PRIVATE EVOLUTION FOR IPM OPTIMIZATION: LIMITS AND ALTERNATIVES

Anonymous authors
Paper under double-blind review

ABSTRACT

Private Evolution (PE) is a framework for differentially private (DP) synthetic data generation. Given a sensitive dataset S and a public dataset S_0, PE *evolves* by generating public variations of S_t and privately selecting S_{t+1} from them to reduce a target discrepancy to S. Prior work proves that PE converges to S in 1-Wasserstein (W_1) distance, but generating DP synthetic data with W_1 accuracy suffers from the curse of dimensionality, even outside the evolution framework. The W_1 distance is an integral probability metric (IPM), defined by the largest discrepancy over a class of queries. It is known that more structured query classes can produce DP synthetic data that converge faster than existing lower bounds on W_1 convergence. We ask whether PE can exploit this structure by optimizing IPMs other than W_1. Our first result is that even simple IPMs induced by linear or convex-quadratic queries can be difficult to optimize directly via evolution. Even without privacy, exponentially many variations may be needed to generate a candidate dataset with strictly smaller target IPM. For linear queries, we recover statistically- and computationally-efficient PE by replacing the target IPM with a surrogate distance that bounds it, and for which polynomial-size Gaussian variation pools decrease the surrogate in expectation. We then use a DP Frank–Wolfe subroutine within PE to efficiently optimize a distribution over the variations and construct S_{t+1} to match the mean of it. The resulting PE achieves IPM error $\widetilde{O}((d^2/(n\varepsilon))^{1/3})$ and avoids the curse of dimensionality. This works in part because a dataset with a prescribed mean is easy to construct; for richer queries, reconstruction can itself be hard: we prove that matching only two bounded convex quadratic queries with a fixed-size dataset is already NP-hard. Overall, our results disentangle statistical and computational aspects of the three pieces of PE: the IPM, the evolution, and privacy. Our lower bounds show how non-private evolution fails (computationally), even for simple IPMs, while our positive result shows that replacing the target IPM by a compatible surrogate can recover efficient PE in some special cases.

1 INTRODUCTION

Synthetic data has been proposed as a proxy for sensitive datasets in downstream tasks like fine-tuning models, running analytics, and debugging implementation pipelines. The goal of *differentially private* (DP) synthetic data is to release an artificial dataset S' that preserves useful properties of a sensitive dataset S while protecting the privacy of the individuals represented in it.

For a synthetic dataset to be useful, it should preserve statistical properties of the sensitive data. This preservation is often formalized through a query class $\mathcal{F}$: given a sensitive dataset S and a synthetic dataset S', define discrepancy $D_{\mathcal{F}}(S, S') := \sup_{f\in\mathcal{F}} |f(S) - f(S')|$ where $f(S) := \frac{1}{|S|}\sum_{z\in S} f(z)$. These discrepancies are Integral Probability Metrics (IPMs) between the empirical measures of S and S' (Müller, 1997), with different choices of $\mathcal{F}$ encoding different notions of utility. If $\mathcal{F} = \mathcal{F}_{\mathrm{Lip}}$ is the class of 1-Lipschitz functions, $D_{\mathcal{F}}$ is the 1-Wasserstein (Villani, 2008); if $\mathcal{F}$ is the unit ball of a reproducing kernel Hilbert space, $D_{\mathcal{F}}$ is the maximum mean discrepancy (Gretton et al., 2012).

The choice of $\mathcal{F}$ can dramatically affect the statistical difficulty of learning a DP synthetic dataset. For example, Boedihardjo et al. (2024) shows that when $\mathcal{F} = \mathcal{F}_{\mathrm{Lip}}$ and the domain is $[0,1]^d$, every ε-DP algorithm that outputs a synthetic dataset S' approximating a sensitive dataset S of size n must incur error $D_{\mathcal{F}_{\mathrm{Lip}}}(S, S') = \Omega((n\varepsilon)^{-1/d})$, which depends unfavorably on problem dimension

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

d. In contrast, several works show that IPMs induced by more structured query classes can achieve substantially more favorable statistical rates. For example, for s-sparse Lipschitz queries, Donhauser et al. (2024) obtain a rate $\tilde{O}_s(\binom{d}{s}/(n\varepsilon))^{1/s})$, while for k-smooth queries, Ding et al. (2026) obtain the minimax rate $\tilde{O}_{d,k}((n\varepsilon)^{-\min\{1,k/d\}})$, improving on Wang et al. (2016). These results show that restricting the query class can remove, or substantially weaken, the statistical curse of dimensionality.

Private Evolution (PE) is a popular DP synthetic data framework that iteratively constructs a synthetic dataset with the aid of a public API (e.g. a pretrained generative model) that can generate *variations* of data points (Lin et al., 2023; Xie et al., 2024; Lin et al., 2025; Hou et al., 2024; Swanberg et al., 2025; Zou et al., 2025; Hou et al., 2025). Specifically, given a sensitive dataset S and a public synthetic dataset S_0, PE evolves by using this API to generate variations of S_t, then privately selecting S_{t+1} from the set of variations. This selection is crafted to reduce a target discrepancy to S. Since PE only queries a public API, it is training-free, which has contributed to its growing popularity and adoption in industry (Afonja et al., 2024; Apple, 2025).

To date, existing theory for PE is tied to Wasserstein distance. González Lara et al. (2026) improve a result from Lin et al. (2023), showing that PE can achieve a worst-case rate of $\mathbb{E}[W_1(S, S')] = \tilde{O}\big(d(n\varepsilon)^{-1/d}\big)$, which exhibits the usual Wasserstein curse of dimensionality. Our main question is:

Can PE be made to optimize IPMs beyond 1*-Wasserstein? If so, do the resulting convergence rates avoid the curse of dimensionality?*

A natural adaptation of evolution (both private and non-private) is *direct evolution:* when selecting the next dataset from the variations of S_t, choose a subset S_{t+1} that minimizes the target IPM. Our first main result shows that direct evolution can fail even without privacy: for some simple IPMs, exponentially many variations may be needed before they contain a next iterate S_{t+1} that decreases the target discrepancy. Theorem 3.1 establishes this for linear and convex-quadratic IPMs and a broad class of variation APIs, including the Gaussian one studied in (González Lara et al., 2026). The reason is that the set of variations that improve the IPM can have very small probability under the variation mechanism. If the variation mechanism (e.g., Gaussian) has a small probability of generating samples in that improvement region, then (exponentially) many variations may be required to improve the overall objective. This could also occur if the improvement region is large, but the variation mechanism generates from it with low probability.

A reasonable alternative to direct evolution is to replace the IPM by a surrogate objective that can be efficiently optimized for the given class of variations The challenge is to ensure that the surrogate upper bounds the IPM uniformly, so the error bounds achieved by evolving the surrogate can transfer back to the IPM of interest. Nevertheless, we present a case where this idea succeeds. An IPM induced by linear queries can be hard for direct evolution, as shown in Theorem 3.1, while minimizing a surrogate of the IPM with Gaussian variations converges at a rate of $\tilde{O}\left(\frac{d^{2/3}\log(1/\delta)^{1/6}}{(n\varepsilon)^{1/3}}\right)$ under (ε, δ)-DP (Theorem 4.2). The main insight is that the linear IPM can be upper bounded between the difference of the means of the sensitive and the synthetic datasets, and in order to decrease this difference we only need a variation to get closer to the mean of the synthetic dataset, which has constant probability under a symmetric additive noise. Another difference with direct evolution is that we use the variations to privately compute a certain condition that S_{t+1} needs to satisfy in order to improve the surrogate, but we do not require S_{t+1} itself to consist of generated variations.

More broadly, our results disentangle the roles of evolution and privacy: the lower bounds isolate failures of the evolution dynamics even without privacy, while the positive result shows how a compatible surrogate and a more flexible rule for constructing S_{t+1} can recover efficient private synthetic data generation.

1.1 Related work

Private synthetic data via evolution algorithms. To the best of our knowledge, the existing theoretical analyses of PE are Lin et al. (2023) and González Lara et al. (2026), and both measure accuracy through Wasserstein distances. The analysis of Lin et al. (2023) uses stronger assumptions on the variation mechanism, while González Lara et al. (2026) proves a $\widetilde{O}(d(n\varepsilon)^{-1/d})$ guarantee only under Gaussian variations. We relax both the discrepancy measure (to IPMs beyond W_1) and the mechanism for generating variations (to include methods that allow for projection onto a desired

domain, for instance). Empirically, PE has been applied with foundation-model APIs for image and text generation (Lin et al., 2023; Xie et al., 2024), in federated settings (Hou et al., 2024), for tabular data (Swanberg et al., 2025), with multiple public APIs (Zou et al., 2025), and with simulators in place of APIs (Lin et al., 2025). Related zeroth-order approaches include the genetic-algorithm of Liu et al. (2023).

DP synthetic data with IPM accuracy guarantees. A broad line of work studies DP synthetic data through accuracy guarantees for classes of statistical queries. Early results focused on linear or counting queries on discrete domains (Hardt et al., 2012; Blum et al., 2013; Dwork et al., 2015). On continuous domains, He et al. (2023); Boedihardjo et al. (2024) study synthetic-data generation under Wasserstein-type accuracy; in particular, 1-Wasserstein accuracy on $[0,1]^d$ has worst-case error scaling as $(n\varepsilon)^{-1/d}$. More structured IPMs can admit substantially faster rates. Balog et al. (2018) use private kernel mean embeddings, which naturally yield maximum mean discrepancy guarantees (Gretton et al., 2012); Donhauser et al. (2024) obtain rates for s-sparse Lipschitz queries scaling as $\tilde{O}_s(\binom{d}{s}/(n\varepsilon))^{1/s})$; and Ding et al. (2026); Wang et al. (2016) study smooth queries, where the former establishes a minimax rate of $\tilde{O}_{d,k}((n\varepsilon)^{-\min\{1,k/d\}})$ for k-smooth queries. Many of these methods follow an estimate-then-reconstruct approach: some statistics of the sensitive dataset are estimated privately, and a synthetic dataset is then constructed to approximately realize them. This reconstruction step can be computationally demanding; for example, the exact construction of Donhauser et al. (2024) is described as practically infeasible, while the runtime of Ding et al. (2026) grows rapidly with dimension and smoothness. Our work asks if these improved statistical rates can instead be realized through efficient evolution; our hardness result (Proposition 5.1) shows that reconstruction itself can present a worst-case computational barrier.

2 Notation and problem setup

We write $[n] := \{1, \dots, n\}$. The data domain is a convex compact set $\mathcal{X} \subset \mathbb{R}^d$ with diameter $D := \sup_{x,y\in\mathcal{X}} \|x - y\|_2$ and radius R, meaning $\mathcal{X} \subseteq B_2(0, R)$. $\Pi_{\mathcal{X}}$ is the Euclidean projection onto $\mathcal{X}$. A sensitive dataset is denoted by $S = \{x_1, \dots, x_n\} \in \mathcal{X}^n$, and a synthetic dataset by $S' = \{y_1, \dots, y_m\} \in \mathcal{X}^m$. For any dataset A, we write $m(A) := \frac{1}{|A|}\sum_{z\in A} z$ for its empirical mean. For a function $f : \mathcal{X} \to \mathbb{R}$, we write $f(A) := |A|^{-1}\sum_{z\in A} f(z)$ for its empirical average over A.

Definition 2.1 (Differential privacy Dwork et al. (2006))**.** *A randomized algorithm $\mathcal{A}$ is (ε, δ)-differentially private if, for every pair of neighboring datasets S, S' differing in at most one data point and every measurable event $\mathcal{E}$ in the output space,* $\Pr[\mathcal{A}(S) \in \mathcal{E}] \le e^{\varepsilon}\Pr[\mathcal{A}(S') \in \mathcal{E}] + \delta$.

Private synthetic data. Given a query class $\mathcal{F}$ of continuous functions $f : \mathcal{X} \to \mathbb{R}$, the goal is to generate an (ε, δ)-differentially private synthetic dataset $S' \subset \mathcal{X}$ with small discrepancy $D_{\mathcal{F}}(S, S') := \sup_{f\in\mathcal{F}} |f(S) - f(S')|$, where S is the sensitive dataset and S' the synthetic one.

Evolution algorithms. An evolution algorithm maintains a synthetic dataset $S_t = (y_1^t, \dots, y_m^t) \in \mathcal{X}^m$. A public variation mechanism provides a prescribed finite menu $Q_1, \dots, Q_L$ of additive noise laws on $\mathbb{R}^d$. At round t, the algorithm chooses law-scale pairs $\{(\ell_{t,r}, \sigma_{t,r})\}_{r=1}^{N_t}$, common to all points in S_t, and generates $\mathrm{Variations}(S_t) := S_t \cup \left\{y_j^t + \sigma_{t,r}\xi_{j,r}^t : j \in [m],\ r \in [N_t]\right\}, \qquad \xi_{j,r}^t \sim Q_{\ell_{t,r}}$, using fresh independent draws. The law-scale pairs may depend on the preceding history. Taking Gaussian laws and K draws at each of L scales recovers the variation model of González Lara et al. (2026); Lin et al. (2023). A generic update rule uses S and $\mathrm{Variations}(S_t)$ to construct $S_{t+1} \in \mathcal{X}^m$. An evolution algorithm is *private* if the resulting procedure is differentially private with respect to S.

In *direct evolution*, variations are projected onto $\mathcal{X}$, which we denote by $\Pi_{\mathcal{X}}(\mathrm{Variations}(S_t))$, and S_{t+1} is selected to minimize the target discrepancy $D_{\mathcal{F}}(S, \cdot)$ over $\Pi_{\mathcal{X}}(\mathrm{Variations}(S_t))^m$. Thus, direct evolution both optimizes the target IPM itself and requires the next dataset to consist of generated variations.

3 SIMPLER IPMS DO NOT IMPLY MORE EFFICIENT DIRECT EVOLUTION

In this section, we show that minimizing an IPM simpler than W_1 can be difficult via direct evolution, even without privacy. We construct simple IPMs induced by linear or convex quadratic queries for which direct evolution may require exponentially many variations before making any progress.

The obstruction is that a variation set containing an improving next dataset may be generated only with exponentially small probability. The following theorem formalizes this phenomenon and shows that a broad class of additive variation mechanisms, including the multiscale Gaussian variations used in previous PE work, can fail to generate variation sets that permit progress on some simple IPMs. Its proof can be found in Appendix A.1.

Theorem 3.1 (Linear and quadratic barriers to direct IPM evolution)**.** *Fix $d \geq 2$, $\mathcal{X} = [-1, 1]^d$, and a menu $Q_1, \ldots, Q_L$ of probability laws on $\mathbb{R}^d$, prescribed before the query class and target are chosen. In each of the following cases there exist a target $y \in \mathcal{X}$, a query class $\mathcal{F}$ with $f : \mathcal{X} \to [0, 1]$, and an initial point $x_0 \in \mathcal{X}$, such that $\Phi(x) := D_{\mathcal{F}}(\{y\}, \{x\})$ and $\min_{x \in \mathcal{X}} \Phi(x) = \Phi(y) = 0$.*

Consider any direct IPM evolution method starting at x_0 and using at most N proposals in total, with arbitrary tie-breaking, adaptive choices from the menu, and arbitrary positive scales.

1. ***Linear queries.*** *The class contains d affine queries (equivalently, linear queries for the IPM). The objective Φ is convex, $\Phi(x_0) = 1/6$, and $\Pr(\text{the method ever leaves } x_0) \leq NL\, 2^{-d}$.*
2. ***Three quadratic queries.*** *The class contains three convex quadratic queries. The point x_0 is a strict local minimum, $\Phi(x_0) = 3/67$, and $\Pr(\text{the method ever leaves } x_0) \leq NLd\, e^{-d/18}$.*

Thus, if L and N scaly polynomially with d, the method retains a constant optimality gap with probability $1 - \mathrm{poly}(d)e^{-\Omega(d)}$.

Theorem 3.1 shows that the difficulty of optimizing certain IPMs by direct evolution can lie in generating useful variations, rather than in private selection. Importantly, however, this is not an intrinsic failure of the variation API itself. Rather, it reflects a compatibility issue between the variation API and the IPM being optimized: datasets that improve a particular IPM may have very low probability under a given variation mechanism. For example, the Gaussian variation model of González Lara et al. (2026) is covered by the lower bound presented above, so there are simple IPMs and initial points for which Gaussian variations are exponentially unlikely to generate an improvement. Yet González Lara et al. (2026) show that the same variation API can generate variation pools that improve W_1 in expectation. The following proposition shows that this discrepancy dependence can arise even deterministically: the same fixed variation pool can allow improvement for one IPM and not another.

Proposition 3.1 (Compatibility depends on the target IPM)**.** *There exist a target dataset S, an initial dataset S_0, and a variation pool V such that some size-two dataset S_1 supported on V strictly improves $W_1(S, S_1) < W_1(S, S_0)$, while no size-two dataset supported on V improves an IPM induced by a single affine query.*

Proposition 3.1 shows that the usefulness of a variation mechanism cannot be considered independently of the discrepancy being optimized. In particular, a variation pool can contain a descent direction for W_1 while containing no descent direction at all for a simpler IPM. We refer to this interaction as *compatibility* between the variation mechanism and the evolution objective.

Together, Theorem 3.1 and proposition 3.1 show that making the target IPM statistically or computationally simpler is not, by itself, sufficient to make direct evolution efficient. The evolution objective must also be compatible with the variation mechanism. This observation motivates the alternative developed in the next section: rather than evolving directly with respect to the target IPM, we seek a surrogate objective that remains useful for the target task while being easier for the variation mechanism to optimize.

4 A PRIVATE EVOLUTION ALGORITHM FOR LINEAR IPMS

Theorem 3.1 shows that direct evolution can be inefficient even for a convex IPM induced by linear queries. We therefore seek a surrogate that is more compatible with variation geometry, bounds the

Algorithm 1 Mean-matching evolution

Require: Target dataset S, initial synthetic dataset $S_0 \in \mathcal{X}^m$, iterations T, Gaussian scales $\sigma_1, \ldots, \sigma_L$, and K draws per scale and point.

1: **for** $t = 0, \ldots, T-1$ **do**
2: $V_{t+1} \leftarrow \text{Variations}(S_t)$.
3: Choose $\widetilde{S}_{t+1} = (\widetilde{y}_1^{t+1}, \ldots, \widetilde{y}_m^{t+1}) \in \arg\min_{S' \in V_{t+1}^m} H(S')$.
4: Choose any $S_{t+1} \in \mathcal{X}^m$ such that $m(S_{t+1}) = \Pi_{\mathcal{X}}(m(\tilde{S}_{t+1}))$.
5: **end for**
6: **return** S_T.

IPM, and has the same optimum as the original problem. We formally explore this idea in the simple case of linear queries. If $f_a(x) = \langle a, x \rangle$, then for any $S, S' \subset \mathcal{X}$,

$$|f_a(S) - f_a(S')| = |\langle a, m(S) - m(S') \rangle| \leq \|a\| \, \|m(S) - m(S')\|,$$

where $m(S)$ is the average of the elements in S. Thus, for any set $\mathcal{A} \subset \mathbb{R}^d$, recalling that $\mathcal{F}_{\mathcal{A}} = \{x \mapsto \langle a, x \rangle : a \in \mathcal{A}\}$, we obtain

$$D_{\mathcal{F}_{\mathcal{A}}}(S, S')^2 \leq \left(\sup_{a \in \mathcal{A}} \|a\| \right)^2 \|m(S') - m(S)\|^2.$$

We denote $H(S') := \frac{1}{2}\|m(S') - m(S)\|^2$. It is easy to see that every minimizer of H minimizes the IPM, and a bound on H immediately gives a bound on the original query error. There are simpler ways to answer linear queries privately, most notably direct private mean estimation. Our aim here is different: linear queries provide the cleanest test of whether the *PE synthesis paradigm itself* can exploit a structured IPM.

We first prove in section 4.1 that evolution on the surrogate H has favorable non-private dynamics when the variations are Gaussian, but a naive algorithm takes exponential time to run. However, the exponential time comes from the selection procedure, not from requiring an exponential number of variations. Thus, we can hope to design a more efficient selection algorithm. Our approach first selects records from the variations of S_t which can potentially lie outside of the data domain $\mathcal{X}$ since they are created by adding Gaussian noise. It then reconstructs a feasible dataset $S_{t+1} \in \mathcal{X}^m$ such that the mean of S_{t+1} coincides with projection of the mean of the selected records. Then, in Section 4.2, we replace the computationally-intensive selection step by an efficient private stochastic Frank–Wolfe relaxation, yielding a computationally-tractable private evolution algorithm.

4.1 Warm-up: Gaussian variation works for mean matching

For the upper bounds in this section, we model variations as additive, independent, isotropic Gaussian noise, as in González Lara et al. (2026); Lin et al. (2023). Each current point has K variations at each of L scales. We first study an idealized non-private update that selects the m-tuple from the (unprojected) variation pool minimizing H. This tuple may contain points outside the feasible region $\mathcal{X}$, so the algorithm then constructs a feasible $S_{t+1} \in \mathcal{X}^m$ from its projected mean. Point selection is then based on minimizing the surrogate function.

The surrogate H has a particularly simple geometry. Whenever the current mean differs from the target mean, moving any synthetic point in the direction opposite to this mean residual decreases H. Among K Gaussian variations of a point, one typically has a component of order $\sqrt{\log K}$ in this favorable direction. The theorem below shows that this yields exponential contraction of H by selecting the intermediate dataset that minimizes $H(S')$ over $S' \in V_{t+1}^m$. This selection requires minimizing over V_{t+1}^m and therefore takes exponential time in m, so it is not a practical algorithm. Its crucial feature is that the exponential time comes from the selection step, and not from the need for an exponentially large variation pool. Moreover, the contraction is established before projecting the variations to the feasible set $\mathcal{X}$; feasibility is restored only afterward through the mean-reconstruction step. This selection step will be made efficient and private in the next subsection.

Theorem 4.1 (Mean matching by Gaussian variations)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be convex and compact with diameter D, let $S = \{x_1, \ldots, x_n\} \subset \mathcal{X}$, and let $S_t = \{y_1^t, \ldots, y_m^t\}$ be the iterates of Algorithm 1.*

Fix $\alpha > 0$, and assume $2 \le K \le e^d$. Suppose that, for every t, the variation pool contains the current points and K independent Gaussian variations per point at each scale

$$\sigma_\ell = \frac{1}{16}\frac{\sqrt{\log K}}{d}\,\alpha 2^{\ell-1}, \qquad \ell = 1, \dots, L,$$

where $\alpha 2^{L-1} \ge D$. Then

$$\mathbb{E}[H(S_t)] \le H(S_0)\exp\left(-\frac{1}{64}\frac{\log K}{d}t\right) + \frac{\alpha^2}{2}.$$

The proof is deferred to Section A.3.

The linear lower-bound instance. For the construction in Theorem 3.1, the coefficient norms are at most $1/3$, the target is 0, and $H(\{x_0\}) = d/2$. Thus, with one synthetic point and $K = 2$, the preceding theorem gives

$$\mathbb{E}[D_{\mathcal{F}}(\{0\}, \{x_t\})^2] \le \frac{1}{9}\left(d\exp\left(-\frac{\log 2}{64d}t\right) + \alpha^2\right).$$

Taking $\alpha \le \eta$ and the smallest grid covering $D = 2\sqrt{d}$, $O(d\log^2(d/\eta))$ Gaussian proposals suffice for expected squared IPM error at most η^2, for $0 < \eta < 1/6$. Hence the surrogate succeeds on the convex hard instance with the same Gaussian variation mechanism for which direct evolution fails, followed by the feasible mean-reconstruction step.

4.2 Private mean-matching via stochastic Frank–Wolfe

The warm-up establishes that the variation pool contains useful descent moves, but selecting the best size-m dataset from a pool of M variations requires searching over M^m possibilities. We instead relax the selection problem to continuous distributions over the variation pool and optimize the resulting smooth convex objective over the simplex using stochastic Frank–Wolfe (Frank & Wolfe, 1956; Jaggi, 2013). The key feature is that each Frank–Wolfe linearized subproblem over the simplex reduces to selecting a single variation, which we privatize using the exponential mechanism, similarly to Asi et al. (2021); Bassily et al. (2021). One could alternatively use mirror descent over the simplex, in the spirit of González et al. (2026), but Frank–Wolfe gives us better theoretical rates.

More concretely, let $V_{t+1} = \{v_1, \dots, v_M\}$ be the Gaussian variation pool and, for $w \in \Delta_M$, define $V_{t+1}w := \sum_{i=1}^M w_i v_i$ and $H_t(w) := \frac{1}{2}\|V_{t+1}w - m(S)\|^2$. We approximately solve

$$\min_{w\in\Delta_M} H_t(w)$$

with Q private Frank–Wolfe steps. This relaxation can only improve upon the original combinatorial problem: every size-m dataset drawn from the variation pool induces an empirical distribution over the same pool with the same mean. Hence

$$\min_{w\in\Delta_M} H_t(w) \le \min_{S'\in V_{t+1}^m} H(S').$$

The per-step improvement achieved by the non-private evolution step in the previous section can therefore be recovered up to the optimization error of private Frank–Wolfe.

Finally, we must turn the Frank–Wolfe last iterate $w^{(Q)}$ into a size-m synthetic dataset. To do so, we project its weighted mean onto $\mathcal{X}$ and apply any post-processing procedure that returns $S_{t+1} \in \mathcal{X}^m$ with $m(S_{t+1}) = \Pi_{\mathcal{X}}(V_{t+1}w^{(Q)})$. Such a reconstruction always exists. In particular, S_{t+1} need not consist of points in the variation pool; this is precisely where the update departs from direct evolution. Beyond this mean constraint, the theory places no restriction on the post-processing. In practice, one would naturally use this freedom to retain information about which variations received weight under $w^{(Q)}$, so that the next population reflects the variations favored by the private selection step rather than only their mean. Since $m(S) \in \mathcal{X}$, projecting the Frank–Wolfe mean cannot increase H; therefore any such reconstruction introduces no additional error in the mean-matching objective. Algorithm 2 summarizes the procedure, and its guarantees are presented in Theorem 4.2.

Algorithm 2 Private mean-matching evolution via stochastic Frank–Wolfe

Require: Target dataset S, public initial dataset $S_0 \in \mathcal{X}^m$, outer iterations T, inner iterations Q, Gaussian scales $\sigma_1, \dots, \sigma_L$, K draws per scale and point, privacy parameters (ε, δ).

Set $\varepsilon_0 := \left(\sqrt{2\log(1/\delta) + 2\varepsilon} - \sqrt{2\log(1/\delta)}\right)/\sqrt{TQ}$.
for $t = 0, \dots, T-1$ **do**
  $V_{t+1} \leftarrow \mathrm{Variations}(S_t)$; enumerate $V_{t+1} = \{v_1, \dots, v_M\}$.
  Set $B_t := \max\{R, \max_{i\in[M]} \|v_i\|_2\}$.
  Initialize $w^{(0)} = \frac{1}{M}(1, ..., 1) \in \Delta_M$.
  for $q = 0, \dots, Q-1$ **do**
    Set $r^{(q)} := V_{t+1}w^{(q)} - m(S)$.
    Sample $I_{q+1} \in [M]$ according to $\Pr(I_{q+1} = i) \propto \exp\left(-\frac{\varepsilon_0 n}{4B_t^2}\langle v_i, r^{(q)}\rangle\right)$.
    Set $\eta_q := 2/(q+2)$, and update $w^{(q+1)} = (1-\eta_q)w^{(q)} + \eta_q e_{I_{q+1}}$.
  end for
  Construct any $S_{t+1} \in \mathcal{X}^m$ such that $m(S_{t+1}) = \mu_{t+1} := \Pi_{\mathcal{X}}(V_{t+1}w^{(Q)})$.
end for
return S_T.

Theorem 4.2 (Private Frank–Wolfe mean-matching evolution). *Assume $\mathcal{X} \subseteq B_2(0, R)$ is convex and compact, $0 < \alpha < R$, $0 < \varepsilon \le \log(1/\delta)$, and $0 < \delta < 1/2$. Let $m \ge 1$, $2 \le K \le e^d$, and let L be the number of Gaussian variation scales in Theorem 4.1. Let S_t be the iterates of Algorithm 2, starting from a public dataset $S_0 \in \mathcal{X}^m$, using K Gaussian variations per point at each of the L scales from Theorem 4.1, and any post-processing reconstruction satisfying the prescribed mean constraint. Set $M := m(1+KL)$, $\gamma := \frac{1}{64}\frac{\log K}{d}$, and $G := R + \frac{\alpha 2^{L-1}\sqrt{\log K}}{16d}\left(\sqrt{d} + \sqrt{2\log\frac{2TmKL}{\delta}}\right)$. Then there exists a setting of T and Q such that, for every target dataset $S \in \mathcal{X}^n$, Algorithm 2, run with privacy parameters (ε, δ), is (ε, δ)-differentially private and outputs a synthetic dataset $S_T \in \mathcal{X}^m$ satisfying, for a universal constant $C > 0$,*

$$\mathbb{E}[H(S_T)] \le C\alpha^2 + CG^2\left(\frac{d^2\sqrt{\log(2R/\alpha)\log(2/\delta)}\log M}{(\log K)^2 n\varepsilon}\right)^{2/3}.$$

The proof is deferred to Section A.4.

5 Reconstruction can be hard beyond linear queries

The linear-query result succeeds for two reasons: the surrogate has geometry that allows Gaussian variations to make progress, and the and the next dataset has to be constructed only with a simple mean constraint. The latter is a reconstruction step. Similar results are harder to obtain beyond linear queries, even when favorable statistical rates are available. Many previous works on structured IPMs follow a related approach: privatize statistics that characterize the IPM and then solve the inverse problem of constructing a valid dataset that approximately matches those noisy statistics. The reconstruction step in these approaches is generally computationally expensive. Thus, reconstruction becomes a potential computational bottleneck.

Our next result isolates this reconstruction problem from both privacy and evolution. The target answers are given exactly, so there is no privacy loss and no estimation error. Nevertheless, for a fixed synthetic dataset size, matching only two bounded convex quadratic queries to constant additive accuracy is already NP-hard. Thus, the computational burden of reconstruction is not merely an artifact of a particular privacy mechanism.

Proposition 5.1 (Hardness of dataset-level convex query matching). *Fix an integer $M \ge 1$. There exists a constant $E_0 \in (0, 1)$ such that the following problem is NP-hard. Given a bounded box $\mathcal{Z} \subset \mathbb{R}^D$, convex smooth functions $f_1, f_2 : \mathcal{Z} \to [0, 1]$, and targets $c_1, c_2 \in [0, 1]$, define*

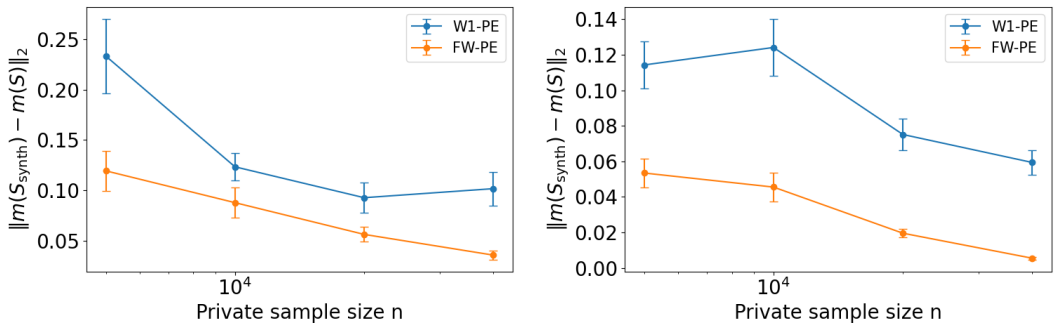

Figure 1: Mean-estimation error for private evolution on $\mathcal{X} = B_2(0, 1.5) \subset \mathbb{R}^{10}$, with ε equal to 1 (left) and 4 (right). Both the sensitive dataset and the public initialization are sampled independently and uniformly from $\mathcal{X}$. The W_1 baseline is the original nearest-neighbor PE update, while FW-PE privately optimizes the quadratic mean-matching surrogate. Curves show means over ten repetitions and error bars show 95% confidence intervals.

$$F_M(S') := \max_{j\in\{1,2\}} \left(\frac{1}{M} \sum_{\ell=1}^{M} f_j(z_\ell) - c_j \right)^2, \qquad S' = (z_1, \ldots, z_M) \in \mathcal{Z}^M.$$

Then, for every fixed $\varepsilon < E_0$, it is NP-hard to find $S' \in \mathcal{Z}^M$ satisfying $F_M(S') \le \inf_{T\in\mathcal{Z}^M} F_M(T) + \varepsilon$. This hardness holds even when f_1 and f_2 are convex quadratics.

The proposition should be interpreted as a worst-case computational barrier for this reconstruction problem, not as a hardness theorem for any particular prior method or for all query classes beyond linear queries. Its role in the paper is complementary to the linear and quadratic evolution obstructions. Section 3 shows that simple query classes can defeat projected direct evolution with prescribed variation laws; the result above shows that, even when the desired query answers are already known, globally realizing them with a fixed-size dataset can itself be computationally intractable. At the same time, this does not rule out efficient algorithms beyond linear queries when additional structure makes either evolution or reconstruction tractable.

6 Simulations

The theory suggests that when utility is measured as a structured statistic, optimizing that statistic directly can be more effective than using an evolution rule designed to match the full data distribution in W_1. We illustrate this for mean estimation. Our goal is not to compete with specialized private mean estimators, but to isolate the effect of the evolution objective within the PE framework.

Setup. We take $\mathcal{X} = B_2(0, 1.5) \subset \mathbb{R}^{10}$. In each repetition, the sensitive dataset S and public initialization S_0 are sampled independently and uniformly from $\mathcal{X}$. We compare the original W_1-based PE update of González Lara et al. (2026) with the mean-matching method from Algorithm 2, which privately optimizes $H(S') = \frac{1}{2}\|m(S') - m(S)\|_2^2$.

Both methods use the same public initialization and synthetic population size, and we evaluate them by the final mean error $\|m(S_T) - m(S)\|_2$. For linear queries, this controls the target IPM through $D_{\mathcal{F}_\mathcal{A}}(S, S_T) \le A_{\max}\|m(S_T) - m(S)\|_2$.

We consider and $\varepsilon \in \{1, 4\}$ with $\delta = 10^{-6}$ and plot the erros for different values of n. The plots show the average over ten independent runs. Further implementation details are given in Section B.

Results. Figure 1 shows that directly optimizing the mean can outperform evolution through W_1. At $n = 40{,}000$, FW-PE has significantly lower mean error for every privacy level considered.

This illustrates the motivation for our positive result. Matching W_1 requires reproducing aspects of the full empirical distribution that are irrelevant to linear-query utility. When the target discrepancy depends only on the mean, optimizing this statistic through the variation pool and then constructing the next dataset from it can yield better accuracy within the same evolution framework.

7 Conclusion

Our results show that statistically simpler IPMs do not necessarily make Private Evolution easier: direct evolution can require exponentially many proposals even for simple linear and quadratic query classes. At the same time, the linear-query case shows that moving beyond direct evolution—by replacing the target IPM with a compatible surrogate and relaxing how the next dataset is constructed—can restore efficient evolution and yield a private algorithm that achieves an error rate with polynomial dimension dependence and runs in polynomial time.

A natural next step is to understand which query classes admit both a surrogate that is compatible with the variation mechanism and an efficient procedure for realizing the resulting statistics as a synthetic dataset. Beyond linear queries, either requirement may fail: our lower bounds expose barriers in the evolution dynamics, while our reconstruction result shows that even realizing a small number of target statistics can be computationally hard in the worst case.

AI use statement

In this work, we used generative AI tools to assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments and implement methods. We have not used generative AI tools for help develop theoretical models or conceptual frameworks, formulate mathematical claims, or interpret results. Our work does not need AI to generate synthetic data sets (at least not for evaluation—the whole topic of the paper is on generating synthetic datasets), assist with translation, and clean and reformat datasets. We have reviewed all AI-assisted work. For example, LLM-generated code and proofs were verified for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

Gbola Afonja, Robert Sim, Zinan Lin, Huseyin Atahan Inan, and Sergey Yekhanin. The crossroads of innovation and privacy: Private synthetic data for generative AI. Microsoft Research Blog, May 2024. URL https://www.microsoft.com/en-us/research/blog/the-crossroads-of-innovation-and-privacy-private-synthetic-data-for-generative-ai/ Published May 29, 2024.

Apple. Understanding aggregate trends for apple intelligence using differential privacy. https://machinelearning.apple.com/research/differential-privacy-aggregate-trends, 2025. Apple Machine Learning Research.

Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: Optimal rates in ℓ_1 geometry. In *International Conference on Machine Learning*, pp. 393–403. PMLR, 2021.

Matej Balog, Ilya Tolstikhin, and Bernhard Schölkopf. Differentially private database release via kernel mean embeddings. In *International Conference on Machine Learning*, pp. 414–422. PMLR, 2018.

Raef Bassily, Cristóbal Guzmán, and Anupama Nandi. Non-euclidean differentially private stochastic convex optimization. In *Conference on Learning Theory*, pp. 474–499. PMLR, 2021.

Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *Journal of the ACM*, 60(2):1–25, 2013.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Metric geometry of the privacy-utility tradeoff. *arXiv preprint arXiv:2405.00329*, 2024.

Rundong Ding, Yiyun He, and Yizhe Zhu. Minimax optimal differentially private synthetic data for smooth queries. *arXiv preprint arXiv:2602.01607*, 2026.

Konstantin Donhauser, Johan Lokna, Amartya Sanyal, March Boedihardjo, Robert Hönig, and Fanny Yang. Certified private data release for sparse lipschitz functions. In *International Conference on Artificial Intelligence and Statistics*, pp. 1396–1404, 2024.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284, 2006.

Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53:650–673, 2015.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

Tomás González, Cristóbal Guzmán, and Courtney Paquette. Mirror descent algorithms with nearly dimension-independent rates for differentially-private stochastic saddle-point problems. *SIAM Journal on Optimization*, 36(1):233–262, 2026.

Tomás González Lara, Giulia Fanti, and Aaditya Ramdas. Private evolution converges. *Advances in Neural Information Processing Systems*, 38:148405–148441, 2026.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Moritz Hardt, Katrina Ligett, and Frank Mcsherry. A simple and practical algorithm for differentially private data release. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

Yiyun He, Roman Vershynin, and Yizhe Zhu. Algorithmically effective differentially private synthetic data. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 3941–3968. PMLR, 2023.

Charlie Hou, Akshat Shrivastava, Hongyuan Zhan, Rylan Conway, Trang Le, Adithya Sagar, Giulia Fanti, and Daniel Lazar. Pre-text: training language models on private federated data in the age of LLMs. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Charlie Hou, Mei-Yu Wang, Yige Zhu, Daniel Lazar, and Giulia Fanti. Private federated learning using preference-optimized synthetic data. *arXiv preprint arXiv:2504.16438*, 2025.

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International conference on machine learning*, pp. 427–435. PMLR, 2013.

Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, Harsha Nori, and Sergey Yekhanin. Differentially Private Synthetic Data via Foundation Model APIs 1: Images. In *ICLR*, 2023.

Zinan Lin, Tadas Baltrusaitis, and Sergey Yekhanin. Differentially private synthetic data via APIs 3: Using simulators instead of foundation models. *arXiv preprint arXiv:2502.05505*, 2025.

Terrance Liu, Jingwu Tang, Giuseppe Vietri, and Steven Wu. Generating private synthetic data with genetic algorithms. In *International Conference on Machine Learning*, pp. 22009–22027, 2023.

Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997. doi: 10.2307/1428011.

Thomas Steinke. Composition of differential privacy & privacy amplification by subsampling. 2025.

Marika Swanberg, Ryan McKenna, Edo Roth, Albert Cheu, and Peter Kairouz. Is API access to LLMs useful for generating private synthetic tabular data? In *ICLR Workshop: Will Synthetic Data Finally Solve the Data Access Problem?*, 2025.

C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. ISBN 9783540710509.

Ziteng Wang, Chi Jin, Kai Fan, Jiaqi Zhang, Junliang Huang, Yiqiao Zhong, and Liwei Wang. Differentially private data releasing for smooth queries. *Journal of Machine Learning Research*, 17(51):1–42, 2016.

Chulin Xie, Zinan Lin, Arturs Backurs, Sivakanth Gopi, Da Yu, Huseyin Inan, Harsha Nori, Haotian Jiang, Huishuai Zhang, Yin Tat Lee, Bo Li, and Sergey Yekhanin. Differentially private synthetic data via foundation model APIs 2: text. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Tianyuan Zou, Yang Liu, Peng Li, Yufei Xiong, Jianqing Zhang, Jingjing Liu, Xiaozhou Ye, Ye Ouyang, and Ya-Qin Zhang. Contrastive Private Data Synthesis via Weighted Multi-PLM Fusion. *arXiv preprint arXiv:2502.00245*, 2025.

A MISSING PROOFS

A.1 PROOF OF THEOREM 3.1

Proof. In both constructions, we average over sign vectors $\theta \in \{-1, 1\}^d$ to select an instance after the noise menu is fixed.

Linear queries. For a sign vector θ, take $y = 0$, $x_0 = \theta$, and define

$$g_j(x) = \theta_j x_j - \frac{1}{2d} \sum_{i=1}^{d} \theta_i x_i, \qquad f_j(x) = \frac{1}{2} + \frac{1}{3} g_j(x), \qquad j \in [d].$$

Since $|g_j(x)| \le 3/2$ on the cube, every f_j takes values in $[0, 1]$. The induced IPM is

$$\Phi(x) = \frac{1}{3} \max_{j \in [d]} |g_j(x)|, \qquad \Phi(0) = 0, \qquad \Phi(\theta) = \frac{1}{6}.$$

It is convex, being a maximum of absolute linear functions. The point θ is not a local minimum: $(1 - u)\theta$ improves for every $0 < u < 1$.

Suppose $z \in \mathcal{X}$, $z \neq \theta$, and $\Phi(z) \le 1/6$. If $\theta_j z_j = 1$ for some j, then $d^{-1} \sum_i \theta_i z_i < 1$, so $g_j(z) > 1/2$, a contradiction. Thus every distinct nonworsening point satisfies $\theta_j z_j < 1$ for all j. For a projected proposal $z = \Pi_{\mathcal{X}}(\theta + \sigma h)$, this requires

$$h \in O_\theta := \{h : \theta_j h_j < 0 \text{ for all } j\},$$

uniformly over $\sigma > 0$. The 2^d open orthants are disjoint, hence

$$\frac{1}{2^d} \sum_{\theta} \sum_{\ell=1}^{L} Q_\ell(O_\theta) \le L 2^{-d}.$$

Choose θ for which the inner sum is at most $L2^{-d}$. Before the first departure, the state remains θ, and the conditional probability that any fresh draw permits departure is at most this sum, whatever the chosen law and scale. A union bound over the N calls gives $NL2^{-d}$. This argument covers ties as well as strict improvements. If each law is invariant under coordinatewise sign flips, every open orthant has mass at most 2^{-d}, giving $N2^{-d}$ at every corner.

Three convex quadratic queries. For a sign vector θ, write $a(x) = \langle \theta, x \rangle / d$, take $x_0 = 0$, $y = -\theta/2$, and set

$$\begin{aligned} g_1(x) &= a(x)^2 + \frac{1}{3} a(x) - \frac{1}{12}, \\ g_2(x) &= a(x)^2 + \frac{2}{3} a(x) + \frac{1}{12}, \\ g_3(x) &= \frac{1}{d} \|x\|_2^2 + \frac{2}{3} a(x) + \frac{1}{12}, \qquad f_i(x) = \frac{36}{67} \left(g_i(x) + \frac{1}{9} \right). \end{aligned}$$

All three queries are convex quadratics. On the cube, $|a(x)| \le 1$ and $a(x)^2 \le \|x\|^2/d \le 1$, so $-1/9 \le g_i(x) \le 7/4$ and $f_i(x) \in [0,1]$. Since $g_i(y) = 0$,

$$\Phi(x) = \frac{36}{67} \max_i |g_i(x)|, \qquad \Phi(y) = 0, \qquad \Phi(0) = \frac{3}{67}.$$

The origin is a strict local minimum. For $a(x) > 0$, $g_2(x) > 1/12$; for $-1/3 < a(x) < 0$, $g_1(x) < -1/12$; and for $a(x) = 0$, $x \neq 0$, $g_3(x) = 1/12 + \|x\|^2/d > 1/12$.

If $z \neq 0$ is nonworsening, then $g_1(z) \ge -1/12$ and $g_2(z) \le 1/12$, which imply

$$a(z)\big(a(z) + 1/3\big) \ge 0, \qquad a(z)\big(a(z) + 2/3\big) \le 0.$$

Thus $a(z) \in [-2/3, -1/3] \cup \{0\}$. The case $a(z) = 0$, together with $g_3(z) \le 1/12$, forces $z = 0$. Every distinct nonworsening point therefore satisfies

$$-\langle \theta, z \rangle \ge d/3.$$

To account for projection uniformly over all scales, define

$$E_\theta := \{h : \text{there exists } \sigma > 0 \text{ with } -\langle \theta, \Pi_{\mathcal{X}}(\sigma h) \rangle \ge d/3\}.$$

Fix $h \neq 0$. Coordinatewise clipping makes $\sigma \mapsto \Pi_{\mathcal{X}}(\sigma h)$ a piecewise affine path with at most d breakpoints, at the values $1/|h_j|$ with $h_j \neq 0$, and the path is constant after its final breakpoint. A linear functional achieves its maximum at one of these breakpoints or at $\sigma = 0$. The latter has value zero. For any fixed breakpoint vector $z \in \mathcal{X}$, a uniform random sign vector Θ satisfies the Rademacher tail bound

$$\Pr_\Theta\{-\langle \Theta, z \rangle \ge d/3\} \le \exp\left(-\frac{d^2}{18\|z\|_2^2}\right) \le e^{-d/18}.$$

A union bound over the breakpoints gives $\Pr_\Theta(h \in E_\Theta) \le de^{-d/18}$; for $h = 0$ the probability is zero. Averaging over the prescribed laws yields

$$\frac{1}{2^d} \sum_\theta \sum_{\ell=1}^{L} Q_\ell(E_\theta) \le Lde^{-d/18}.$$

Choose θ whose inner sum is at most this bound. Until the first departure the state is zero, and every distinct nonworsening projected proposal requires its noise draw to belong to E_θ, regardless of the scale. The same conditional union bound as above gives $NLde^{-d/18}$, completing the proof. □

A.2 Proof of Proposition 3.1

Proof. Let $S = \{-1/2, 1/2\}$, $S_0 = \{-1, 9/10\}$, $V = \{-1, 9/10, -3/5, 7/10\}$. Selecting $S_1 = \{-3/5, 7/10\}$ gives

$$W_1(S, S_0) = \frac{9}{20}, \qquad W_1(S, S_1) = \frac{3}{20},$$

so the variation pool contains a strict improvement in W_1.

Now consider the affine query $f(x) = (x+1)/2$ on $[-1, 1]$. Since S has mean zero, for any size-two multiset $T = \{u, v\}$ supported on V,

$$D_{\{f\}}(S, T) = \frac{1}{4}|u + v|.$$

Moreover, $\min_{u,v \in V} |u + v| = \frac{1}{10}$, while $D_{\{f\}}(S, S_0) = 1/40$. Hence every size-two dataset T supported on V satisfies

$$D_{\{f\}}(S, T) \ge D_{\{f\}}(S, S_0),$$

so no dataset in the variation pool strictly improves this IPM. □

A.3 Proof of Theorem 4.1

Proof. We start proving that the objective shrinks in expectation after one step:

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le \left(1 - \frac{1}{64}\frac{\log K}{d}\right) H(S_t).$$

To obtain the final result, we iterate this inequality over T steps. We first prove that

$$H(S_{t+1}) \le H(\tilde{S}_{t+1}).$$

Recall that by construction, $m(S_{t+1}) = \Pi_{\mathcal{X}}(m(\widetilde{S}_{t+1}))$. Since $m(S) \in \mathcal{X}$, Euclidean projection onto the closed convex set $\mathcal{X}$ cannot increase distance to $m(S)$. Hence

$$\|m(S_{t+1}) - m(S)\| = \|\Pi_{\mathcal{X}}(m(\widetilde{S}_{t+1})) - m(S)\| \le \|m(\widetilde{S}_{t+1}) - m(S)\|.$$

Consequently, $H(S_{t+1}) \le H(\widetilde{S}_{t+1})$. Hence, for the one-step contraction it suffices to show that

$$\mathbb{E}[H(\tilde{S}_{t+1}) \mid S_t] \le \left(1 - \frac{1}{64}\frac{\log K}{d}\right) H(S_t).$$

Next, we construct a dataset S_{t+1}^{cand} containing m elements from the variation set V_{t+1}. Since $\tilde{S}_{t+1}$ minimizes H over V_{t+1}^m, it follows that $H(\widetilde{S}_{t+1}) \le H(S_{t+1}^{\text{cand}})$. In what follows, we actually prove that

$$\mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] \le \left(1 - \frac{1}{64}\frac{\log K}{d}\right) H(S_t),$$

and since $H(S_{t+1}) \le H(\tilde{S}_{t+1}) \le H(S_{t+1}^{\text{cand}})$, the one-step improvement follows. S_{t+1}^{cand} consists of a dataset where each element in S_t is added some noise by the variation API. We now describe how to construct it. Fix S_t, and write

$$r_t := m(S_t) - m(S), \qquad g_{t,j} := \nabla_{y_j} H(S_t) = \frac{1}{m} r_t.$$

Assume first that $\|r_t\| \ge \alpha$. Since $\|r_t\| \le D$ because the domain is bounded, and $\alpha 2^{L-1} \ge D$, there exists $\ell \in [L]$ such that

$$\alpha 2^{\ell-1} \le \|r_t\| \le 2\alpha 2^{\ell-1}.$$

For this level, set

$$\sigma := \sigma_\ell = \frac{1}{16}\frac{\sqrt{\log K}}{d}\alpha 2^{\ell-1}.$$

Then

$$\frac{1}{32}\frac{\sqrt{\log K}}{d}\|r_t\| \le \sigma \le \frac{1}{16}\frac{\sqrt{\log K}}{d}\|r_t\|.$$

For each j, consider the K Gaussian perturbations

$$h_{j,k} := \sigma \xi_{j,k}, \qquad \xi_{j,k} \sim N(0, I_d),$$

and choose $k_j^\star \in [K]$ minimizing $\langle g_{t,j}, h_{j,k}\rangle$. Let

$$M_K := \max_{k\in[K]} Z_k, \qquad Z_1, \dots, Z_K \overset{\text{i.i.d.}}{\sim} N(0,1).$$

We use the elementary Gaussian maximum bounds

$$\mathbb{E}[M_K] \ge \frac{1}{2}\sqrt{\log K}, \qquad \mathbb{E}[M_K^2] \le 2\log K + 1.$$

By decomposing the selected Gaussian into its component along $g_{t,j}$ and its orthogonal component,

$$\mathbb{E}[\langle g_{t,j}, h_{j,k_j^\star}\rangle \mid S_t] = -\sigma\|g_{t,j}\|\mathbb{E}[M_K] \le -\frac{1}{2}\sigma\sqrt{\log K}\|g_{t,j}\|,$$

and

$$\mathbb{E}[\|h_{j,k_j^\star}\|^2 \mid S_t] \le \sigma^2\big(\mathbb{E}[M_K^2] + d - 1\big) \le \sigma^2(2\log K + d).$$

Since we assumed $K \le e^d$, we have $\log K \le d$, and therefore

$$\mathbb{E}[\|h_{j,k_j^\star}\|^2 \mid S_t] \le 3d\sigma^2.$$

Now take

$$S_{t+1}^{\text{cand}} = (y_1^t + h_{1,k_1^\star}, \dots, y_m^t + h_{m,k_m^\star}).$$

To prove that the objective improves with this perturbations, we need to understand how H behaves when perturbing all the elements of a dataset. Fix $S' = \{y_1, \dots, y_m\}$. Since $m(S') = \frac{1}{m}\sum_{j=1}^m y_j$, we have

$$\nabla_{y_j} H(S') = \frac{1}{m}(m(S') - m(S)).$$

Hence, for simultaneous perturbations $h_1, \dots, h_m$, the quadratic identity and Jensen's inequality give

$$\begin{aligned} H(y_1 + h_1, \dots, y_m + h_m) &= H(S') + \sum_{j=1}^m \langle \nabla_{y_j} H(S'), h_j \rangle + \frac{1}{2}\left\| \frac{1}{m}\sum_{j=1}^m h_j \right\|^2 \\ &\le H(S') + \sum_{j=1}^m \left(\langle \nabla_{y_j} H(S'), h_j \rangle + \frac{1}{2m}\|h_j\|^2 \right). \end{aligned}$$

We apply this inequality to $S' = S_t$ and the perturbations play the role of the Gaussian noises from the variation API:

$$\begin{aligned} \mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] &\le H(S_t) + \sum_{j=1}^m \mathbb{E}\left[\langle g_{t,j}, h_{j,k_j^\star} \rangle + \frac{1}{2m}\|h_{j,k_j^\star}\|^2 \middle| S_t \right] \\ &\le H(S_t) - \frac{1}{2}\sigma\sqrt{\log K}\sum_{j=1}^m \|g_{t,j}\| + \frac{3d\sigma^2}{2}. \end{aligned}$$

Since $\|g_{t,j}\| = \|r_t\|/m$, we have

$$\sum_{j=1}^m \|g_{t,j}\| = \|r_t\|.$$

Hence

$$\mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] \le H(S_t) - \frac{1}{2}\sigma\sqrt{\log K}\|r_t\| + \frac{3d\sigma^2}{2}.$$

Using the bounds on σ,

$$\frac{1}{2}\sigma\sqrt{\log K}\|r_t\| \ge \frac{1}{64}\frac{\log K}{d}\|r_t\|^2,$$

and

$$\frac{3d\sigma^2}{2} \le \frac{3}{512}\frac{\log K}{d}\|r_t\|^2.$$

Thus

$$\mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] \le H(S_t) - \left(\frac{1}{64} - \frac{3}{512}\right)\frac{\log K}{d}\|r_t\|^2.$$

Since $1/64 - 3/512 = 5/512$, and a quick calculation shows that $H(S_t) = \frac{1}{2}\|r_t\|^2$, we get

$$\mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] \le \left(1 - \frac{5}{256}\frac{\log K}{d}\right) H(S_t).$$

In particular,

$$\mathbb{E}[H(S_{t+1}^{\text{cand}}) \mid S_t] \le \left(1 - \frac{1}{64}\frac{\log K}{d}\right) H(S_t).$$

Now consider the other case, $\|r_t\| < \alpha$. Then

$$H(S_t) = \frac{1}{2}\|r_t\|^2 < \frac{\alpha^2}{2}.$$

Since S_t itself is included in the variation pool, the minimization step satisfies

$$H(\widetilde{S}_{t+1}) \le H(S_t).$$

The final projection step cannot increase H, so

$$H(S_{t+1}) \le H(\widetilde{S}_{t+1}) \le H(S_t).$$

Define

$$q := \frac{1}{64}\frac{\log K}{d}, \qquad A := \frac{\alpha^2}{2}.$$

We claim that in both cases,

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le (1-q)H(S_t) + qA.$$

Indeed, if $\|r_t\| \ge \alpha$, then we proved

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le (1-q)H(S_t),$$

and therefore the claimed bound follows because $qA \ge 0$. If $\|r_t\| < \alpha$, then $H(S_t) < A$, and monotonicity gives

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le H(S_t).$$

But

$$H(S_t) = (1-q)H(S_t) + qH(S_t) < (1-q)H(S_t) + qA.$$

Thus the unified recursion holds:

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le (1-q)H(S_t) + qA.$$

Taking expectations gives

$$\mathbb{E}[H(S_{t+1})] \le (1-q)\mathbb{E}[H(S_t)] + qA.$$

Subtracting A from both sides,

$$\mathbb{E}[H(S_{t+1})] - A \le (1-q)\big(\mathbb{E}[H(S_t)] - A\big).$$

Iterating,

$$\mathbb{E}[H(S_t)] - A \le (1-q)^t(H(S_0) - A) \le (1-q)^t H(S_0).$$

Therefore

$$\mathbb{E}[H(S_t)] \le (1-q)^t H(S_0) + A.$$

Finally, using $1 - q \le e^{-q}$,

$$\mathbb{E}[H(S_t)] \le H(S_0)\exp\left(-\frac{1}{64}\frac{\log K}{d}t\right) + \frac{\alpha^2}{2}.$$

□

A.4 Proof of Theorem 4.2

Proof. We prove privacy and accuracy guarantees separately.

Privacy. Condition on the preceding private outputs, the current synthetic dataset, and the public variation randomness. The pool $V = \{v_1, \dots, v_M\}$, its radius B_t, and the current Frank–Wolfe iterate w are then fixed. For neighboring target datasets, the score

$$u_i(S) = -\langle v_i, Vw - m(S)\rangle$$

has sensitivity at most

$$\|v_i\|\frac{2R}{n} \le \frac{2B_t^2}{n}.$$

Thus every private selection step is ε_0-DP for every realized pool. Composition of TQ pure ε_0-DP calls gives (ε, δ)-DP when

$$\varepsilon = \frac{1}{2}TQ\varepsilon_0^2 + \sqrt{2TQ\varepsilon_0^2\log(1/\delta)},$$

according to Steinke (2025). This is exactly the algorithm's choice of ε_0. The mean projection and any subsequent reconstruction from the variation pool, the Frank–Wolfe output, and the preceding private outputs are post-processing and incur no additional privacy loss.

Expected error on one pool. Write

$$H_t(w) = \frac{1}{2}\|Vw - m(S)\|^2,$$

and let $w_t^\star$ minimize H_t over Δ_M. Conditional on the pool and the current Frank–Wolfe iterate, the exponential mechanism has expected linear-selection error at most

$$C\frac{B_t^2 \log M}{n\varepsilon_0}.$$

The quadratic objective also satisfies, for $w \in \Delta_M$, $i \in [M]$, and $0 \le \eta \le 1$,

$$H_t((1-\eta)w + \eta e_i) \le H_t(w) + \eta\langle \nabla H_t(w), e_i - w\rangle + 2\eta^2 B_t^2.$$

Convexity and the preceding selection bound therefore give the usual Frank–Wolfe recursion. With step sizes $\eta_q = 2/(q+2)$, and since $\eta_0 = 1$, unrolling gives, for any initialization $w^{(0)} \in \Delta_M$,

$$\mathbb{E}[H_t(w^{(Q)}) \mid S_t, V] \le H_t(w_t^\star) + CB_t^2\left(\frac{1}{Q} + \frac{\log M}{n\varepsilon_0}\right).$$

The algorithm sets

$$\mu_{t+1} := \Pi_{\mathcal{X}}(Vw^{(Q)})$$

and then applies an arbitrary post-processing procedure producing $S_{t+1} \in \mathcal{X}^m$ such that

$$m(S_{t+1}) = \mu_{t+1}.$$

Since $m(S) \in \mathcal{X}$, Euclidean projection onto the closed convex set $\mathcal{X}$ cannot increase distance to $m(S)$. Therefore, pointwise,

$$\begin{aligned} H(S_{t+1}) &= \frac{1}{2}\|m(S_{t+1}) - m(S)\|^2 \\ &= \frac{1}{2}\|\mu_{t+1} - m(S)\|^2 \\ &\le \frac{1}{2}\|Vw^{(Q)} - m(S)\|^2 \\ &= H_t(w^{(Q)}). \end{aligned}$$

Consequently,

$$\mathbb{E}[H(S_{t+1}) \mid S_t, V] \le H_t(w_t^\star) + CB_t^2\left(\frac{1}{Q} + \frac{\log M}{n\varepsilon_0}\right).$$

Averaging over Gaussian variations. All current points lie in $B_2(0, R)$. For the mKL fresh Gaussian vectors generated at an evolution step,

$$B_t \le R + \sigma_L \max_{j,k,\ell} \|\xi_{j,k,\ell}^t\|.$$

The Gaussian norm tail bound

$$\Pr\Big(\|\xi\| \ge \sqrt{d} + \sqrt{2u}\Big) \le e^{-u},$$

integrated after a union bound, gives

$$\mathbb{E}[B_t^2 \mid S_t] \le C\left(R^2 + \sigma_L^2\big(d + \log(2mKL)\big)\right) \le CG^2.$$

Every tuple in V^m induces a feasible point of the simplex relaxation with the same empirical mean. Hence

$$\min_{w\in\Delta_M} H_t(w) \le \min_{S'\in V^m} H(S').$$

The one-step bound from the proof of Theorem 4.1 therefore gives

$$\mathbb{E}[H_t(w_t^\star) \mid S_t] \le (1-\gamma)H(S_t) + \gamma\frac{\alpha^2}{2}.$$

Since $\varepsilon \le \log(1/\delta)$, the definition of ε_0 gives

$$\varepsilon_0 \asymp \frac{\varepsilon}{\sqrt{TQ\log(1/\delta)}}.$$

Combining the preceding bounds yields

$$\mathbb{E}[H(S_{t+1}) \mid S_t] \le (1-\gamma)H(S_t) + \gamma\frac{\alpha^2}{2} + CG^2\left[\frac{1}{Q} + \frac{\sqrt{TQ\log(2/\delta)}\log M}{n\varepsilon}\right].$$

Unrolling the recursion gives

$$\mathbb{E}[H(S_T)] \le (1-\gamma)^T H(S_0) + \frac{\alpha^2}{2} + \frac{CG^2}{\gamma}\left[\frac{1}{Q} + \frac{\sqrt{TQ\log(2/\delta)}\log M}{n\varepsilon}\right].$$

Using the public bound $H(S_0) \le 2R^2$, choose

$$T = \left\lceil \gamma^{-1}\log\frac{4R^2}{\alpha^2}\right\rceil.$$

Balancing the two Q-dependent terms gives

$$Q \asymp \left(\frac{n\varepsilon}{\sqrt{T\log(2/\delta)}\log M}\right)^{2/3},$$

rounded to an integer at least one. If the displayed balance is below one, the claimed bound already follows from $H(S_T) \le 2R^2$. Otherwise,

$$\mathbb{E}[H(S_T)] \le C\alpha^2 + \frac{CG^2}{\gamma}\left(\frac{\sqrt{T\log(2/\delta)}\log M}{n\varepsilon}\right)^{2/3}.$$

Finally, substituting

$$\gamma = \frac{\log K}{64d}, \qquad T = O\left(\frac{d}{\log K}\log\frac{2R}{\alpha}\right)$$

gives

$$\mathbb{E}[H(S_T)] \le C\alpha^2 + CG^2\left(\frac{d^2\sqrt{\log(2R/\alpha)\log(2/\delta)}\log M}{(\log K)^2 n\varepsilon}\right)^{2/3},$$

after absorbing universal constants, which proves the theorem. □

A.5 Proof of Proposition 5.1

Proof. We reduce from a constant-gap version of Max-Cut on bounded-degree graphs. Namely, for some constants $\Delta \in \mathbb{N}$ and $0 < \beta < \alpha < 1$, it is NP-hard to distinguish between the following two cases for a Δ-regular graph $G = (V, E)$ with $|V| = d$ and $|E| = m = \Delta d/2$:

$$\mathrm{MaxCut}(G) \ge \alpha m \qquad \text{or} \qquad \mathrm{MaxCut}(G) \le \beta m.$$

Let L be the graph Laplacian of G. For $s \in \{\pm 1\}^d$,

$$s^\top L s = \sum_{(i,j)\in E}(s_i - s_j)^2 = 4\,\mathrm{Cut}(s).$$

Since G is Δ-regular, $\lambda_{\max}(L) \le 2\Delta$. Set

$$\lambda := 2\Delta, \qquad Q := \lambda I - L \succeq 0.$$

The per-record domain is

$$\mathcal{Z} = [-1, 1]^d \times [0, R], \qquad R^2 := 2\Delta.$$

A record is denoted $z = (x, r) \in \mathcal{Z}$. Define

$$f_1(x, r) := \frac{1}{d} \sum_{i=1}^{d} x_i^2,$$

and

$$f_2(x, r) := \frac{1}{4\Delta} \left(\frac{1}{d} x^\top Q x + r^2 \right).$$

Both functions are convex smooth quadratics. Moreover,

$$0 \le f_1(x, r) \le 1.$$

Also, since $Q \succeq 0$ and $Q \preceq \lambda I = 2\Delta I$, for all $x \in [-1, 1]^d$,

$$0 \le \frac{1}{d} x^\top Q x \le 2\Delta,$$

and $0 \le r^2 \le 2\Delta$. Hence

$$0 \le f_2(x, r) \le 1.$$

Set

$$c_1 := 1, \qquad c_2 := \frac{1}{4\Delta} \left(\lambda - \frac{4\alpha m}{d} \right).$$

Since $m/d = \Delta/2$ and $\lambda = 2\Delta$,

$$c_2 = \frac{1}{4\Delta} (2\Delta - 2\alpha\Delta) = \frac{1 - \alpha}{2} \in [0, 1].$$

We first prove completeness. Suppose G has a cut $s \in \{\pm 1\}^d$ with $\mathrm{Cut}(s) \ge \alpha m$. Define the same record $z_\ell = (s, r)$ for every $\ell \in [M]$, where

$$r^2 := \frac{4}{d} (\mathrm{Cut}(s) - \alpha m).$$

This is nonnegative. Also, since $\mathrm{Cut}(s) \le m$,

$$r^2 \le \frac{4m}{d} = 2\Delta = R^2,$$

so each $z_\ell \in \mathcal{Z}$. For every ℓ,

$$f_1(z_\ell) = 1 = c_1.$$

Furthermore,

$$\begin{aligned} f_2(z_\ell) &= \frac{1}{4\Delta} \left(\frac{1}{d} s^\top Q s + r^2 \right) \\ &= \frac{1}{4\Delta} \left(\lambda - \frac{1}{d} s^\top L s + r^2 \right) \\ &= \frac{1}{4\Delta} \left(\lambda - \frac{4\,\mathrm{Cut}(s)}{d} + r^2 \right) \\ &= \frac{1}{4\Delta} \left(\lambda - \frac{4\alpha m}{d} \right) = c_2. \end{aligned}$$

Therefore both empirical averages equal their targets, and hence

$$F_M(z_1, \ldots, z_M) = 0.$$

Thus, in the YES case,

$$\inf_{\mathcal{Z}^M} F_M = 0.$$

We now prove soundness. Suppose instead that $\mathrm{MaxCut}(G) \le \beta m$. Let $S' = (z_1, \dots, z_M) \in \mathcal{Z}^M$ be arbitrary, with $z_\ell = (x_\ell, r_\ell)$, and suppose

$$F_M(S') \le \eta.$$

Set

$$\delta := \sqrt{\eta}.$$

The first residual gives

$$\left| \frac{1}{M} \sum_{\ell=1}^{M} f_1(z_\ell) - 1 \right| \le \delta.$$

Since each $f_1(z_\ell) \le 1$, this implies

$$\sum_{\ell=1}^{M} (1 - f_1(z_\ell)) \le M\delta.$$

Equivalently,

$$\sum_{\ell=1}^{M} \sum_{i=1}^{d} (1 - x_{\ell i}^2) \le dM\delta.$$

For each ℓ, let $s_\ell \in \{\pm 1\}^d$ be the coordinate-wise rounding of x_ℓ, i.e. $(s_\ell)_i = \mathrm{sign}(x_{\ell i})$, with arbitrary tie-breaking when $x_{\ell i} = 0$. For every $a \in [-1, 1]$,

$$|a - \mathrm{sign}(a)| = 1 - |a| \le 1 - a^2.$$

Therefore

$$\sum_{\ell=1}^{M} \sum_{i=1}^{d} |x_{\ell i} - (s_\ell)_i| \le \sum_{\ell=1}^{M} \sum_{i=1}^{d} (1 - x_{\ell i}^2) \le dM\delta.$$

Using

$$x^\top L x = \sum_{(i,j) \in E} (x_i - x_j)^2,$$

and the fact that $(a, b) \mapsto (a - b)^2$ is 4-Lipschitz in each coordinate on $[-1, 1]^2$, we get

$$\begin{aligned} \left| \sum_{\ell=1}^{M} x_\ell^\top L x_\ell - \sum_{\ell=1}^{M} s_\ell^\top L s_\ell \right| &\le 4 \sum_{\ell=1}^{M} \sum_{(i,j) \in E} \left(|x_{\ell i} - (s_\ell)_i| + |x_{\ell j} - (s_\ell)_j| \right) \\ &\le 4\Delta \sum_{\ell=1}^{M} \sum_{i=1}^{d} |x_{\ell i} - (s_\ell)_i| \\ &\le 4\Delta dM\delta. \end{aligned}$$

Since every cut has size at most βm,

$$s_\ell^\top L s_\ell = 4\,\mathrm{Cut}(s_\ell) \le 4\beta m \qquad \text{for every } \ell.$$

Hence

$$\frac{1}{Md} \sum_{\ell=1}^{M} x_\ell^\top L x_\ell \le \frac{4\beta m}{d} + 4\Delta\delta.$$

Now use the second residual. Define the unnormalized per-record quantity

$$g_2(x, r) := \frac{1}{d} x^\top Q x + r^2, \qquad \gamma_2 := \lambda - \frac{4\alpha m}{d}.$$

Thus $f_2 = g_2/(4\Delta)$ and $c_2 = \gamma_2/(4\Delta)$. Since $F_M(S') \le \eta$, we have

$$\left| \frac{1}{M} \sum_{\ell=1}^{M} f_2(z_\ell) - c_2 \right| \le \delta,$$

and therefore

$$\left|\frac{1}{M}\sum_{\ell=1}^{M} g_2(z_\ell) - \gamma_2\right| \le 4\Delta\delta.$$

On the other hand,

$$g_2(x,r) = \lambda f_1(x,r) - \frac{1}{d}x^\top Lx + r^2.$$

Using $r_\ell^2 \ge 0$, the first residual lower bound

$$\frac{1}{M}\sum_{\ell=1}^{M} f_1(z_\ell) \ge 1 - \delta,$$

and the bound above on the average Laplacian quadratic, we obtain

$$\begin{aligned}\frac{1}{M}\sum_{\ell=1}^{M} g_2(z_\ell) - \gamma_2 &= \lambda\frac{1}{M}\sum_{\ell=1}^{M} f_1(z_\ell) - \frac{1}{Md}\sum_{\ell=1}^{M} x_\ell^\top Lx_\ell + \frac{1}{M}\sum_{\ell=1}^{M} r_\ell^2 - \left(\lambda - \frac{4\alpha m}{d}\right)\\ &\ge \lambda(1-\delta) - \left(\frac{4\beta m}{d} + 4\Delta\delta\right) - \lambda + \frac{4\alpha m}{d}\\ &= \frac{4(\alpha-\beta)m}{d} - (\lambda + 4\Delta)\delta.\end{aligned}$$

Combining this lower bound with

$$\left|\frac{1}{M}\sum_{\ell=1}^{M} g_2(z_\ell) - \gamma_2\right| \le 4\Delta\delta$$

gives

$$4\Delta\delta \ge \frac{4(\alpha-\beta)m}{d} - (\lambda + 4\Delta)\delta.$$

Hence

$$\delta \ge \frac{4(\alpha-\beta)m/d}{\lambda + 8\Delta}.$$

Using $m/d = \Delta/2$ and $\lambda = 2\Delta$, we get

$$\delta \ge \frac{2\Delta(\alpha-\beta)}{10\Delta} = \frac{\alpha-\beta}{5} =: \delta_0.$$

Therefore, in the NO case, every dataset $S' \in \mathcal{Z}^M$ satisfies

$$F_M(S') \ge \delta_0^2.$$

Set

$$E_0 := \delta_0^2 = \left(\frac{\alpha-\beta}{5}\right)^2.$$

This is a fixed constant in $(0,1)$ depending only on the fixed gap-Max-Cut hardness constants, and crucially it does not depend on M or d.

We have constructed a gap:

$$\text{YES case:} \qquad \inf_{\mathcal{Z}^M} F_M = 0,$$

whereas

$$\text{NO case:} \qquad \inf_{\mathcal{Z}^M} F_M \ge E_0.$$

Now suppose there were a polynomial-time algorithm which, for some fixed $\varepsilon < E_0$, always returned $S' \in \mathcal{Z}^M$ with

$$F_M(S') \le \inf_{T\in\mathcal{Z}^M} F_M(T) + \varepsilon.$$

In the YES case it would return a dataset with $F_M(S') \le \varepsilon < E_0$, while in the NO case every dataset satisfies $F_M(S') \ge E_0$. Thus we could distinguish the two cases in polynomial time by evaluating $F_M(S')$ and checking whether it is smaller than E_0. This would solve the NP-hard gap-Max-Cut problem. Therefore additive-ε minimization is NP-hard for every fixed $\varepsilon < E_0$. □

B SIMULATION DETAILS

For the W_1 baseline, we use the original PE update of González Lara et al. (2026): at each iteration, it forms the nearest-neighbor histogram of the sensitive data on the variation pool, adds Gaussian noise, and resamples the next synthetic population. The mean-matching method uses the private stochastic Frank–Wolfe selection from Algorithm 2. Both methods use $K = 10$ Gaussian variations per point and scale and the same public initialization and synthetic population size m in each repetition. We set m in our method according to the original PE prescription from González Lara et al. (2026).

For mean matching, we use the Gaussian scale grid from Theorem 4.1,

$$\sigma_\ell = \frac{1}{16}\frac{\sqrt{\log K}}{d}\,\alpha 2^{\ell-1}.$$

We use fewer outer iterations and a more accurate inner Frank–Wolfe solve than the conservative theoretical prescription, setting

$$T = \lceil T_{\mathrm{th}}/64 \rceil\,, \qquad Q = 32 Q_{\mathrm{bal}},$$

where

$$T_{\mathrm{th}} = \left\lceil \gamma^{-1} \log \frac{4R^2}{\alpha^2} \right\rceil, \qquad Q_{\mathrm{bal}} = \left\lceil \left(\frac{n\varepsilon}{\sqrt{T \log(2/\delta)} \log M} \right)^{2/3} \right\rceil,$$

with $\gamma = (\log K)/(64d)$ and $M = m(1 + KL)$. In the inner loop we use step-size $\eta_q = 2/(q + 3)$, so that the previous synthetic population retains positive weight after the first Frank–Wolfe step.

After the private inner loop, we project the private weighted mean onto $\mathcal{X}$ and reconstruct an m-point dataset with exactly that mean. This is pure post-processing and removes the additional Monte Carlo error introduced by sampling from the final weighted measure.

We consider $n \in \{5{,}000, 10{,}000, 20{,}000, 40{,}000\}$ and $\varepsilon \in \{1, 4\}$, with $\delta = 10^{-6}$. Each configuration is repeated ten times, and the error bars report 95% confidence intervals based on the standard error across repetitions.