST when varying both the *noise multiplier* and the number of retained *subspace dimensions* (horizontal axis in each subplot). Solid curves denote our method with SER (*w/ SER*); dashed curves are the vanilla DP baseline (*w/o SER*). A clear pattern, consistent with the residual decomposition in Theorem 1, emerges:

Low-noise regime (noise multiplier $\lesssim 0.4 \times 10^{-3}$). Here, the DP noise injected per coordinate is small, so the total error is dominated by the *projection error* introduced by compressing and reconstructing the data. In this regime, SER can even *increase* MSE if the bottleneck is too tight; the loss of information outweighs the modest noise reduction. Consequently, retaining more subspace dimensions monotonically lowers the error, and the gap between "w/" and "w/o" SER narrows.

High-noise regime (noise multiplier $\gtrsim 0.9 \times 10^{-3}$). When the privacy budget is tight, the additive Gaussian noise dominates. Dimensionality reduction now acts as a signal-to-noise enhancer: a lower-rank subspace filters out much of the high-dimensional noise before reconstruction. As a result, SER yields a pronounced MSE drop relative to the baseline, particularly when only a few hundred components are kept. Beyond this point, adding more dimensions simply reintroduces noise and the benefit diminishes.

Intermediate-noise regime ($\sim 0.5 \times 10^{-3}$ to 0.8×10^{-3}). At moderate noise levels, the two error sources balance each other. The MSE curves adopt a classic U-shape, indicative of a trade-off: MSE first decreases as noise is tamed by projection, reaches a minimum at an *optimal* dimensionality (typically 300–800 components), then increases again as projection bias begins to dominate. This turning point aligns with the crossover predicted by the dimensional-reduction effect term in Theorem 1.

Together, these three regimes offer actionable insight into how SER should be tuned in practice:

- When privacy is **loose**, favor a larger subspace or skip SER entirely.
- When privacy is **tight**, reduce dimensionality aggressively to suppress noise.
- For **intermediate** privacy budgets, select the number of

subspace dimensions that minimizes MSE.

We apply this same strategy to FashionMNIST and CIFAR-10 (Figures 6 and 7), and observe analogous trends.

C. Qualitative Results

In Fig. 8, we present distilled samples from the CIFAR-10, FashionMNIST, and MNIST datasets. Each row corresponds to a distinct class, with all samples generated using an IPC of 10 and a privacy budget of $(1, 10^{-5})$.

D. Settings for Generating Auxiliary Datasets

D.1. Auxiliary Data Generation with Stable Diffusion (SD) [22]

For the CIFAR-10 dataset, we generate auxiliary images using Stable Diffusion version 1.4 (SD-v1-4). The generation process employs the following prompt for each category:

“A photo of a $\{$category$\}$”.

SD-v1-4 was trained on LAION-5B [25], a dataset that contains no information related to CIFAR-10. Therefore, using it to train CIFAR-10 is not considered a privacy leakage. Representative image samples are illustrated in Fig. 9a.

D.2. Auxiliary Data Generation using Differentially Private Diffusion Model

For MNIST and FashionMNIST we generate auxiliary images with the Differentially Private Diffusion Model (DPDM). Concretely, we train a Noise Conditional Score Network (NCSN++)[28] for 50 epochs using Adam [13] (no weight decay), a batch size of 64, and a learning rate of 3×10^{-4}. The trained network is then sampled with a deterministic DDIM sampler [27] for 500 inference steps, ensuring the entire procedure conforms to the prescribed differential-privacy budget. We sample random images from the auxiliary dataset in Fig. 9b and Fig. 9c.

D.3. Other Models as Auxiliary Data Generator

Beyond SD and DPDM, we evaluate DP-Diffusion and DP-LDM as alternative generative models for producing auxiliary datasets under differential privacy constraints. In this experiment, we generate synthetic data for MNIST, FashionMNIST, and CIFAR-10 using each model while maintaining a fixed privacy budget $(1, 10^{-5})$. To assess the impact of dataset size, we vary the number of images per class (IPC) between 10 and 50. The generated datasets are then used to train downstream models, following the same evaluation protocol as in previous experiments. The results are shown in Appendix D.3.

Table 4. Comparison of DP-based generative models for SER.

	DPDM		DP-Diffusion		DP-LDM	
Dataset	IPC=10	IPC=50	IPC=10	IPC=50	IPC=10	IPC=50
MNIST	**96.4**	96.7	96.3	96.7	96.3	**96.8**
FashionMNIST	80.1	83.1	80.5	83.0	**80.8**	**83.4**
CIFAR-10	47.8	**51.5**	47.5	51.0	**48.2**	51.2

D.4. Controlling for the Impact of Extra Information in DP-Based Generative Models

To isolate the effect of additional information introduced by different generative models, we compare our method with DPDM, DP-Diffusion, and DP-LDM under the same privacy budget of $(1, 10^{-5})$. This comparison helps determine whether simply using these models for downstream training provides sufficient utility or if our method introduces meaningful improvements beyond what these baselines achieve. The results are shown in Appendix D.4.

Table 5. DP-based enerative models as baselines.

	Dosser		DPDM		DP-Diffusion		DP-LDM	
Dataset	IPC=10	IPC=50	IPC=10	IPC=50	IPC=10	IPC=50	IPC=10	IPC=50
MNIST	**96.4**	**96.7**	69.1	70.5	72.3	74.8	71.5	73.6
FashionMNIST	**80.1**	**83.1**	59.7	63.6	60.2	65.1	61.1	64.9
CIFAR-10	**50.6**	**52.3**	10.0	9.9	10.0	10.0	10.0	10.2

E. Discussion

E.1. Why Not DP-PCA for SER?

One may argue that, instead of learning a fixed projection from auxiliary data, we could simply run DP–PCA at every iteration to discover the informative subspace on-the-fly. However, because the extractor is randomly re-initialized each iteration, its output lies in a fresh feature space. Performing DP-PCA on *every* feature batch would therefore require an independent DP query each time. With a total budget (ϵ, δ) split across I iterations, each PCA call receives only ϵ/I privacy, which completely drowns the signal and devastates downstream accuracy.

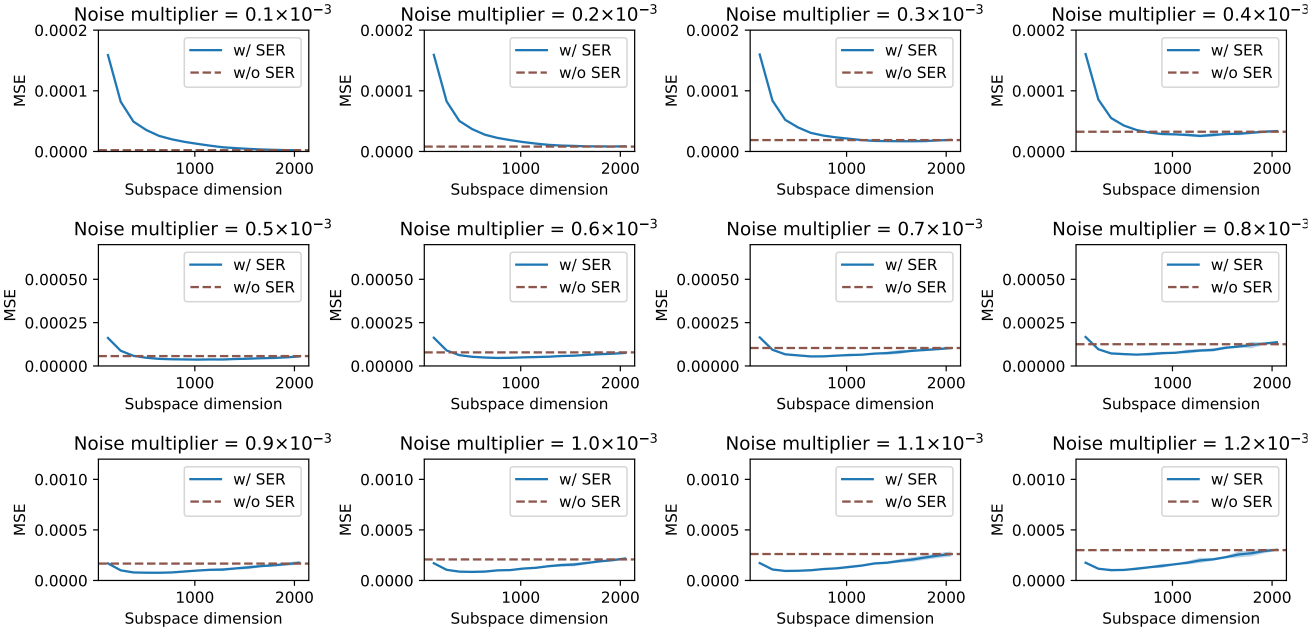


Figure 5. **MNIST: MSE of mean estimation as a function of retained subspace dimensions and noise multiplier.** Each subplot corresponds to a different *noise multiplier* (privacy level). The horizontal axis shows the number of retained subspace dimensions; the vertical axis shows mean-squared error (MSE). Solid curves are our method with *SER*; dashed curves are the vanilla DP baseline.

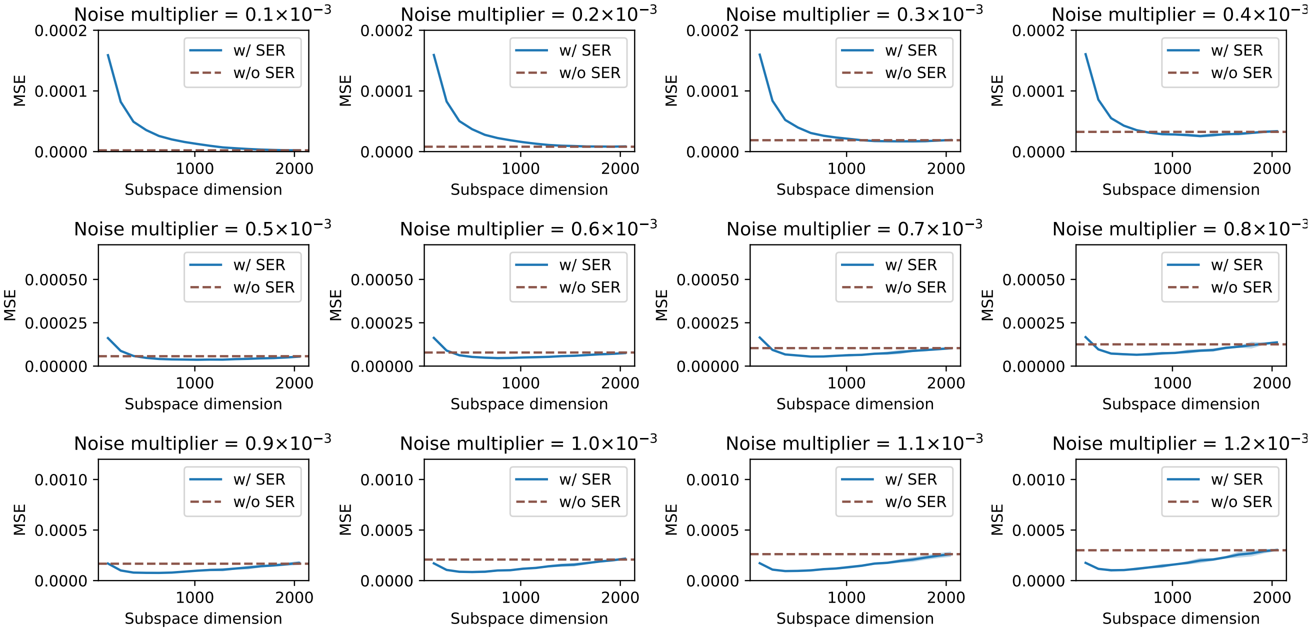


Figure 6. **FashionMNIST: MSE of mean estimation versus subspace dimension and noise multiplier.** Plot settings match Fig. 5. FashionMNIST exhibits the same qualitative behavior: SER offers little benefit in the low-noise regime, achieves a clear optimum in the intermediate regime (300–800 components), and substantially reduces MSE under tight privacy budgets (high noise multipliers).

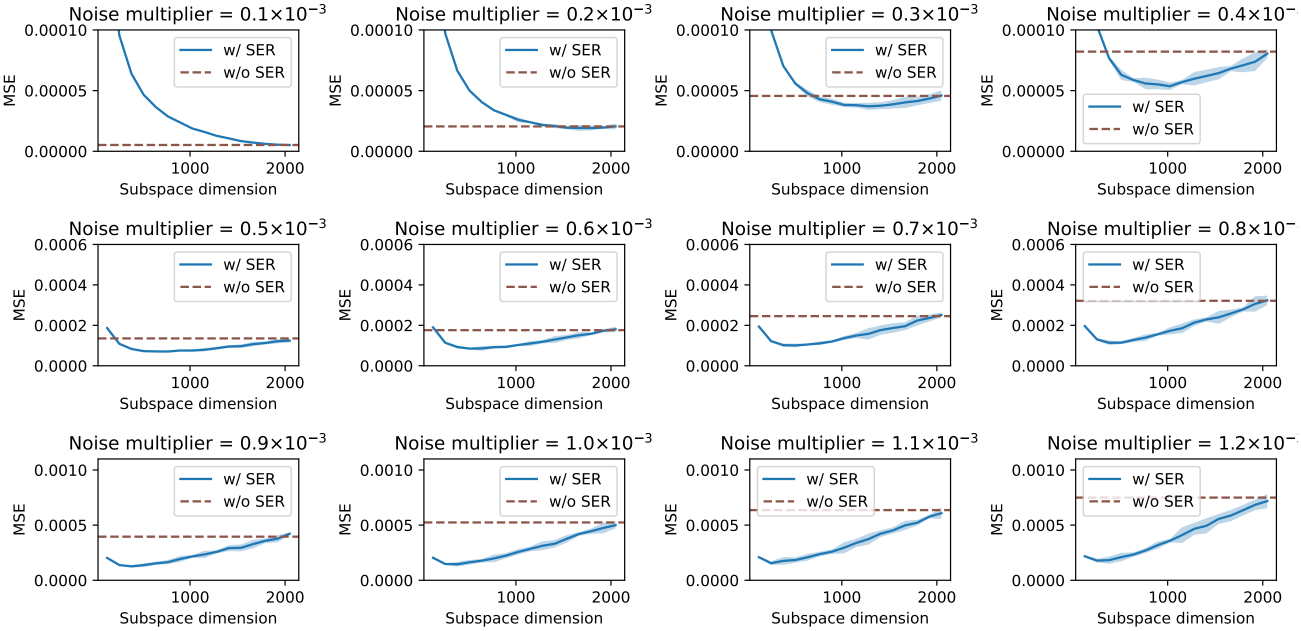


Figure 7. **CIFAR-10: MSE of mean estimation versus subspace dimension and noise multiplier.** Despite the higher input dimensionality of CIFAR-10, the same trends appear: SER markedly lowers the MSE when privacy is tight (high noise), has diminishing returns as more subspace dimensions are added, and converges to the baseline when privacy is loose.

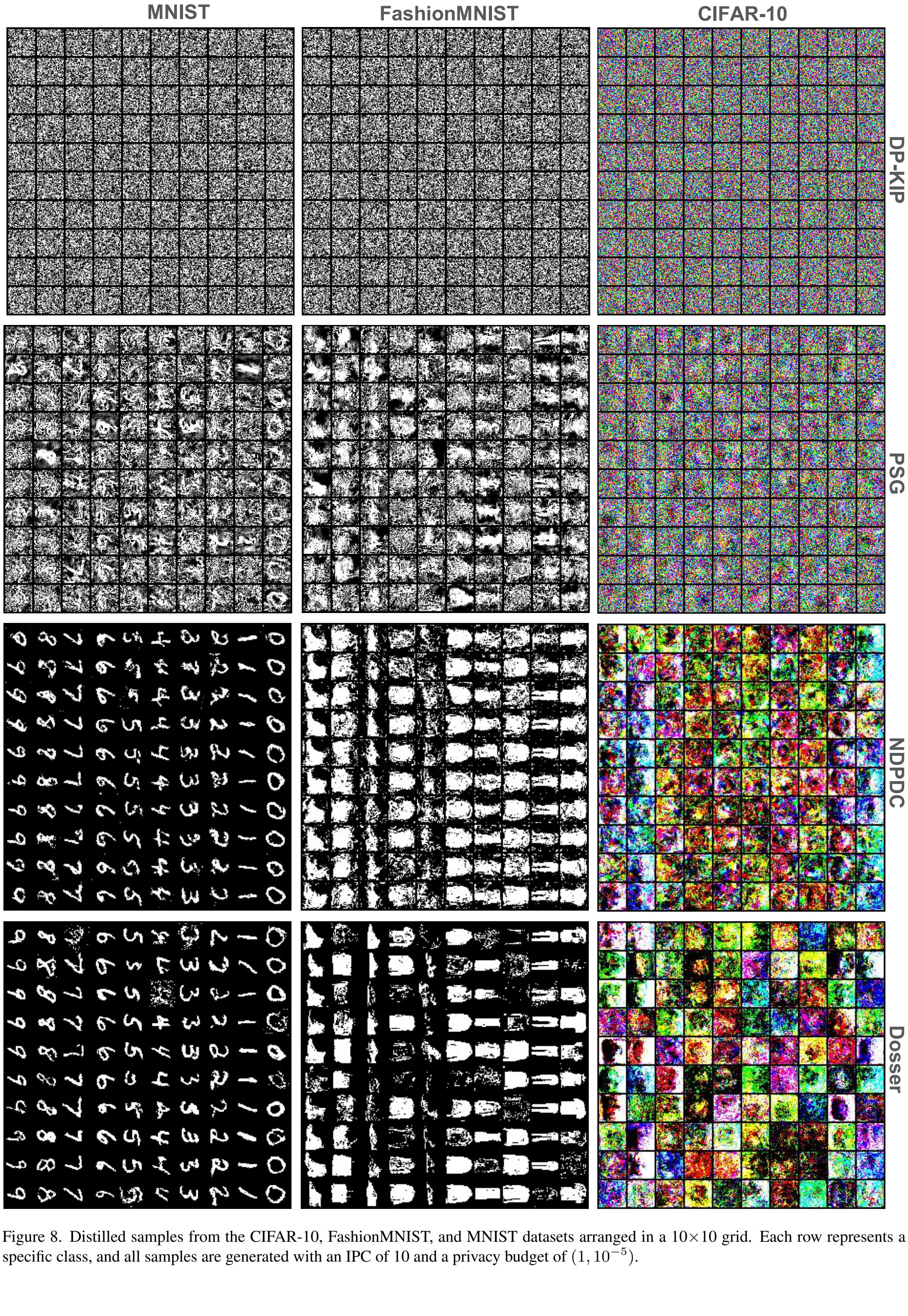
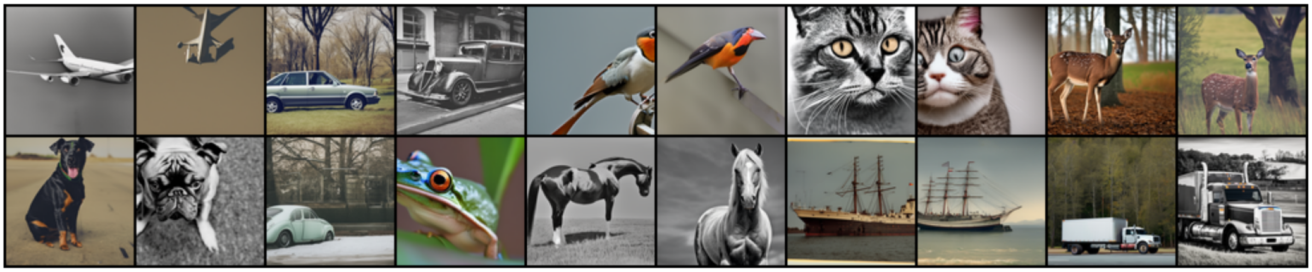
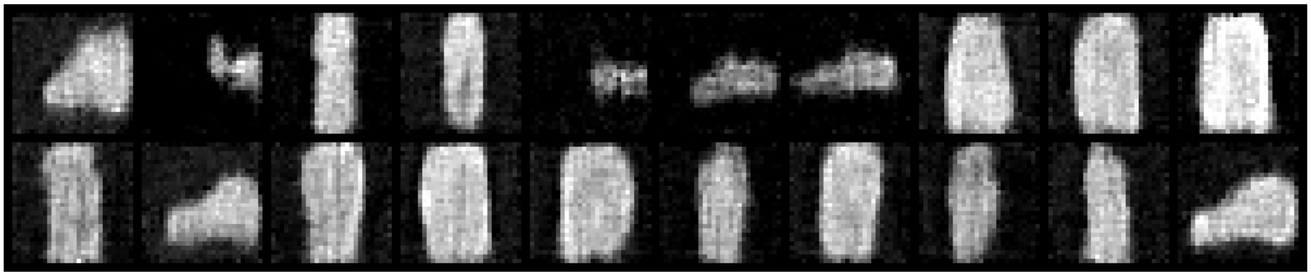


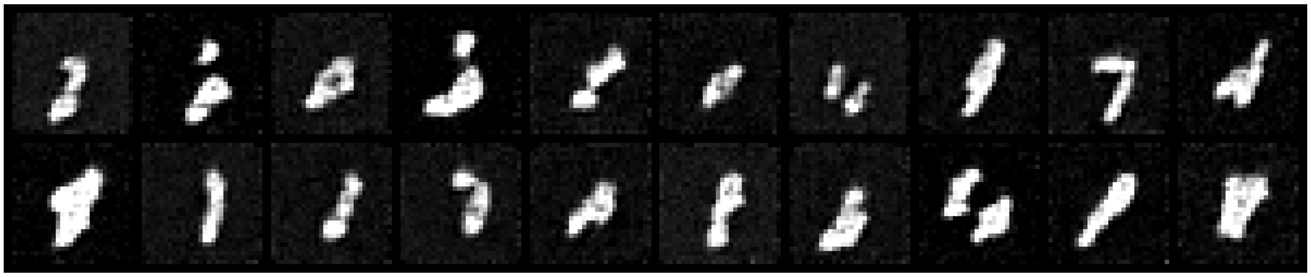
Figure 8. Distilled samples from the CIFAR-10, FashionMNIST, and MNIST datasets arranged in a 10×10 grid. Each row represents a specific class, and all samples are generated with an IPC of 10 and a privacy budget of $(1, 10^{-5})$.



(a) Sampled images from the CIFAR-10 auxiliary dataset.



(b) Sampled images from the FashionMNIST auxiliary dataset.



(c) Sampled images from the MNIST auxiliary dataset.

Figure 9. Sample auxiliary datasets.