Mixture-Greedy for Online Generative Model Selection: Is UCB Necessary in Diversity-Aware Multi-Armed Bandits?

Anonymous Authors[1]

Abstract

Efficient selection among multiple generative models is increasingly important in modern generative AI, where sampling from suboptimal models is costly. This problem can be formulated as a multi-armed bandit task. Under diversity-aware evaluation metrics, a non-degenerate mixture of generators can outperform any individual model, distinguishing this setting from classical best-arm identification. Prior approaches therefore incorporate an Upper Confidence Bound (UCB) exploration bonus into the mixture objective. However, across multiple datasets and evaluation metrics, we observe that the UCB term consistently slows convergence and often reduces sample efficiency. In contrast, a simple *Mixture-Greedy* strategy without explicit UCB-type optimism converges faster and achieves even better performance, particularly for widely used metrics such as FID and Vendi where tight confidence bounds are difficult to construct. We provide theoretical insight explaining this behavior: under transparent structural conditions, diversity-aware objectives induce implicit exploration by favoring interior mixtures, leading to linear sampling of all arms and sublinear regret guarantees for entropy-based, kernel-based, and FID-type objectives. These results suggest that in diversity-aware multi-armed bandits for generative model selection, exploration can arise intrinsically from the objective geometry, questioning the necessity of explicit confidence bonuses.

1. Introduction

The past decade has witnessed remarkable advances in generative modeling, leading to a wide availability of high-quality pretrained generators across diverse architectures and training paradigms. Early approaches such as variational autoencoders (VAEs) (Kingma & Welling, 2013) and generative adversarial networks (GANs) (Goodfellow et al., 2014) established scalable frameworks for learning complex data distributions. Subsequent architectural and training advances (Karras et al., 2019; Brock et al., 2019; Vahdat & Kautz, 2020), together with the emergence of diffusion-based generative models (Ho et al., 2020; Song et al., 2021; Rombach et al., 2022), have substantially improved both sample fidelity and diversity. As a result, practitioners often have access to multiple well-trained generative models that differ in architecture, training data, and inference cost.

In such settings, a fundamental practical question arises: how can one select or combine these generators in a sample-efficient manner? Generating samples from suboptimal models can be computationally expensive, and evaluation budgets are typically limited. Suppose we are given m candidate generators, where the ith generator $\mathcal{G}_i$ induces a distribution $P_{\mathcal{G}_i}$. At each round $t = 1, 2, \ldots, T$, we select a generator index $I_t \in [m]$, observe a sample $X_t \sim P_{\mathcal{G}_{I_t}}$, and update an estimate of a target evaluation score. The objective is to allocate samples so as to approach the best achievable evaluation performance while minimizing the number of samples drawn from suboptimal generators.

A natural formalization of this task is as a stochastic multi-armed bandit problem (Russo et al., 2018; Slivkins, 2019; Lattimore & Szepesvári, 2020). Standard bandit algorithms aim to balance exploration and exploitation in the online selection process. To do this, a well-established approach is the Upper Confidence Bound (UCB) principle (Auer et al., 2002), which augments empirical estimates with optimism bonuses to encourage exploration of uncertain arms. In the context of generative model selection, however, the evaluation criteria are typically non-linear functionals of the underlying distribution rather than simple expected rewards. For example, the Fréchet Inception Distance (FID) (Heusel et al., 2017) compares mean and covariance statistics of two distributions, and other widely used metrics such as Kernel Inception Distance (KID) (Binkowski et al., 2018) and precision-recall based measures (Sajjadi et al., 2018; Kynkäänniemi et al., 2019) also depend on higher-order

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the FoGen Workshop at ICML 2026. Do not distribute.

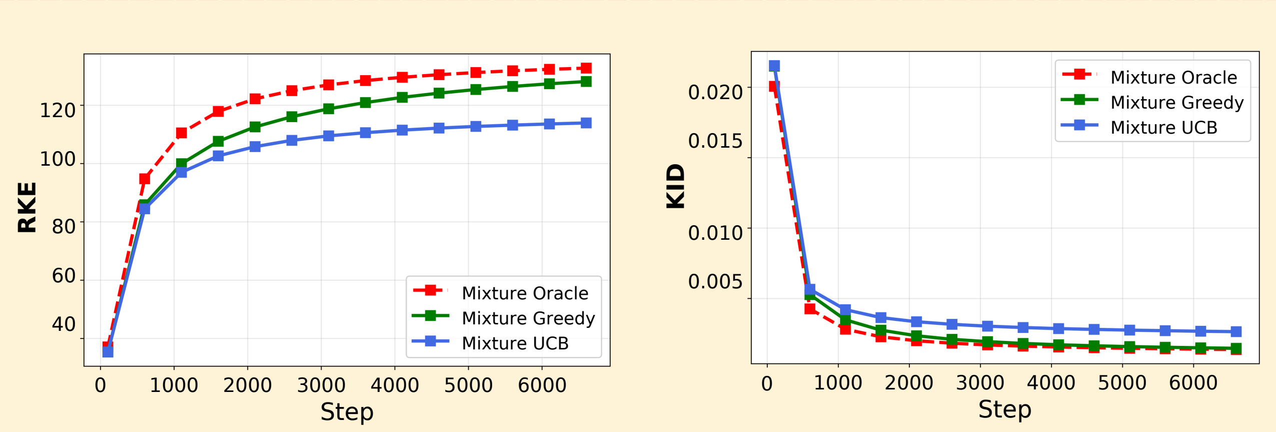

Figure 1. Comparison of mixture selection strategies over five ImageNet pretrained generative models from Stein et al. (2023)'s repository in terms of RKE↑ (left) and KID↓ (right) of generated samples from step 1 to t (plots' x-axis). Mixture-Greedy performed better than Mixture-UCB (Rezaei et al., 2025) (with the UCB exploration term) in both the experiments with RKE and KID objectives. Mixture-Oracle always generates samples from the population-optimal mixture of models.

distributional properties. To address this, Hu et al. (2025) proposed a bandit framework tailored to generative model evaluation and derived UCB-type confidence bounds for the FID metric, enabling online best-arm identification despite the non-linearity of the objective.

A notable finding in generative model selection is the recognition that the optimal solution need not be a single generator, i.e., the best individual arm in the bandit formulation considered by Hu et al. (2025). Rather, as shown by Rezaei et al. (2025), under several diversity-aware evaluation metrics, a non-degenerate mixture of generators can strictly outperform every individual model. Let $\alpha \in \Delta_m := \{\alpha \in \mathbb{R}^m : \alpha \succeq 0,\ \mathbf{1}^\top\alpha = 1\}$ denote mixture weights and define the mixture distribution $P_\alpha = \sum_{i=1}^m \alpha_i P_{\mathcal{G}_i}$. The mixture-model selection problem can then be formulated as

$$\alpha^\star \in \operatorname*{argmax}_{\alpha\in\Delta_m} \text{Score}\big(P_\alpha\big) \tag{1}$$

Unlike classical best-arm identification, the optimizer of (1) may lie in the interior of the simplex, implying that combining generators can strictly improve performance. To address this, Rezaei et al. (2025) proposed *Mixture-UCB* algorithms that operate over the simplex and incorporate *Upper Confidence Bound (UCB)* bonuses. For quadratic kernel-based objectives such as KID (Binkowski et al., 2018) and RKE (Jalali et al., 2023), they showed that uniform high-probability guarantees can be obtained by estimating a finite number of pairwise expectations, making UCB-style exploration analytically tractable in that regime. Extending such guarantees to non-quadratic metrics, including FID and entropy-based diversity scores such as the Vendi score (Friedman & Dieng, 2023), remains significantly more challenging, as the resulting UCB bound would be loose due to the non-quadratic structure, leading to over-exploration and substantially slow convergence to the optimal solution.

In this work, we revisit a foundational design choice in bandit-based generative model selection: Is an explicit UCB exploration bonus actually necessary? In classical best-arm bandits with linear rewards, purely greedy strategies are known to be brittle: early estimation noise can cause the algorithm to over-commit to an apparently good arm, stop collecting data from others, and never recover. UCB-style optimism was introduced precisely to prevent this collapse by forcing continued exploration (Auer et al., 2002). Surprisingly, we empirically find that this classical intuition does not necessarily carry over to diversity-aware *mixture* selection. Across extensive experiments on multiple datasets and evaluation metrics, adding the UCB bonus consistently slows convergence and often reduces sample efficiency. In contrast, a simple *Mixture-Greedy* strategy that repeatedly optimizes the empirical objective over the simplex Δ_m (without any explicit exploration bonus) converges faster and achieves comparable or superior performance, especially for widely used non-quadratic metrics such as FID and Vendi, where constructing UCB bounds is challenging.

Our explanation is that diversity-aware mixture objectives can induce *implicit exploration*. The main distinction from best-arm identification is that the decision space here is not the discrete set of arms but the continuous simplex of mixtures. For many diversity-aware metrics, the optimum is attained by a non-degenerate mixture, and the objective itself tends to disfavor degenerate solutions that place nearly all mass on a single generator. As a result, optimizing the empirical objective can naturally produce interior mixture weights, which in turn ensures that *every* generator continues to receive samples. In other words, the mixture geometry plus the diversity-aware objective can create an exploration effect that is internal to the optimization problem, rather than an external confidence bonus appended to it.

We formalize this phenomenon by providing theoretical results showing that, under explicit structural conditions, Mixture-Greedy stays uniformly away from the boundary of Δ_m, which yields linear sampling of all arms and sublinear regret without UCB. Our analyses cover representative families of objectives that matter in practice and were difficult to handle within prior UCB-based mixture frameworks: entropy-based objectives related to log-Vendi, FID-type Fréchet objectives under suitable interiority margin conditions, and kernel-based quadratic objectives including inverse-RKE-type formulations. Taken together with the empirical evidence, these results highlight a structural gap between diversity-aware mixture selection and classical linear-reward bandits: in the former, exploration may arise from the objective and the mixture structure themselves, suggesting that bandit methods for generative model selection should be designed around objective-induced geometry rather than relying solely on generic optimism bonuses. Here is a summary of this work's main contributions:

- Demonstrating empirically that removing the UCB exploration bonus accelerates convergence and improves sample efficiency in mixture-based generative model multi-armed bandits.
- Studying the theoretical nature of exploration in diversity-aware generative model selection and characterizing when explicit confidence bonuses are unnecessary.
- Developing sublinear regret guarantees for Mixture-Greedy under entropy-based, kernel-based, and FID-type objectives via implicit exploration induced by objective structure.

2. Related Works

Evaluation metrics: fidelity, diversity, and memorization/novelty. Generative models are routinely evaluated with embedding-based sample metrics, including Inception Score (IS) (Salimans et al., 2016), Fréchet Inception Distance (FID) (Heusel et al., 2017), and MMD/KID (Gretton et al., 2012; Binkowski et al., 2018), as well as precision–recall style decompositions that separate quality from coverage (Sajjadi et al., 2018; Kynkäänniemi et al., 2019; Naeem et al., 2020). Beyond distribution-level scores, recent work emphasizes sample-level diagnostics for memorization and novelty, including authenticity-style auditing (Alaa et al., 2022), feature-likelihood based generalization measures (Jiralerspong et al., 2023), and rarity score (Han et al., 2023).

For diversity-aware evaluation, kernel/entropy-based criteria, including the Vendi score (Friedman & Dieng, 2023) and RKE (Jalali et al., 2023), quantify diversity via similarity structure. Also, two extensions of the Vendi-based metrics are discussed in the literature. Quality-weighted variants incorporate per-item utility to trade off quality and diversity within a unified similarity-based objective (Nguyen & Dieng, 2024). Also, Pasarkar & Dieng (2024) generalize Vendi to a broader Renyi-entropy family, offering different sensitivity to prevalence and enabling application-specific diversity control. Additionally, (Ospanov & Farnia, 2025) study statistical convergence and proposes truncated variants with provable finite-sample convergence guarantees.

Another line of work highlights that these metrics inherit biases from the embedding space. Kynkäänniemi et al. (2023) show that FID is tightly coupled to ImageNet class geometry; Stein et al. (2023) demonstrate systematic mismatches between embedding-based scores and human evaluation. In our numerical analysis, we use CLIP (Radford et al., 2021b) and DINOv2 (Oquab et al., 2023) backbone embeddings following the results and suggestions of these references.

Online selection of generators, mixtures, and hyperparameters. Online allocation of sampling/evaluation budgets across pretrained generators can be cast as a stochastic bandit problem. Hu et al. (2025) derive UCB-style guarantees for online selection under FID-type objectives, and Rezaei et al. (2025) show that diversity-aware objectives may be optimized by *interior* mixtures, proposing Mixture-UCB methods with guarantees for quadratic kernel objectives. Chen & Ghosh (2024) view generative model hyperparameter search as adaptive resource allocation, proposing successive-Halving-style procedures guided by distributional tests. Our work focuses on a different design question: in diversity-aware *mixture* selection, is explicit optimism necessary, or can the objective geometry enforce sampling of all arms?

Greedy bandits: failures and implicit exploration under structure. In classical stochastic bandits with linear rewards, purely greedy policies can incur linear regret due to early mis-ordering, motivating explicit exploration principles such as UCB (Auer et al., 2002; Lattimore & Szepesvári, 2020). However, positive results show that greedy (or exploration-light) algorithms can achieve sublinear regret when the problem structure induces implicit exploration, including many-armed regimes (Bayati et al., 2020) and contextual/linear bandits under diversity or smoothing assumptions (Bastani et al., 2021; Kannan et al., 2018). Our analysis establishes an analogous phenomenon for diversity-aware *mixture* objectives in generative model selection.

3. Preliminaries

3.1. Generative model evaluation

A generative model $\mathcal{G}$ induces a probability distribution $P_{\mathcal{G}}$ over a sample space $\mathcal{X}$. We write $x \sim P_{\mathcal{G}}$ for a generated sample. In practice, evaluation is performed in a fixed embedding space. Throughout our analysis, we assume $x \in \mathbb{R}^d$ denotes the *embedded representation* of a sample. Mathematically, given i.i.d. generated samples $x_1, \ldots, x_n \sim P_{\mathcal{G}}$ and real samples $y_1, \ldots, y_{n_{\text{real}}} \sim P_{\text{data}}$ in the same embedding space, an evaluation metric computes a scalar score $\text{Score}(x_1, \ldots, x_n)$ reflecting fidelity and/or diversity.

Fréchet Distance (FD). Let $\widehat{\mu}_x, \widehat{\Sigma}_x$ denote the empirical mean and covariance matrix of $\{x_i\}_{i=1}^n$, and similarly $\widehat{\mu}_y, \widehat{\Sigma}_y$ for real data. The empirical Fréchet distance is

$$\text{FD}(\widehat{P}_G, \widehat{P}_{\text{data}}) = \|\widehat{\mu}_x - \widehat{\mu}_y\|_2^2 + \text{Tr}(\widehat{\Sigma}_x + \widehat{\Sigma}_y) - 2\,\text{Tr}\Big[(\widehat{\Sigma}_x^{1/2}\widehat{\Sigma}_y\widehat{\Sigma}_x^{1/2})^{1/2}\Big]. \quad (2)$$

3.2. Kernel-based scores

A kernel is a symmetric positive semidefinite function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Equivalently, there exists a Hilbert space $\mathcal{H}$ and feature map $\phi : \mathcal{X} \to \mathcal{H}$ such that

$$k(x, y) = \big\langle \phi(x), \phi(y) \big\rangle_{\mathcal{H}}.$$

Throughout, we assume the kernel is normalized: $k(x, x) = 1$ for every $x \in \mathcal{X}$. Note that even if a kernel function k is not normalized, we can replace k with the normalized $\widetilde{k}(x, y) = k(x, y)/\sqrt{k(x, x)k(y, y)}$ that remains a valid kernel function. Given samples $x_1, \ldots, x_n$, let $K \in \mathbb{R}^{n \times n}$ denote the kernel (Gram) matrix with entries defined as $K_{ij} = k(x_i, x_j)$ for every $1 \le i, j \le n$.

Kernel Distance (KD). KD is the squared maximum mean discrepancy (MMD) between P and Q:

$$\text{KD}(P, Q) = \mathbb{E}[k(X, X')] + \mathbb{E}[k(Y, Y')] - 2\,\mathbb{E}[k(X, Y)]\,. \quad (3)$$

where $X, X' \sim P$ and $Y, Y' \sim Q$ are independent. Empirical estimators replace expectations with sample averages.

Von Neumann entropy and Vendi. Define the normalized Gram matrix $\rho = \frac{1}{n}K$, that is PSD and unit-trace given a normalized kerk function. Let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues, which form a probability model as they are non-negative and add to 1. The von Neumann entropy (VNE) of this density matrix is defined as

$$\text{VNE}(\rho) = \sum_{i=1}^{n} \lambda_i \log \frac{1}{\lambda_i}$$

The Vendi score (Friedman & Dieng, 2023) is defined as the exponential of the VNE of $\frac{1}{n}K$

$$\text{Vendi}(x_1, \ldots, x_n) = \exp\Big(\text{VNE}\big(\tfrac{1}{n}K\big)\Big). \quad (4)$$

RKE and inverse-RKE. For a distribution P, (Jalali et al., 2023) define the RKE mode-count as follows:

$$\text{RKE}(P) = \frac{1}{\mathbb{E}_{X, X' \overset{\text{iid}}{\sim} P}\big[k^2(X, X')\big]},$$

which gives the Inverse-RKE definition as $\text{InvRKE}(P) = \mathbb{E}_{X, X' \sim P}[k(X, X')^2]$. For the empirical distribution $\widehat{P}_n$ supported on $\{x_i\}_{i=1}^n$, the above definitions reduce to $\text{InvRKE}(\widehat{P}_n) = \frac{1}{n^2}\|K\|_F^2$, and $\text{RKE}(\widehat{P}_n) = n^2/\|K\|_F^2$.

3.3. Multi-armed bandits for online generative model selection

We consider m pretrained generators $\mathcal{G}_1, \ldots, \mathcal{G}_m$, with corresponding distributions $P_{\mathcal{G}_1}, \ldots, P_{\mathcal{G}_m}$. At each round $t = 1, \ldots, T$, the learner selects $I_t \in [m]$ and observes $X_t \sim P_{\mathcal{G}_{I_t}}$. The objective is to maximize an evaluation score computed from the accumulated generated samples. This task can be naturally viewed as a stochastic multi-armed bandit (Slivkins, 2019; Lattimore & Szepesvári, 2020; Hu et al., 2025), where the online learner aims to identify the best model by successive queries to the models.

Beyond selecting a single generator, we consider identifying a mixture of the models. Let $\Delta_m := \{\alpha \in \mathbb{R}^m : \alpha_i \ge 0,\ \sum_{i=1}^m \alpha_i = 1\}$ be the probability simplex of weights assigned to the m models, and define the mixture distribution $P_\alpha = \sum_{i=1}^m \alpha_i P_{\mathcal{G}_i}$. The optimization problem is therefore (1), where standard choices for Score can be negative-FD, negative-KD, Vendi, RKE, or a composite score combining multiple scores, especially combining the pure-diversity scores with quality precision or density scores.

To address the task, Rezaei et al. (2025) propose the Mixture-UCB algorithm that adapts the established upper-confidence-bound (UCB) approach to the mixture setting. Note that the proposed Mixture-UCB in (Rezaei et al., 2025) is designed exclusively for quadratic evaluation scores, including KD, RKE, and their linear combinations with Precision and Density fidelity scores. Furthermore, another natural algorithm that we will analyze in our work is *Mixture-Greedy*, which assigns no UCB bonus term and at each round optimizes the empirical objective over Δ_m and samples I_t from the resulting mixture α_t. We highlight that the Mixture-Greedy approach is directly applicable to all the evaluation scores discussed in this section.

4. Methodology: Mixture-Greedy & score-based simplex program

We consider m pretrained generators $\mathcal{G}_1, \ldots, \mathcal{G}_m$ with induced distributions $P_{\mathcal{G}_i}$ over embedded samples in $\mathbb{R}^d$. At round t, Mixture-Greedy forms an empirical loss $\widehat{\mathcal{L}}_{t-1}(\alpha)$ over the simplex $\Delta_m := \{\alpha \in \mathbb{R}^m : \alpha \succeq 0,\ \mathbf{1}^\top \alpha = 1\}$ and selects

$$\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{\mathcal{L}}_{t-1}(\alpha). \tag{5}$$

It then samples an arm $I_t \sim \alpha_t$ and draws $X_t \sim P_{\mathcal{G}_{I_t}}$.

We focus on three objective families used throughout the paper: (i) Fréchet distance (FD) via Gaussian moment matching, (ii) negative log-Vendi (equivalently, negative von Neumann entropy), and (iii) convex quadratic scores (KD/MMD, inverse-RKE), possibly combined with a linear fidelity term (Precision/Density-type).

(i) FD via Gaussian moment matching. Fix the empirical real-data moments $(\widehat{\mu}_{\text{data}}, \widehat{\Sigma}_{\text{data}})$ in the embedding space. Let $\widehat{\mu}_i(t) \in \mathbb{R}^d$ and $\widehat{S}_i(t) \in \mathbb{R}^{d\times d}$ denote, respectively, the empirical mean and uncentered second moment of samples from arm i up to time t: $\widehat{S}_i(t) = \frac{1}{n_i(t)} \sum_{r=1}^{n_i(t)} x_{i,r} x_{i,r}^\top \succeq 0$. For $\alpha \in \Delta_m$, define the mixture mean and covariance via the exact mixture identities

$$\widehat{\mu}_t(\alpha) := \sum_{i=1}^m \alpha_i \widehat{\mu}_i(t), \tag{6}$$

$$\widehat{\Sigma}_t(\alpha) := \sum_{i=1}^m \alpha_i \widehat{S}_i(t) - \widehat{\mu}_t(\alpha)\widehat{\mu}_t(\alpha)^\top. \tag{7}$$

The FD-type loss is then the Gaussian Fréchet functional

$$\widehat{\mathcal{L}}_t^{\text{FD}}(\alpha) := \|\widehat{\mu}_t(\alpha) - \widehat{\mu}_{\text{data}}\|_2^2 + \operatorname{Tr}\Big(\widehat{\Sigma}_t(\alpha)\Big) + \operatorname{Tr}\Big(\widehat{\Sigma}_{\text{data}}\Big) - 2\operatorname{Tr}\left[\Big(\widehat{\Sigma}_{\text{data}}^{1/2} \widehat{\Sigma}_t(\alpha) \widehat{\Sigma}_{\text{data}}^{1/2}\Big)^{1/2}\right]. \tag{8}$$

(ii) Negative log-Vendi from pooled samples. Let $\{x_j\}_{j=1}^N$ be the pooled generated samples available at time t, where each x_j came from arm $I_j \in [m]$, and let $K \in \mathbb{R}^{N\times N}$ be the normalized kernel Gram matrix $K_{rs} = k(x_r, x_s)$. For $\alpha \in \Delta_m$, define per-sample weights

$$q_j(\alpha) := \frac{\alpha_{I_j}}{n_{I_j}(t)}, \qquad j \in [N], \tag{9}$$

so $q(\alpha) \succeq 0$ and $\sum_{j=1}^N q_j(\alpha) = 1$. Consider the following unit-trace PSD matrix

$$\rho_t(\alpha) := \operatorname{diag}(q(\alpha))^{1/2}\, K \operatorname{diag}(q(\alpha))^{1/2}, \tag{10}$$

where we note that $\operatorname{Tr}(\rho_t(\alpha)) = 1$ holds and we set the negative log-Vendi loss as

$$\widehat{\mathcal{L}}_t^{\text{Vendi}}(\alpha) := \operatorname{Tr}\big(\rho_t(\alpha) \log \rho_t(\alpha)\big). \tag{11}$$

Minimizing (11) is equivalent to maximizing $\operatorname{VNE}(\rho_t(\alpha))$ and hence the Vendi score of the generated samples.

(iii) Convex quadratics and linear fidelity. Let $\widehat{K}(t) \in \mathbb{R}^{m\times m}$ be symmetric PSD and $\widehat{b}(t) \in \mathbb{R}^m$ estimate a linear functional $b_i = \mathbb{E}[\psi_i(X)]$ with $\psi_i \in [0,1]$. We consider

$$\widehat{\mathcal{L}}_t^{\text{quad}}(\alpha) := \alpha^\top \widehat{K}(t)\alpha + w\,\alpha^\top \widehat{b}(t), \qquad w \ge 0. \tag{12}$$

The following proposition shows that the above loss functions yield convex functions of the mixture weights, which can be optimized by convex optimization algorithms.

Proposition 4.1. *For each fixed t, the update* (5) *is a convex optimization problem over Δ_m for: (i) $\widehat{\mathcal{L}}_t^{\text{FD}}(\alpha)$ in* (8)*, (ii) $\widehat{\mathcal{L}}_t^{\text{Vendi}}(\alpha)$ in* (11)*, and (iii) $\widehat{\mathcal{L}}_t^{\text{quad}}(\alpha)$ in* (12) *whenever $\widehat{K}(t) \succeq 0$.*

Proof. We provide the proof in the Appendix. □

Exponentiated-gradient (EG) solver and practical instantiation. To implement the Mixture-Greedy update (5) in an efficient manner on the simplex, we apply the exponentiated gradient (EG) descent, a mirror-descent method with KL geometry that maintains $\alpha \in \Delta_m$ by construction. Starting from a feasible iterate $\alpha^{(0)}$ (we use the uniform distribution $\frac{1}{m}\mathbf{1}$ in our experiments), every EG step updates

$$\widetilde{\alpha}_i^{(s+1)} = \alpha_i^{(s)} \exp\Big(-\eta g_i^{(s)}\Big), \qquad \alpha^{(s+1)} = \frac{\widetilde{\alpha}^{(s+1)}}{\big\|\widetilde{\alpha}^{(s+1)}\big\|_1}. \tag{13}$$

where $g^{(s)} = \nabla_\alpha \widehat{\mathcal{L}}_{t-1}(\alpha^{(s)})$ and $\eta > 0$ is a stepsize. In our setting, the objective families in this section lead to smooth convex losses over Δ_m (Proposition 4.1); therefore, a fixed number of EG steps per round would be sufficient. Algorithm 1 summarizes the main steps.

Score-specific gradients. For quadratic objectives (12), the gradient is explicit: $\nabla_\alpha \widehat{\mathcal{L}}_t^{\text{quad}}(\alpha) = 2\widehat{K}(t)\alpha + w\,\widehat{b}(t)$. For FD (8) and log-Vendi (11), gradients require matrix-function evaluations (e.g., a matrix square-root/inverse-square-root for FD, and a matrix logarithm for log-Vendi). As we discuss, one can compute these via eigendecomposition. To reduce cost for log-Vendi when the pooled sample size is significantly large, we also use a finite-dimensional feature representation (e.g., random Fourier features for shift-invariant kernels), replacing the $N \times N$ kernel-matrix logarithm by a $D \times D$ feature-covariance logarithm when $D \ll N$. The full derivations are provided in Appendix A.

5. Implicit exploration and regret for Diversity Scores

We analyze Mixture-Greedy for two diversity-aware objectives: (i) negative log-Vendi (negative von Neumann entropy

Algorithm 1 Mixture-Greedy via Exponentiated Gradient

```
Require: Models {G_i}_{i=1}^m, horizon T, warm start M ≥ 1, EG stepsize η > 0, EG steps S ≥ 1
Initialize counts n_i(0) = M for all i ∈ [m], and draw M warm-start samples from each model to initialize statistics
for t = 1 to T do
    Form the empirical loss \hat{L}_{t−1}(α) from samples up to round t − 1
    Initialize α^(0) ← α_{t−1} (warm start) or α^(0) ← (1/m) 1
    for s = 0 to S − 1 do
        Compute g^(s) ← ∇_α \hat{L}_{t−1}(α^(s))
        \tilde{α}_i^(s+1) ← α_i^(s) exp(−η g_i^(s)) for all i ∈ [m]
        α^(s+1) ← \tilde{α}^(s+1) / Σ_{j=1}^m \tilde{α}_j^(s+1)
    end for
    Set α_t ← α^(S), sample I_t ∼ α_t, draw X_t ∼ P_{G_{I_t}}
    Update count n_{I_t}(t) ← n_{I_t}(t − 1) + 1, and update the empirical statistics needed for \hat{L}_t
end for
```

of the kernel matrix), and (ii) the Fréchet distance functional under mean and covariance matching. Our goal is to characterize structural conditions under which Mixture-Greedy, without any explicit UCB bonus, achieves sublinear regret.

Regarding the bandit protocol and regret definition, we consider a warm start of $M \geq 1$ samples per arm (i.e., $n_i(0) = M$). At each round $t = 1, \dots, T$, Mixture-Greedy computes

$$\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{F}_{t-1}(\alpha), \qquad I_t \sim \alpha_t, \tag{14}$$

where $\widehat{F}_{t-1}$ is the plug-in empirical objective formed from samples up to time $t-1$. We measure regret against the best fixed mixture:

$$\mathrm{Reg}_T := \sum_{t=1}^{T} \Big[F(\alpha_t) - \min_{\alpha \in \Delta_m} F(\alpha) \Big]. \tag{15}$$

5.1. Negative log-Vendi induces implicit exploration

Let $\phi : \mathcal{X} \to \mathbb{R}^d$ be a finite-dimensional feature map. Define $S_i := \mathbb{E}[\phi(X)\phi(X)^\top]$ for $X \sim P_{\mathcal{G}_i}$ and $S_\alpha := \sum_{i=1}^m \alpha_i S_i$. The negative log-Vendi objective is

$$F_{\mathrm{NLV}}(\alpha) := \mathrm{Tr}(S_\alpha \log S_\alpha).$$

Assumption 5.1 (Normalized kernel)**.** $\|\phi(x)\|_2 = 1$ for all x.

Assumption 5.2 (Population innovation for log-Vendi)**.** There exist constants $\nu_0 \in (0,1]$ and $\varepsilon_0 \in [0, \nu_0/8)$ such that for every $i \in [m]$ there exists a unit vector v_i satisfying $v_i^\top S_i v_i \geq \nu_0$ and $\sum_{j \neq i} v_i^\top S_j v_i \leq \varepsilon_0$.

Under these structural conditions, the entropy geometry enforces a uniform interiority property for empirical minimizers. This yields linear sampling of all arms and enables time-uniform concentration.

Theorem 5.3 (Mixture-Greedy regret for negative log-Vendi)**.** *Fix $T \geq 1$ and $\delta \in (0,1)$. Under Assumptions 5.1–5.2, there exists a warm-start size $M = M(d, m, \nu_0, \varepsilon_0, T, \delta)$ such that, with probability at least $1-\delta$ the following holds, where C_{NLV} depends only on $(d, m, \nu_0, \varepsilon_0)$:*

$$\mathrm{Reg}_T \;\leq\; C_{\mathrm{NLV}} \Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}} \Big) \sqrt{T}\,(1 + \log T),$$

Proof. We provide the proof in the Appendix. □

5.2. Fréchet Distance: regret under a population interiority margin

The Fréchet Distance functional is smooth at the simplex boundary and may admit boundary minimizers. We therefore state a population interiority condition ensuring that optimal mixtures lie in the interior.

Let arm i produce i.i.d. embeddings $Z \in \mathbb{R}^d$ with mean μ_i and covariance Σ_i. For $\alpha \in \Delta_m$ define

$$\mu_\alpha = \sum_{i=1}^m \alpha_i \mu_i, \;\; \Sigma_\alpha = \sum_{i=1}^m \alpha_i \Big(\Sigma_i + (\mu_i - \mu_\alpha)(\mu_i - \mu_\alpha)^\top \Big)$$

Fix (μ_0, Σ_0) with $\Sigma_0 \succeq \lambda_0 I$ and define

$$F_{\mathrm{FD}}(\alpha) := \|\mu_\alpha - \mu_0\|_2^2 + \mathrm{Tr}\Big(\Sigma_\alpha + \Sigma_0 - 2(\Sigma_0^{1/2} \Sigma_\alpha \Sigma_0^{1/2})^{1/2} \Big)$$

Assumption 5.4 (Bounded embeddings)**.** There exists $B < \infty$ such that $\|Z\|_2 \leq B$ almost surely for every arm.

Assumption 5.5 (Uniform positive definiteness)**.** There exists $\nu > 0$ such that $\Sigma_i \succeq \nu I$ for all $i \in [m]$.

Assumption 5.6 (Population interiority margin)**.** There exist $\gamma_0 \in (0, 1/m]$ and $\Delta_0 > 0$ such that

$$\inf_{\alpha \in \Delta_m :\, \min_i \alpha_i \leq \gamma_0} F_{\mathrm{FD}}(\alpha) \;\geq\; \min_{\alpha \in \Delta_m} F_{\mathrm{FD}}(\alpha) + \Delta_0.$$

Theorem 5.7 (Mixture-Greedy regret for Fréchet Distance)**.** *Fix $T \geq 1$ and $\delta \in (0,1)$. Under Assumptions 5.4–5.6, there exists $M = M(B, \lambda_0, \nu, m, d, \gamma_0, \Delta_0, T, \delta)$ such that, with probability at least $1-\delta$, where C_{FD} depends only on $(B, \lambda_0, \nu, m, d, \gamma_0)$:*

$$\mathrm{Reg}_T \;\leq\; C_{\mathrm{FD}} \Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}} \Big) \sqrt{T}$$

Proof. We provide the proof in the Appendix. □

Remark 5.8. Our analysis extends to the RKE diversity objective, $\mathrm{RKE}(\alpha) = 1/\mathrm{Tr}(S_\alpha^2)$, optimized via the equivalent convex functional $F_{\mathrm{RKE}}(\alpha) = \mathrm{Tr}(S_\alpha^2)$. Under a standard population margin condition analogous to Assumption 5.6, Mixture-Greedy achieves an $\widetilde{\mathcal{O}}(\sqrt{T})$ regret bound. The same rate holds when the objective is augmented with linear fidelity terms. The statements and proofs are in Appendix D.

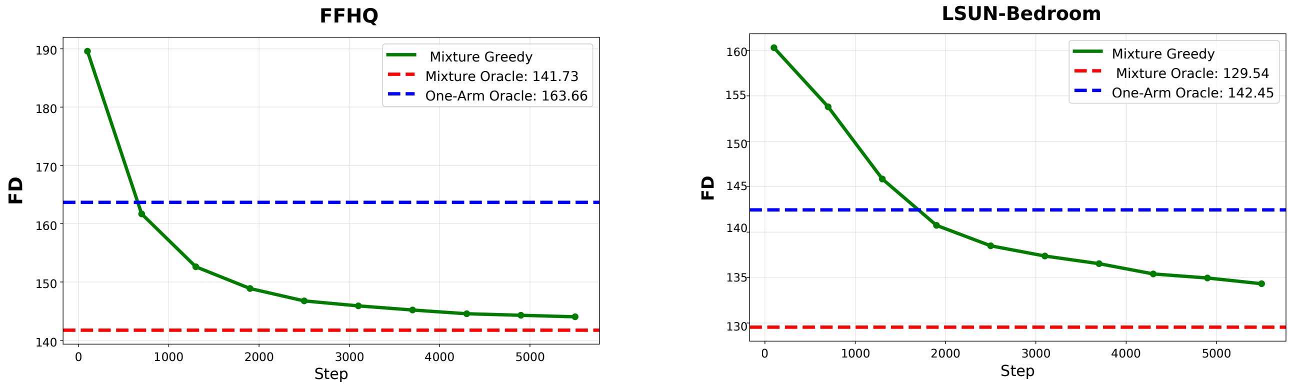

Figure 2. Convergence of Mixture Greedy to the Mixture Oracle in terms of Fréchet Distance (FD) across image selection steps on FFHQ (left) and LSUN-Bedroom (right).

5.3. Including a linear fidelity term

The fidelity scores for generative models, including Precision (Kynkäänniemi et al., 2019) and Density (Naeem et al., 2020) are linear functionals of the generator distribution. We model this using bounded functions $\psi_i : \mathcal{X} \to [0,1]$:

$$G(\alpha) := \sum_{i=1}^{m} \alpha_i \theta_i, \quad \theta_i := \mathbb{E}_{X\sim P_{\mathcal{G}_i}}[\psi_i(X)] \in [0,1],$$

and for a base diversity-aware objective $F(\alpha)$, we consider $H(\alpha) := F(\alpha) + w\,G(\alpha)$ for given coefficient $w > 0$. Given this objective, Mixture-Greedy minimizes the plug-in empirical objective $\widehat{H}_{t-1}(\alpha) = \widehat{F}_{t-1}(\alpha) + w\,\widehat{G}_{t-1}(\alpha)$.

Negative log-Vendi. The linear increment along any simplex direction is uniformly bounded, thus the entropy-induced interiority mechanism remains intact, with the quantitative floor adjusted by a controlled w-dependent term.

Corollary 5.9 (Mixture-Greedy regret for negative log-Vendi + linear term)**.** *Assume the conditions of Theorem 5.3 and $0 \le \psi_i \le 1$. For M sufficiently large (see Appendix C), with probability at least $1-\delta$,*

$$\sum_{t=1}^{T}\Big(H(\alpha_t) - \min_{\alpha\in\Delta_m} H(\alpha)\Big) \le C_{\mathrm{NLV},w}\sqrt{T\log\frac{m(T+1)}{\delta}},$$

where $C_{\mathrm{NLV},w}$ depends only on $(d, m, \nu_0, \varepsilon_0, w)$.

Fréchet Distance. Since $G(\alpha) \in [0,1]$, adding wG perturbs any interiority margin by at most w, hence the Fréchet Distance guarantee extends under the corresponding margin condition for H.

Corollary 5.10 (Mixture-Greedy regret for Fréchet Distance + linear term)**.** *Assume the conditions of Theorem 5.7 and $0 \le \psi_i \le 1$. If $H(\alpha) = F_{\mathrm{FD}}(\alpha) + wG(\alpha)$ satisfies the population interiority margin (Assumption 5.6 with F_{FD} replaced by H), then with probability at least $1-\delta$,*

$$\sum_{t=1}^{T}\Big(H(\alpha_t) - \min_{\alpha\in\Delta_m} H(\alpha)\Big) \le C_{\mathrm{FD},w}\sqrt{T\log\Big(\frac{m(T+1)}{\delta}\Big)},$$

where $C_{\mathrm{FD},w}$ depends only on $(B, \lambda_0, \nu, m, d, \gamma_0, w)$.

6. Numerical Results

We assess the performance of the proposed approach in both real-data and synthetic environments. For the real-data experiments, we adopt benchmark datasets standard in (Stein et al., 2023). Complementarily, we constructed synthetic scenarios with controlled data-generating mechanisms. Please refer to the Appendix E.1 for details and results. For image feature extraction, we use DINOv2-ViT-L/14 (Oquab et al., 2023) following the study by Stein et al. (2023). For text encoding, we use SBERT (Reimers & Gurevych, 2019).

6.1. Real-World Settings

In the real-data setting, we evaluate our method on samples from pretrained generative models provided by the evaluation benchmark in (Stein et al., 2023). Each generative model is treated as an arm in our experimental framework. Here, we consider three datasets: FFHQ, LSUN-Bedroom, and ImageNet. For each dataset, we treat pretrained generative models as arms in our framework. All models are pre-trained on the dataset provided by Stein et al. (2023).

FFHQ. For the FFHQ dataset (Karras et al., 2019), we considered five distinct generative models: LDM (Rombach et al., 2022), StyleGAN-XL (Sauer et al., 2022), Efficient-VDVAE (Hazami et al., 2022), InsGen (Yang et al., 2021), and StyleNAT (Walton et al., 2022).

LSUN-Bedroom. For the LSUN-Bedroom dataset, we used generated samples from four models: StyleGAN (Karras

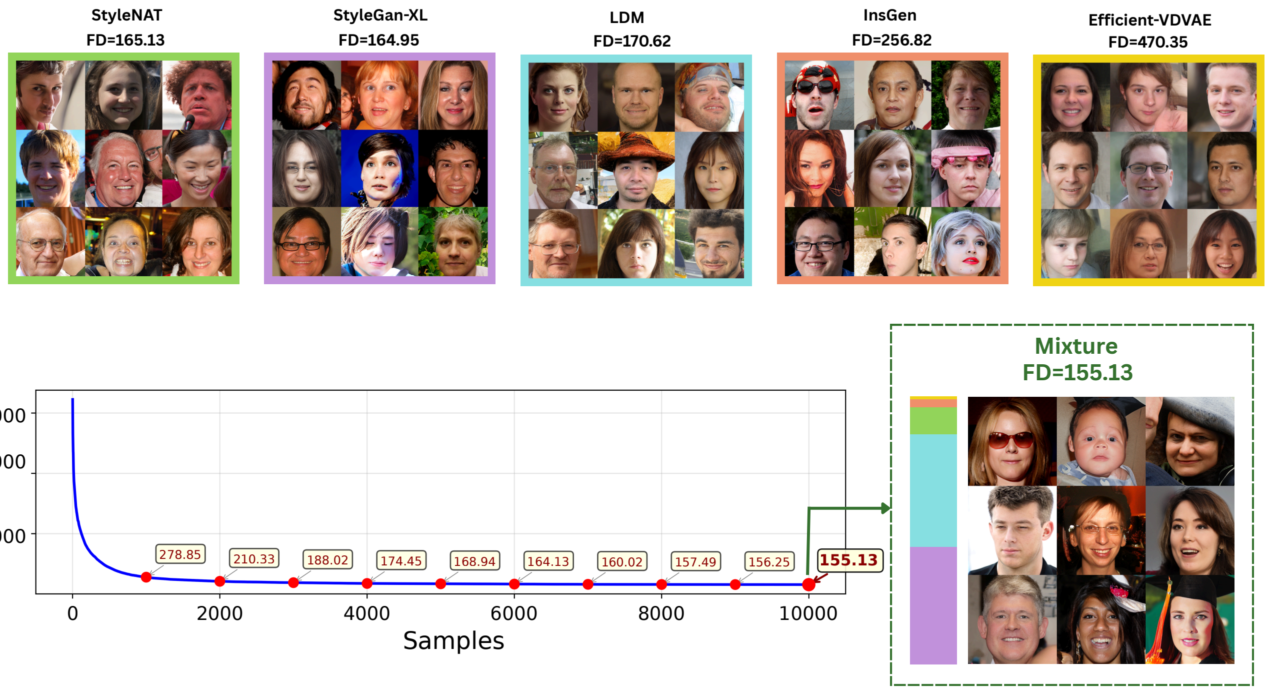

Figure 3. Mixture Greedy, minimizing Frechet Distance using exponentiated gradient descent on FFHQ generated samples.

et al., 2019), Projected GAN (Sauer & Geiger, 2021), iDDPM (Nichol & Dhariwal, 2021), and Unleashing Transformers (Bond-Taylor et al., 2022).

ImageNet. For ImageNet, we evaluate four pretrained generative models: DiT-XL-2-guided (Peebles & Xie, 2023), LDM (Rombach et al., 2022), RQ-Transformer (Lee et al., 2022), and StyleGAN-XL (Sauer et al., 2022).

Mixture Improves Fréchet Distance. Figure 2 shows the evolution of FD across selection rounds on two datasets. Alongside the online Mixture Greedy algorithm, we report two oracle baselines: the One-Arm Oracle, which samples exclusively from the single generator with the lowest standalone FD, and the Mixture Oracle, which assumes knowledge of the optimal mixture weights $\alpha^\star$ and provides the corresponding lower bound on FD.

Across both datasets, Mixture Greedy rapidly decreases FD and converges to the Mixture Oracle within a few rounds, consistently outperforming the One-Arm Oracle. This highlights the advantage of optimizing over generator mixtures rather than selecting a single best generator. See Figure 4 in Appendix for ImageNet results. Figure 3 illustrates the FD values obtained from individual generative models (arms) trained on FFHQ, as well as the FD achieved by their optimized mixture. While the best single model attains an FD of 164.13 (StyleGAN-XL), the samples from the mixture achieves a substantially lower FD of 155.13, outperforming every individual arm. The optimization is performed using exponentiated gradient descent.

7. Conclusion

We studied mixture-based online selection of generative models, and our analysis indicates that, for a broad class of diversity-aware objectives, explicit optimism is not always necessary. Conceptually, our results highlight that exploration can arise intrinsically from the structure of the objective itself, with no need for externally imposed confidence bonuses. In regimes where diversity metrics reward interior mixtures or penalize degeneracy, the optimization landscape naturally prevents collapse onto a single arm and promotes sufficient sampling of all generators. Our numerical experiments support this message: across multiple datasets and metrics, Mixture-Greedy could converge faster than the Mixture-UCB baseline. Indeed, a Mixture-Greedy strategy could achieve sublinear regret under transparent structural conditions. We note that our analysis is conducted in a specific stochastic setting with a fixed pool of generators and objective-specific regularity assumptions, and it focuses on regret with respect to the chosen evaluation metric rather than broader notions of downstream utility. Extending these guarantees to more dynamic and nonstationary settings remains a relevant direction for future work.

References

Aaron Grattafiori, A. D. et al. The llama 3 herd of models, 2024.

Alaa, A., van Breugel, B., Saveliev, E. S., and van der

Schaar, M. How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *Proceedings of Machine Learning Research*, pp. 290–306. PMLR, 2022. URL https://proceedings.mlr.press/v162/alaa22a.html.

Audenaert, K. M. R. A sharp fannes-type inequality for the von neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007. doi: 10.1088/1751-8113/40/28/S18.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002.

Bastani, H., Bayati, M., and Khosravi, K. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 2021. doi: 10.1287/mnsc.2020.3605.

Bayati, M., Hamidi, N., Johari, R., and Khosravi, K. Unreasonable effectiveness of greedy algorithms in multi-armed bandit with many arms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Binkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. Demystifying MMD GANs. In *International Conference on Learning Representations (ICLR)*, 2018.

Bond-Taylor, S., Leach, A., Long, Y., and Willcocks, C. G. Unleashing transformers for image generation. In *International Conference on Learning Representations (ICLR)*, 2022.

Brock, A., Donahue, J., and Simonyan, K. Large scale gan training for high fidelity natural image synthesis. In *International Conference on Learning Representations (ICLR)*, 2019.

Chen, J., Yu, J., Ge, C., Yao, L., Xie, E., Wu, Y., Wang, Z., Kwok, J., Luo, P., Lu, H., and Li, Z. Pixart-α: Fast training of diffusion transformer for photorealistic text-to-image synthesis. *arXiv preprint arXiv:2310.00426*, 2023. URL https://arxiv.org/abs/2310.00426.

Chen, L. and Ghosh, S. K. Fast model selection and hyperparameter tuning for generative models. *Entropy*, 26 (2):150, 2024. doi: 10.3390/e26020150. URL https://www.mdpi.com/1099-4300/26/2/150.

Friedman, D. and Dieng, A. B. The vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023.

Gemma-Team. Gemma 2 technical report, 2025.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2014.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

Han, J., Choi, H., Choi, Y., Kim, J., Ha, J.-W., and Choi, J. Rarity score: A new metric to evaluate the uncommonness of synthesized images. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=JTGimap_-F.

Hazami, L., Comon, P., et al. Efficient-vdvae: Less is more. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Hu, X., Leung, H., and Farnia, F. A multi-armed bandit approach to online selection and evaluation of generative models. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Proceedings of Machine Learning Research. PMLR, 2025. URL https://proceedings.mlr.press/v258/hu25a.html.

Jalali, M., Li, C. T., and Farnia, F. An information-theoretic evaluation of generative models in learning multi-modal distributions. *Advances in Neural Information Processing Systems*, 36:9931–9943, 2023.

Jiralerspong, M., Bose, A. J., Gemp, I., Qin, C., Bachrach, Y., and Gidel, G. Feature likelihood divergence: Evaluating the generalization of generative models using samples. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Kannan, S., Morgenstern, J., Roth, A., Waggoner, B., and Wu, Z. S. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Kynkäänniemi, T., Karras, T., Aittala, M., Aila, T., and Lehtinen, J. The role of imagenet classes in fréchet inception distance. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=4oXTQ6m_ws8.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Lee, D., Kim, J., and Kim, J. Autoregressive image generation using residual quantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Naeem, M., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J.-J. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning (ICML)*, 2020.

Nguyen, Q. and Dieng, A. B. Quality-weighted vendi scores and their application to diverse experimental design. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 37667–37682. PMLR, 2024.

Nichol, A. Q. and Dhariwal, P. Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P. Dinov2: Learning robust visual features without supervision, 2024.

Ospanov, A. and Farnia, F. Do vendi scores converge with finite samples? truncated vendi score for finite-sample convergence guarantees. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 3272–3299. PMLR, 2025.

Pasarkar, A. P. and Dieng, A. B. Cousins of the Vendi score: A family of similarity-based diversity metrics for science and machine learning. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3808–3816. PMLR, 2024.

Peebles, W. and Xie, S. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023a. URL https://arxiv.org/abs/2307.01952.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023b.

Qwen-Team. Qwen2 technical report, 2024. URL https://arxiv.org/abs/2407.10671.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. Learning transferable visual models from natural language supervision, 2021a.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021b.

Razzhigaev, A., Arkhipkin, V., et al. Kandinsky: An improved text-to-image synthesis with image prior and latent diffusion. *arXiv preprint arXiv:2310.03502*, 2023.

Reimers, N. and Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks, 2019. URL https://arxiv.org/abs/1908.10084.

Rezaei, P., Farnia, F., and Li, C. T. Be more diverse than the most diverse: Optimal mixtures of generative models via mixture-ucb bandit algorithms. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=2Chkk5Ye2s.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. A tutorial on thompson sampling. *Foundations* *and Trends® in Machine Learning*, 11(1):1–96, 2018. doi: 10.1561/2200000070.

Sajjadi, M. S., Bachem, O., Lucic, M., Bousquet, O., and Gelly, S. Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31, 2018.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved techniques for training gans. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Sauer, A. and Geiger, A. Projected gans converge faster. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Sauer, A., Schwarz, K., and Geiger, A. Stylegan-xl: Scaling stylegan to large diverse datasets. In *Proceedings of the ACM SIGGRAPH Conference*, 2022.

Slivkins, A. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Stein, G., Cresswell, J. C., Hosseinzadeh, R., Sui, Y., Ross, B. L., Villecroze, V., Liu, Z., Caterini, A. L., Taylor, J. E. T., and Loaiza-Ganem, G. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Sutherland, D. J., Strathmann, H., Arbel, M., and Gretton, A. Efficient and principled score estimation with nyström kernel exponential families. In *International Conference on Artificial Intelligence and Statistics*, pp. 652–660. PMLR, 2018.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Vahdat, A. and Kautz, J. Nvae: A deep hierarchical variational autoencoder. *Advances in neural information processing systems*, 33:19667–19679, 2020.

Walton, S. et al. Stylenat: Giving each head a new perspective. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2022.

Yang, T. et al. Instance-conditioned gan. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

A. Proofs and gradient computations for Section 4

A.1. Proof of Proposition 4.1

In the following, we prove convexity for each of the three objective families.

Part (i): FD convexity in α. Fix $(\widehat{\mu}_{\text{data}}, \widehat{\Sigma}_{\text{data}})$ and the per-arm empirical moments $\{\widehat{\mu}_i(t), \widehat{S}_i(t)\}_{i=1}^m$. Define $\widehat{\mu}(\alpha) = \widehat{\mu}_t(\alpha)$ and $\widehat{\Sigma}(\alpha) = \widehat{\Sigma}_t(\alpha)$ as in (6)–(7). Let $\Sigma_0 := \widehat{\Sigma}_{\text{data}} \succeq 0$ and $\mu_0 := \widehat{\mu}_{\text{data}}$.

We rewrite the FD loss (8) as

$$\widehat{\mathcal{L}}_t^{\text{FD}}(\alpha) = \|\widehat{\mu}(\alpha) - \mu_0\|_2^2 + \text{Tr}(\widehat{\Sigma}(\alpha)) + \text{Tr}(\Sigma_0) - 2\,\text{Tr}\Big((\Sigma_0^{1/2}\widehat{\Sigma}(\alpha)\Sigma_0^{1/2})^{1/2}\Big). \tag{16}$$

First, we note that the mean plus trace part is affine in α. Using $\widehat{\Sigma}(\alpha) = \sum_{i=1}^m \alpha_i \widehat{S}_i - \widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top$ and $\text{Tr}(\widehat{\mu}\widehat{\mu}^\top) = \|\widehat{\mu}\|_2^2$,

$$\begin{aligned}\|\widehat{\mu}(\alpha) - \mu_0\|_2^2 + \text{Tr}(\widehat{\Sigma}(\alpha)) &= \|\mu_0\|_2^2 - 2\mu_0^\top \widehat{\mu}(\alpha) + \|\widehat{\mu}(\alpha)\|_2^2 \\ &\quad + \text{Tr}\Big(\sum_{i=1}^m \alpha_i \widehat{S}_i\Big) - \text{Tr}(\widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top) \\ &= \|\mu_0\|_2^2 - 2\mu_0^\top \widehat{\mu}(\alpha) + \text{Tr}\Big(\sum_{i=1}^m \alpha_i \widehat{S}_i\Big),\end{aligned}$$

where the $\|\widehat{\mu}(\alpha)\|_2^2$ terms cancel exactly. Since $\widehat{\mu}(\alpha)$ and $\sum_i \alpha_i \widehat{S}_i$ are affine in α, this entire expression is affine in α.

Next, we show that $\widehat{\Sigma}(\alpha)$ is concave in Loewner order. To do this, let $\alpha, \beta \in \Delta_m$ and $\theta \in [0,1]$. Since $\sum_i \alpha_i \widehat{S}_i$ is affine, it suffices to show that $\widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top$ is convex in Loewner order. Because $\widehat{\mu}(\cdot)$ is affine, we have

$$\begin{aligned}&\theta\widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top + (1-\theta)\widehat{\mu}(\beta)\widehat{\mu}(\beta)^\top - \widehat{\mu}(\theta\alpha + (1-\theta)\beta)\widehat{\mu}(\theta\alpha + (1-\theta)\beta)^\top \\ &= \theta(1-\theta)\big(\widehat{\mu}(\alpha) - \widehat{\mu}(\beta)\big)\big(\widehat{\mu}(\alpha) - \widehat{\mu}(\beta)\big)^\top \succeq 0.\end{aligned}$$

Thus $\widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top$ is Loewner-convex, and therefore $\widehat{\Sigma}(\alpha) = \sum_i \alpha_i \widehat{S}_i - \widehat{\mu}(\alpha)\widehat{\mu}(\alpha)^\top$ is Loewner-concave:

$$\widehat{\Sigma}(\theta\alpha + (1-\theta)\beta) \;\succeq\; \theta\widehat{\Sigma}(\alpha) + (1-\theta)\widehat{\Sigma}(\beta).$$

Then, we show the concavity of the square-root trace term. We define $Z(\alpha) := \Sigma_0^{1/2}\widehat{\Sigma}(\alpha)\Sigma_0^{1/2}$. Pre- and post-multiplication by a fixed PSD matrix preserves Loewner inequalities, hence $Z(\alpha)$ is Loewner-concave. The matrix square-root map $M \mapsto M^{1/2}$ is operator concave and operator monotone on the PSD cone. Therefore, for $\theta \in [0,1]$,

$$Z(\theta\alpha + (1-\theta)\beta)^{1/2} \succeq \big(\theta Z(\alpha) + (1-\theta)Z(\beta)\big)^{1/2} \succeq \theta Z(\alpha)^{1/2} + (1-\theta)Z(\beta)^{1/2}.$$

Taking traces (trace is a positive linear functional on PSD matrices) yields that $H(\alpha) := \text{Tr}\big(Z(\alpha)^{1/2}\big)$ is concave in α. Finally, by (16), $\widehat{\mathcal{L}}_t^{\text{FD}}(\alpha)$ is, up to affine term in α, $-2H(\alpha)$, and is therefore convex on Δ_m.

Part (ii): log-Vendi convexity in α (kernel-matrix form). We fix the pooled samples $\{x_j\}_{j=1}^N$ and the Gram matrix $K_{rs} = k(x_r, x_s)$, where k is PSD and normalized, thus we have $K \succeq 0$ and $K_{jj} = 1$ for every j. For a weight vector $q \in \mathbb{R}^N$ with $q \succeq 0$ and $\sum_j q_j = 1$, we define

$$\rho(q) := \text{diag}(q)^{1/2} K \,\text{diag}(q)^{1/2} \succeq 0, \qquad \text{Tr}(\rho(q)) = \sum_{j=1}^N q_j K_{jj} = 1.$$

Note that the log-Vendi loss is defined to be $f(q) := \text{Tr}(\rho(q)\log\rho(q))$.

Let $\mathcal{H}$ be a Hilbert space and $\phi : \mathcal{X} \to \mathcal{H}$ such that $k(x,y) = \langle \phi(x), \phi(y)\rangle_{\mathcal{H}}$ (Section 2). Define the weighted covariance operator

$$C(q) := \sum_{j=1}^N q_j\, \phi(x_j) \otimes \phi(x_j) \;\succeq\; 0. \tag{17}$$

This operator is affine in q. Moreover, $C(q)$ and $\rho(q)$ share the same multiset of non-zero eigenvalues. One can see this by noting the linear map $A_q : \mathbb{R}^N \to \mathcal{H}$ defined by $A_q e_j = \sqrt{q_j}\,\phi(x_j)$. Then, $A_q^* A_q = \rho(q)$ (in the $\mathbb{R}^N$ basis) and $A_q A_q^* = C(q)$ as an operator on $\mathcal{H}$. It follows that $A_q^* A_q$ and $A_q A_q^*$ have identical nonzero spectra, hence

$$\operatorname{Tr}\big(\rho(q)\log\rho(q)\big) = \operatorname{Tr}\big(C(q)\log C(q)\big), \tag{18}$$

where the trace on the right is over the (finite-rank) operator $C(q)$.

Now, we note that the negative von Neumann entropy $X \mapsto \operatorname{Tr}(X \log X)$ is convex on the cone of trace-one PSD (finite-rank) operators/matrices. Since $C(q)$ is affine in q and $\operatorname{Tr}(C(q)) = \sum_j q_j \|\phi(x_j)\|_{\mathcal{H}}^2 = \sum_j q_j k(x_j, x_j) = 1$ (using kernel normalization), we conclude that $q \mapsto \operatorname{Tr}(C(q)\log C(q))$ is convex on the simplex. By (18), $q \mapsto \operatorname{Tr}(\rho(q)\log\rho(q))$ is convex as well. Finally, in our setting $q = q(\alpha)$ defined by (9) is affine in α (counts are fixed during the update), thus $\alpha \mapsto \widehat{\mathcal{L}}_t^{\text{Vendi}}(\alpha)$ is convex on Δ_m.

Part (iii): convex quadratics. If $\widehat{K}(t) \succeq 0$, then $\alpha \mapsto \alpha^\top \widehat{K}(t)\alpha$ is convex on $\mathbb{R}^m$ and $\alpha \mapsto w\,\alpha^\top \widehat{b}(t)$ is linear. Hence $\widehat{\mathcal{L}}_t^{\text{quad}}$ is convex on Δ_m.

The above completes the proof of Proposition 4.1.

A.2. Gradients for EG optimization

Algorithm 1 requires $g = \nabla_\alpha \widehat{\mathcal{L}}_{t-1}(\alpha)$. Below we provide the gradients used in our implementation. In all cases, counts $n_i(t-1)$ and pooled indices I_j are treated as fixed while solving (5) at round t.

Quadratic + linear fidelity. For $\widehat{\mathcal{L}}_t^{\text{quad}}(\alpha) = \alpha^\top \widehat{K}(t)\alpha + w\,\alpha^\top \widehat{b}(t)$ with $\widehat{K}(t)$ symmetric,

$$\nabla_\alpha \widehat{\mathcal{L}}_t^{\text{quad}}(\alpha) = 2\widehat{K}(t)\alpha + w\,\widehat{b}(t).$$

FD gradient (moment-matching form). Let $\mu(\alpha) = \widehat{\mu}_t(\alpha)$ and $\Sigma(\alpha) = \widehat{\Sigma}_t(\alpha)$ as in (6)–(7). Then

$$\frac{\partial \mu(\alpha)}{\partial \alpha_i} = \widehat{\mu}_i(t), \qquad \frac{\partial \Sigma(\alpha)}{\partial \alpha_i} = \widehat{S}_i(t) - \widehat{\mu}_i(t)\mu(\alpha)^\top - \mu(\alpha)\widehat{\mu}_i(t)^\top.$$

We write $\widehat{\mathcal{L}}_t^{\text{FD}}(\alpha) = \mathcal{F}(\mu(\alpha), \Sigma(\alpha))$, where

$$\mathcal{F}(\mu, \Sigma) = \|\mu - \mu_0\|_2^2 + \operatorname{Tr}(\Sigma + \Sigma_0) - 2\operatorname{Tr}\big((\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2})^{1/2}\big),$$

with $(\mu_0, \Sigma_0) = (\widehat{\mu}_{\text{data}}, \widehat{\Sigma}_{\text{data}})$. Then by the chain rule,

$$\frac{\partial \widehat{\mathcal{L}}_t^{\text{FD}}(\alpha)}{\partial \alpha_i} = 2(\mu(\alpha) - \mu_0)^\top \widehat{\mu}_i(t) + \Big\langle G(\Sigma(\alpha)),\ \widehat{S}_i(t) - \widehat{\mu}_i(t)\mu(\alpha)^\top - \mu(\alpha)\widehat{\mu}_i(t)^\top \Big\rangle, \tag{19}$$

where $\langle A, B\rangle := \operatorname{Tr}(A^\top B)$ and

$$G(\Sigma) := \nabla_\Sigma \mathcal{F}(\mu, \Sigma) = I - \Sigma_0^{1/2}\,(\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2})^{-1/2}\,\Sigma_0^{1/2}. \tag{20}$$

Note that we can compute the square root $(\cdot)^{-1/2}$ via eigendecomposition.

Negative log-Vendi gradient (kernel matrix form). At time t, let pooled samples $\{x_j\}_{j=1}^N$ with indices I_j, Gram matrix K, and $q_j(\alpha) = \alpha_{I_j}/n_{I_j}(t)$ as in (9). Define $\rho(\alpha) = \operatorname{diag}(q(\alpha))^{1/2} K \operatorname{diag}(q(\alpha))^{1/2}$. For $\rho \succ 0$, the differential identity

$$d\operatorname{Tr}(\rho\log\rho) = \langle \log\rho + I,\ d\rho\rangle$$

implies $\nabla_\rho \widehat{\mathcal{L}}_t^{\text{Vendi}}(\alpha) = \log\rho(\alpha) + I$. Moreover,

$$\frac{\partial q_j(\alpha)}{\partial \alpha_i} = \frac{1}{n_i(t)}\,\mathbf{1}\{I_j = i\}.$$

Let $D := \mathrm{diag}(q)^{1/2}$. For $q_j > 0$,

$$\frac{\partial D}{\partial q_j} = \frac{1}{2\sqrt{q_j}}\, e_j e_j^\top, \qquad \frac{\partial \rho}{\partial q_j} = \frac{\partial D}{\partial q_j} K D + D K \frac{\partial D}{\partial q_j}.$$

Therefore,

$$\frac{\partial \widehat{\mathcal{L}}_t^{\mathrm{Vendi}}(\alpha)}{\partial \alpha_i} = \sum_{j: I_j = i} \frac{1}{n_i(t)} \Big\langle \log \rho(\alpha) + I,\ \frac{\partial \rho}{\partial q_j} \Big\rangle. \tag{21}$$

We compute $\log \rho$ via eigendecomposition $\rho = U \Lambda U^\top$ and $\log \rho = U (\log \Lambda) U^\top$.

Negative log-Vendi gradient (finite-dimensional feature / random-feature form). If we use a finite-dimensional feature map $\varphi(x) \in \mathbb{R}^D$ (e.g. random Fourier features) and define

$$C(\alpha) := \sum_{j=1}^{N} q_j(\alpha)\, \varphi(x_j) \varphi(x_j)^\top \in \mathbb{R}^{D \times D},$$

then the objective can be equivalently computed as $\mathrm{Tr}(C(\alpha) \log C(\alpha))$ (finite-rank case), and $\nabla_C \mathrm{Tr}(C \log C) = \log C + I$. Since $\partial C / \partial q_j = \varphi(x_j) \varphi(x_j)^\top$,

$$\begin{aligned}
\frac{\partial \widehat{\mathcal{L}}_t^{\mathrm{Vendi}}(\alpha)}{\partial \alpha_i} &= \sum_{j: I_j = i} \frac{1}{n_i(t)} \Big\langle \log C(\alpha) + I,\ \varphi(x_j) \varphi(x_j)^\top \Big\rangle \\
&= \sum_{j: I_j = i} \frac{1}{n_i(t)}\, \varphi(x_j)^\top (\log C(\alpha) + I) \varphi(x_j).
\end{aligned} \tag{22}$$

This replaces the $O(N^3)$ cost of eigendecomposing ρ by an $O(D^3)$ eigendecomposition of C when $D < N$.

B. Proofs for Section 5: log-Vendi and Fréchet Distance regret

B.1. Bandit protocol and ERM-to-regret reduction

Let $\Delta_m = \{\alpha \in \mathbb{R}^m : \alpha_i \ge 0,\ \sum_{i=1}^m \alpha_i = 1\}$. For the warm start, we suppose $n_i(0) = M$ for all i. At round $t = 1, \dots, T$, Mixture-Greedy selects

$$\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{F}_{t-1}(\alpha), \qquad I_t \sim \alpha_t.$$

The regret at iteration T is defined as $\mathrm{Reg}_T = \sum_{t=1}^T (F(\alpha_t) - \min_\alpha F(\alpha))$.

Lemma B.1. *For every $t \ge 1$, let $\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{F}_{t-1}(\alpha)$. Then, the following holds*

$$F(\alpha_t) - \min_{\alpha \in \Delta_m} F(\alpha) \le 2 \sup_{\alpha \in \Delta_m} \big| \widehat{F}_{t-1}(\alpha) - F(\alpha) \big|.$$

Proof. Let $\alpha^\star \in \arg\min_\alpha F(\alpha)$. Since α_t minimizes $\widehat{F}_{t-1}$,

$$\begin{aligned}
F(\alpha_t) - F(\alpha^\star) &= \big(F(\alpha_t) - \widehat{F}_{t-1}(\alpha_t)\big) + \big(\widehat{F}_{t-1}(\alpha_t) - \widehat{F}_{t-1}(\alpha^\star)\big) + \big(\widehat{F}_{t-1}(\alpha^\star) - F(\alpha^\star)\big) \\
&\le |F(\alpha_t) - \widehat{F}_{t-1}(\alpha_t)| + |\widehat{F}_{t-1}(\alpha^\star) - F(\alpha^\star)| \le 2 \sup_\alpha |\widehat{F}_{t-1}(\alpha) - F(\alpha)|.
\end{aligned}$$

The lemma's proof is hence complete. □

Lemma B.2. *Assume there exists $\gamma \in (0, 1]$ such that $\alpha_{t,i} \ge \gamma$ for all i and all $t \le T$. Then, for every $\delta > 0$, with probability at least $1 - \delta$, the following holds simultaneously for all $i \in [m]$ and all $t \le T$,*

$$n_i(t) \ge M + \gamma t - \sqrt{2t \log \frac{mT}{\delta}}.$$

Proof. Consider integer i and let $N_i(t) = \sum_{s=1}^t \mathbf{1}\{I_s = i\}$, hence $n_i(t) = M + N_i(t)$. Define $D_s := \mathbf{1}\{I_s = i\} - \mathbb{E}[\mathbf{1}\{I_s = i\} \mid \mathcal{F}_{s-1}]$, therefore $|D_s| \le 1$ and $\sum_{s=1}^t D_s = N_i(t) - \sum_{s=1}^t \alpha_{s,i}$. Azuma–Hoeffding and union bound over i, t yields $N_i(t) \ge \sum_{s=1}^t \alpha_{s,i} - \sqrt{2t \log \frac{mT}{\delta}} \ge \gamma t - \sqrt{2t \log \frac{mT}{\delta}}$. □

B.2. Negative log-Vendi: concentration and entropy continuity

Assumption B.3 (Normalized features)**.** $\|\phi(x)\|_2 = 1$ for all x.

For arm i, define $S_i = \mathbb{E}[\phi(X)\phi(X)^\top]$ and $S_\alpha = \sum_i \alpha_i S_i$. Define $F_{\mathrm{NLV}}(\alpha) = \mathrm{Tr}(S_\alpha \log S_\alpha)$ and its plug-in estimator $\widehat{F}_t^{\mathrm{NLV}}(\alpha) = \mathrm{Tr}(\widehat{S}_\alpha(t) \log \widehat{S}_\alpha(t))$.

Assumption B.4 (Population innovation)**.** There exist $\nu_0 \in (0, 1]$ and $\varepsilon_0 \in [0, \nu_0/8)$ such that for each $i \in [m]$ there exists a unit vector v_i with $v_i^\top S_i v_i \geq \nu_0$ and $\sum_{j \neq i} v_i^\top S_j v_i \leq \varepsilon_0$.

Lemma B.5 (from (Sutherland et al., 2018))**.** *Let $(Y_r)_{r=1}^n$ be i.i.d. random elements in a real Hilbert space with $\mathbb{E}[Y_r] = 0$ and $\|Y_r\| \leq L$ a.s. Then, for every $\delta > 0$, with probability at least $1 - \delta$, we have the following*

$$\left\| \frac{1}{n} \sum_{r=1}^n Y_r \right\| \leq \frac{L}{\sqrt{n}} \left(1 + \sqrt{2 \log \frac{1}{\delta}} \right).$$

Lemma B.6. *Fix $T \geq 1$ and $\delta \in (0, 1)$. Under Assumption B.3, with probability at least $1 - \delta$, simultaneously for all $i \in [m]$ and all $n \in \{M, \ldots, M + T\}$,*

$$\|\widehat{S}_i(n) - S_i\|_F \leq \frac{2}{\sqrt{n}} \left(1 + \sqrt{2 \log \frac{m(T+1)}{\delta}} \right).$$

Proof. Fix (i, n) and set $Y_{i,r} = \phi(X_{i,r})\phi(X_{i,r})^\top - S_i$. Then $\mathbb{E}Y_{i,r} = 0$ and $\|Y_{i,r}\|_F \leq \|\phi\phi^\top\|_F + \|S_i\|_F \leq 1 + 1 = 2$. The application of Lemma B.5 in Frobenius space with failure probability $\delta/(m(T+1))$ and applying the union bound over (i, n)'s proves the result. □

Fannes–Audenaert inequality. Let $\|\cdot\|_1$ be the trace norm. We use Fannes–Audenaert inequality (Audenaert, 2007), which shows for every two density matrices $S, S' \succeq 0$ with $\mathrm{Tr}(S) = \mathrm{Tr}(S') = 1$ and $T = \frac{1}{2}\|S - S'\|_1 \in [0, 1 - 1/d]$, the following holds

$$|\mathrm{Tr}(S \log S) - \mathrm{Tr}(S' \log S')| \leq T \log(d-1) + h(T), \quad h(T) = -T \log T - (1-T) \log(1-T).$$

Lemma B.7. *Let $S, S' \succeq 0$ with $\mathrm{Tr}(S) = \mathrm{Tr}(S') = 1$ in dimension $d \geq 2$, and assume $\|S - S'\|_F \leq \varepsilon$. Let $T := \frac{1}{2}\|S - S'\|_1$. Then $T \leq \frac{1}{2}\sqrt{d}\,\varepsilon$ and whenever $T \leq 1 - 1/d$,*

$$|\mathrm{Tr}(S \log S) - \mathrm{Tr}(S' \log S')| \leq T \log(d-1) + h(T).$$

Moreover, if $\varepsilon \leq \frac{2}{e\sqrt{d}}$ (thus $T \leq 1/e$), then

$$|\mathrm{Tr}(S \log S) - \mathrm{Tr}(S' \log S')| \leq \frac{1}{2}\sqrt{d}\,\varepsilon \Big(\log(d-1) + \log \frac{e\sqrt{d}}{2} + 1 + \log \frac{1}{\varepsilon} \Big). \tag{23}$$

In particular, in this regime,

$$|\mathrm{Tr}(S \log S) - \mathrm{Tr}(S' \log S')| \leq C_d\, \varepsilon \log \frac{1}{\varepsilon} \quad \textit{with} \quad C_d := \frac{1}{2}\sqrt{d} \Big(2 + \log(d-1) + \log \frac{e\sqrt{d}}{2} \Big).$$

Proof. Fro Frobenius ($\|\cdot\|_F$) and trace norms ($\|\cdot\|_1$) of a matrix $A \in \mathbb{R}^{d \times d}$, a well-known inequality is that $\|A\|_1 \leq \sqrt{d}\|A\|_F$. This implies that $T \leq \frac{1}{2}\sqrt{d}\,\varepsilon$. We then apply Fannes–Audenaert to directly obtain the first bound. For the simplified second bound, note that if $T \leq 1/e$, then $-x \log x$ is increasing on $[0, T]$, and one can see $h(T) \leq T \log(1/T) + T$. Since $T \leq \frac{1}{2}\sqrt{d}\,\varepsilon$, we will have $\log(1/T) \leq \log \frac{2}{\sqrt{d}} + \log \frac{1}{\varepsilon}$, and a substitution yields (23). □

B.3. Quantitative interiority for log-Vendi Mixture-Greedy (patched)

Lemma B.8. *Fix $T \geq 1$ and $\delta \in (0, 1)$. Assume Assumptions B.3 and B.4, and assume the event of Lemma B.6 holds. Define*

$$\eta := \frac{2}{\sqrt{M}} \left(1 + \sqrt{2 \log \frac{m(T+1)}{\delta}} \right), \qquad \nu_{\mathrm{eff}} := \nu_0 - \eta, \qquad \varepsilon_{\mathrm{eff}} := \varepsilon_0 + (m-1)\eta,$$

and assume $\eta \le \nu_0/4$ (thus $\nu_{\text{eff}} \ge 3\nu_0/4$). Define

$$\gamma_{\min} := \max\left\{0,\ \exp\left(-1 - \frac{1}{\nu_{\text{eff}}}\left(\frac{d}{e} + \log m\right)\right) - \varepsilon_{\text{eff}}\right\}.$$

Then for every $t \in \{1, \dots, T\}$, every minimizer $\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{F}^{\text{NLV}}_{t-1}(\alpha)$ satisfies

$$\alpha_{t,i} \ge \gamma_{\min} \qquad \forall i \in [m].$$

Proof. Fix t and let α minimize $\widehat{F}^{\text{NLV}}_{t-1}$. Write $\widehat{S}_i := \widehat{S}_i(t-1)$ and $\widehat{S}_\alpha := \sum_{i=1}^m \alpha_i \widehat{S}_i$.

First, we analyze the empirical innovation along each v_i. To do this, fix index i and consider the population innovation direction v_i from Assumption B.4. By Lemma B.6 and $\|A\|_{\text{op}} \le \|A\|_F$,

$$|v_i^\top (\widehat{S}_j - S_j) v_i| \le \eta \qquad \forall j.$$

As a result, we can write

$$v_i^\top \widehat{S}_i v_i \ge \nu_0 - \eta = \nu_{\text{eff}}, \quad \sum_{j \ne i} v_i^\top \widehat{S}_j v_i \le \varepsilon_0 + (m-1)\eta = \varepsilon_{\text{eff}}.$$

Also, $\|\widehat{S}_j\|_{\text{op}} \le \operatorname{Tr}(\widehat{S}_j) = 1$. Therefore, we have

$$v_i^\top \widehat{S}_\alpha v_i = \alpha_i\, v_i^\top \widehat{S}_i v_i + \sum_{j \ne i} \alpha_j\, v_i^\top \widehat{S}_j v_i \le \alpha_i \cdot 1 + \varepsilon_{\text{eff}}. \tag{24}$$

Next, we derive the directional optimality condition. Note that the map $S \mapsto \operatorname{Tr}(S \log S)$ is convex on unit-trace PSD matrices (i.e., density matrices). Therefore, $\alpha \mapsto \widehat{F}^{\text{NLV}}_{t-1}(\alpha)$ is convex on Δ_m. Let j be an index with $\alpha_j \ge 1/m$. For every $i \in [m]$ and $\tau \in (0, \alpha_j]$, we define $\alpha^{(\tau)} = \alpha + \tau(e_i - e_j) \in \Delta_m$. Optimality of α implies

$$\frac{\widehat{F}^{\text{NLV}}_{t-1}(\alpha^{(\tau)}) - \widehat{F}^{\text{NLV}}_{t-1}(\alpha)}{\tau} \ \ge\ 0. \tag{25}$$

Then, we let $\mathcal{G}(S) = \operatorname{Tr}(S \log S)$. By convexity,

$$\frac{\widehat{F}^{\text{NLV}}_{t-1}(\alpha^{(\tau)}) - \widehat{F}^{\text{NLV}}_{t-1}(\alpha)}{\tau} \le \left\langle \log \widehat{S}_\alpha + I,\ \widehat{S}_i - \widehat{S}_j \right\rangle. \tag{26}$$

Since $\operatorname{Tr}(\widehat{S}_i) = \operatorname{Tr}(\widehat{S}_j) = 1$, the identity terms cancel:

$$\left\langle \log \widehat{S}_\alpha + I,\ \widehat{S}_i - \widehat{S}_j \right\rangle = \operatorname{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) - \operatorname{Tr}(\widehat{S}_j \log \widehat{S}_\alpha). \tag{27}$$

We bound the $\widehat{S}_i$ term using v_i. Since $\widehat{S}_i \succeq 0$ and $\operatorname{Tr}(\widehat{S}_i) = 1$,

$$\operatorname{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) \le (v_i^\top \widehat{S}_i v_i) \cdot v_i^\top (\log \widehat{S}_\alpha) v_i.$$

By Jensen's inequality for the concave log function (on the spectral measure of $\widehat{S}_\alpha$), we obtain $v_i^\top (\log \widehat{S}_\alpha) v_i \le \log(v_i^\top \widehat{S}_\alpha v_i)$. Thus, by (24) we attain

$$\operatorname{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) \le (v_i^\top \widehat{S}_i v_i) \log(\alpha_i + \varepsilon_{\text{eff}}) \le \nu_{\text{eff}} \log(\alpha_i + \varepsilon_{\text{eff}}).$$

Combining this with (27) yields the following

$$\frac{\widehat{F}^{\text{NLV}}_{t-1}(\alpha^{(\tau)}) - \widehat{F}^{\text{NLV}}_{t-1}(\alpha)}{\tau} \le \nu_{\text{eff}} \log(\alpha_i + \varepsilon_{\text{eff}}) - \operatorname{Tr}(\widehat{S}_j \log \widehat{S}_\alpha). \tag{28}$$

Subsequently, we derive a lower bound for the $\widehat{S}_j$ term. Since $\alpha_j \geq 1/m$, $\widehat{S}_\alpha \succeq \alpha_j \widehat{S}_j \succeq \frac{1}{m}\widehat{S}_j$. Using an ϵ-regularization argument to apply operator monotonicity of log on PD matrices and then letting $\epsilon \downarrow 0$, we obtain

$$\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha) \geq \mathrm{Tr}\left(\widehat{S}_j \log\left(\tfrac{1}{m}\widehat{S}_j\right)\right) = \mathrm{Tr}(\widehat{S}_j \log \widehat{S}_j) - (\log m)\mathrm{Tr}(\widehat{S}_j).$$

Since $\mathrm{Tr}(\widehat{S}_j) = 1$ and $x \log x \geq -1/e$ for $x \geq 0$, $\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_j) \geq -d/e$, and then

$$\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha) \geq -\frac{d}{e} - \log m. \tag{29}$$

Combining the above, using (25) and (28), we find that

$$0 \leq \frac{\widehat{F}(\alpha^{(\tau)}) - \widehat{F}(\alpha)}{\tau} \leq \nu_{\mathrm{eff}} \log(\alpha_i + \varepsilon_{\mathrm{eff}}) - \mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha).$$

Using (29) gives

$$0 \leq \nu_{\mathrm{eff}} \log(\alpha_i + \varepsilon_{\mathrm{eff}}) + \frac{d}{e} + \log m,$$

which can be written as

$$\alpha_i + \varepsilon_{\mathrm{eff}} \geq \exp\left(-\frac{1}{\nu_{\mathrm{eff}}}\left(\frac{d}{e} + \log m\right)\right),$$

which yields the stated $\gamma_{\min}$ after subtracting $\varepsilon_{\mathrm{eff}}$ and truncating below at 0. □

B.4. Negative log-Vendi regret (dimension-explicit warm-start condition)

Theorem B.9 (Negative log-Vendi regret)**.** *Fix $T \geq 1$ and $\delta \in (0, 1)$. Assume Assumptions B.3 and B.4. Assume M is large enough such that: (i) $\eta \leq \nu_0/4$ and $\gamma_{\min} > 0$ in Lemma B.8, and (ii) along the analysis event, the Frobenius deviation $\varepsilon_t := \sup_\alpha \|\widehat{S}_\alpha(t) - S_\alpha\|_F$ satisfies $\frac{1}{2}\sqrt{d}\,\varepsilon_t \leq 1/e$ for all $t \in \{0, \ldots, T-1\}$ (equivalently, it suffices that $\varepsilon_0 \leq 2/(e\sqrt{d})$ since ε_t decreases with t under linear sampling). Then with probability at least $1 - \delta$,*

$$\mathrm{Reg}_T \leq C_{\mathrm{NLV}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big)\sqrt{T}\,(1 + \log T),$$

where C_{NLV} depends only on $(d, m, \nu_0, \varepsilon_0)$.

Proof. We consider the intersection of these events: (A) Lemma B.6 (failure probability below $\delta/3$), (B) Lemma B.2 with $\gamma = \gamma_{\min}$ (failure probability below $\leq \delta/3$), (C) the deterministic interiority conclusion of Lemma B.8 (holds on event A). On this intersection, $n_{\min}(t) \geq M + \gamma_{\min} t - \sqrt{2t \log \frac{3mT}{\delta}} \geq ct$ for an explicit $c = c(\gamma_{\min}) > 0$. Therefore, by Lemma B.6,

$$\max_i \|\widehat{S}_i(t) - S_i\|_F \leq \frac{2}{\sqrt{n_{\min}(t)}}\left(1 + \sqrt{2 \log \frac{3m(T+1)}{\delta}}\right) \leq \frac{C_1}{\sqrt{t}}\left(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\right).$$

For every $\alpha \in \Delta_m$, the convexity of norm functions shows that $\|\widehat{S}_\alpha(t) - S_\alpha\|_F \leq \max_i \|\widehat{S}_i(t) - S_i\|_F =: \varepsilon_t$. By assumption (ii), we have $\varepsilon_t \leq 2/(e\sqrt{d})$, and therefore Lemma B.7 implies

$$\sup_{\alpha \in \Delta_m} \left|\widehat{F}_t^{\mathrm{NLV}}(\alpha) - F_{\mathrm{NLV}}(\alpha)\right| \leq C_d\, \varepsilon_t \log \frac{1}{\varepsilon_t} \leq \frac{C_2}{\sqrt{t}}\left(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\right)(1 + \log t),$$

where the last step uses $\varepsilon_t \asymp t^{-1/2}$. Hence, we obtain $\log(1/\varepsilon_t) \lesssim 1 + \log t$. Lemma B.1 then yields instantaneous regret bounded by twice the uniform deviation, and summing $\sum_{t=2}^{T} t^{-1/2}(1 + \log t) \leq 2\sqrt{T}(1 + \log T)$ gives the stated bound. The application of union bound over events A and B thereof yields a success probability of at least $1 - \delta$. □

C. Including a bounded linear term: Regret for NLV/FD + linear score

C.1. Linear term: definition, estimator, and uniform concentration

Fix measurable functions $\psi_i : \mathcal{X} \to [0,1]$ for each arm $i \in [m]$ and define

$$\theta_i := \mathbb{E}[\psi_i(X)] \in [0,1], \quad X \sim P_{\mathcal{G}_i}, \qquad G(\alpha) := \sum_{i=1}^{m} \alpha_i \theta_i.$$

For a base objective $F(\alpha)$ (negative log-Vendi or Fréchet Distance), define

$$H(\alpha) := F(\alpha) + w\, G(\alpha), \qquad w \geq 0.$$

At time t, define empirical estimates

$$\widehat{\theta}_i(t) := \frac{1}{n_i(t)} \sum_{r=1}^{n_i(t)} \psi_i(X_{i,r}), \qquad \widehat{G}_t(\alpha) := \sum_{i=1}^{m} \alpha_i \widehat{\theta}_i(t), \qquad \widehat{H}_t(\alpha) := \widehat{F}_t(\alpha) + w\, \widehat{G}_t(\alpha).$$

Mixture-Greedy with the combined objective selects

$$\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{H}_{t-1}(\alpha), \qquad I_t \sim \alpha_t. \tag{30}$$

We analyze combined-objective regret

$$\mathrm{Reg}_T(H) := \sum_{t=1}^{T} \Big(H(\alpha_t) - \min_{\alpha \in \Delta_m} H(\alpha) \Big).$$

Lemma C.1. *Fix $T \geq 1$ and $\delta \in (0,1)$. With probability at least $1-\delta$, simultaneously for all $i \in [m]$ and all $n \in \{M, M+1, \ldots, M+T\}$,*

$$|\widehat{\theta}_i(n) - \theta_i| \leq \frac{1}{\sqrt{2n}} \sqrt{\log \frac{2m(T+1)}{\delta}}.$$

Proof. Consider a fixed (i,n). Since $\psi_i(X_{i,r}) \in [0,1]$ and samples are i.i.d., Hoeffding's inequality gives $\mathbb{P}(|\widehat{\theta}_i(n) - \theta_i| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$. If we set $\varepsilon = \frac{1}{\sqrt{2n}}\sqrt{\log \frac{2m(T+1)}{\delta}}$, then the right hand side will be bounded by $\delta/(m(T+1))$, and applying union bound over (i,n)'s proves the result. □

Lemma C.2 (Uniform deviation of the linear objective)**.** *Assuming the event of Lemma C.1 occurs, for every $t \leq T$ we have the following where $n_{\min}(t) := \min_i n_i(t)$:*

$$\sup_{\alpha \in \Delta_m} |\widehat{G}_t(\alpha) - G(\alpha)| \leq \max_{i \in [m]} |\widehat{\theta}_i(t) - \theta_i| \leq \frac{1}{\sqrt{2n_{\min}(t)}} \sqrt{\log \frac{2m(T+1)}{\delta}}$$

Proof. For every $\alpha \in \Delta_m$, we can write the following

$$|\widehat{G}_t(\alpha) - G(\alpha)| = \left| \sum_{i=1}^{m} \alpha_i (\widehat{\theta}_i(t) - \theta_i) \right| \leq \sum_{i=1}^{m} \alpha_i \max_j |\widehat{\theta}_j(t) - \theta_j| = \max_j |\widehat{\theta}_j(t) - \theta_j|.$$

Then, we apply Lemma C.1 with $n = n_{\min}(t)$ to prove the result. □

Lemma C.3. *For every $t \geq 1$, if α_t minimizes $\widehat{H}_{t-1}$ over Δ_m, then the following holds*

$$H(\alpha_t) - \min_{\alpha \in \Delta_m} H(\alpha) \leq 2 \sup_{\alpha \in \Delta_m} |\widehat{H}_{t-1}(\alpha) - H(\alpha)|.$$

Moreover, we will have

$$\sup_{\alpha \in \Delta_m} |\widehat{H}_{t-1}(\alpha) - H(\alpha)| \leq \sup_{\alpha \in \Delta_m} |\widehat{F}_{t-1}(\alpha) - F(\alpha)| + w \sup_{\alpha \in \Delta_m} |\widehat{G}_{t-1}(\alpha) - G(\alpha)|.$$

Proof. Let $\alpha^\star \in \arg\min_{\alpha\in\Delta_m} H(\alpha)$. Since α_t minimizes $\widehat{H}_{t-1}$, we can write

$$\begin{aligned} H(\alpha_t) - H(\alpha^\star) &= \big(H(\alpha_t) - \widehat{H}_{t-1}(\alpha_t)\big) + \big(\widehat{H}_{t-1}(\alpha_t) - \widehat{H}_{t-1}(\alpha^\star)\big) + \big(\widehat{H}_{t-1}(\alpha^\star) - H(\alpha^\star)\big) \\ &\le 2 \sup_{\alpha\in\Delta_m} |\widehat{H}_{t-1}(\alpha) - H(\alpha)|. \end{aligned}$$

The second inequality follows from the triangle inequality and linearity of the combination as $H - \widehat{H} = (F - \widehat{F}) + w(G - \widehat{G})$. □

C.2. NLV + linear term: corrected interiority floor and regret

We use the NLV setup and notation from the base NLV appendix: normalized features (Assumption B.3), population innovation (Assumption B.4), uniform concentration of $\widehat{S}_i$ (Lemma B.6), and the resulting quantities $\eta, \nu_{\text{eff}}, \varepsilon_{\text{eff}}$.

Lemma C.4 (Updated interiority floor for NLV + linear term (tight; no -1))**.** *Fix $T \ge 1$ and $\delta \in (0,1)$. Assume Assumptions B.3 and B.4, and assume the event of Lemma B.6 holds. Define*

$$\eta := \frac{2}{\sqrt{M}}\left(1 + \sqrt{2\log\frac{m(T+1)}{\delta}}\right), \qquad \nu_{\text{eff}} := \nu_0 - \eta, \qquad \varepsilon_{\text{eff}} := \varepsilon_0 + (m-1)\eta,$$

and assume $\eta \le \nu_0/4$ to obtain $\nu_{\text{eff}} > 0$. Then, for every $t \in \{1,\dots,T\}$, every minimizer $\alpha_t \in \arg\min_{\alpha\in\Delta_m} \widehat{H}_{t-1}(\alpha)$ satisfies

$$\alpha_{t,i} \ge \gamma_{\min}^{(w)} \qquad \forall i \in [m],$$

where

$$\gamma_{\min}^{(w)} := \max\left\{0,\ \exp\left(-\frac{1}{\nu_{\text{eff}}}\left(\frac{d}{e} + \log m + w\right)\right) - \varepsilon_{\text{eff}}\right\}. \tag{31}$$

Proof. Fix t and let α minimize $\widehat{H}_{t-1}(\cdot) = \widehat{F}_{t-1}(\cdot) + w\widehat{G}_{t-1}(\cdot)$. Let j be an index with $\alpha_j \ge 1/m$ (exists since $\sum_i \alpha_i = 1$). For any i and $\tau \in (0, \alpha_j]$, define the feasible perturbation $\alpha^{(\tau)} = \alpha + \tau(e_i - e_j)$.

First, we analyze directional optimality. Because $\widehat{H}_{t-1}$ is convex and α is a minimizer over the convex set Δ_m, we have

$$\frac{\widehat{H}_{t-1}(\alpha^{(\tau)}) - \widehat{H}_{t-1}(\alpha)}{\tau} \ge 0. \tag{32}$$

Then, we show a convex upper bound for the increment term. By convexity of $\widehat{F}_{t-1}$ and linearity of $\widehat{G}_{t-1}$, we have

$$\begin{aligned} \frac{\widehat{H}_{t-1}(\alpha^{(\tau)}) - \widehat{H}_{t-1}(\alpha)}{\tau} &= \frac{\widehat{F}_{t-1}(\alpha^{(\tau)}) - \widehat{F}_{t-1}(\alpha)}{\tau} + w(\widehat{\theta}_i - \widehat{\theta}_j) \\ &\le \left\langle \log \widehat{S}_\alpha + I,\ \widehat{S}_i - \widehat{S}_j \right\rangle + w(\widehat{\theta}_i - \widehat{\theta}_j), \end{aligned} \tag{33}$$

where $\widehat{S}_\alpha = \sum_k \alpha_k \widehat{S}_k$. Since $\mathrm{Tr}(\widehat{S}_i) = \mathrm{Tr}(\widehat{S}_j) = 1$, the identity components cancel exactly:

$$\left\langle \log \widehat{S}_\alpha + I,\ \widehat{S}_i - \widehat{S}_j \right\rangle = \mathrm{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) - \mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha).$$

Now, we bound the $\widehat{S}_i$ term using the innovation direction v_i. By Lemma B.6 and the fact that $\|A\|_{\text{op}} \le \|A\|_F$, we attain

$$v_i^\top \widehat{S}_i v_i \ge \nu_0 - \eta = \nu_{\text{eff}}, \quad \sum_{k\ne i} v_i^\top \widehat{S}_k v_i \le \varepsilon_0 + (m-1)\eta = \varepsilon_{\text{eff}}.$$

Also, $\|\widehat{S}_k\|_{\text{op}} \le \mathrm{Tr}(\widehat{S}_k) = 1$. Hence

$$v_i^\top \widehat{S}_\alpha v_i = \alpha_i\, v_i^\top \widehat{S}_i v_i + \sum_{k\ne i} \alpha_k\, v_i^\top \widehat{S}_k v_i \le \alpha_i \cdot 1 + \varepsilon_{\text{eff}}.$$

Since $\widehat{S}_i \succeq 0$ and $\mathrm{Tr}(\widehat{S}_i) = 1$, we have $\mathrm{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) \le (v_i^\top \widehat{S}_i v_i)\, v_i^\top (\log \widehat{S}_\alpha) v_i$. By concavity of log and using Jensen's inequality on the spectral measure of $\widehat{S}_\alpha$ induced by v_i, $v_i^\top (\log \widehat{S}_\alpha) v_i \le \log(v_i^\top \widehat{S}_\alpha v_i)$, hence

$$\mathrm{Tr}(\widehat{S}_i \log \widehat{S}_\alpha) \le (v_i^\top \widehat{S}_i v_i)\ \log(v_i^\top \widehat{S}_\alpha v_i) \le \nu_{\mathrm{eff}} \log(\alpha_i + \varepsilon_{\mathrm{eff}}).$$

To obtain a lower bound for the $\widehat{S}_j$ term, note that since $\alpha_j \ge 1/m$, we have $\widehat{S}_\alpha \succeq \alpha_j \widehat{S}_j \succeq \frac{1}{m}\widehat{S}_j$. By operator monotonicity of log on the support,

$$\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha) \ge \mathrm{Tr}\left(\widehat{S}_j \log\left(\tfrac{1}{m}\widehat{S}_j\right)\right) = \mathrm{Tr}(\widehat{S}_j \log \widehat{S}_j) - (\log m)\mathrm{Tr}(\widehat{S}_j).$$

Let $\lambda_1, \dots, \lambda_d$ be eigenvalues of $\widehat{S}_j$; then $\sum_k \lambda_k = 1$ and $\lambda_k \ge 0$. Since $x \log x \ge -1/e$ for all $x \ge 0$, we get $\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_j) = \sum_k \lambda_k \log \lambda_k \ge -d/e$. Also, $\mathrm{Tr}(\widehat{S}_j) = 1$. Therefore,

$$\mathrm{Tr}(\widehat{S}_j \log \widehat{S}_\alpha) \ge -\frac{d}{e} - \log m.$$

Because $0 \le \widehat{\theta}_i, \widehat{\theta}_j \le 1$, we have $\widehat{\theta}_i - \widehat{\theta}_j \le 1$. Substituting the bounds from Steps 3–4 into (33) yields

$$\frac{\widehat{H}(\alpha^{(\tau)}) - \widehat{H}(\alpha)}{\tau} \le \nu_{\mathrm{eff}} \log(\alpha_i + \varepsilon_{\mathrm{eff}}) - \left(-\frac{d}{e} - \log m\right) + w.$$

Combining with (32) gives

$$0 \le \nu_{\mathrm{eff}} \log(\alpha_i + \varepsilon_{\mathrm{eff}}) + \frac{d}{e} + \log m + w.$$

Rearranging yields

$$\alpha_i + \varepsilon_{\mathrm{eff}} \ge \exp\left(-\frac{1}{\nu_{\mathrm{eff}}}\left(\frac{d}{e} + \log m + w\right)\right),$$

and the claimed floor follows after subtracting $\varepsilon_{\mathrm{eff}}$ and truncating at 0, giving (31). ☐

Theorem C.5 (NLV + linear term: Mixture-Greedy regret)**.** *Fix $T \ge 1$ and $\delta \in (0,1)$. Assume the NLV conditions (Assumptions B.3, B.4) and $0 \le \psi_i \le 1$. Assume M is sufficiently large to satisfy the following: (i) $\eta \le \nu_0/4$ and $\gamma_{\min}^{(w)} > 0$ in Lemma C.4, and (ii) the dimension-explicit continuity regime holds along the analysis event, namely $\frac{1}{2}\sqrt{d}\,\varepsilon_t \le 1/e$ for all $t \in \{0, \dots, T-1\}$, where $\varepsilon_t := \sup_{\alpha \in \Delta_m} \|\widehat{S}_\alpha(t) - S_\alpha\|_F$. Then with probability at least $1 - \delta$,*

$$\mathrm{Reg}_T(H) \le C_{\mathrm{NLV},w}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big)\sqrt{T}\,(1 + \log T),$$

where $C_{\mathrm{NLV},w}$ depends only on $(d, m, \nu_0, \varepsilon_0, w)$.

Proof. Work on the intersection of the following events: (A) Lemma B.6 (uniform concentration of $\widehat{S}_i$), (B) Lemma C.1 (uniform concentration of $\widehat{\theta}_i$), (C) Lemma B.2 with $\gamma = \gamma_{\min}^{(w)}$ (linear sampling under the floor), and allocate failure probabilities so that the intersection holds with probability at least $1 - \delta$.

On this intersection, Lemma C.4 ensures $\alpha_{t,i} \ge \gamma_{\min}^{(w)}$ for all $i, t \le T$. Thus $n_{\min}(t) = \Omega(t)$ by Lemma B.2. As in the base NLV proof,

$$\varepsilon_t := \sup_\alpha \|\widehat{S}_\alpha(t) - S_\alpha\|_F \le \max_i \|\widehat{S}_i(t) - S_i\|_F \le \frac{C_1}{\sqrt{t}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big).$$

By assumption (ii) we may apply the dimension-explicit continuity modulus (Lemma B.7) to obtain

$$\sup_\alpha \big|\widehat{F}_t(\alpha) - F(\alpha)\big| \le C_2\, \varepsilon_t \log \frac{1}{\varepsilon_t} \le \frac{C_3}{\sqrt{t}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big)(1 + \log t).$$

For the linear part, Lemma C.2 gives

$$\sup_\alpha |\widehat{G}_t(\alpha) - G(\alpha)| \le \frac{1}{\sqrt{2n_{\min}(t)}}\sqrt{\log \frac{2m(T+1)}{\delta}} \le \frac{C_4}{\sqrt{t}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big).$$

Therefore,

$$\sup_\alpha |\widehat{H}_t(\alpha) - H(\alpha)| \le \sup_\alpha |\widehat{F}_t(\alpha) - F(\alpha)| + w \sup_\alpha |\widehat{G}_t(\alpha) - G(\alpha)| \le \frac{C_5}{\sqrt{t}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big)(1 + \log t).$$

Lemma C.3 gives

$$H(\alpha_{t+1}) - \min_\alpha H(\alpha) \le 2 \sup_\alpha |\widehat{H}_t(\alpha) - H(\alpha)|,$$

and summing $\sum_{t=1}^{T-1} t^{-1/2}(1 + \log t) \le 2\sqrt{T}(1 + \log T)$ completes the proof. □

C.3. Fréchet Distance + linear term: margin transfer and regret

Let F_{FD} be the Fréchet Distance objective and $\widehat{F}$ its plug-in estimator from the base FD appendix. Define $H(\alpha) = F_{\mathrm{FD}}(\alpha) + wG(\alpha)$, where $G(\alpha) \in [0, 1]$.

Assumption C.6 (Population interiority margin for H)**.** There exist $\gamma_0 \in (0, 1/m]$ and $\Delta_0^{(H)} > 0$ such that

$$\inf_{\alpha \in \Delta_m:\ \min_i \alpha_i \le \gamma_0} H(\alpha) \ \ge \ \min_{\alpha \in \Delta_m} H(\alpha) + \Delta_0^{(H)}.$$

Lemma C.7. *Assume F_{FD} satisfies Assumption 5.6 with gap Δ_0 and the same γ_0. Then $H = F_{\mathrm{FD}} + wG$ satisfies Assumption C.6 with gap $\Delta_0^{(H)} = \Delta_0 - w$ provided $\Delta_0 > w$.*

Proof. Since $G(\alpha) \in [0, 1]$ for all $\alpha \in \Delta_m$,

$$\inf_{\min_i \alpha_i \le \gamma_0} H(\alpha) \ge \inf_{\min_i \alpha_i \le \gamma_0} F_{\mathrm{FD}}(\alpha) \ge \min_\alpha F_{\mathrm{FD}}(\alpha) + \Delta_0 \ge \min_\alpha H(\alpha) + (\Delta_0 - w),$$

because $\min_\alpha H(\alpha) \le \min_\alpha F_{\mathrm{FD}}(\alpha) + w$. □

Theorem C.8 (Fréchet Distance + linear term: Mixture-Greedy regret)**.** *Fix $T \ge 1$ and $\delta \in (0, 1)$. Assume the Fréchet Distance conditions from Theorem 5.7 (boundedness and PD) and $0 \le \psi_i \le 1$. Assume the population margin Assumption C.6. Assume M is large enough to provide the following: On an event of probability at least $1 - \delta/3$,*

$$\sup_{\alpha \in \Delta_m} |\widehat{H}_t(\alpha) - H(\alpha)| \le \Delta_0^{(H)}/4 \quad \textit{for all } t \in \{0, 1, \ldots, T-1\}.$$

Then, with probability at least $1 - \delta$, the following regret bound holds

$$\mathrm{Reg}_T(H) \le C_{\mathrm{FD},w}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big)\sqrt{T},$$

where $C_{\mathrm{FD},w}$ depends only on $(B, \lambda_0, \nu, m, d, \gamma_0, w)$.

Proof. The proof follows the non-linear-term Fréchet Distance regret proof with H by replacing F_{FD}. Under the supposed uniform deviation, we know $\sup_\alpha |\widehat{H}_t - H| \le \Delta_0^{(H)}/4$ for all $t \le T - 1$. Therefore, the same margin-transfer argument implies that every empirical minimizer satisfies $\min_i \alpha_{t,i} \ge \gamma_0$ for all $t \le T$. Then, Lemma B.2 shows that $n_{\min}(t) = \Omega(t)$.

On this interior region, the Lipschitz analysis for the Fréchet term yields $\sup_\alpha |\widehat{F}_{t-1}(\alpha) - F_{\mathrm{FD}}(\alpha)| \lesssim t^{-1/2}(1 + \sqrt{\log(\cdot)})$, and Lemma C.2 yields $w \sup_\alpha |\widehat{G}_{t-1} - G| \lesssim w\, t^{-1/2}(1 + \sqrt{\log(\cdot)})$. Therefore,

$$\sup_\alpha |\widehat{H}_{t-1}(\alpha) - H(\alpha)| \le \frac{C}{\sqrt{t}}\Big(1 + \sqrt{\log \frac{m(T+1)}{\delta}}\Big).$$

Lemma C.3 converts this into instantaneous regret, and summing $\sum_{t=1}^{T} t^{-1/2} \le 2\sqrt{T}$ completes the bound. A union bound yields an overall probability of at least $1 - \delta$. □

D. Mixture-Greedy regret for the RKE score and RKE + linear fidelity term

D.1. Setup: feature covariances and the RKE objective

Let $\phi : \mathcal{X} \to \mathbb{R}^d$ be a feature map and assume normalized features.

Assumption D.1 (Normalized features)**.** For all $x \in \mathcal{X}$, $\|\phi(x)\|_2 = 1$.

Each arm $i \in [m]$ produces i.i.d. samples $X_{i,r} \sim P_{\mathcal{G}_i}$ when pulled. Define the (population) feature covariance

$$S_i := \mathbb{E}\big[\phi(X_i)\phi(X_i)^\top\big] \in \mathbb{R}^{d\times d}, \qquad X_i \sim P_{\mathcal{G}_i}.$$

Then $S_i \succeq 0$ and $\mathrm{Tr}(S_i) = \mathbb{E}\|\phi(X_i)\|_2^2 = 1$. For $\alpha \in \Delta_m$, define the mixture covariance $S_\alpha := \sum_{i=1}^m \alpha_i S_i$ which by default satisfies the requirements for being a density matrix, i.e., $\mathrm{Tr}(S_\alpha) = 1,\ S_\alpha \succeq 0$. Consider the following function whose minimization is equivalent to optimizing RKE:

$$F_{\mathrm{RKE}}(\alpha) := \mathrm{Tr}(S_\alpha^2). \tag{34}$$

D.2. Empirical estimators

Given $n_i(t)$ samples observed from arm i by time t, define the empirical covariance

$$\widehat{S}_i(t) := \frac{1}{n_i(t)} \sum_{r=1}^{n_i(t)} \phi(X_{i,r})\phi(X_{i,r})^\top, \qquad \widehat{S}_\alpha(t) := \sum_{i=1}^m \alpha_i \widehat{S}_i(t).$$

Define the plug-in estimator of the RKE-minimization objective:

$$\widehat{F}_t^{\mathrm{RKE}}(\alpha) := \mathrm{Tr}\big(\widehat{S}_\alpha(t)^2\big). \tag{35}$$

Mixture-Greedy for RKE uses the same protocol as in the main paper:

$$\alpha_t \in \arg\min_{\alpha\in\Delta_m} \widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha), \qquad I_t \sim \alpha_t.$$

The cumulative regret is

$$\mathrm{Reg}_T(F_{\mathrm{RKE}}) := \sum_{t=1}^T \Big(F_{\mathrm{RKE}}(\alpha_t) - \min_{\alpha\in\Delta_m} F_{\mathrm{RKE}}(\alpha)\Big).$$

D.3. Uniform concentration for $\widehat{S}_i$ and a uniform deviation bound for $\widehat{F}^{\mathrm{RKE}}$

We use the following concentration statement, proved elsewhere in the appendix (it matches the uniform matrix Hoeffding bound used in the NLV section); we restate it here for readability.

Lemma D.2 (from (Sutherland et al., 2018))**.** *Fix $T \geq 1$ and $\delta \in (0, 1)$. Under Assumption D.1, with probability at least $1 - \delta$, simultaneously for all $i \in [m]$ and all $n \in \{M, M+1, \ldots, M+T\}$,*

$$\|\widehat{S}_i(n) - S_i\|_{\mathrm{F}} \leq \frac{2}{\sqrt{n}}\left(1 + \sqrt{2\log\frac{m(T+1)}{\delta}}\right).$$

Lemma D.3. *Fix $T \geq 1$ and $\delta \in (0, 1)$. On the event of Lemma D.2, for every $t \in \{0, 1, \ldots, T\}$,*

$$\sup_{\alpha\in\Delta_m}\Big|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\Big| \leq 2\,\varepsilon_t + \varepsilon_t^2, \qquad \text{where } \ \varepsilon_t := \max_{i\in[m]} \|\widehat{S}_i(t) - S_i\|_{\mathrm{F}}.$$

In particular, for $t \geq 1$,

$$\sup_{\alpha\in\Delta_m}\Big|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\Big| \leq \frac{C_{\mathrm{rke}}}{\sqrt{n_{\min}(t)}}\left(1 + \sqrt{\log\frac{m(T+1)}{\delta}}\right),$$

where $n_{\min}(t) := \min_i n_i(t)$ and C_{rke} is a universal numerical constant.

Proof. Fix t and $\alpha \in \Delta_m$. Let $\Delta_\alpha(t) := \widehat{S}_\alpha(t) - S_\alpha$. By convexity of the Frobenius norm,

$$\|\Delta_\alpha(t)\|_F = \left\|\sum_{i=1}^m \alpha_i(\widehat{S}_i(t) - S_i)\right\|_F \le \sum_{i=1}^m \alpha_i \|\widehat{S}_i(t) - S_i\|_F \le \varepsilon_t.$$

Now we can expand this as

$$\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha) = \mathrm{Tr}\big((S_\alpha + \Delta_\alpha)^2 - S_\alpha^2\big) = 2\,\mathrm{Tr}(S_\alpha \Delta_\alpha) + \mathrm{Tr}(\Delta_\alpha^2).$$

Using $|\mathrm{Tr}(AB)| \le \|A\|_F\|B\|_F$ and $\mathrm{Tr}(\Delta_\alpha^2) = \|\Delta_\alpha\|_F^2$ gives

$$\left|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right| \le 2\|S_\alpha\|_F\|\Delta_\alpha\|_F + \|\Delta_\alpha\|_F^2.$$

Since $S_\alpha \succeq 0$ and $\mathrm{Tr}(S_\alpha) = 1$, we have $\|S_\alpha\|_F \le \mathrm{Tr}(S_\alpha) = 1$. Therefore, we can write

$$\left|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right| \le 2 \cdot 1 \cdot \varepsilon_t + \varepsilon_t^2 = 2\,\varepsilon_t + \varepsilon_t^2.$$

Taking supremum over $\alpha \in \Delta_m$ yields the first claim. For the second claim, apply Lemma D.2 with $n = n_{\min}(t)$ to obtain $\varepsilon_t \le \frac{2}{\sqrt{n_{\min}(t)}}\big(1 + \sqrt{2\log\frac{m(T+1)}{\delta}}\big)$, and substitute into $2\varepsilon_t + \varepsilon_t^2$, absorbing the ε_t^2 term into the same rate. □

The need for explicit population margin. Because $F_{\mathrm{RKE}}(\alpha) = \mathrm{Tr}(S_\alpha^2)$ is smooth on Δ_m and can admit boundary minimizers, pure Mixture-Greedy may legitimately assign vanishing weights to some arms unless additional structure is imposed. We therefore assume a population interiority margin, exactly as in the Fréchet Distance analysis.

Assumption D.4 (Population interiority margin for RKE-minimization)**.** There exist $\gamma_0 \in (0, 1/m]$ and $\Delta_0 > 0$ such that

$$\inf_{\alpha \in \Delta_m:\, \min_i \alpha_i \le \gamma_0} F_{\mathrm{RKE}}(\alpha) \;\ge\; \min_{\alpha \in \Delta_m} F_{\mathrm{RKE}}(\alpha) + \Delta_0.$$

Lemma D.5. *Fix $T \ge 1$. Suppose that for all $t \in \{0, 1, \dots, T-1\}$,*

$$\sup_{\alpha \in \Delta_m}\left|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right| \le \frac{\Delta_0}{4}.$$

Then for every $t \in \{0, 1, \dots, T-1\}$, every minimizer $\alpha_{t+1} \in \arg\min_{\alpha \in \Delta_m} \widehat{F}_t^{\mathrm{RKE}}(\alpha)$ satisfies $\min_i \alpha_{t+1,i} \ge \gamma_0$.

Proof. Fix t and let α_{t+1} minimize $\widehat{F}_t^{\mathrm{RKE}}$. Let $\alpha^\star$ minimize F_{RKE}. Then

$$F_{\mathrm{RKE}}(\alpha_{t+1}) \le \widehat{F}_t^{\mathrm{RKE}}(\alpha_{t+1}) + \frac{\Delta_0}{4} \le \widehat{F}_t^{\mathrm{RKE}}(\alpha^\star) + \frac{\Delta_0}{4} \le F_{\mathrm{RKE}}(\alpha^\star) + \frac{\Delta_0}{2}.$$

Thus $F_{\mathrm{RKE}}(\alpha_{t+1}) \le \min_\alpha F_{\mathrm{RKE}}(\alpha) + \Delta_0/2$. By Assumption D.4, any α with $\min_i \alpha_i \le \gamma_0$ has $F_{\mathrm{RKE}}(\alpha) \ge \min_\alpha F_{\mathrm{RKE}}(\alpha) + \Delta_0$, hence α_{t+1} cannot lie in that boundary layer. Therefore, $\min_i \alpha_{t+1,i} \ge \gamma_0$. □

D.4. RKE regret bound (pure RKE-minimization)

We use the same ERM-to-regret comparison lemma as elsewhere in the appendix, but stated here for completeness.

Lemma D.6. *If $\alpha_t \in \arg\min_{\alpha \in \Delta_m} \widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha)$, then*

$$F_{\mathrm{RKE}}(\alpha_t) - \min_{\alpha \in \Delta_m} F_{\mathrm{RKE}}(\alpha) \le 2 \sup_{\alpha \in \Delta_m}\left|\widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right|.$$

Proof. This is the standard ERM comparison: let $\alpha^\star \in \arg\min_\alpha F_{\mathrm{RKE}}(\alpha)$ and use $\widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha_t) \le \widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha^\star)$, then add and subtract. □

Theorem D.7 (Mixture-Greedy regret for RKE-minimization)**.** *Fix $T \geq 1$ and $\delta \in (0,1)$. Assume Assumption D.1 and the population margin Assumption D.4. Assume the warm start M is adequately large so that the uniform deviation event*

$$\sup_{\alpha\in\Delta_m}\left|\widehat{F}_t^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right| \leq \frac{\Delta_0}{4} \quad \textit{holds for all } t \in \{0,1,\dots,T-1\}$$

on an event of probability at least $1-\delta/3$. Then with probability at least $1-\delta$,

$$\mathrm{Reg}_T(F_{\mathrm{RKE}}) \leq C_{\mathrm{RKE}}\Big(1+\sqrt{\log\frac{m(T+1)}{\delta}}\Big)\sqrt{T},$$

where C_{RKE} depends only on (m,γ_0) (and universal numerical constants).

Proof. On the event that the uniform deviation bound is $\leq \Delta_0/4$ for all $t \leq T-1$, Lemma D.5 implies $\min_i \alpha_{t,i} \geq \gamma_0$ for all $t \leq T$. Applying the standard sampling-count lemma (Lemma B.2) with $\gamma = \gamma_0$ yields $n_{\min}(t) = \Omega(t)$ with probability at least $1-\delta/3$ after allocating failure probability and union bounding over arms and times.

On the intersection of this count event and the concentration event of Lemma D.2, Lemma D.3 yields for all $t \geq 2$,

$$\sup_{\alpha\in\Delta_m}\left|\widehat{F}_{t-1}^{\mathrm{RKE}}(\alpha) - F_{\mathrm{RKE}}(\alpha)\right| \leq \frac{C}{\sqrt{t-1}}\Big(1+\sqrt{\log\frac{m(T+1)}{\delta}}\Big),$$

where we used $n_{\min}(t-1) = \Omega(t-1)$. Lemma D.6 converts this into instantaneous regret for rounds $t \geq 2$, and summing

$$\sum_{t=2}^{T}(t-1)^{-1/2} \leq 2\sqrt{T}$$

gives the stated $\mathcal{O}(\sqrt{T})$ bound (with the initial $t=1$ warm-start round absorbed into the constant). A union bound over the deviation and count events yields probability at least $1-\delta$. ☐

D.5. RKE + bounded linear term

We now combine F_{RKE} with the bounded linear fidelity term exactly as in Appendix C. We reuse your definitions

$$G(\alpha) = \sum_{i=1}^{m}\alpha_i\theta_i, \quad \theta_i = \mathbb{E}[\psi_i(X)] \in [0,1], \quad \psi_i : \mathcal{X} \to [0,1], \quad H(\alpha) = F_{\mathrm{RKE}}(\alpha) + wG(\alpha),$$

and the plug-in estimator $\widehat{H}_t(\alpha) = \widehat{F}_t^{\mathrm{RKE}}(\alpha) + w\widehat{G}_t(\alpha)$, with $\widehat{G}_t$ and $\widehat{\theta}_i$ as in Appendix C.

Assumption D.8 (Population interiority margin for $H = F_{\mathrm{RKE}} + wG$)**.** There exist $\gamma_0 \in (0, 1/m]$ and $\Delta_0^{(H)} > 0$ such that

$$\inf_{\alpha\in\Delta_m:\,\min_i \alpha_i \leq \gamma_0} H(\alpha) \;\geq\; \min_{\alpha\in\Delta_m} H(\alpha) + \Delta_0^{(H)}.$$

Lemma D.9. *If Assumption D.4 holds with gap Δ_0 and $\Delta_0 > w$, then Assumption D.8 holds with $\Delta_0^{(H)} = \Delta_0 - w$.*

Proof. Since $G(\alpha) \in [0,1]$ for all $\alpha \in \Delta_m$,

$$\inf_{\min_i \alpha_i \leq \gamma_0} H(\alpha) \geq \inf_{\min_i \alpha_i \leq \gamma_0} F_{\mathrm{RKE}}(\alpha) \geq \min_\alpha F_{\mathrm{RKE}}(\alpha) + \Delta_0 \geq \min_\alpha H(\alpha) + (\Delta_0 - w),$$

because $\min_\alpha H(\alpha) \leq \min_\alpha F_{\mathrm{RKE}}(\alpha) + w$. ☐

Theorem D.10 (Mixture-Greedy regret for RKE + linear term)**.** *Fix $T \geq 1$ and $\delta \in (0,1)$. Assume Assumption D.1, $0 \leq \psi_i \leq 1$, and the population margin Assumption D.8. Assume the warm start M is sufficiently large so that on an event of probability at least $1-\delta/3$,*

$$\sup_{\alpha\in\Delta_m}\left|\widehat{H}_t(\alpha) - H(\alpha)\right| \leq \frac{\Delta_0^{(H)}}{4} \quad \textit{for all } t \in \{0,1,\dots,T-1\}.$$

Then with probability at least $1-\delta$*,*

$$\mathrm{Reg}_T(H) \le C_{\mathrm{RKE},w}\Big(1+\sqrt{\log\frac{m(T+1)}{\delta}}\Big)\sqrt{T},$$

where $C_{\mathrm{RKE},w}$ *depends only on* (m,γ_0,w) *(and universal numerical constants).*

Proof. On the event $\sup_\alpha|\widehat{H}_t - H| \le \Delta_0^{(H)}/4$ for all $t \le T-1$, the same argument as Lemma D.5 (with H in place of F_{RKE}) implies $\min_i \alpha_{t,i} \ge \gamma_0$ for all $t \le T$. Then Lemma B.2 yields $n_{\min}(t) = \Omega(t)$ with high probability.

By Lemma C.3 (already in Appendix C),

$$H(\alpha_t) - \min_\alpha H(\alpha) \le 2\sup_\alpha|\widehat{H}_{t-1}(\alpha) - H(\alpha)|.$$

Moreover, by triangle inequality and linearity,

$$\sup_\alpha|\widehat{H}_{t-1}(\alpha) - H(\alpha)| \le \sup_\alpha\Big|\widehat{F}^{\mathrm{RKE}}_{t-1}(\alpha) - F_{\mathrm{RKE}}(\alpha)\Big| + w\sup_\alpha\Big|\widehat{G}_{t-1}(\alpha) - G(\alpha)\Big|.$$

The first term is bounded by Lemma D.3 and $n_{\min}(t-1) = \Omega(t-1)$. The second term is bounded by Lemma C.2. Therefore, on the intersection event, for all $t \ge 2$,

$$\sup_\alpha|\widehat{H}_{t-1}(\alpha) - H(\alpha)| \le \frac{C}{\sqrt{t-1}}\Big(1+\sqrt{\log\frac{m(T+1)}{\delta}}\Big).$$

Summing the instantaneous regret over $t = 2,\dots,T$ yields

$$\sum_{t=2}^{T}(t-1)^{-1/2} \le 2\sqrt{T},$$

which completes the $\mathcal{O}(\sqrt{T})$ bound (with the initial $t=1$ warm-start round absorbed into the constant). A union bound yields an overall probability of at least $1-\delta$. □

E. Additional Numerical Results

Effect of the UCB Coefficient. In Figure 6 For each dataset, we conducted two independent experiments corresponding to two distinct optimization objectives as (Rezaei et al., 2025) suggests. In the first experiment, the objective is to *minimize* the KD metric. In the second experiment, the objective is to *maximize* the RKE metric. As a performance upper bound, the *Mixture Oracle* is introduces as a baseline. For a given objective, the oracle provides the corresponding optimal mixture weights $\alpha^\star$ in advance, and arms are sampled according to this fixed distribution. The optimal mixture $\alpha^\star$ is obtained by solving the quadratic program using a large number of samples from each arm. We analyze the impact of the exploration coefficient δ_L in the UCB term for both objectives. Empirically, decreasing δ_L leads to faster convergence toward the corresponding Mixture Oracle. When $\delta_L = 0$, corresponding to the Mixture Greedy strategy, convergence is fastest.

E.1. Synthetic Settings

text-to-image setting. In Figure 7, we constructed a controlled text-to-image generation scenario using Stable Diffusion XL (SDXL) (Podell et al., 2023a). We generated samples corresponding to five dog breeds: `poodle`, `bulldog`, `german shepherd`, `golden retriever`, and `havanese`. Each breed-specific prompt defines one arm in our framework. The objective in this experiment is to optimize a diversity metric, namely the Vendi Score (Friedman & Dieng, 2023). Rather than optimizing the Vendi Score directly, we optimize its logarithm, which is a convex function of the mixture weights using sing exponentiated gradient descent (EGD). We additionally report results for the *One-Arm Greedy* algorithm, which selects at each round the arm with the highest current Vendi score, as well as an ϵ-greedy baseline with $\epsilon = 0.1$.

In Figure 5, we evaluate three text-to-image generators, Kandinsky (Razzhigaev et al., 2023), PixArt-α (Chen et al., 2023), and SDXL (Podell et al., 2023b), using the prompt: "Generate a red cartoony bird." We then apply the Mixture Greedy algorithm to adaptively select among these models in order to maximize the Vendi Score of the generated samples.

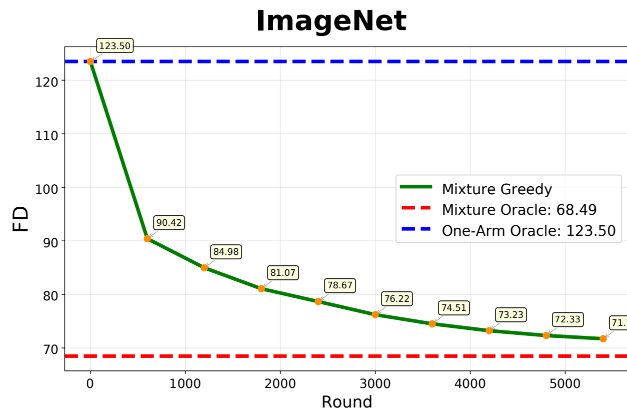

Figure 4. Convergence of FD in ImageNet dataset generative models.

As in our previous experiments, the algorithm rapidly converges toward the optimal mixture, leading to a substantial increase in Vendi Score compared to any individual model. This further supports the benefit of mixture optimization for enhancing generative diversity.

text-to-text setting. In Figure 8, we consider three large language models, Qwen 2(Qwen-Team, 2024), Gemma 3 (Gemma-Team, 2025), and Llama 3.2 (Aaron Grattafiori et al., 2024), prompted with: "Write a short sentence about a vibrant city in the U.S." We then apply the Mixture Greedy algorithm to adaptively select among the models so as to maximize the Vendi Score of the aggregated outputs.

The resulting mixture achieves a higher Vendi Score than any individual model evaluated in isolation. This demonstrates that adaptively combining multiple LLMs can yield greater diversity, as measured by the Vendi Score, than relying on a single model alone.

E.2. Ablation Studies

E.2.1. Number of Random Fourier Features and Cosine Kernel

To evaluate the robustness of Vendi Score optimization with respect to kernel approximation, we use the Gaussian kernel with Random Fourier Features (RFF) at $R \in \{256, 512, 1024\}$, and additionally compare against the cosine kernel. As shown in Figure 9, the results are consistent across all settings. On both datasets, *Mixture Greedy* consistently obtains the highest Vendi Score, outperforming *One Arm Greedy* and *One Arm Epsilon Greedy* for all values of R. Increasing R changes the curves only slightly and does not affect the relative ranking of the methods. The same trend also holds for the cosine kernel. These results indicate that our method is robust to both the number of random Fourier features and the choice of kernel. Its advantage is preserved for low- and high-dimensional RFF approximations, as well as when using the cosine kernel instead of the Gaussian kernel.

E.2.2. Kernel Bandwidth

We evaluate three kernel bandwidths, $\sigma \in \{20, 30, 40\}$, to examine the sensitivity of RKE and KID to this hyperparameter. Figure 10 shows that the overall conclusions are consistent across all bandwidths. In all cases, RKE increases with training while KID decreases, indicating stable convergence behavior. Across the three bandwidths, the relative ordering of the methods is unchanged: *Mixture Oracle* performs best, *Mixture Greedy* is consistently close to the oracle, and *Mixture UCB* performs worst. For RKE, all methods improve rapidly in the early stage and then plateau, while for KID they decrease sharply at first and then gradually stabilize. Although the absolute values of RKE and KID vary with σ, the same qualitative trend is preserved. This ablation suggests that our method is robust to kernel bandwidth choice. In particular, the strong performance of *Mixture Greedy* relative to *Mixture UCB*, and its small gap to *Mixture Oracle*, holds uniformly for $\sigma = 20$, 30, and 40.

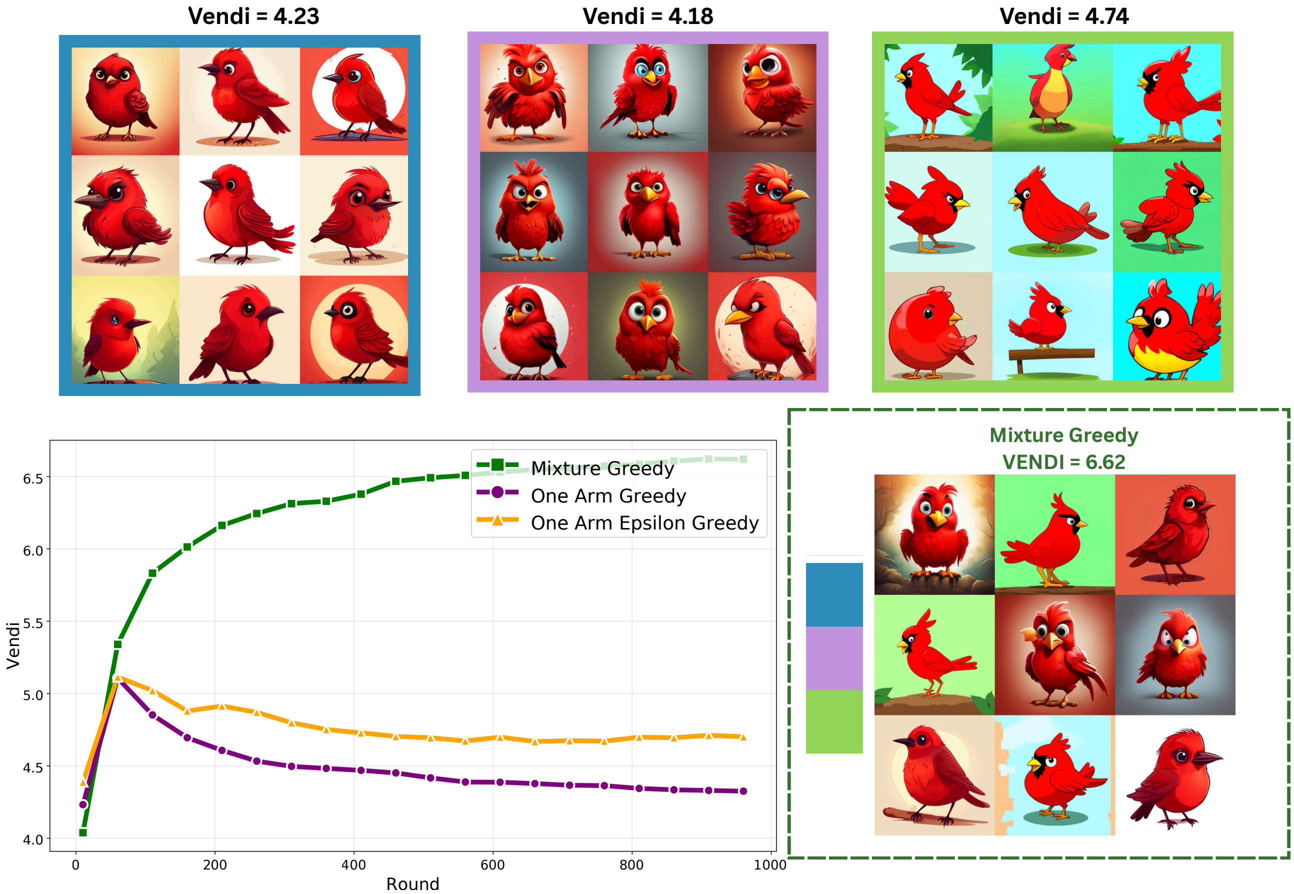

Figure 5. Application of Mixture Greedy on the generated red birds dataset to optimize Vendi Score.

E.2.3. Feature Extractors

To test the robustness of our results to the underlying representation, we compute RKE and KID using three different feature extractors: *Inception* (Szegedy et al., 2016), *CLIP* (Radford et al., 2021a), and *DINO-v2* (Oquab et al., 2024). Figure 11 shows that the conclusions are consistent across all three feature spaces. In every case, RKE increases and KID decreases over training, indicating stable convergence. The relative ranking of the methods is also unchanged: *Mixture Oracle* performs best, *Mixture Greedy* remains very close to it, and *Mixture UCB* performs worst. Although the absolute metric values vary across feature extractors, the same qualitative behavior is preserved. This demonstrates that our method is robust to the choice of feature representation, and that the strong performance of *Mixture Greedy* is not tied to a particular feature extractor.

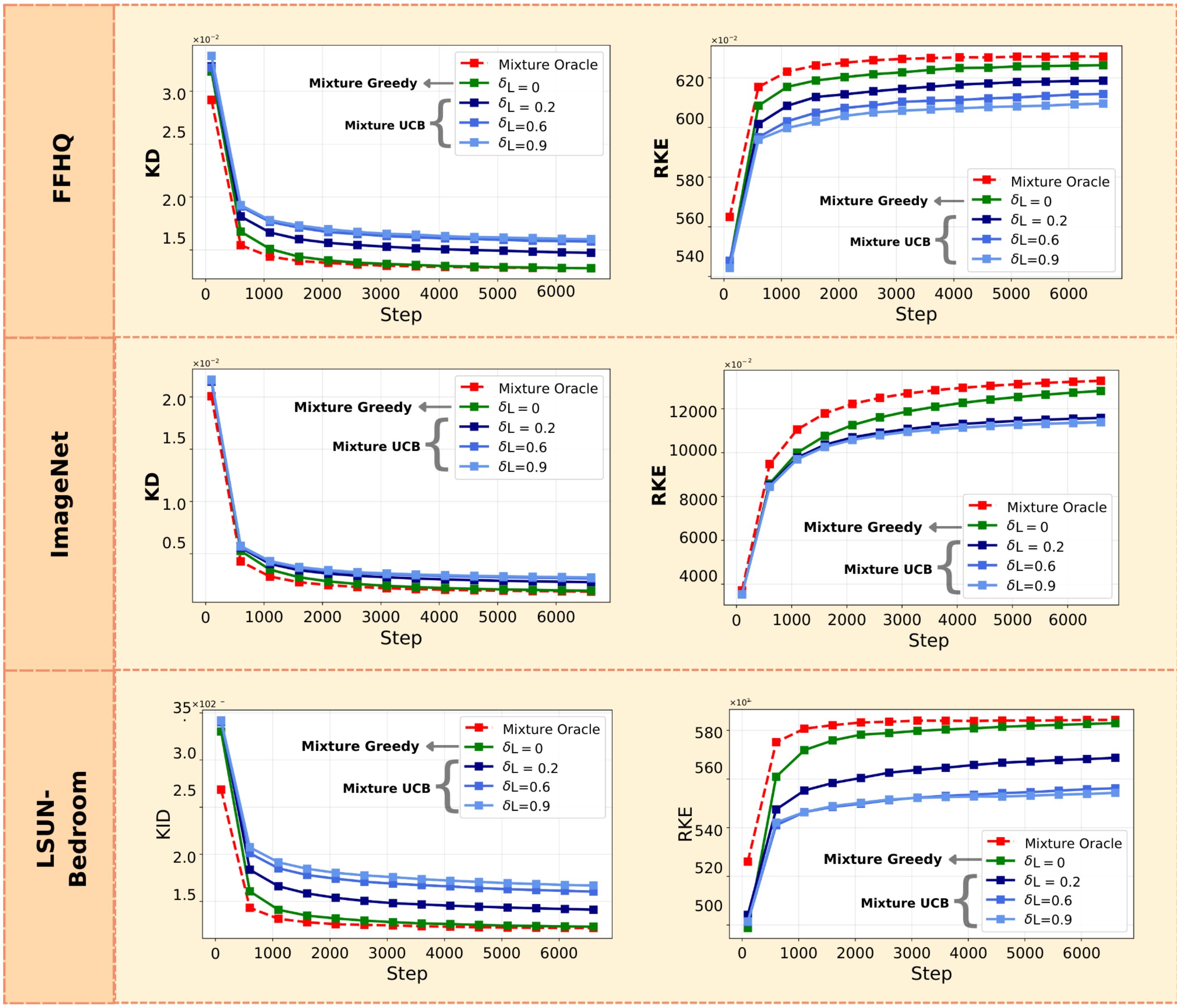

Figure 6. Comparison of KD and RKE convergence among different datasets.

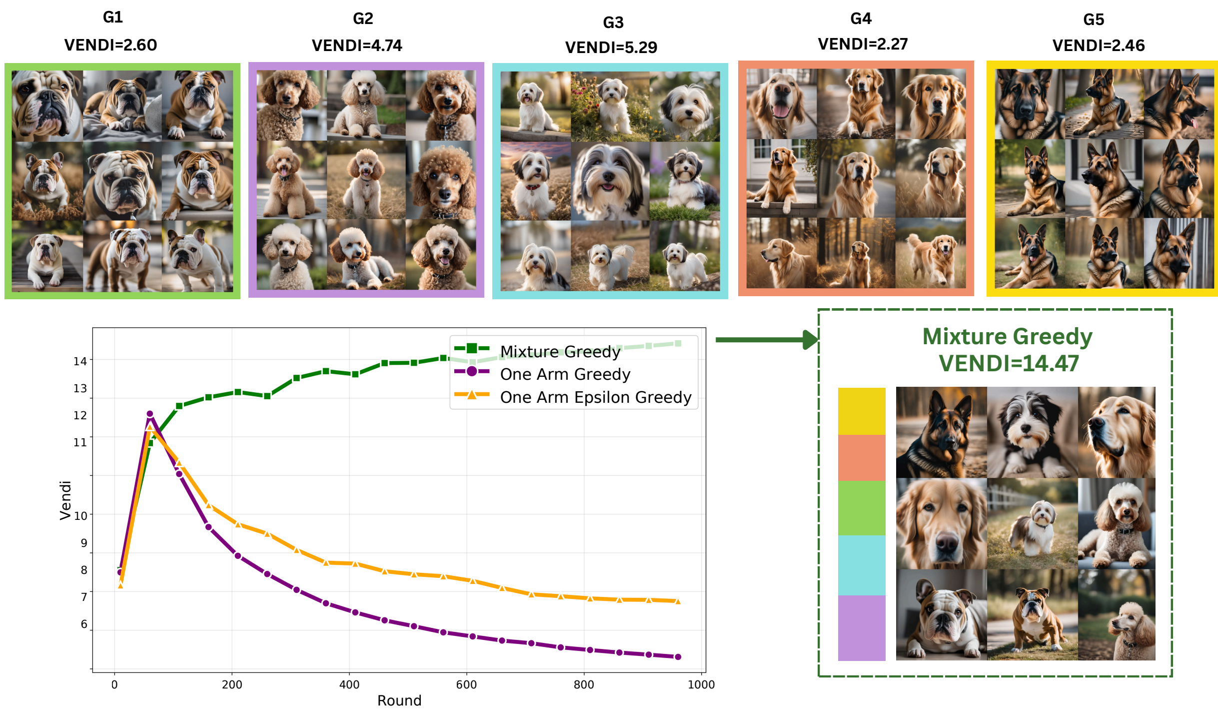

Figure 7. Application of Mixture Greedy on generated dog breeds dataset to optimize Vendi Score.

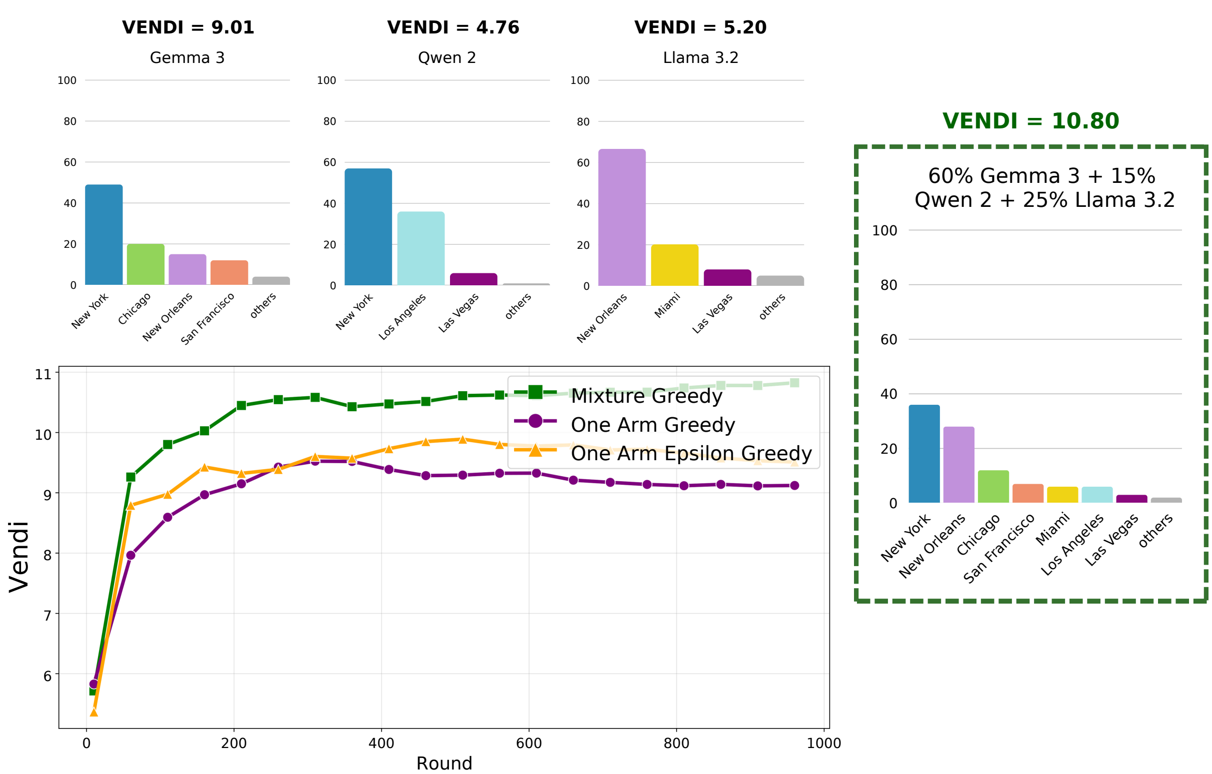

Figure 8. Application of Mixture Greedy on LLM generated texts to optimize Vendi Score.

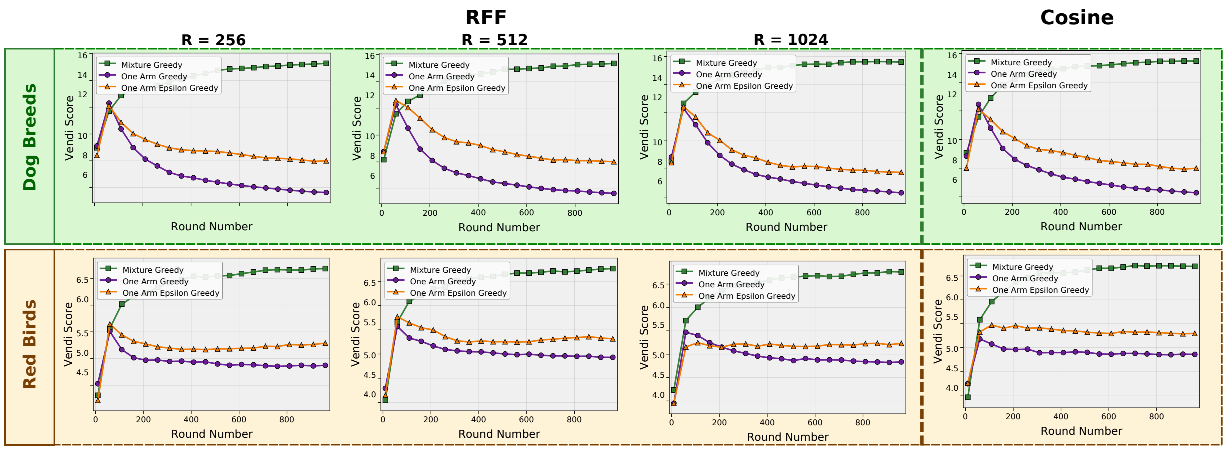

Figure 9. Vendi score over rounds for different using RFF (R=256/512/1024) and cosine kernel on two generated datasets.

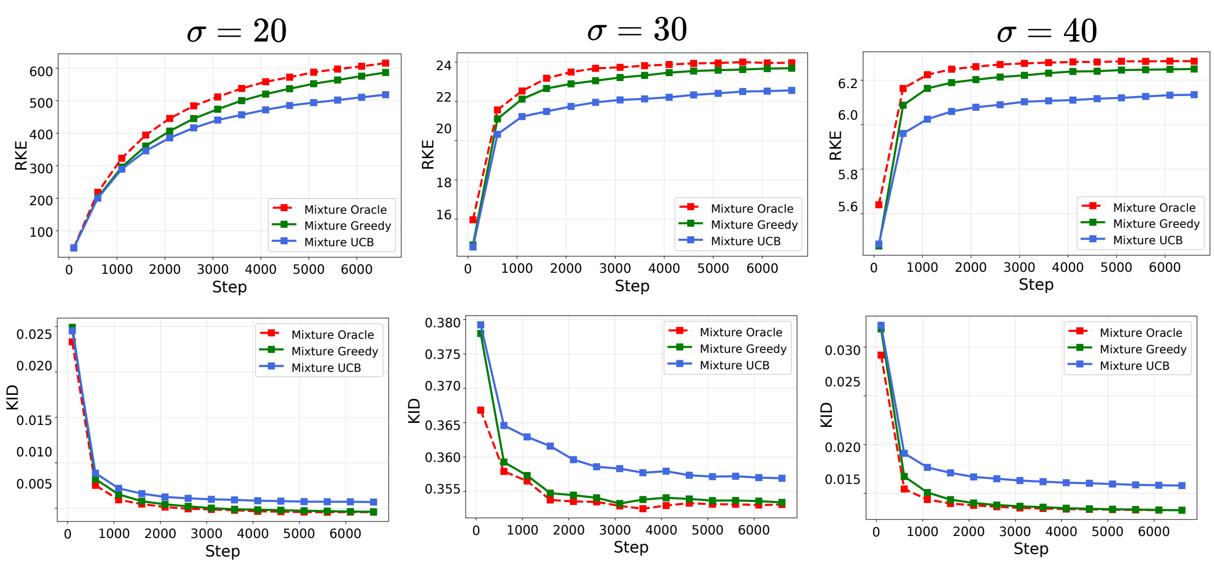

Figure 10. Convergence of KD and RKE for generative models trained on FFHQ dataset using three bandwidths

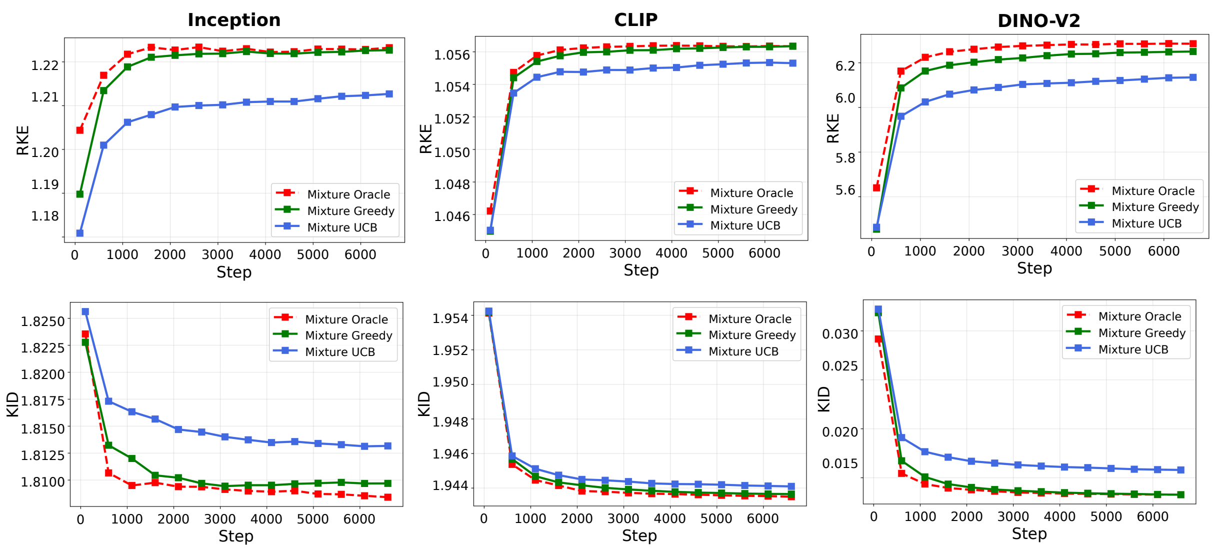

Figure 11. Convergence of RKE and KD on FFHQ generative models, using different feature extractors.