EXACT PRIVACY CALIBRATION FOR RANDOM-FEATURE KERNEL MEANS

Anonymous authors
Paper under double-blind review

ABSTRACT

Random Fourier features enable compact releases of private kernel means for querying and learning. Uniform-sensitivity calibration ignores variation in feature distances across random maps and can require more Gaussian noise than necessary. For scale mixtures of Gaussian kernels, we derive exact calibration of map-independent isotropic Gaussian noise for public Fourier feature means with independent or orthogonal frequencies. With independent frequencies, the required variance approaches half the uniform-sensitivity variance as the feature count grows on unbounded domains, with further reductions under public input bounds. Independent sampling also requires the least noise on the full input space among nonresonant designs with the same marginal frequency distribution and feature count. These noise reductions improve query accuracy most when privacy noise dominates feature approximation. Experiments across five kernels and four datasets show lower query error. With protected labels, our calibration improves data reconstruction and downstream SVM accuracy, with gains of 1–10 percentage points at $\varepsilon = 1$ over uniform-sensitivity calibration.

1 INTRODUCTION

Kernel mean embeddings summarize a dataset by averaging its feature vectors, supporting density queries, distribution comparison, and learning (Muandet et al., 2017; Balog et al., 2018). Random Fourier features approximate these embeddings with finite-dimensional vectors (Rahimi & Recht, 2007). Adding calibrated Gaussian noise to the feature mean gives a private summary (Harder et al., 2021). The feature map is drawn independently of a fixed dataset and published together with the noisy mean.

For n records with unit-norm features, standard calibration bounds the change in the mean under one record replacement by $2/n$ (Dwork et al., 2006). We show that accounting for variation across random maps allows the same privacy guarantee with less Gaussian noise. For a fixed pair of neighboring datasets, changing the map changes the distance between their feature means. The Gaussian privacy divergence measures how distinguishable their noisy means are for a given map (Balle & Wang, 2018). Averaging this divergence over maps before taking the worst neighboring pair gives the exact privacy guarantee for the joint release.

We derive exact calibration (Exact) for independently sampled Fourier features represented by sine–cosine pairs, with map-independent isotropic Gaussian noise. The analysis covers mixtures of Gaussian kernels with different bandwidths, including Gaussian (RBF), rational-quadratic, and Matérn kernels. Although individual features oscillate, we prove that larger input distances make larger feature distances more likely. The public input diameter determines the worst-case distance distribution and the minimum required noise. On unbounded domains, the required variance approaches half that of uniform-sensitivity analytic Gaussian calibration as the feature count grows. Public bounds on the input domain allow further reductions (Figure 1).

Independent sampling also requires the least noise on the full input space among nonresonant designs with the same marginal frequency distributions and feature count. Orthogonal random features (ORF) can reduce approximation error by introducing dependence among frequencies (Yu et al., 2016). We derive the exact privacy profile for block ORF in the same kernel family, reducing its calibration to a search over input distance.

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

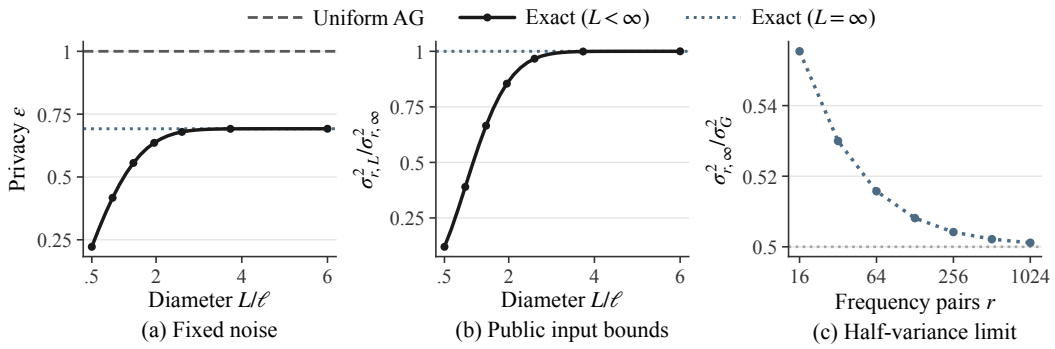


Figure 1: **Stronger privacy at fixed noise, or less noise at fixed privacy.** (a) With noise fixed at the Uniform AG scale for $\varepsilon = 1$, Exact gives a smaller valid ε. (b) At fixed privacy, a finite public diameter L reduces variance relative to Exact with $L = \infty$. (c) On the full input space, variance relative to Uniform AG approaches $1/2$ as r increases (gray dotted line). All panels use iid paired features and $\delta = 10^{-6}$, with $\varepsilon = 1$ in (b,c). Panels (a,b) use the RBF kernel with $r = 256$.

Experiments on Adult, Covertype, USPS, and ISOLET examine how the balance between approximation and privacy noise affects query error. At $\varepsilon = 1$, Exact reduces ORF query MSE by 12–57% relative to uniform-sensitivity calibration across the four datasets.

For learning, we release class means that protect both inputs and labels and reconstruct weighted data for support vector machines (SVMs). With ORF at $\varepsilon = 1$, Exact improves SVM accuracy by 1.1–9.8 percentage points relative to uniform-sensitivity calibration.

Relation to prior work. DP-MERF adds Gaussian noise to random feature means for private data generation (Harder et al., 2021). Chatalic et al. (2021) establish sharp sensitivity bounds for fixed Fourier maps. Analytic Gaussian calibration computes the minimum noise for a given sensitivity (Balle & Wang, 2018). We average the Gaussian privacy divergence over maps and identify the worst neighboring pair to obtain exact calibration for the map and noisy mean released together.

2 Public random-feature mean release

The goal is to release a noisy feature mean from which one can estimate the average kernel similarity between a query point and the dataset. We consider a dataset $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ of fixed, public size n, where $\mathcal{X} \subseteq \mathbb{R}^p$ is a public input domain with diameter $L \in (0, \infty]$.

2.1 Random Fourier features

Random Fourier features represent an input by evaluating sine and cosine functions at sampled frequencies (Rahimi & Recht, 2007). For the kernel family studied here, the frequencies are Gaussian vectors multiplied by random scales. Let ν be a public probability distribution on $(0, \infty)$, chosen independently of the data. For $j = 1, \ldots, r$, draw

$$S_j \sim \nu, \quad G_j \sim N(0, I_p), \quad \omega_j = \sqrt{S_j} G_j, \tag{1}$$

with all scales and Gaussian vectors mutually independent. Let $\Omega \in \mathbb{R}^{r \times p}$ have rows $\omega_j^\top$ and define the paired feature map

$$\phi_\Omega(x) = \frac{1}{\sqrt{r}} (\cos(\omega_j^\top x), \sin(\omega_j^\top x))_{j=1}^r. \tag{2}$$

Each frequency contributes two coordinates, and $\|\phi_\Omega(x)\| = 1$ for every input. Taking the expectation of the feature inner product over the random map gives the kernel

$$\mathbb{E}_\Omega [\phi_\Omega(x)^\top \phi_\Omega(y)] = k(x, y) := \int_{(0,\infty)} \exp\left(-\frac{s\|x - y\|^2}{2}\right) \nu(ds). \tag{3}$$

For each scale s, the integrand is a Gaussian kernel with bandwidth $s^{-1/2}$. The distribution ν determines how these kernels are averaged (Rasmussen & Williams, 2005). For the RBF kernel with bandwidth ℓ, ν is a point mass at ℓ^{-2}, so every S_j equals ℓ^{-2}. Choosing ν to be a Gamma distribution gives a rational-quadratic kernel, while an inverse-Gamma distribution gives Matérn kernels, including the Euclidean-distance Laplacian. Appendix F.2 gives the mixing laws and parameter conventions.

2.2 Releasing and querying the mean

The mechanism draws a fresh map independently of the fixed dataset, then releases it together with a noisy empirical feature mean:

$$\mu_\Omega(D) = \frac{1}{n}\sum_{i=1}^{n}\phi_\Omega(x_i), \tag{4}$$

$$M(D) = (\Omega, \widetilde{\mu}_D), \quad \widetilde{\mu}_D = \mu_\Omega(D) + \sigma Z, \quad Z \sim N(0, I_{2r}). \tag{5}$$

Here Z is independent of the map and the records, and $\sigma > 0$ is chosen independently of the realized map and the records. A recipient can use the public map to compute the feature vector of a new query $x_\star$ and return

$$\widehat{f}_D(x_\star) = \phi_\Omega(x_\star)^\top \widetilde{\mu}_D \quad \text{to estimate} \quad f_D(x_\star) = \frac{1}{n}\sum_{i=1}^{n} k(x_i, x_\star). \tag{6}$$

For a fixed query, the estimate is unbiased over the map and noise. Its error comes from feature approximation and Gaussian perturbation.

2.3 Privacy profile

We use central differential privacy under replacement adjacency, where $D \sim D'$ means that the datasets differ in one record (Dwork & Roth, 2014). The guarantee covers the joint output $(\Omega, \widetilde{\mu}_D)$, with probabilities taken over both the map and the Gaussian noise. For output distributions P, Q, the hockey-stick divergence measures the largest gap in the DP inequality over measurable events A. The privacy profile takes its worst case over ordered neighboring datasets:

$$D_{e^\varepsilon}(P\|Q) = \sup_A\{P(A) - e^\varepsilon Q(A)\}, \tag{7}$$

$$\delta_M(\varepsilon;\sigma) = \sup_{D\sim D'} D_{e^\varepsilon}(M(D)\|M(D')). \tag{8}$$

The release is (ε, δ)-DP exactly when $\delta_M(\varepsilon;\sigma) \le \delta$.

For Gaussian distributions with covariance $\sigma^2 I$ and mean separation d, the divergence is $g_\varepsilon(d/\sigma)$, where (Balle & Wang, 2018)

$$g_\varepsilon(t) = \Phi(t/2 - \varepsilon/t) - e^\varepsilon \Phi(-t/2 - \varepsilon/t), \quad g_\varepsilon(0) = 0. \tag{9}$$

Here Φ is the standard normal CDF. The function increases with t, since more widely separated means are easier to distinguish at the same noise scale.

3 Privacy calibration and sharp limits

3.1 The joint profile and the worst input distance

Consider fixed neighboring datasets whose replaced records are x and x'. Their common records cancel in the difference between feature means, so with $z = x - x'$ the squared distance is

$$d_\Omega(z)^2 := \|\mu_\Omega(D) - \mu_\Omega(D')\|^2 = \frac{2}{n^2 r}\sum_{j=1}^{r}[1 - \cos(\omega_j^\top z)]. \tag{10}$$

Because the map is sampled independently of the records, it has the same distribution under both datasets. The joint divergence is the average conditional Gaussian divergence over the map:

$$D_{e^\varepsilon}(M(D)\|M(D')) = \mathbb{E}_\Omega g_\varepsilon\left(\frac{d_\Omega(z)}{\sigma}\right). \tag{11}$$

For each pair, Equation (11) accounts for all events that distinguish the outputs using the public map. Taking the worst case over pairs only after this average gives the exact profile. By contrast, applying the uniform sensitivity bound $2/n$ used in prior work (Harder et al., 2021) gives

$$\sup_z \mathbb{E}_\Omega g_\varepsilon(d_\Omega(z)/\sigma) \;\le\; \mathbb{E}_\Omega \sup_z g_\varepsilon(d_\Omega(z)/\sigma) \;\le\; g_\varepsilon(2/(n\sigma)), \tag{12}$$

where $z \in \mathcal{X}-\mathcal{X}$. The middle term takes the worst displacement separately for each map, producing an upper bound on the fixed-pair quantity on the left. We now find the input distance $\|z\|$ that maximizes the map-averaged divergence on the left.

For a single frequency with scale S fixed, the angle $\omega^\top z$ is Gaussian with mean zero and variance $S\|z\|^2$. Since cosine is 2π-periodic, adding or subtracting multiples of 2π to place the angle in $[-\pi, \pi]$ leaves its contribution to Equation (10) unchanged. The next lemma shows that larger Gaussian variance makes larger absolute angles more likely even after this reduction. We write $X \le_{\mathrm{st}} Y$ when $\Pr(X > t) \le \Pr(Y > t)$ for every threshold t.

Lemma 1 (Wrapped-angle order). *Let Θ_v be $N(0, v)$ reduced modulo 2π to $[-\pi, \pi]$. For $0 \le v \le w$, $|\Theta_v| \le_{\mathrm{st}} |\Theta_w|$. As $v \to \infty$, Θ_v converges in distribution to $U \sim \mathrm{Unif}[-\pi, \pi]$.*

Since $1 - \cos(a)$ increases on $[0, \pi]$, the lemma orders each summand in Equation (10) by input distance. Averaging over S and summing independent terms preserve this order, so larger $\|z\|$ makes larger feature distances more likely. Since g_ε is increasing, the expected divergence also increases with input distance. The worst case is obtained as input distances approach L. Appendix A.1 proves the lemma.

Theorem 2 (Exact profile for any input diameter). *For any public domain of diameter $L \in (0, \infty]$, the mechanism in Equation* (5) *has exact privacy profile*

$$\delta_M(\varepsilon; \sigma) = H_{r,L}(\varepsilon, n\sigma) := \mathbb{E} g_\varepsilon\left(\frac{\sqrt{2\bar{A}_{r,L}}}{n\sigma}\right). \tag{13}$$

Here $\bar{A}_{r,L} = r^{-1}\sum_{j=1}^r [1 - \cos(L\sqrt{S_j}B_j)]$ for finite L, with mutually independent $S_j \sim \nu$ and $B_j \sim N(0,1)$. For $L = \infty$, set $\bar{A}_{r,\infty} = r^{-1}\sum_{j=1}^r [1 - \cos U_j]$ with iid $U_j \sim \mathrm{Unif}[-\pi, \pi]$.

The proof is given in Appendix A.

At fixed mixing distribution and diameter, the profile is independent of the input dimension p. For $L = \infty$, the uniform-angle law also removes dependence on the mixing distribution.

3.2 Noise calibration and variance reduction

Theorem 2 determines the smallest Gaussian noise standard deviation for a target (ε, δ), with $\varepsilon \ge 0$ and $0 < \delta < 1$. We call this calibration *Exact* and denote its scale by $\sigma_{r,L}$. For comparison, uniform-sensitivity analytic Gaussian calibration (Uniform AG) uses the bound $2/n$ and has scale σ_G (Balle & Wang, 2018). The two scales solve

$$H_{r,L}(\varepsilon, n\sigma_{r,L}) = \delta, \quad g_\varepsilon\left(\frac{2}{n\sigma_G}\right) = \delta. \tag{14}$$

The calibration uses finite L when a public diameter bound is available and $L = \infty$ otherwise. We evaluate the expectation and solve Equation (14) for the noise scale (Appendix D).

Corollary 3 (Variance reduction). *Let $\kappa_L = \mathbb{E}_S e^{-SL^2/2}$ be the kernel similarity at distance L, and set $\kappa_\infty = 0$. For every fixed r, $0 < \sigma_{r,L} \le \sigma_{r,\infty} < \sigma_G$, and*

$$\lim_{r\to\infty} \frac{\sigma_{r,L}^2}{\sigma_G^2} = \frac{1 - \kappa_L}{2}. \tag{15}$$

Exact uses less noise than Uniform AG at every feature count, approaching half its variance when $L = \infty$ (Figure 1c). A finite public diameter permits further variance reduction because even the most distant admissible inputs retain positive kernel similarity κ_L (Figure 1b).

3.3 Optimality of independent frequencies

Dependent frequencies can reduce kernel approximation variance while preserving the expected kernel, as in ORF (Yu et al., 2016) and joint sampling of frequency norms (Reid et al., 2025). They also change the feature distance distribution that determines privacy. We ask whether this dependence can also reduce privacy noise at the same feature count r and marginal frequency distribution.

We compare designs satisfying nonresonance, which rules out exact integer relations among the frequencies (Chatalic et al., 2021):

$$\Pr\left(\sum_{j=1}^{r} a_j \omega_j = 0\right) = 0 \quad \text{for every } a \in \mathbb{Z}^r \setminus \{0\}. \tag{16}$$

The integer restriction allows real linear dependence, including $r > p$. The block ORF construction with positive scale mixtures in Section 3.4 satisfies this condition.

Theorem 4 (Optimality of independent sampling)**.** *On $\mathcal{X} = \mathbb{R}^p$, consider the joint release in Equation* (5) *with fixed, map-independent isotropic Gaussian noise of scale $\sigma > 0$. Among nonresonant designs with r frequencies, each having the marginal distribution in Equation* (1)*, let $\delta_{\mathcal{E}}(\varepsilon;\sigma)$ denote the joint privacy profile of design $\mathcal{E}$. For every $\varepsilon \geq 0$, independent sampling attains the infimum over these designs:*

$$\inf_{\mathcal{E}} \delta_{\mathcal{E}}(\varepsilon;\sigma) = H_{r,\infty}(\varepsilon, n\sigma). \tag{17}$$

Its noise scale $\sigma_{r,\infty}$ is minimal among these designs for every target (ε,δ) with $0 < \delta < 1$.

The lower bound follows by averaging divergence over Gaussian input differences with increasing variance. Under nonresonance, the wrapped angles converge jointly to independent uniform angles, recovering the iid profile $H_{r,\infty}$ (Appendix B).

Dependent designs can still improve query accuracy through lower approximation error. Their privacy noise requires a separate profile, which we derive next for ORF.

3.4 Calibrating orthogonal random features

We extend block ORF (Yu et al., 2016) to the kernel family in Equation (3). The r frequencies form independent blocks of sizes $m_b \leq p$ with $\sum_b m_b = r$. Each block multiplies uniformly rotated orthogonal directions by independent radii $R_{bj} \sim \chi_p$ and scales $\sqrt{S_{bj}}$, $S_{bj} \sim \nu$. Here χ_p is the distribution of the norm of a standard Gaussian vector in $\mathbb{R}^p$. Each frequency has the distribution in Equation (1), so the expected kernel is unchanged. Rotation invariance then reduces the worst case to a search over input distance.

Theorem 5 (Exact calibration for orthogonal features)**.** *Let $\mathcal{X}$ be convex with diameter $L \in (0,\infty]$. For the block ORF map above, the joint release in Equation* (5) *has exact privacy profile*

$$\delta_{\mathrm{O}}(\varepsilon;\sigma) = \sup_{0 \leq \rho \leq L} \mathbb{E} g_\varepsilon\left(\frac{\sqrt{2\bar{A}_{\mathrm{O}}(\rho)}}{n\sigma}\right). \tag{18}$$

Here $\bar{A}_{\mathrm{O}}(\rho) = r^{-1}\sum_b \sum_{j=1}^{m_b}[1 - \cos(\rho\sqrt{S_{bj}}R_{bj}V_{bj})]$. The V_{bj} are coordinates of independent uniform unit vectors $V_b \in \mathbb{R}^p$, sampled independently of the radii and scales. For $L = \infty$, the supremum is over all finite $\rho \geq 0$. For every $\varepsilon \geq 0$ and $0 < \delta < 1$, the equation $\delta_{\mathrm{O}}(\varepsilon;\sigma) = \delta$ has a unique solution σ_{O}, the minimum fixed, map-independent isotropic noise scale for this release. It satisfies $0 < \sigma_{\mathrm{O}} < \sigma_G$, and on $\mathcal{X} = \mathbb{R}^p$, $\sigma_{r,\infty} \leq \sigma_{\mathrm{O}}$.

Unlike the independent frequencies in Theorem 2, ORF need not have its worst input distance at the diameter. Appendix C proves the profile, and Appendix D.1 gives a computable upper bound for RBF calibration that accounts for integration and distance-search error.

4 Kernel queries and means with protected labels

4.1 Error of a kernel query

For a fixed dataset D and query $x_\star$, let A be the mean squared error of the unperturbed query estimate, averaged over the public map. The query vector has unit norm, so Gaussian noise adds

variance σ^2:

$$\mathbb{E}_{\Omega,Z}\left[(\widehat{f}_D(x_\star) - f_D(x_\star))^2\right] = \underbrace{A}_{\text{approximation error}} + \underbrace{\sigma^2}_{\text{privacy noise}}. \tag{19}$$

For iid frequencies, A decreases as $1/r$ under standard RFF variance analysis (Sutherland & Schneider, 2015, Section 2.1). For ORF, correlations between frequencies also contribute to the approximation variance (Yu et al., 2016).

At the same kernel, feature count, and frequency design, Exact changes only the noise term. Let σ_E denote its scale, given by $\sigma_{r,L}$ for iid features and σ_{O} for ORF. The relative MSE reduction over Uniform AG is

$$1 - \frac{A + \sigma_E^2}{A + \sigma_G^2} = \left(1 - \frac{\sigma_E^2}{\sigma_G^2}\right) \frac{\sigma_G^2}{A + \sigma_G^2}. \tag{20}$$

This is the noise variance reduction multiplied by the fraction of error due to privacy noise. When privacy noise dominates, most of its relative variance reduction carries over to query MSE. Approximation limits the gain when noise is small.

4.2 Means with protected labels

Reconstructing labeled data and training a classifier require information about the association between inputs and labels. We retain this information by releasing a separate feature mean for each class. For records (x_i, y_i) and a public label set of size $C \geq 2$, define

$$\mu_c = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{y_i = c\} \phi_\Omega(x_i), \quad c = 1, \ldots, C. \tag{21}$$

Normalizing every block by n retains each class's share of the data. We release the stacked blocks with the map and isotropic Gaussian noise as in Equation (5). A same-label replacement follows the unlabeled distance law of the chosen feature design. A cross-label replacement changes two orthogonal blocks and has distance $\sqrt{2}/n$, since each feature vector has unit norm. The minimum fixed, map-independent isotropic noise scale is

$$\sigma_{\mathrm{lab}} = \max\left\{\sigma_E, \frac{\sigma_G}{\sqrt{2}}\right\} \tag{22}$$

(Appendices A.5 and C). A finite public diameter can reduce the same-label requirement, while cross-label replacements impose a floor independent of the input domain.

5 Experiments

We evaluate how calibration affects kernel queries and learning from reconstructed data on Adult, Covertype, USPS, and ISOLET.

The query experiments compare the independently sampled frequencies of Equation (1), denoted iid, with ORF from Section 3.4. Each feature design is calibrated with both Uniform AG and its own Exact profile. For ORF, we calibrate using the profile upper bound in Appendix D.1. Within each feature design, we keep the records and kernel fixed and use the same map for both calibrations.

All reference settings use $n = 5000$ protected records and $r = 2p$ frequency pairs, where p is the public preprocessed input dimension. This gives every dataset the same record budget and two complete ORF blocks. The RBF bandwidth is $\gamma = 1/p$, equivalently $\ell = \sqrt{p/2}$. We evaluate other bandwidth choices in Appendices G.2 and J. Inputs are clipped to fixed coordinate ranges, and Exact uses their box's diagonal length as the public diameter L. We vary $\varepsilon \in \{0.1, 0.2, 0.5, 1, 2, 5\}$ at $\delta = 10^{-6}$. Each query condition uses 100 independently sampled feature maps. Reconstruction and classification use 30 trials with independent maps, noise, and initialization. Record splits remain fixed across trials. Appendix F specifies preprocessing and evaluation.

5.1 Kernel queries across privacy budgets

Query MSE compares the estimate $\widehat{f}_D(x_\star)$ with the exact kernel average $f_D(x_\star)$ in Equation (6), averaging squared error over evaluation queries and the randomness of the release. Figure 2 reports absolute MSE and the reduction from Exact for iid and ORF features separately.

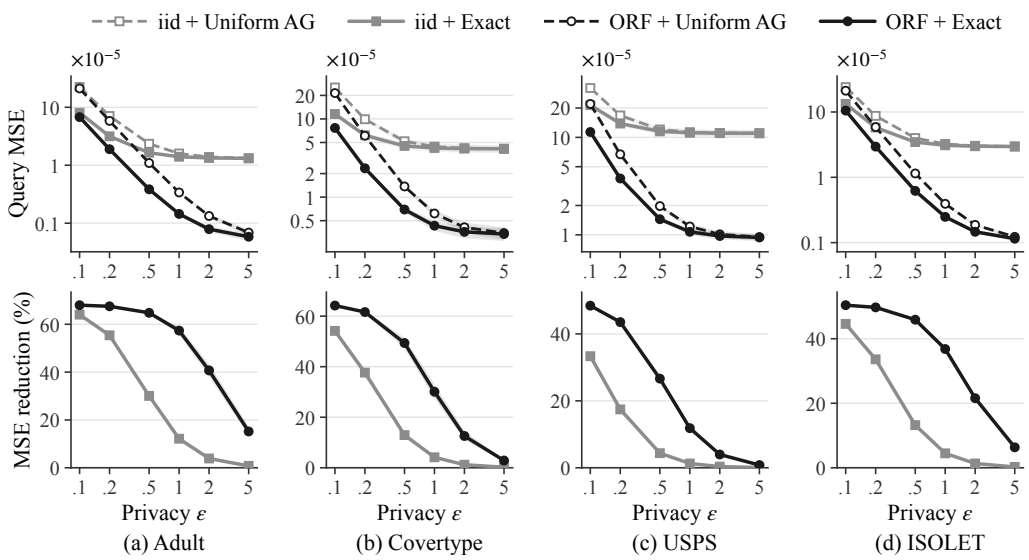


Figure 2: **Kernel-query error across privacy budgets.** Top: absolute RBF query MSE on logarithmic axes at $n = 5000$, $r = 2p$. Bottom: percentage MSE reduction from Exact relative to Uniform AG for each feature design. Bands are pointwise 95% confidence intervals over 100 independently sampled maps, with paired bootstrap resampling for the reductions and fixed record splits.

At $\varepsilon = 1$, Exact reduces ORF query MSE by 57%, 30%, 12%, and 37% on Adult, Covertype, USPS, and ISOLET, respectively. The iid reductions range from 1% to 12%. The larger ORF gains agree with Equation (20), since reducing approximation error leaves privacy noise as a larger fraction of total error. At $\varepsilon = 0.1$, Exact reduces MSE by 33–68% across both designs and all four datasets, and iid Exact outperforms ORF Uniform AG. As ε grows, the iid curves flatten while the ORF curves continue to decrease.

5.2 Sample size and feature count

We vary feature count and sample size separately to examine the balance between approximation and privacy noise. Figure 3 varies $r = \lceil ap \rceil$ for $a \in \{0.5, 1, 2, 4, 8\}$ at $n = 5000$, and uses nested record subsets of size 500, 2000, and 5000 at $r = 2p$. Both sweeps fix $\varepsilon = 1$.

Increasing the feature count lowers MSE for both designs, with the largest ORF drop between a partial block and the first complete block at $r = p$. Additional blocks give smaller absolute improvements. The unit-norm query keeps the noise term equal to σ^2 even as the released vector grows (Equation (19)).

Increasing n reduces the calibrated noise variance as n^{-2} and narrows the absolute MSE gap between calibrations. At $n = 500$, iid Exact has lower MSE than ORF Uniform AG on all four datasets, while the ordering reverses at $n = 5000$.

5.3 Kernel families

Figure 4 extends the iid comparison to rational-quadratic, Matérn, and Euclidean Laplacian kernels, keeping the same bandwidth rule and resource budgets. Exact lowers MSE throughout the comparison, but the reduction varies across kernels. At $\varepsilon = 0.1$, it ranges from 9% to 64%, with larger gains for RBF and rational-quadratic kernels than for Laplacian on Adult and Covertype. By $\varepsilon = 5$, every reduction is below 1%.

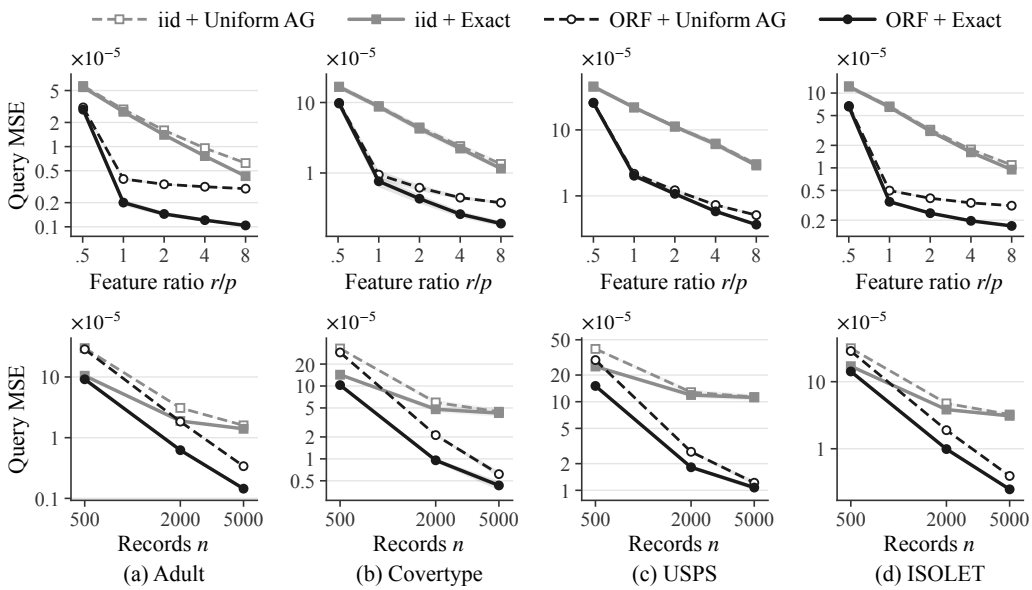


Figure 3: **Absolute error across feature and record budgets.** RBF query MSE on logarithmic axes at $\varepsilon = 1$. Top: vary r/p with $n = 5000$. Bottom: vary n with $r = 2p$. Each condition uses 100 maps, with pointwise 95% intervals. The query target is the empirical kernel mean of the corresponding record subset.

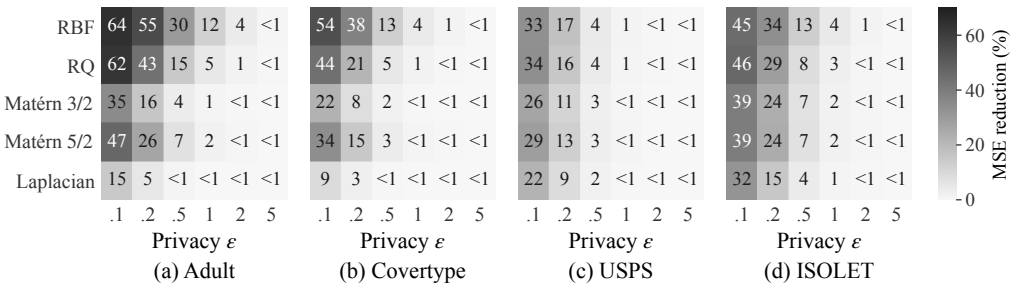


Figure 4: **Calibration gains across kernel families.** Percentage reduction in iid query MSE relative to Uniform AG, at $n = 5000$, $r = 2p$, averaged over 100 maps. RQ has shape one, and Laplacian uses Euclidean distance. Each comparison holds its target kernel fixed. < 1 denotes a positive reduction below 1%. Absolute MSE values appear in Table 3.

5.4 Reconstruction and downstream learning

We reconstruct $m = 5000$ weighted labeled inputs from the ORF release in Section 4.2. Following reduced-set reconstruction (Balog et al., 2018), we fix their labels and adjust their locations within the public box and nonnegative weights to approximate the released class means (Appendix H). Squared maximum mean discrepancy (MMD) compares the reconstruction with the protected data using the original kernel. Inputs with different labels have zero similarity in this metric.

We train SVMs on the weighted reconstructions, using binary hinge loss on Adult and Crammer–Singer multiclass hinge loss (Crammer & Singer, 2002) on Covertype, USPS, and ISOLET (7, 10, and 26 classes). Reconstruction and SVM training are post-processing of the released means and require no auxiliary target-domain examples. As a baseline for classification, direct DP-SVM trains on the protected records and perturbs the model weights W. We use Gaussian output perturbation for binary classification (Chaudhuri et al., 2011; Rubinstein et al., 2012) and the multiclass method of Park et al. (2025). All methods share the records, ORF map, bandwidth, and privacy budget. With

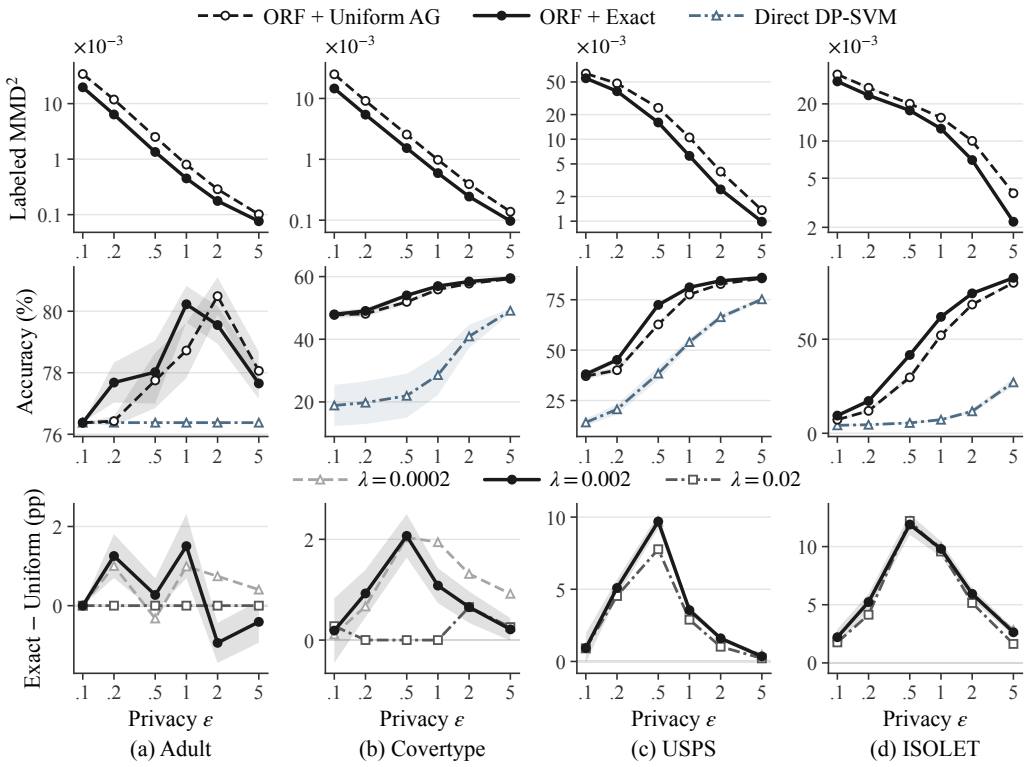


Figure 5: **Labeled reconstruction and learning.** Top: original-kernel joint input–label MMD squared (log axes). Middle: SVM accuracy from the same reconstructions ($\lambda = 0.002$) and direct DP-SVM ($\lambda = 0.5$). Bottom: paired Exact–Uniform accuracy differences at three shared regularizations. Here $n = m = 5000$, $r = 2p$. Bands are pointwise 95% intervals over 30 trials.

average hinge loss and penalty $\lambda\|W\|_F^2/2$, we use $\lambda = 0.002$ for reconstructions and $\lambda = 0.5$ for direct DP-SVM. These settings are fixed across datasets and budgets after preliminary benchmark exploration, whose privacy cost is outside this evaluation (Appendix I).

In these reference conditions, the cross-label term in Equation (22) sets the Exact scale, halving variance relative to Uniform AG. Exact lowers mean reconstruction MMD squared at every displayed budget and dataset (Figure 5, top). At $\varepsilon = 1$, the reductions range from 19% on ISOLET to 44% on Adult. For SVMs on the same reconstructions, Exact improves accuracy by 1.5, 1.1, 3.6, and 9.8 percentage points on Adult, Covertype, USPS, and ISOLET, respectively (middle). Both calibrations yield higher SVM accuracy than direct DP-SVM on all three multiclass datasets throughout the displayed grid.

On Adult at $\varepsilon = 2$, Exact lowers labeled MMD, but its SVM accuracy gain is positive for $\lambda = 0.0002$ and negative for $\lambda = 0.002$ (bottom). USPS and ISOLET have positive mean gains at all six budgets and all three displayed regularizations.

6 Discussion

Exact calibration improves kernel queries most when privacy noise dominates feature approximation. With protected labels, cross-label replacements cap the variance reduction at 50% relative to Uniform AG, even under tighter input bounds. Lower reconstruction MMD need not improve classification at every regularization, as the Adult results show.

The analysis assumes Gaussian scale-mixture kernels, data-independent maps, and map-independent isotropic noise. Other kernels and noise covariances require further analysis. The optimal frequency design on bounded domains also remains open.

AI USE STATEMENT

Generative AI tools assisted with research ideation and execution, drafting, and polishing manuscript text. The authors take responsibility for the final text, methods, claims, and results.

REFERENCES

Borja Balle and Yu-Xiang Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 394–403, 2018.

Matej Balog, Ilya Tolstikhin, and Bernhard Schölkopf. Differentially private database release via kernel mean embeddings. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 414–422, 2018.

Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996. DOI: https://doi.org/10.24432/C5XW20.

Jock Blackard. Covertype. UCI Machine Learning Repository, 1998. DOI: https://doi.org/10.24432/C50K5N.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-Lopez, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5230–5242, 2022.

A Chatalic, V Schellekens, F Houssiau, Y A de Montjoye, L Jacques, and R Gribonval. Compressive learning with privacy guarantees. *Information and Inference: A Journal of the IMA*, 11(1): 251–305, 2021.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(29):1069–1109, 2011.

Yujin Choi, Jinseong Park, Youngjoo Park, Jaewook Lee, and Junyoung Byun. Differentially private upsampling for enhanced anomaly detection in imbalanced data. *Engineering Applications of Artificial Intelligence*, 163:113124, 2026.

Ron Cole and Mark Fanty. ISOLET. UCI Machine Learning Repository, 1991. DOI: https://doi.org/10.24432/C51G69.

Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2:265–292, 2002.

DLMF. *NIST Digital Library of Mathematical Functions*. `https://dlmf.nist.gov/`, Release 1.2.8 of 2026-09-15.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–487, 2014.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pp. 265–284, 2006.

Frederik Harder, Kamil Adamczewski, and Mijung Park. Dp-merf: Differentially private mean embeddings with randomfeatures for practical privacy-preserving data generation. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pp. 1819–1827, 2021.

Matthew Joseph and Alexander Yu. Some constructions of private, efficient, and optimal k-norm and elliptic gaussian noise. In *Proceedings of Thirty Seventh Conference on Learning Theory*, pp. 2723–2766, 2024.

Bryna Kra, Nimish A. Shah, and Wenbo Sun. Equidistribution of dilated curves on nilmanifolds. *Journal of the London Mathematical Society*, 98(3):708–732, 2018.

Christian Janos Lebeda. Better gaussian mechanism using correlated noise. In *2025 Symposium on Simplicity in Algorithms (SOSA)*, pp. 119–133, 2025.

Bonwoo Lee, Jeongyoun Ahn, and Cheolwoo Park. Differential privacy in scalable general kernel learning via k-means nyström random features. In *Advances in Neural Information Processing Systems*, volume 37, pp. 21141–21171, 2024.

Bonwoo Lee, Cheolwoo Park, and Jeongyoun Ahn. Optimal differentially private kernel learning with random projection. *arXiv preprint arXiv:2507.17544*, 2026.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization, 2009. URL https://arxiv.org/abs/0907.3740.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141, 2017.

Aleksandar Nikolov and Haohua Tang. General Gaussian Noise Mechanisms and Their Optimality for Unbiased Mean Estimation. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pp. 85:1–85:23, 2024.

Jinseong Park, Yujin Choi, Junyoung Byun, Jaewook Lee, and Saerom Park. Efficient differentially private kernel support vector classifier for multi-class classification. *Information Sciences*, 619: 889–907, 2023.

Jinseong Park, Yujin Choi, and Jaewook Lee. Multi-class support vector machine with differential privacy. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 73044–73071, 2025.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

Isaac Reid, Stratis Markou, Krzysztof Choromanski, Richard E Turner, and Adrian Weller. Variance-reducing couplings for random features. In *International Conference on Learning Representations*, pp. 21749–21788, 2025.

Benjamin I. P. Rubinstein, Peter L. Bartlett, Ling Huang, and Nina Taft. Learning in a large function space: Privacy-preserving mechanisms for svm learning. *Journal of Privacy and Confidentiality*, 4(1), 2012.

Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(50):1517–1561, 2010.

Danica J. Sutherland and Jeff Schneider. On the error of random fourier features. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 862–871, 2015.

Filip Tronarp, Toni Karvonen, and Simo Sarkka. Mixture representation of the matÉrn class with applications in state space approximations and bayesian quadrature. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2018.

Margarita Vinaroz, Mohammad-Amin Charusaie, Frederik Harder, Kamil Adamczewski, and Mi Jung Park. Hermite polynomial features for private data generation. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 22300–22324, 2022.

Tal Wagner, Yonatan Naamad, and Nina Mishra. Fast private kernel density estimation via locality sensitive quantization. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 35339–35367, 2023.

Felix Xinnan X Yu, Ananda Theertha Suresh, Krzysztof M Choromanski, Daniel N Holtmann-Rice, and Sanjiv Kumar. Orthogonal random features. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

A EXACT PRIVACY PROFILE AND CALIBRATION

The proof has three steps: order the single-frequency distance, average the conditional Gaussian privacy profile over the public map, and construct input pairs approaching the bound. We then derive the minimum noise scale, its limiting value, and the extension to protected labels.

A.1 WRAPPED-NORMAL STOCHASTIC ORDER

We prove Lemma 1 by showing that every interval centered at angle zero loses probability as the Gaussian variance increases. The wrapped density decreases away from zero, and the heat equation converts that spatial decrease into the required change in probability. For $v > 0$, reduction modulo 2π gives the smooth periodic density

$$p_v(\theta) = \sum_{k\in\mathbb{Z}} \frac{e^{-(\theta+2\pi k)^2/(2v)}}{\sqrt{2\pi v}} = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} e^{-k^2 v/2} e^{ik\theta} = \frac{1}{2\pi}\theta_3(\theta/2, e^{-v/2}). \tag{23}$$

Here θ_3 is the Jacobi theta function. Its product formula (DLMF, Eq. 20.5.3) states that, for $q \in (0,1)$,

$$\theta_3(z,q) = \prod_{m=1}^{\infty}(1-q^{2m})(1+2q^{2m-1}\cos(2z)+q^{4m-2}). \tag{24}$$

Every factor depending on $\theta = 2z$ is positive for real θ. Logarithmic differentiation gives

$$\partial_\theta \log p_v(\theta) = -2\sin\theta \sum_{m=1}^{\infty} \frac{q^{2m-1}}{1+2q^{2m-1}\cos\theta + q^{4m-2}}, \quad q = e^{-v/2}. \tag{25}$$

For fixed $q < 1$, the denominator is at least $(1-q^{2m-1})^2$. After finitely many terms this lower bound is uniformly bounded away from zero, and the numerator has a summable geometric tail. The derivative series converges uniformly in θ. The logarithmic derivative in Equation (25) is strictly negative on $(0,\pi)$. p_v is also even.

On every compact interval of positive v, the Fourier series in Equation (23) and its derivatives of first order in v and second order in θ converge uniformly. Termwise differentiation establishes $\partial_v p_v = \frac{1}{2}\partial_\theta^2 p_v$. Consequently, for $0 < a < \pi$,

$$\partial_v F_v(a) = \frac{1}{2}[p_v'(a) - p_v'(-a)] = p_v'(a) < 0, \quad F_v(a) = \Pr(|\Theta_v| \le a). \tag{26}$$

Thus $|\Theta_v|$ increases in first-order stochastic order with v. At $v = 0$, $\Theta_0 = 0$ almost surely, which extends the order to zero. For $v \to \infty$, all nonzero Fourier coefficients vanish. On the compact circle this implies weak convergence to uniform. Since $F_v(a)$ decreases to the uniform limit a/π, it satisfies $F_v(a) \ge a/\pi$.

A.2 THE PUBLIC-MAP IDENTITY

The joint release makes the map observable. We derive Equation (11) by integrating the conditional Gaussian divergence over the map distribution. For fixed neighbors D, D', let R be the common map law and let $p_\Omega(y)$ and $q_\Omega(y)$ be the conditional Gaussian densities of their noisy means. Both joint laws are dominated by $R(d\Omega)\,dy$, with densities p_Ω and q_Ω. Writing $[u]_+ = \max\{u, 0\}$, Tonelli's theorem gives

$$\begin{aligned} D_{e^\varepsilon}(M(D)\|M(D')) &= \int R(d\Omega) \int [p_\Omega(y) - e^\varepsilon q_\Omega(y)]_+ dy \\ &= \int R(d\Omega)\, g_\varepsilon(\|\mu_\Omega(D) - \mu_\Omega(D')\|/\sigma). \end{aligned} \tag{27}$$

This expression includes all measurable events depending jointly on Ω and y. Reversing the ordered datasets leaves the conditional Gaussian expression unchanged.

We also need the monotonicity of the Gaussian profile. Let φ denote the standard normal density. To differentiate Equation (9), set $a = t/2 - \varepsilon/t$ and $b = -t/2 - \varepsilon/t$. Since $e^\varepsilon\varphi(b) = \varphi(a)$, the derivative simplifies to $\varphi(a) > 0$. The profile itself extends continuously to $g_\varepsilon(0) = 0$ for all $\varepsilon \ge 0$.

A.3 From one frequency to the exact profile

We complete Theorem 2 by transferring the single-frequency order to the mean and showing that fixed input pairs approach the resulting bound. Write A_L for one of the iid $1-\cos(\cdot)$ terms averaged in $\bar{A}_{r,L}$ in Theorem 2. For a displacement of length h, rotational invariance of G_j gives $\omega_j^\top z \overset{d}{=} h\sqrt{S_j}B_j$, where the B_j are independent standard normals. For $0 \le h \le H$ and each $s > 0$, Lemma 1 gives

$$1-\cos(h\sqrt{s}B) \le_{\mathrm{st}} 1-\cos(H\sqrt{s}B). \tag{28}$$

Integrating the CDF inequality over ν preserves the order. For each coordinate, apply the two inverse distribution functions to the same uniform variable. Use independent uniforms across coordinates. The resulting coordinatewise inequalities also order their sums. Applying the increasing bounded function $a \mapsto g_\varepsilon(\sqrt{2a}/(n\sigma))$ bounds the divergence of every realizable neighboring pair by $H_{r,L}(\varepsilon, n\sigma)$.

If $0 < L < \infty$, the definition of diameter provides pairs (x_t, x_t') with $h_t = \|x_t - x_t'\| \to L$. Using the same (S_j, B_j) for every h_t, their summands converge almost surely to $1-\cos(L\sqrt{S_j}B_j)$ because S_j is finite almost surely. Bounded convergence then gives convergence of the divergence. If $L = \infty$, choose $h_t \to \infty$. For any nonzero integer k, the kth wrapped-angle Fourier coefficient is

$$\mathbb{E}_S e^{-k^2 S h_t^2/2} \longrightarrow 0. \tag{29}$$

This follows from $S > 0$ almost surely and bounded convergence, without any moment assumption on S. Weak convergence to uniform and finite-product convergence give the limiting expectation in Theorem 2. Appending the same $n-1$ arbitrary records turns every such input pair into a valid neighboring pair. Fixed neighboring pairs approach the claimed supremum. An atom at $S = 0$ would leave positive mass at angle zero, which explains the positivity assumption in Equation (1). For $L = \infty$, the one-coordinate contribution has density

$$f_A(a) = \frac{1}{\pi\sqrt{a(2-a)}}, \quad 0 < a < 2. \tag{30}$$

A.4 The noise scale and its limiting value

We first show that the profile crosses each target exactly once, then use concentration of the squared distance to identify the limiting scale. For $\varepsilon \ge 0$ and $0 < \delta < 1$, let $\tau_{\varepsilon,\delta}$ solve $g_\varepsilon(\tau) = \delta$, so $\sigma_G = 2/(n\tau_{\varepsilon,\delta})$. For $L > 0$ and finite r, $0 < \bar{A}_{r,L} < 2$ almost surely. The claim holds both for finite L and for uniform limiting angles. Thus $H_{r,L}(\varepsilon, n\sigma)$ is continuous and strictly decreasing on $\sigma > 0$ by bounded convergence and the strict monotonicity of g_ε. Its limits at zero and infinity are 1 and 0. This proves existence and uniqueness of the positive root. At $\sigma = \sigma_G$, each profile argument is strictly less than $2/(n\sigma_G) = \tau_{\varepsilon,\delta}$ almost surely. The profile is strictly below δ and $\sigma_{r,L} < \sigma_G$.

Since $\mathbb{E}A_L = 1 - \kappa_L > 0$, the law of large numbers gives $\bar{A}_{r,L} \to 1-\kappa_L$ almost surely. For each fixed normalized scale $s > 0$,

$$H_{r,L}(\varepsilon, s) \longrightarrow g_\varepsilon(\sqrt{2(1-\kappa_L)}/s). \tag{31}$$

The limiting function crosses δ uniquely at $s_* = \sqrt{2(1-\kappa_L)}/\tau_{\varepsilon,\delta}$. Choose any $0 < s_- < s_* < s_+$. Pointwise convergence makes the finite-r profile exceed δ at s_- and fall below δ at s_+ for all sufficiently large r. Monotonicity places the root $n\sigma_{r,L}$ between these two scales. Letting them approach s_* proves Equation (15). For a singleton input domain, $L = 0$ and the unlabeled mean is determined by the public map and the sole admissible input, so no noise is needed.

A.5 Protected labels

The classification release must protect changes in the label as well as the input. We separate replacements that preserve the label from those that change it. Suppose each record contains an input and a protected label c in a public set of size $C \ge 2$. Let e_c be the cth standard basis vector and form the mean of $e_c \otimes \phi_\Omega(x)$, normalized by the full sample size n. Release this $2rC$-dimensional mean with independent isotropic Gaussian noise of coordinate variance σ^2, together with the map. Its exact profile is

$$\max\{H_{r,L}(\varepsilon, n\sigma),\ g_\varepsilon(\sqrt{2}/(n\sigma))\}. \tag{32}$$

Same-label replacements recover the unlabeled profile. Cross-label replacements occupy orthogonal blocks, so their distance is exactly $\sqrt{2}/n$ regardless of the inputs or map. Taking the maximum proves Equation (32).

The minimum noise scale is $\max\{\sigma_{r,L}, \sqrt{2}/(n\tau_{\varepsilon,\delta})\}$. Its variance lies in $[\sigma_G^2/2, \sigma_G^2)$ and converges to $\sigma_G^2/2$ as $r \to \infty$. Protected labels retain a cross-label noise requirement even when the public input domain is small.

B A SHARP FLOOR ACROSS FREQUENCY ENSEMBLES

We prove Theorem 4 by averaging over input displacements to obtain a lower bound, then showing that iid frequencies attain it.

B.1 MECHANISM AND THE JOINT PROFILE

We use the release and Gaussian profile in Equations (5) and (9), allowing any joint frequency law $\mathcal{E}$ independent of the records. The noise scale is fixed independently of the map. Write $\delta_{\mathcal{E}}(\varepsilon;\sigma)$ for the supremum of the joint output privacy divergence over neighboring datasets on $\mathbb{R}^p$. Put $s = n\sigma$ and

$$F_{\varepsilon,s}(\theta) = g_\varepsilon\left(\frac{\sqrt{(2/r)\sum_{j=1}^r(1-\cos\theta_j)}}{s}\right), \quad \theta \in \mathbb{T}^r = (\mathbb{R}/2\pi\mathbb{Z})^r. \tag{33}$$

Denote the uniform probability measure (normalized Haar measure) on this torus by m_r and define

$$H_{r,\infty}(\varepsilon, s) = \int_{\mathbb{T}^r} F_{\varepsilon,s}(\theta)\, dm_r(\theta). \tag{34}$$

This is the iid uniform-angle profile from Theorem 2.

Equations (10) and (11) hold for any joint frequency law, so

$$\delta_{\mathcal{E}}(\varepsilon;\sigma) = \sup_{z\in\mathbb{R}^p} \Psi_{\mathcal{E}}(z), \quad \Psi_{\mathcal{E}}(z) = \mathbb{E}_\Omega F_{\varepsilon,n\sigma}((\omega_j^\top z)_{j=1}^r). \tag{35}$$

Every displacement $z \in \mathbb{R}^p$ is realized by a replacement pair, which gives the supremum above.

B.2 THE LOWER BOUND UNDER INTEGER NONRESONANCE

The required geometric condition is

$$\Pr_{\mathcal{E}}\left(\sum_{j=1}^r a_j\omega_j = 0\right) = 0 \quad \text{for every } a \in \mathbb{Z}^r \setminus \{0\}. \tag{36}$$

Because $\mathbb{Z}^r$ is countable, this is equivalent to almost surely having no nonzero integer relation among the rows. It does not require real linear independence: it can hold when $r > p$.

Fixed-map sensitivity maximizes over displacements separately for each map (Chatalic et al., 2021). Here we bound Equation (35), which averages over maps before taking this supremum.

Lemma 6 (Displacement averaging)**.** *Assume* (36). *Let $V \sim N(0, I_p)$ independently of Ω. Then the unconditional law of $(t\omega_j^\top V)_{j=1}^r$ modulo 2π converges weakly to m_r as $t \to \infty$.*

Proof. For $a \in \mathbb{Z}^r$, the corresponding Fourier coefficient equals

$$\mathbb{E}_{\Omega,V} \exp\left(it\Big\langle \sum_j a_j\omega_j, V\Big\rangle\right) = \mathbb{E}_\Omega \exp\left(-\frac{t^2}{2}\Big\|\sum_j a_j\omega_j\Big\|^2\right) \tag{37}$$

$$\longrightarrow \Pr\left(\sum_j a_j\omega_j = 0\right).$$

The conditional expectation is the Gaussian characteristic function, and the limit follows from dominated convergence. Under (36), the limits are one at $a = 0$ and zero otherwise, exactly the Fourier coefficients of m_r. Hence integrals of every trigonometric polynomial converge. Trigonometric polynomials are uniformly dense in continuous functions on the compact torus, and all measures have total mass one, so integrals of every continuous function converge as well. □

Theorem 7 (Universal floor for nonresonant ensembles)**.** *For every $\mathcal{E}$ satisfying* (36)*, every $\varepsilon \geq 0$, and every $\sigma > 0$,*

$$\delta_{\mathcal{E}}(\varepsilon; \sigma) \geq H_{r,\infty}(\varepsilon, n\sigma). \tag{38}$$

No moment, density, isotropy, or marginal-spectrum assumption on $\mathcal{E}$ is needed for this inequality.

Proof. For every $t > 0$, Equation (35) implies

$$\delta_{\mathcal{E}}(\varepsilon; \sigma) \geq \mathbb{E}_V \Psi_{\mathcal{E}}(tV) = \mathbb{E}_{V,\Omega} F_{\varepsilon,n\sigma}((t\omega_j^\top V)_j). \tag{39}$$

The auxiliary variable V averages divergences for input pairs chosen independently of the map. Their supremum is at least this average for every t. By Lemma 6 and continuity of $F_{\varepsilon,n\sigma}$, the average tends to $H_{r,\infty}(\varepsilon, n\sigma)$. This gives the claimed lower bound. □

The argument uses convergence of Fourier coefficients on the torus. Related dilation and equidistribution results are studied by Kra et al. (2018).

B.3 Attainability in Theorem 4

Let Λ_ν be the frequency marginal in Equation (1). Conditional on the positive scales, every nonzero linear combination of iid frequencies is a nondegenerate Gaussian vector, so the iid design is nonresonant. Theorem 2 gives, on $\mathbb{R}^p$,

$$\delta_{\Lambda_\nu^{\otimes r}}(\varepsilon; \sigma) = H_{r,\infty}(\varepsilon, n\sigma). \tag{40}$$

Iid frequencies attain the lower bound of Theorem 7. Their profile decreases continuously and strictly from one to zero as σ grows. The lower bound excludes every noise scale below the iid root for the target δ. Since iid attains that root, it also minimizes the required noise within the comparison class, as claimed in Theorem 4.

B.4 Frequency ensembles covered by the bound

Proposition 8 (Gaussian-marginal block ORF and positive scale mixtures)**.** *Partition the r rows into blocks of at most p rows. In each block, take the corresponding rows q_j of a Haar-distributed orthogonal matrix in $O(p)$. Let $R_j \sim \chi_p$ be independent across rows and independent of all orthogonal matrices and scales. Let $S_j \sim \nu$ be positive finite scales, independent of the matrices and radii. Dependence among the S_j is allowed. Set*

$$\omega_j = \sqrt{S_j} R_j q_j. \tag{41}$$

Then every row has marginal Λ_ν and the ensemble satisfies (36)*, for arbitrary finite r and $p \geq 1$.*

Proof. A Haar row is uniform on the unit sphere. Its product with an independent χ_p radius has law $N(0, I_p)$, including $p = 1$ where the direction is a uniform sign. Multiplication by $\sqrt{S_j}$ gives the required marginal.

Fix $a \in \mathbb{Z}^r \setminus \{0\}$ and choose j with $a_j \neq 0$. Condition on all directions, all scales, and every radius except R_j. The relation $\sum_i a_i \omega_i = 0$ has the form $a_j \sqrt{S_j} q_j R_j + b = 0$. Its coefficient is a nonzero vector, so at most one real value of R_j can solve it. The independent radius has an atomless distribution, so the conditional probability is zero. Integrating and then using countability proves (36). □

For constant $S_j = \ell^{-2}$ this is the Gaussian-marginal ORF construction of Yu et al. (2016). Independent blocks cover $r > p$. The random radii preserve the exact Gaussian spectrum. Fixed-radius Haar features and SORF generally have different spectra and therefore fall outside this fixed-marginal comparison.

B.5 WHY THE ASSUMPTIONS MATTER

B.5.1 ROLE OF NONRESONANCE

Take $p = 1, r = 2$ and $\omega_1 = \omega_2 = G \sim N(0, 1)$. Both marginals are Gaussian, but $(1, -1)$ is an integer relation. After an orthogonal change of output coordinates, the informative part is exactly the $r = 1$ release. The other coordinates are independent noise. Hence its exact profile is $H_{1,\infty}$, at the same $s = n\sigma$. For $\varepsilon = 0$, let $f_s(a) = g_0(\sqrt{2a}/s)$ on $[0, 2]$. It is strictly concave: the square root is strictly concave and $g_0(t) = 2\Phi(t/2) - 1$ is increasing and concave on $[0, \infty)$. If A_1, A_2 are independent copies of $1 - \cos U$ for $U \sim \mathrm{Unif}[-\pi, \pi]$, strict Jensen's inequality gives

$$H_{2,\infty}(0, s) = \mathbb{E} f_s((A_1 + A_2)/2) > \mathbb{E} f_s(A_1) = H_{1,\infty}(0, s). \tag{42}$$

This design has the same marginals as iid sampling but a smaller privacy profile, showing why nonresonance is needed. Continuity in ε preserves the strict inequality for sufficiently small positive ε at any fixed $s > 0$.

B.5.2 ROLE OF THE DOMAIN AND NOISE MODEL

The input domain in Theorem 7 is all of $\mathbb{R}^p$. An arbitrary unbounded subset is insufficient for the general theorem: for deterministic $\omega_1 = 1$ and domain $2\pi\mathbb{Z}$, the feature is constant, although (36) holds. Bounded domains likewise require separate analysis. Theorem 4 compares data-independent maps with equal feature weights and fixed isotropic Gaussian noise. Map-dependent scales, anisotropic covariance, or a different statistic require a new comparison class.

C EXACT CALIBRATION FOR BLOCK-ORTHOGONAL FEATURES

Theorem 5 calibrates ORF by maximizing a radial privacy profile. We derive this profile for the scale-mixture construction in Proposition 8, with independent scales $S_{bj} \sim \nu$, and establish its noise bounds. The block sizes are public and sum to r. Independent blocks allow $r > p$, and the last block may contain fewer than p rows. Let V_b be independent uniform vectors on the unit sphere in $\mathbb{R}^p$, with coordinates V_{bj}. The radii $R_{bj} \sim \chi_p$ and scales $S_{bj} \sim \nu$ are mutually independent and independent of these vectors. The scalar law used to evaluate the distance in Theorem 5 is

$$\bar{A}_{\mathrm{O}}(\rho) := \frac{1}{r} \sum_b \sum_{j=1}^{m_b} [1 - \cos(\rho \sqrt{S_{bj}} R_{bj} V_{bj})], \quad d_\Omega(\rho e_1)^2 \overset{d}{=} \frac{2\bar{A}_{\mathrm{O}}(\rho)}{n^2}. \tag{43}$$

Write $s = n\sigma$ and $J(\rho; s) = \mathbb{E} g_\varepsilon(\sqrt{2\bar{A}_{\mathrm{O}}(\rho)}/s)$, with ε fixed throughout each argument.

C.1 THE RADIAL PROFILE AND ITS MINIMUM NOISE

Let Q_b denote the Haar-distributed orthogonal matrix for block b. For a nonzero input difference z, Haar invariance gives $Q_b z/\|z\| \sim \mathrm{Unif}(S^{p-1})$. Consequently, the projections in block b have the joint law

$$(\omega_{bj}^\top z)_{j \le m_b} \overset{d}{=} \|z\| (\sqrt{S_{bj}} R_{bj} V_{bj})_{j \le m_b}. \tag{44}$$

The paired feature distance identity and public-map divergence identity do not require independent frequencies. Applying them to Equation (44) gives $J(\|z\|; s)$ for each neighboring pair. Every pair of admissible inputs can be completed to neighboring datasets with the same remaining records, so the profile is the supremum over realizable distances. In a convex domain, all distances smaller than L occur along line segments. Bounded convergence makes J continuous in distance, giving Equation (18), including domains whose diameter is approached but not attained. For a nonconvex domain, the supremum is instead over its realizable distance set. Using a containing public box gives a valid calibration for that box.

With independent blocks of size one, the frequencies are iid with spectrum Λ_ν, so the joint release is the mechanism of Theorem 2 for the same mixing distribution. Its monotone radial profile reduces the distance supremum to the diameter, or to the uniform-angle limit when $L = \infty$.

To handle unbounded distances, condition on all scales and sphere vectors. The independent radii then make the projected angles conditionally independent. If R has a density f on $\mathbb{R}$ with total variation $V_f < \infty$, wrapping aR onto the circle for $a \neq 0$ gives

$$\mathrm{TV}(\mathcal{L}(aR \bmod 2\pi), \mathrm{Unif}[-\pi, \pi]) \leq \min\{1, \pi V_f/|a|\}. \tag{45}$$

Indeed, the wrapped density is $|a|^{-1} \sum_k f((\theta + 2\pi k)/|a|)$ up to reflection. The rectangle-rule error for a density of bounded variation on a grid of spacing h is at most hV_f. Taking $h = 2\pi/|a|$ bounds the pointwise density error by $V_f/|a|$, which gives Equation (45) after integration. The χ_p density, extended by zero to the negative half-line, has finite variation. Each sphere coordinate is nonzero almost surely, including the two-point sphere when $p = 1$. Since $S_{bj} > 0$ almost surely, $|\rho\sqrt{S_{bj}}V_{bj}| \to \infty$ under this conditioning. The product-measure bound for total variation and bounded convergence over the scales and sphere vectors show that all r wrapped angles converge jointly to independent uniform angles as $\rho \to \infty$. No moment assumption on ν is needed. Thus

$$\lim_{\rho \to \infty} J(\rho; s) = H_{r,\infty}(\varepsilon, s). \tag{46}$$

We next show that maximizing the radial profile gives a continuous, strictly decreasing function of the noise scale. Extend J to $\rho = \infty$ using the preceding limit. The same total-variation bound controls the tail uniformly in s because the privacy integrand lies in $[0, 1]$. Also, $g'_\varepsilon(t) = \varphi(t/2 - \varepsilon/t) \leq 1/\sqrt{2\pi}$ and the normalized feature distance is at most two, so

$$|J(\rho; s) - J(\rho; s')| \leq \frac{2}{\sqrt{2\pi}} |s^{-1} - (s')^{-1}|. \tag{47}$$

Here φ is the standard normal density. The distance interval is compact after adjoining infinity, and J is jointly continuous for positive s. Its maximum is attained. At each positive distance, the feature distance is positive almost surely, so J strictly decreases with s. Evaluating at a maximizing distance for the larger scale shows that the maximum also strictly decreases. It tends to one as $s \downarrow 0$, by considering one fixed positive distance, and tends to zero as $s \to \infty$, since it is bounded by $g_\varepsilon(2/s)$. This proves existence and uniqueness of the root.

At every finite distance and at the uniform-angle limit, the normalized feature distance is strictly smaller than two almost surely. The maximum over the compactified interval consequently satisfies

$$\sup_\rho J(\rho; s) < g_\varepsilon(2/s). \tag{48}$$

At the Uniform AG scale, the right side equals δ, so the unique ORF root is strictly smaller. On the full space, the supremum is at least its limit in Equation (46), proving the comparison with the iid root.

For class block releases, the unit-norm argument in Appendix A.5 applies unchanged, giving Equation (22) with $\sigma_E = \sigma_{\mathrm{O}}$.

C.2 Why the largest distance need not be worst

We construct a two-frequency example whose profile exceeds its infinite-distance limit at a finite distance. This shows why ORF calibration retains the distance maximization. Consider the RBF specialization $S_1 = S_2 = 1$ with one full ORF block and $p = r = 2$. With $V = (\cos\theta, \sin\theta)$ uniform on the circle and independent $R_1, R_2 \sim \chi_2$, put $Y = (R_1 V_1, R_2 V_2)$. Integrating the two conditional Rayleigh densities over θ gives

$$f(y_1, y_2) = \frac{|y_1| + |y_2|}{2\sqrt{2\pi}} \exp\left[-\frac{(|y_1| + |y_2|)^2}{2}\right]. \tag{49}$$

For example, in the positive quadrant the integral is

$$\frac{y_1 y_2}{2\pi} \int_0^{\pi/2} \frac{\exp[-y_1^2/(2\cos^2\theta) - y_2^2/(2\sin^2\theta)]}{\cos^2\theta \sin^2\theta}\, d\theta. \tag{50}$$

Substituting $v = \tan\theta$ and using $\int_0^\infty e^{-av^2 - b/v^2} dv = \sqrt{\pi} e^{-2\sqrt{ab}}/(2\sqrt{a})$ yields Equation (49).

Let h_ρ be the continuous density of ρY wrapped onto $[-\pi, \pi]^2$. Counting the $4m$ lattice points with $|k_1| + |k_2| = m$ gives

$$h_\rho(0) = \frac{4\pi}{\sqrt{2\pi}\rho^3} \sum_{m=1}^{\infty} m^2 e^{-2\pi^2 m^2/\rho^2}. \tag{51}$$

In particular, $h_1(0) < 1/(4\pi^2)$, the density of the uniform torus. For $q = e^{-2\pi^2}$, an upper bound on the sum is $q(1+q)/(1-q)^3$, which verifies this strict inequality. For small noise, the divergence deficit is determined by the wrapped density near zero. The normalized chord is $d(\theta) = \sqrt{2\sin^2(\theta_1/2) + 2\sin^2(\theta_2/2)}$. It satisfies $d(\theta) = \|\theta\|/\sqrt{2} + o(\|\theta\|)$ near zero and $d(\theta) \geq \sqrt{2}\|\theta\|/\pi$ on the square. For any continuous bounded torus density h, the change of variables $\theta = s\sqrt{2}u$ and dominated convergence yield

$$\lim_{s \downarrow 0} \frac{1}{s^2} \int [1 - g_\varepsilon(d(\theta)/s)] h(\theta)\, d\theta = 2h(0) \int_{\mathbb{R}^2} [1 - g_\varepsilon(\|u\|)]\, du. \tag{52}$$

The last integral is positive and finite because the divergence deficit has a Gaussian tail. The rescaled integrand is dominated by $\|h\|_\infty[1 - g_\varepsilon(2\|u\|/\pi)]$. Comparing h_1 with the uniform density proves $J(1; s) > H_{2,\infty}(\varepsilon, s)$ for sufficiently small s at every fixed $\varepsilon \geq 0$. Equation (46) then implies a decrease at some larger distance. It also gives privacy targets for which the ORF full-space root strictly exceeds the iid root.

D Evaluating the calibration

Theorem 2 reduces calibration to an expectation over a scalar distance law and a monotone equation for the noise scale. We divide the range $[0, 2]$ of $1 - \cos(\cdot)$ into 4096 bins and convolve the bin probabilities to approximate the distribution of the average over r frequencies. Using the upper bin endpoints gives a conservative discretization of the distance. We then evaluate the Gaussian profile and solve Equation (14) numerically. For labeled means, we also impose Equation (22).

These calculations use the declared kernel, public domain, feature count, and privacy target. The distance distribution can be reused across privacy targets, and the normalized scale $n\sigma$ is divided by the sample size for each release. At $L = \infty$, the same law applies to every positive Gaussian scale mixture. Once calibrated, Exact and Uniform AG have identical online costs: computing the mean takes $O(npr)$ work and $O(pr + r)$ working memory, and evaluating a query takes $O(pr)$ work.

D.1 An upper bound for ORF calibration

For the RBF experiments, we bound both the distance search and the integration error in Equation (18). Write $t = \rho/\ell$, $T = L/\ell$, and $s = n\sigma$. With the radii and sphere vectors from Theorem 5, define

$$F(t) = \frac{2}{r} \sum_{b,j} [1 - \cos(tR_{bj}V_{bj})], \qquad B_\Omega^2 = \frac{1}{r} \sum_{b,j} R_{bj}^2 V_{bj}^2. \tag{53}$$

The profile is $\sup_{0 \leq t \leq T} \mathbb{E} g_\varepsilon(\sqrt{F(t)}/s)$. All integration draws below are independent draws of the whole collection of blocks. Coordinates within a block retain their dependence.

Covering the distance interval. Partition $[0, T]$ at $0 = t_0 < \cdots < t_J = T$, with $h_i = t_{i+1} - t_i$. Since $|F''(t)| \leq 2B_\Omega^2$, linear interpolation gives, throughout the ith interval,

$$\sqrt{F(t)} \leq W_i := \min\left\{2, \sqrt{\max\{F(t_i), F(t_{i+1})\} + \tfrac{1}{4}h_i^2 B_\Omega^2}\right\}. \tag{54}$$

This bound covers every distance between grid points without assuming a monotone or unimodal profile.

Using the ORF distribution. The bound $W_i \le 2$ can make integration error bounds unnecessarily large. Let B be the number of blocks, fix $0 < \beta < 1$, and put $u = \log(3/\beta)$. Define

$$\begin{aligned} a(t) &= \frac{2}{r}\sum_b m_b\left[1-\cos\Big(\min\{t\sqrt{p/m_b},\pi\}\Big)\right], \\ q(t) &= a(t) + \frac{2t\sqrt{2Bu}}{r}, \qquad v = \frac{pB + 2\sqrt{pBu} + 2u}{r}, \\ c_i &= \min\left\{2, \sqrt{\max\{q(t_i), q(t_{i+1})\} + \tfrac{1}{4}h_i^2 v}\right\}. \end{aligned} \tag{55}$$

Then $\Pr\{W_i > c_i\} \le \beta$. Clipping W_i at c_i gives the following bound on the expected divergence:

$$\mathbb{E}g_\varepsilon(W_i/s) \le \mathbb{E}g_\varepsilon(\min\{W_i, c_i\}/s) + \beta g_\varepsilon(2/s). \tag{56}$$

To prove the tail bound, condition on the sphere vectors. The function $h(x) = 1-\cos(\min\{\sqrt{x},\pi\})$ is increasing and concave on $[0,\infty)$. Using $\mathbb{E}R_{bj}^2 = p$ and $\sum_j V_{bj}^2 \le 1$, Jensen's inequality gives $\mathbb{E}[F(t) \mid (V_b)_b] \le a(t)$. Represent each radius as $R_{bj} = \|G_{bj}\|$ for independent standard Gaussian vectors. As a function of these vectors, $F(t)$ has squared gradient norm at most $4t^2B/r^2$. Gaussian concentration yields $\Pr\{F(t) > q(t) \mid (V_b)_b\} \le e^{-u}$. Also, B_Ω^2 is a weighted sum of independent squared Gaussians, with weight sum at most pB/r, squared-weight sum at most pB/r^2, and maximum weight at most $1/r$. The Gaussian quadratic-form tail bound gives $\Pr\{B_\Omega^2 > v \mid (V_b)_b\} \le e^{-u}$. This bound follows from $\log \mathbb{E}e^{\theta\sum_k w_k(G_k^2-1)} \le \theta^2\sum_k w_k^2/(1-2\theta\max_k w_k)$ and Chernoff's inequality. A union bound over the two endpoints and B_Ω^2, followed by averaging over the sphere vectors, proves the claim.

A sufficient noise scale. For $N \ge 2$ independent integration draws, let $\widehat{\mu}_i(s)$ and $\widehat{v}_i(s)$ be the sample mean and unbiased sample variance of $g_\varepsilon(\min\{W_i, c_i\}/s)$. Fix a finite set $\mathcal{S}$ of K candidate scales before drawing these samples. For $0 < \eta < \delta$, set $\lambda_0 = \log(2JK/\eta)$ and

$$\begin{aligned} U(s) = \min\Bigg\{ g_\varepsilon(2/s),\ \max_i \Bigg[\widehat{\mu}_i(s) + \sqrt{\frac{2\widehat{v}_i(s)\lambda_0}{N}} \\ + \frac{7g_\varepsilon(c_i/s)\lambda_0}{3(N-1)} + \beta g_\varepsilon(2/s) \Bigg] \Bigg\}. \end{aligned} \tag{57}$$

The empirical Bernstein inequality (Maurer & Pontil, 2009), applied to all JK pairs, implies that with probability at least $1-\eta$, $\delta_{\mathrm{O}}(\varepsilon; s/n) \le U(s)$ simultaneously for every $s \in \mathcal{S}$. The same integration draws can be shared across intervals and candidate scales. For a simultaneous bound over C calibration settings, replace λ_0 by $\log(2JKC/\eta)$.

Select a candidate satisfying

$$U(s) \le \delta - \eta, \qquad \sigma = s/n, \tag{58}$$

with Uniform AG calibrated at $(\varepsilon, \delta-\eta)$ as a fallback. The integration randomness is sampled independently of the protected records and the released map. Including it in the mechanism gives (ε,δ)-DP: the divergence is at most $\delta-\eta$ on the simultaneous good event and at most one otherwise. This is a guarantee for the randomized calibration procedure, not a deterministic guarantee conditional on every realized integration sample. For labeled means, take the maximum of σ and the cross-label scale in Equation (22).

The reference protocol uses $J = 256$, $N = 2^{18}$, $\beta = 10^{-10}$, and $\eta = 10^{-8}$ at $\delta = 10^{-6}$. The 16,385 candidate scales are equally spaced between one quarter of the normalized fallback scale and the full fallback scale. Upward rounding of W_i to 8192 distance bins preserves the bound when applied before clipping. The sample-variance bound need not be monotone in s, so each selected candidate must satisfy Equation (58) directly. These bounds control integration and distance discretization under exact arithmetic. They do not bound floating-point roundoff.

E KERNEL QUERY ERROR

A kernel query projects the released vector onto $\phi_\Omega(x_\star)$. Its noise error differs from the error of the entire released mean. For any fixed map, the released mean has squared vector error $\mathbb{E}_Z\|\widetilde{\mu}_D -$

Table 1: Dataset dimensions and sample sizes. Here p is the encoded input dimension, C the class count, n the protected sample size, and r the number of sine–cosine frequency pairs.

Dataset	p	C	n	Evaluation	r
Adult	119	2	5000	16281	238
Covertype	55	7	5000	116203	110
USPS	256	10	5000	1860	512
ISOLET	617	26	5000	1559	1234

$\mu_\Omega(D)\|^2 = 2r\sigma^2$. For any fixed query vector w, the Gaussian contribution to $w^\top \widetilde{\mu}_D$ has variance $\sigma^2\|w\|^2$ conditional on the map.

E.1 Proof of the query error decomposition

For a fixed query $x_\star$, define the contribution of one frequency by

$$\xi_\omega(x_\star) = \frac{1}{n}\sum_{i=1}^{n} \cos(\omega^\top(x_i - x_\star)). \tag{59}$$

The released answer then expands as

$$\widehat{f}_D(x_\star) = \frac{1}{r}\sum_{j=1}^{r} \xi_{\omega_j}(x_\star) + \sigma\phi_\Omega(x_\star)^\top Z. \tag{60}$$

The first term is unbiased for $f_D(x_\star)$ whenever the frequency marginals match the target kernel. Its variance is

$$A = \frac{1}{r^2}\sum_{j,k=1}^{r} \mathrm{Cov}(\xi_{\omega_j}(x_\star), \xi_{\omega_k}(x_\star)). \tag{61}$$

For independent frequencies, the off-diagonal covariances vanish and $A = \mathrm{Var}(\xi_\omega(x_\star))/r$. For ORF, the same identity retains the within-block covariances. Conditional on the map, the second term has mean zero and variance $\sigma^2\|\phi_\Omega(x_\star)\|^2 = \sigma^2$. The cross term vanishes by conditioning on Ω. This proves Equation (19). Averaging over a fixed list of queries preserves the identity without assuming that their errors are independent.

Subtracting the Exact and Uniform AG mean squared errors cancels their common approximation term. Dividing by the Uniform AG error gives Equation (20).

F Experimental protocol

The experiments compare noise calibrations while holding the random-feature design and downstream procedure fixed. Within each design and trial, Uniform AG and Exact share the protected records, kernel, map, and random draws used for reconstruction initialization. Downstream releases also share the standard Gaussian draw.

F.1 Datasets and public input domains

Adult is binary, while Covertype, USPS, and ISOLET have 7, 10, and 26 classes, respectively. Adult, Covertype, and ISOLET are UCI datasets (Becker & Kohavi, 1996; Blackard, 1998; Cole & Fanty, 1991), and USPS contains handwritten digits. Adult and ISOLET retain their official train/test splits. Covertype and USPS use fixed, disjoint training and evaluation pools. A fixed random permutation of each training pool supplies 5000 protected records, selected independently of labels and performance. Reconstruction uses only the resulting private release, with no remaining training records supplied as auxiliary inputs.

Adult categorical fields are one-hot encoded from the published schema, including an unknown category. Its six numeric fields, in their published order, are divided by $(100, 10^6, 16, 10^5, 10^4, 100)$ and clipped to $[0, 1]$. Missing-value indicators are included.

Covertype's ten numeric scales are $(4000, 360, 90, 1000, 1000, 10000, 255, 255, 255, 10000)$. Aspect is represented by sine and cosine, and vertical hydrology distance lies in $[-1, 1]$. The remaining scaled fields and 44 binary indicators lie in $[0, 1]$.

USPS pixels and ISOLET coordinates are clipped to $[-1, 1]$. The containing boxes cover every admissible encoded input. Their diagonal lengths give $L = \sqrt{119}$ for Adult, $L = 8$ for Covertype, $L = 32$ for USPS, and $L = 2\sqrt{617}$ for ISOLET.

The reference setting uses $\gamma = 1/p$, equivalently $\ell = \sqrt{p/2}$, and $r = 2p$ frequency pairs. The released unlabeled vector has $4p$ coordinates and ORF uses two complete blocks. The separate feature count sweep varies this common ratio. ORF uses Haar-orthogonal directions and independent χ_p radii, preserving the Gaussian marginal spectrum. The iid and ORF mechanisms are calibrated separately for the same RBF kernel. Appendix J considers a privately estimated bandwidth.

F.2 Kernel conventions

The kernels in this study are scale mixtures of Gaussian kernels: averaging $\exp(-S\|x-y\|^2/2)$ over a positive scale S gives the target kernel. This kernel mixture corresponds to sampling a Gaussian scale-mixture frequency $\omega = \sqrt{S}G$. With $t = \|x - y\|/\ell$, Table 2 gives the kernel formulas and mixing laws (Rasmussen & Williams, 2005; Tronarp et al., 2018). Each frequency uses one scalar scale shared by all its coordinates, independently across frequencies. The Gamma distribution is parameterized by shape and rate.

Table 2: **Kernel families and their spectral scale laws.** All kernels satisfy $k(x, x) = 1$. The rational-quadratic experiments use $a = 1$. The Laplacian kernel uses Euclidean distance.

Kernel	$k(x, y)$	Scale S
RBF	$e^{-t^2/2}$	ℓ^{-2}
RQ	$(1 + t^2/(2a))^{-a}$	$U/\ell^2, U \sim \mathrm{Gamma}(a, a)$
Matérn $3/2$	$(1 + \sqrt{3}t)e^{-\sqrt{3}t}$	$3/(2\ell^2 U), U \sim \mathrm{Gamma}(3/2, 1)$
Matérn $5/2$	$(1 + \sqrt{5}t + 5t^2/3)e^{-\sqrt{5}t}$	$5/(2\ell^2 U), U \sim \mathrm{Gamma}(5/2, 1)$
Laplacian	e^{-t}	$1/(2\ell^2 U), U \sim \mathrm{Gamma}(1/2, 1)$

The scales are positive and finite almost surely. The privacy analysis does not require a finite mean for S.

F.3 Paired trials and uncertainty

We fix $\delta = 10^{-6}$ and vary $\varepsilon \in \{0.1, 0.2, 0.5, 1, 2, 5\}$. The query study has 100 independent map trials per condition. Its feature sweep uses $r = \lceil ap \rceil$ for $a \in \{0.5, 1, 2, 4, 8\}$ at $n = 5000$ and all six privacy budgets. The sample size sweep uses the nested prefixes $n \in \{500, 2000, 5000\}$ at $r = 2p$. Smaller feature maps are prefixes of the larger sampled map, renormalized by their own feature count. Every comparison at a given n, r uses the same map.

The target $f_D(x_\star) = n^{-1} \sum_i k(x_i, x_\star)$ uses the original kernel without a density-normalization factor. Query MSE averages over all evaluation inputs in Table 1 and then over maps. We integrate out Gaussian noise by adding σ^2 to each map's approximation MSE. The Exact–Uniform absolute difference is consequently the difference of their noise variances, with no additional sampling uncertainty from Gaussian draws.

Reconstruction and classification use 30 complete trials with independently sampled maps, noise, and decoder initialization. The same reconstructed dataset supplies all nine SVM regularizations. Additional reconstruction comparisons use the five feature ratios at $\varepsilon \in \{0.1, 1, 5\}$. The main $r = 2p$ comparison and DP-scale use all six budgets. Each application keeps $n = m = 5000$.

Means and paired accuracy differences have pointwise 95% Student-t intervals over trials. Relative error reductions use 5000 paired bootstrap resamples, recomputing the ratio of mean errors in each resample. These intervals quantify variation across mechanism draws, conditional on the fixed training and evaluation records.

Table 3: Absolute iid query MSE (units of 10^{-5}) at $\varepsilon = 1$. Different kernels define different query targets.

Dataset	Kernel	Uniform AG	Exact	Gain (%)
Adult	RBF	1.59 ± 0.08	1.40 ± 0.08	12.2
	RQ	4.40 ± 0.72	4.19 ± 0.72	4.8
	Matérn $3/2$	18.22 ± 3.53	18.03 ± 3.53	1.0
	Matérn $5/2$	8.80 ± 1.66	8.62 ± 1.66	2.1
	Laplacian	66.22 ± 16.43	66.05 ± 16.43	0.3
Covertype	RBF	4.43 ± 0.47	4.25 ± 0.47	4.2
	RQ	14.09 ± 2.45	13.89 ± 2.45	1.5
	Matérn $3/2$	38.16 ± 6.46	37.99 ± 6.46	0.5
	Matérn $5/2$	18.32 ± 3.00	18.14 ± 3.00	1.0
	Laplacian	122.90 ± 26.14	122.73 ± 26.14	0.1
USPS	RBF	11.30 ± 0.86	11.15 ± 0.86	1.3
	RQ	16.70 ± 2.82	16.53 ± 2.82	1.0
	Matérn $3/2$	20.81 ± 3.62	20.66 ± 3.62	0.7
	Matérn $5/2$	17.31 ± 2.55	17.16 ± 2.55	0.9
	Laplacian	29.53 ± 5.80	29.38 ± 5.80	0.5
ISOLET	RBF	3.24 ± 0.20	3.09 ± 0.20	4.5
	RQ	6.62 ± 1.11	6.45 ± 1.11	2.6
	Matérn $3/2$	7.18 ± 1.27	7.04 ± 1.27	2.1
	Matérn $5/2$	6.91 ± 1.12	6.76 ± 1.12	2.1
	Laplacian	14.24 ± 3.21	14.09 ± 3.21	1.1

Table 4: Design and calibration terms (units of 10^{-5}) at $\varepsilon = 1$. The final column isolates the change in Exact noise variance from iid to ORF.

Dataset	A_{iid}	A_{O}	σ_G^2	$\sigma_{r,L}^2$	σ_{O}^2	$\sigma_{\mathrm{O}}^2 - \sigma_{r,L}^2$
Adult	1.31	0.05	0.29	0.09	0.09	-0.001
Covertype	4.15	0.33	0.29	0.10	0.10	-0.002
USPS	11.01	0.93	0.29	0.14	0.14	0.001
ISOLET	2.95	0.11	0.29	0.14	0.14	0.001

G Additional kernel query results

Table 3 complements the kernel family comparison with absolute MSE at $n = 5000$, $r = 2p$, and $\varepsilon = 1$. The RBF decomposition below separates approximation error from calibrated noise.

At this reference budget, the relative reduction ranges from 0.1% to 12.2% across the entries in Table 3. The smaller gains occur when approximation contributes most of the total MSE. Since the target query changes with the kernel, absolute MSE supports calibration comparisons within each row.

G.1 Feature design and calibration

Changing the feature design can alter both approximation error and the Exact noise scale. Table 4 separates these contributions at $\varepsilon = 1$. The difference $\sigma_{\mathrm{O}}^2 - \sigma_{r,L}^2$ measures the change in Exact noise variance from iid to ORF, while Uniform AG has the same scale for both unit-norm maps.

In all four datasets, ORF substantially reduces the approximation term. The change in Exact noise variance is much smaller and has either sign. The observed ORF advantage at these bounded-domain settings comes mainly from approximation.

G.2 Fixed bandwidth sensitivity

The RBF comparison also uses $\gamma \in \{1/(2p), 1/p, 2/p\}$ at $n = 5000$, $r = 2p$, and $\varepsilon = 1$. Bandwidth affects both approximation variance and the public-domain distance law. Errors across bandwidths concern different query targets.

Table 5: Fixed-bandwidth iid query MSE (units of 10^{-5}) at $\varepsilon = 1$.

Dataset	$p\gamma$	Uniform AG	Exact	Gain (%)
Adult	0.5	0.64 ± 0.02	0.41 ± 0.02	35.5
	1	1.59 ± 0.08	1.40 ± 0.08	12.2
	2	4.68 ± 0.26	4.52 ± 0.26	3.4
Covertype	0.5	1.45 ± 0.13	1.23 ± 0.13	15.2
	1	4.43 ± 0.47	4.25 ± 0.47	4.2
	2	13.55 ± 1.53	13.40 ± 1.53	1.1
USPS	0.5	5.28 ± 0.39	5.12 ± 0.39	3.1
	1	11.30 ± 0.86	11.15 ± 0.86	1.3
	2	15.88 ± 1.16	15.74 ± 1.16	0.9
ISOLET	0.5	1.41 ± 0.08	1.25 ± 0.08	11.5
	1	3.24 ± 0.20	3.09 ± 0.20	4.5
	2	5.72 ± 0.31	5.57 ± 0.31	2.5

Table 5 shows a smaller relative calibration gain as γ increases on every dataset. Increasing γ narrows the kernel and changes both terms of the query error decomposition. The sweep measures sensitivity of the calibration gain to the kernel choice.

H Reconstructing a weighted dataset

Reduced-set reconstruction approximates an embedding by weighted inputs (Balog et al., 2018). The decoder receives only the released mean, public map, public box, and public label set. It uses no auxiliary target-domain examples. For m support points u_j with assigned labels c_j, let $\Delta_m = \{q \in \mathbb{R}^m : q_j \geq 0,\ \sum_j q_j = 1\}$ be the probability simplex. We minimize

$$\min_{u_j \in \mathcal{X},\, q \in \Delta_m} \frac{1}{2} \sum_{c=1}^{C} \left\| \sum_{j:c_j=c} q_j \phi_\Omega(u_j) - \widetilde{\mu}_c \right\|^2 + \frac{\zeta}{2}\|q\|^2, \qquad \zeta = 10^{-4}. \tag{62}$$

We use $m = n = 5000$ total support points. For unlabeled reconstruction, $C = 1$ and $\widetilde{\mu}_1 = \widetilde{\mu}_D$. For labeled reconstruction, approximately equal numbers of support points are assigned to each public class. Their optimized total weights may differ. The solver does not access private class counts. For tabular data, the support lies in the continuous encoded box, so categorical coordinates can be fractional. These points represent the feature mean and need not be valid raw database rows.

We alternate accelerated projected-gradient updates of the simplex weights with projected Adam updates of the locations. JAXopt supplies the weight solver (Blondel et al., 2022). One initialization is uniform in the public box. The other perturbs centers inferred from the phases of the released sine–cosine means, using the same public random draw. Both initializations depend only on public quantities and the released mean.

Each initialization uses at most 12 alternating blocks, with 300 Adam steps per block at learning rate 0.1 and early stopping for relative objective improvements below 10^{-4}. The weight solver targets an optimality gap of 10^{-5} within 30,000 iterations. All calibrations use the same stopping rules. Among completed iterates from both initializations, we select the one with the smallest objective in Equation (62).

H.1 Measuring reconstruction quality

Our primary metric is the squared original-kernel MMD between the protected empirical distribution and the weighted reconstruction:

$$\mathrm{MMD}_k^2 = \frac{1}{n^2} \sum_{i,i'} k(x_i, x_{i'}) + \sum_{j,j'} q_j q_{j'} k(u_j, u_{j'}) - \frac{2}{n} \sum_{i,j} q_j k(x_i, u_j). \tag{63}$$

For labeled data, we use the product kernel $k((x,y),(x',y')) = k(x,x')\mathbf{1}\{y = y'\}$. The sums include diagonal terms. We evaluate this metric against all 5000 protected records after reconstruction. It is used only for evaluation, not for selecting a decoder iterate.

Table 6: Unperturbed reconstruction references. MMD squared is in units of 10^{-3}, with 95% interval half-widths. Effective support is $m_{\text{eff}} = 1/\sum_j q_j^2$. U and L denote unlabeled and labeled releases.

Dataset	Design	Unlabeled MMD	Labeled MMD	m_{eff} (U)	m_{eff} (L)
Adult	iid	0.10 ± 0.00	0.06 ± 0.00	29	46
	ORF	0.08 ± 0.00	0.05 ± 0.00	32	50
Covertype	iid	0.16 ± 0.01	0.05 ± 0.00	22	52
	ORF	0.18 ± 0.01	0.05 ± 0.00	25	56
USPS	iid	9.85 ± 0.11	0.77 ± 0.03	66	276
	ORF	8.78 ± 0.08	0.65 ± 0.00	84	247
ISOLET	iid	4.36 ± 0.02	0.04 ± 0.00	152	2577
	ORF	4.56 ± 0.02	0.04 ± 0.00	195	2826

Unperturbed reconstruction uses the same decoder without privacy noise as a diagnostic reference. It does not guide decoder selection. The effective support size, $1/\sum_j q_j^2$, measures the concentration of the learned weights.

H.2 Unlabeled reconstruction and feature-design sensitivity

Figure 6 evaluates ORF reconstruction from the unlabeled mean. At $\varepsilon = 1$, the reduction in mean original-kernel MMD squared under Exact relative to Uniform AG ranges from 5% on USPS to 54% on Adult. The USPS and ISOLET curves nearly meet for $\varepsilon \geq 2$, so the smaller release noise need not preserve the same relative gain after nonlinear reconstruction. These reconstructions are separate from the labeled data used for SVM training in Figure 5.

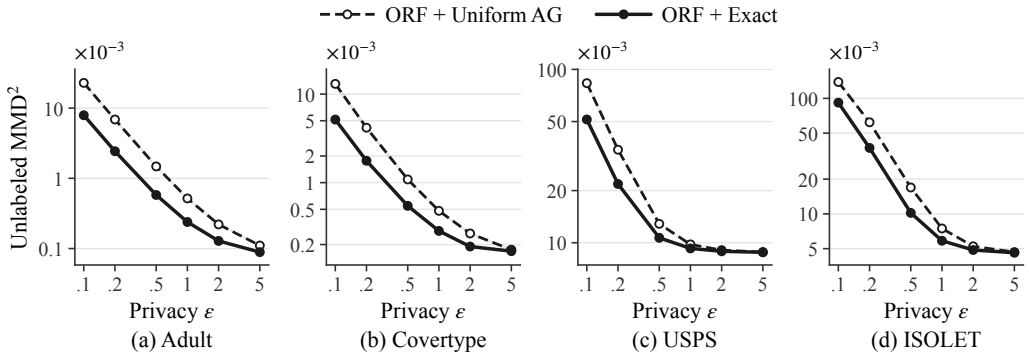


Figure 6: **Unlabeled reconstruction from ORF means.** Original-kernel MMD squared at $n = m = 5000$ and $r = 2p$. Axes are logarithmic, with pointwise 95% intervals over 30 paired trials.

Figure 7 gives the corresponding iid results for both release types, using the same reconstruction procedure. The labeled MMD uses a product kernel and has a different scale from the unlabeled MMD.

Table 6 measures the error remaining without privacy noise. Although the decoder has 5000 support points, its average effective support ranges from 22 to 2826 across these conditions. The weight concentration varies substantially across datasets and release types.

Table 7 summarizes the feature count comparisons. Across all fixed-bandwidth settings, Exact lowers labeled MMD in 141 of 144 comparisons and unlabeled MMD in 133 of 144. The clearest reversals occur for unlabeled USPS and ISOLET at $r = 8p$, $\varepsilon = 5$, so the ordering at the reference feature count does not extend to every reconstruction setting.

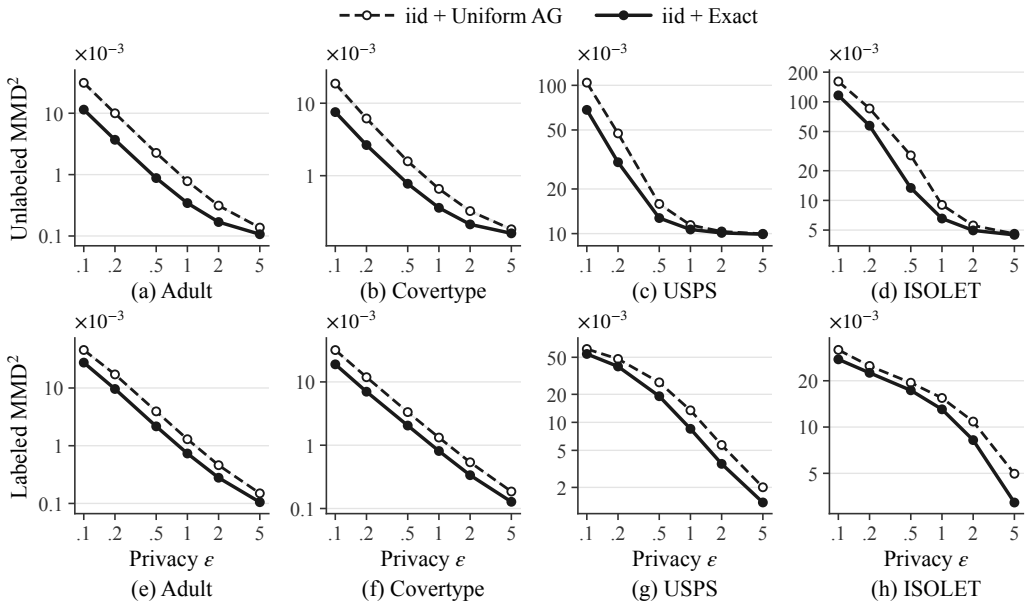


Figure 7: Squared reconstruction MMD from iid means. Top: unlabeled. Bottom: labeled. Here $n = m = 5000$, $r = 2p$. Bands are 95% intervals over 30 trials.

Table 7: Reconstruction across feature counts and privacy budgets. Each design and task has 18 conditions: $r/p \in \{0.5, 1, 2, 4, 8\}$ at $\varepsilon \in \{0.1, 1, 5\}$, plus the other three budgets at $r = 2p$. Columns report the number of conditions with lower squared MMD under Exact and the minimum–maximum mean reduction (%). Ranges describe variation across conditions, not confidence intervals. Negative reductions favor Uniform AG.

Dataset	Design	Unlabeled: lower	Gain range	Labeled: lower	Gain range
Adult	iid	17/18	-0.3 to 64.3	18/18	2.6 to 45.0
	ORF	17/18	-3.2 to 65.4	18/18	1.8 to 46.4
Covertype	iid	17/18	-0.8 to 60.4	17/18	-2.0 to 41.0
	ORF	18/18	0.3 to 60.8	17/18	-1.6 to 42.1
USPS	iid	16/18	-5.1 to 37.3	17/18	-0.2 to 38.2
	ORF	15/18	-5.2 to 38.4	18/18	0.9 to 40.5
ISOLET	iid	17/18	-7.6 to 53.3	18/18	4.6 to 37.6
	ORF	16/18	-16.3 to 45.2	18/18	3.7 to 44.4

I Binary and multiclass classification

The labeled reconstructions provide weighted SVM training examples. Adult uses binary hinge loss. Covertype, USPS, and ISOLET use Crammer–Singer multiclass loss. Both learners minimize weighted average loss plus $\lambda\|W\|_F^2/2$, without an intercept. Synthetic weights sum to one, while direct learners assign weight $1/n$ to each protected record. Synthetic fits use LIBLINEAR with tolerance 10^{-3} and at most 2000 iterations. All trials remain in the evaluation, including those that reach this cap.

I.1 Direct private baselines

The binary baseline adds analytic Gaussian output noise to a regularized SVM using sensitivity $2/(n\lambda)$ for unit-norm features (Chaudhuri et al., 2011; Rubinstein et al., 2012). The multiclass baseline follows the weight-perturbation method of Park et al. (2025), with sensitivity $4\sqrt{2}/(n\lambda)$ and its reference Crammer–Singer implementation. The public class dictionary remains fixed even if a class is absent from a training subset.

Direct baselines release task-specific model weights. The mean-based route releases a class block mean, reconstructs weighted inputs, and trains the nonprivate learner by post-processing. The routes use the same protected records, feature map, kernel, feature count, and total privacy budget. Direct nonprivate SVMs and SVMs trained on reconstructions of unperturbed means provide references for the losses due to the summary and decoder.

I.2 Regularization and complete privacy sweeps

The main mean-based regularization is $\lambda = 0.002$, and the direct private baseline uses $\lambda = 0.5$. These common settings follow preliminary benchmark exploration and are fixed across datasets and privacy budgets. They are not a private hyperparameter-selection procedure. All methods are also evaluated at

$$\lambda \in \{0.0002, 0.0005, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5\}. \tag{64}$$

The lower row of Figure 5 uses $\lambda = 0.002$ and values one decade below and above it, 0.0002 and 0.02. These values are fixed across privacy budgets. The complete panels below report accuracy and macro-F1. Macro-F1 averages over the fixed public class set, assigning zero to undefined classwise scores.

The nonprivate references help separate regularization from the effect of privacy noise. For ORF on Adult, reducing λ from 0.002 to 0.0002 raises direct nonprivate accuracy from 76.4% to 83.7%. The corresponding unperturbed decoded accuracies are 77.1% and 79.6%. Figures 8–11 show how regularization changes the accuracy gap between calibrations, in increasing order of class count.

The nonprivate references also reveal reconstruction losses before privacy noise is added (Table 8). At $\lambda = 0.002$, the ORF decoded SVM trails direct training by 6.9 percentage points on USPS and 8.3 on ISOLET. These gaps reflect the combined effect of the mean summary and reconstruction on the classifier.

J A privately estimated scale bandwidth

We evaluate whether the calibration gains persist when a privately estimated data variance determines the bandwidth. We call this setting DP-scale.

J.1 Bandwidth release and privacy budget

Let v be the variance of all entries of the clipped $n \times p$ input matrix, with denominator np. A public interval $[a, b]$ contains every coordinate, and $B = b - a$. Then

$$v(D) = \min_{t \in [a,b]} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{p} \sum_{j=1}^{p} (x_{ij} - t)^2, \qquad |v(D) - v(D')| \le \frac{B^2}{n}. \tag{65}$$

Replacing one record changes the objective by at most B^2/n at every fixed t, and hence changes its minimum by at most the same amount. We reserve $\varepsilon_b = 0.1\varepsilon$ for a Laplace release and set

$$\widehat{v} = v + \mathrm{Lap}\left(\frac{B^2}{n\varepsilon_b}\right), \qquad \overline{v} = \mathrm{clip}_{[B^2/400,\, B^2/4]}(\widehat{v}), \qquad \gamma = \frac{1}{p\overline{v}}. \tag{66}$$

The public clipping interval prevents degenerate bandwidths. Given the released bandwidth, fresh spectral draws produce the corresponding RBF map. All methods share this bandwidth and the same map within each feature design. Both designs share the bandwidth draw. The mean or SVM release receives the remaining $(0.9\varepsilon, \delta)$ budget, so adaptive composition with the pure-DP bandwidth release gives total (ε, δ) privacy.

We recalculate each Exact profile and its cross-label constraint at the released bandwidth. Sample size, reconstruction support, optimization rules, all nine regularizations, and 30 trials remain unchanged. A separate iid nonprivate reference uses $\gamma = 1/(pv)$ without variance noise. It is not a DP competitor. Classification scores concern the same prediction task across bandwidth rules. MMD at a different bandwidth uses a different kernel, so raw MMD values across these rules do not measure improvement against a common target.

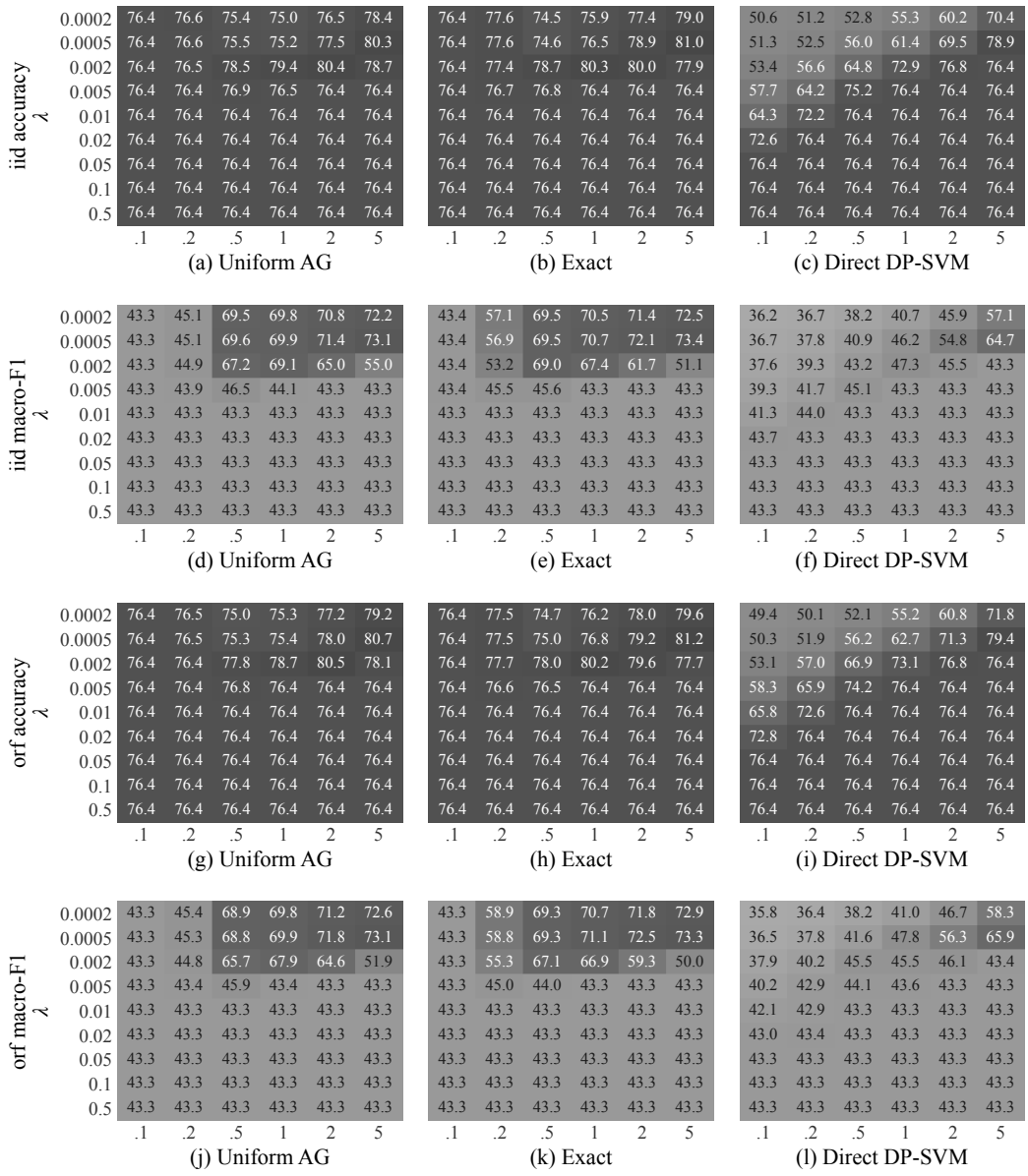


Figure 8: Adult: complete regularization grid with the fixed bandwidth. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

J.2 Classification and reconstruction results

At the main regularization, Exact has higher mean accuracy than Uniform AG on USPS and ISOLET under DP-scale, with the largest differences at intermediate privacy budgets. Adult and Covertype have reversals, including Adult at $\varepsilon = 0.5$ and Covertype at $\varepsilon = 2$. Figures 12–15 report both accuracy and macro-F1 over the full regularization grid.

The ordering of reconstruction MMD also reverses in several conditions (Figure 16). For both designs, Uniform AG has lower MMD on Adult at $\varepsilon = 0.5$ and Covertype at $\varepsilon = 2$. Smaller release noise does not always yield smaller error after nonlinear reconstruction.

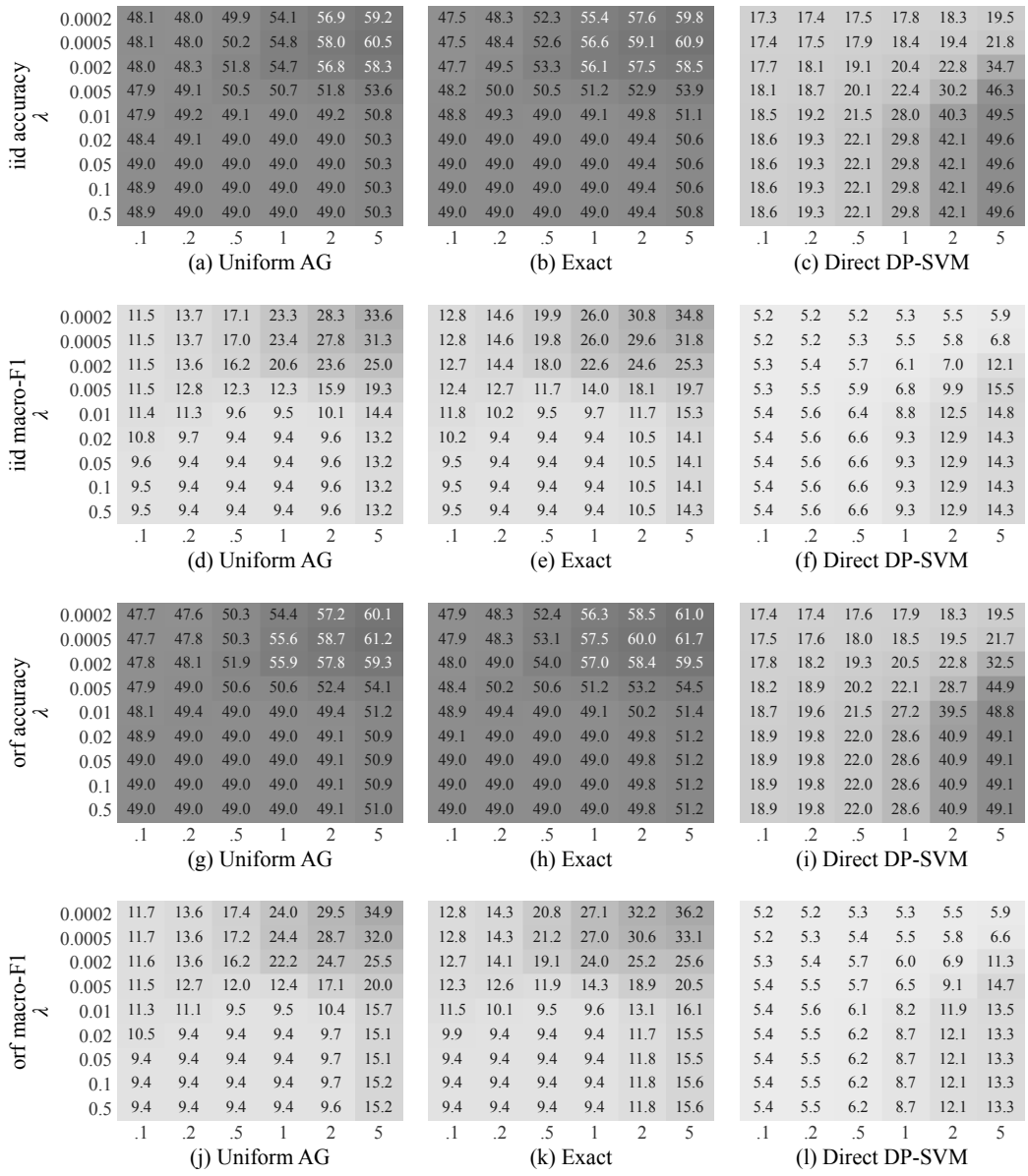


Figure 9: Covertype: complete regularization grid with the fixed bandwidth. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

Table 9 isolates the nonprivate scale-bandwidth reference for direct iid SVMs. At $\lambda = 0.002$, it improves accuracy on all four datasets relative to the fixed-bandwidth direct references in Table 8. The DP-scale results additionally account for variance noise and the budget reserved for bandwidth estimation.

K Additional related work

Private kernel means and representations. Kernel mean embeddings represent probability measures by RKHS elements (Muandet et al., 2017), and characteristic kernels distinguish distributions through these embeddings (Sriperumbudur et al., 2010). Private database release reconstructs

Table 8: Nonprivate classification references (%). Direct trains on the protected records. Decoded trains on data reconstructed from the unperturbed mean. The $\pm$ term is the 95% interval half-width.

Dataset	Design	λ	Direct acc.	Decoded acc.	Direct F1	Decoded F1
Adult	iid	0.0002	83.4 ± 0.1	79.6 ± 0.4	73.9	73.2
		0.002	76.4 ± 0.0	77.0 ± 0.3	43.3	46.7
		0.02	76.4 ± 0.0	76.4 ± 0.0	43.3	43.3
		0.5	76.4 ± 0.0	76.4 ± 0.0	43.3	43.3
	ORF	0.0002	83.7 ± 0.0	79.6 ± 0.4	74.3	73.2
		0.002	76.4 ± 0.0	77.1 ± 0.3	43.3	47.4
		0.02	76.4 ± 0.0	76.4 ± 0.0	43.3	43.3
		0.5	76.4 ± 0.0	76.4 ± 0.0	43.3	43.3
Covertype	iid	0.0002	62.5 ± 0.1	61.5 ± 0.3	28.0	35.8
		0.002	59.4 ± 0.3	59.4 ± 0.3	24.5	26.2
		0.02	51.2 ± 0.2	50.7 ± 0.3	15.2	13.9
		0.5	51.2 ± 0.2	50.8 ± 0.3	15.2	14.4
	ORF	0.0002	62.6 ± 0.0	61.5 ± 0.4	28.4	36.9
		0.002	60.1 ± 0.1	60.0 ± 0.2	25.0	26.5
		0.02	51.6 ± 0.1	51.2 ± 0.2	15.9	15.2
		0.5	51.6 ± 0.1	51.2 ± 0.2	15.9	15.3
USPS	iid	0.0002	95.7 ± 0.1	86.6 ± 0.2	95.4	85.8
		0.002	92.8 ± 0.1	86.2 ± 0.2	92.1	85.3
		0.02	82.3 ± 0.2	77.1 ± 0.3	81.0	74.9
		0.5	77.6 ± 0.1	70.9 ± 0.3	76.3	68.5
	ORF	0.0002	95.8 ± 0.1	86.5 ± 0.2	95.4	85.6
		0.002	93.0 ± 0.1	86.1 ± 0.2	92.4	85.3
		0.02	82.4 ± 0.1	77.7 ± 0.2	81.1	75.6
		0.5	77.6 ± 0.1	71.2 ± 0.2	76.4	68.8
ISOLET	iid	0.0002	95.1 ± 0.1	87.8 ± 0.1	95.0	87.8
		0.002	88.4 ± 0.1	80.6 ± 0.2	88.0	79.6
		0.02	87.0 ± 0.1	78.8 ± 0.1	86.5	77.7
		0.5	87.0 ± 0.1	78.7 ± 0.2	86.5	77.6
	ORF	0.0002	95.4 ± 0.0	87.8 ± 0.1	95.4	87.8
		0.002	88.9 ± 0.1	80.6 ± 0.1	88.5	79.7
		0.02	87.4 ± 0.1	79.0 ± 0.1	87.0	77.9
		0.5	87.4 ± 0.1	78.8 ± 0.2	87.0	77.7

Table 9: Direct iid SVM references with the nonprivate scale bandwidth (%). This bandwidth uses the unperturbed data variance and is not a private competitor.

Dataset	λ	Accuracy	Macro-F1
Adult	0.0002	83.5 ± 0.0	74.8 ± 0.1
	0.002	82.9 ± 0.1	72.3 ± 0.2
	0.02	76.4 ± 0.0	43.3 ± 0.0
	0.5	76.4 ± 0.0	43.3 ± 0.0
Covertype	0.0002	66.0 ± 0.1	39.5 ± 0.2
	0.002	63.2 ± 0.1	30.4 ± 0.5
	0.02	57.3 ± 0.3	22.5 ± 0.3
	0.5	53.9 ± 0.2	18.9 ± 0.2
USPS	0.0002	96.7 ± 0.1	96.4 ± 0.1
	0.002	94.6 ± 0.1	94.1 ± 0.1
	0.02	86.9 ± 0.2	85.7 ± 0.2
	0.5	81.2 ± 0.2	80.3 ± 0.2
ISOLET	0.0002	95.8 ± 0.1	95.8 ± 0.1
	0.002	91.9 ± 0.1	91.8 ± 0.1
	0.02	88.1 ± 0.1	87.8 ± 0.1
	0.5	88.1 ± 0.1	87.8 ± 0.1

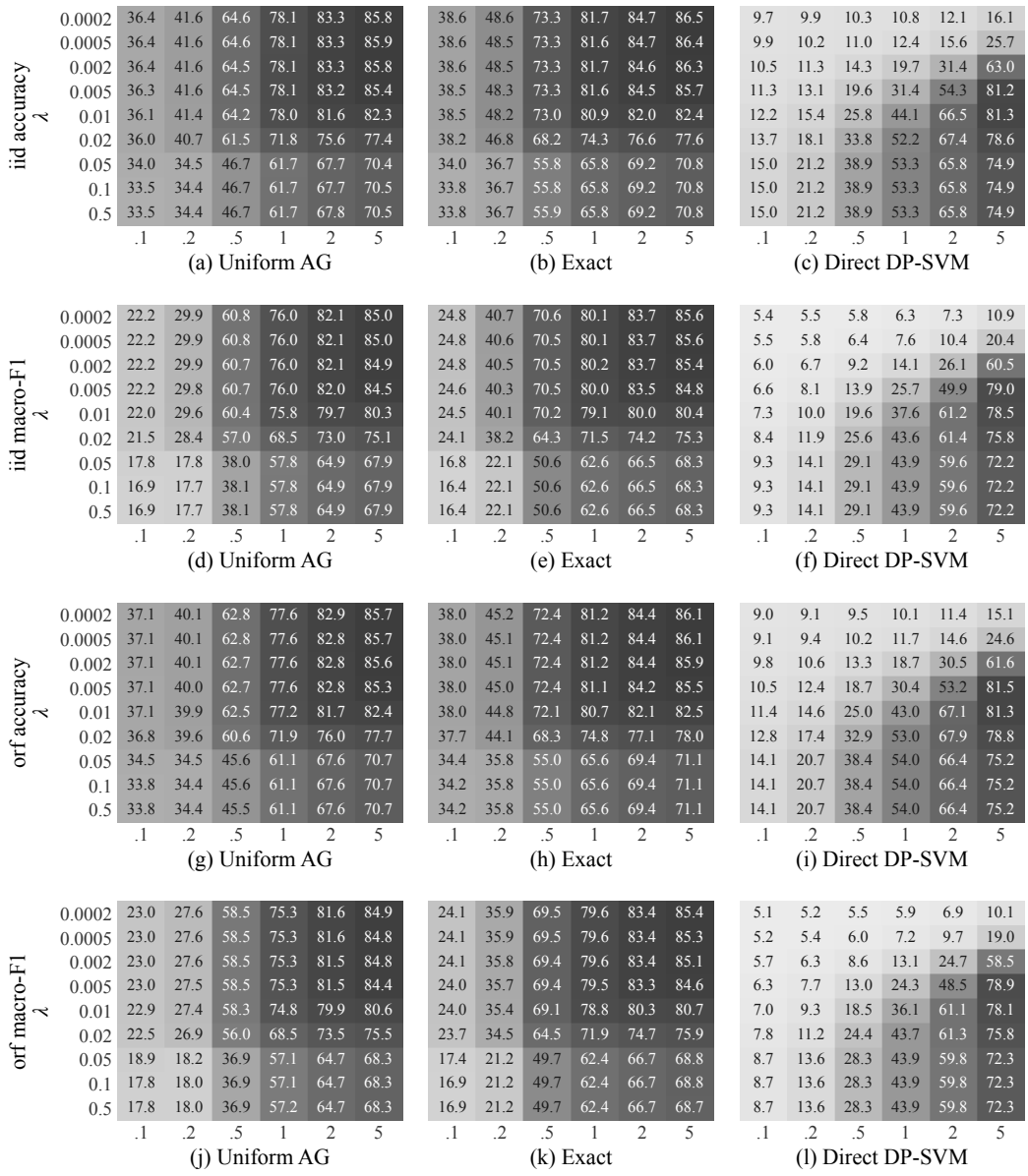


Figure 10: USPS: complete regularization grid with the fixed bandwidth. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

weighted data from perturbed embeddings (Balog et al., 2018). DP-MERF uses Gaussian-perturbed paired feature means for private generative modeling (Harder et al., 2021). We determine the exact joint privacy profile of this public map and noisy mean. Hermite features (Vinaroz et al., 2022), private Nyström features (Lee et al., 2024), and locality-sensitive quantization (Wagner et al., 2023) offer other representations or query mechanisms. Our calibration comparisons keep the feature map fixed within each trial, isolating the effect of the noise scale.

Differentially private kernel learning. Output and objective perturbation target regularized private ERM (Chaudhuri et al., 2011), while finite random-feature encodings support private kernel SVMs (Rubinstein et al., 2012). Park et al. (2023) perturb a sphere center in kernel feature space,

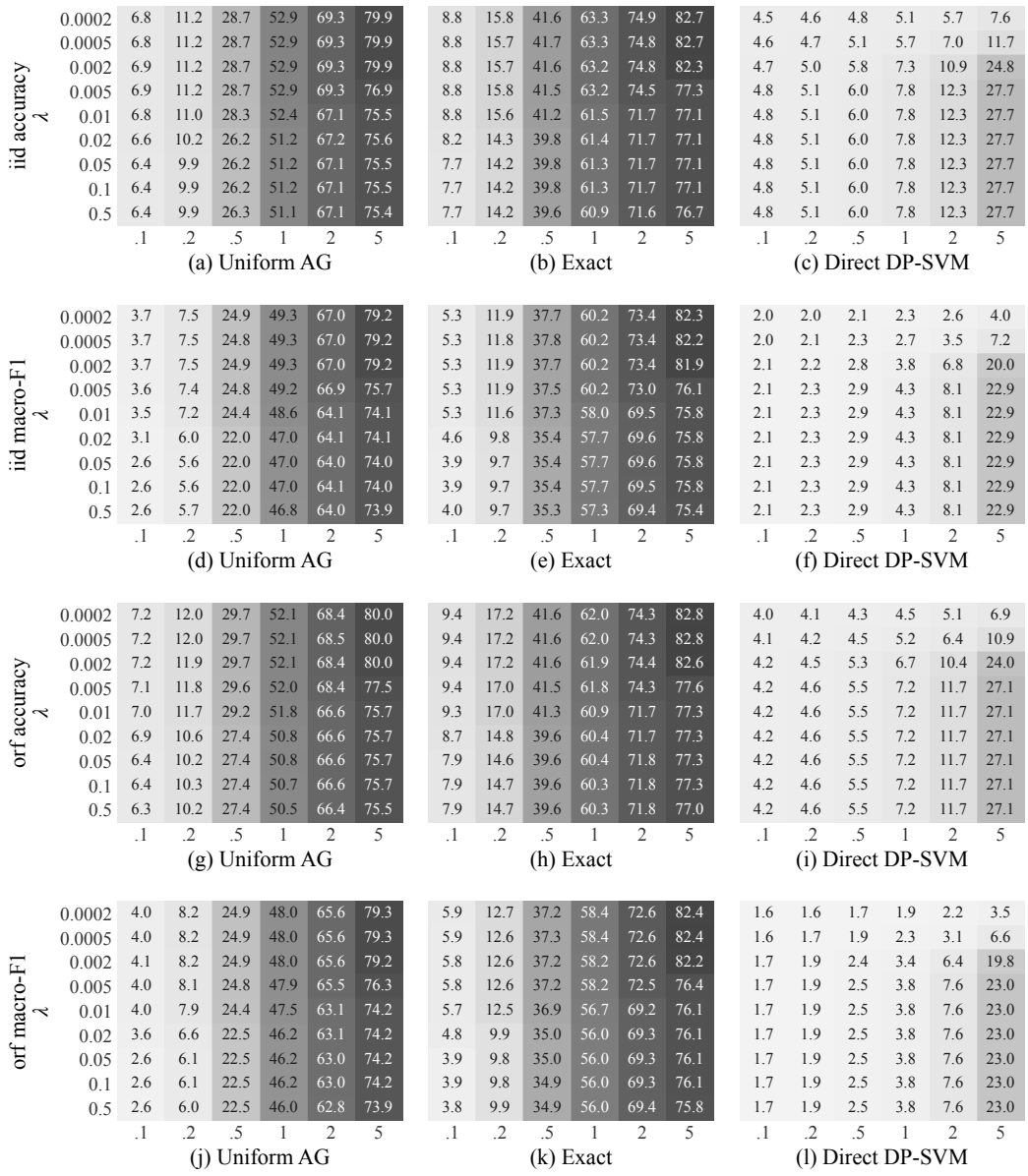


Figure 11: ISOLET: complete regularization grid with the fixed bandwidth. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

partition the input space, and assign private region labels for multiclass classification. Private kernel support functions also enable synthesis of minority-class examples for imbalanced learning (Choi et al., 2026). Approaches focused on predictors include weight or gradient perturbation for all-in-one multiclass SVMs (Park et al., 2025) and Gaussian-process random projection for private kernel ERM (Lee et al., 2026). These methods tailor the private release to a learned support function or predictor. We calibrate an empirical feature mean for subsequent queries and learning, and measure its effect on SVMs through a shared reconstruction and learner.

Frequency design. Random Fourier features approximate stationary kernels (Rahimi & Recht, 2007), with variance determined by the feature construction (Sutherland & Schneider, 2015). Or-

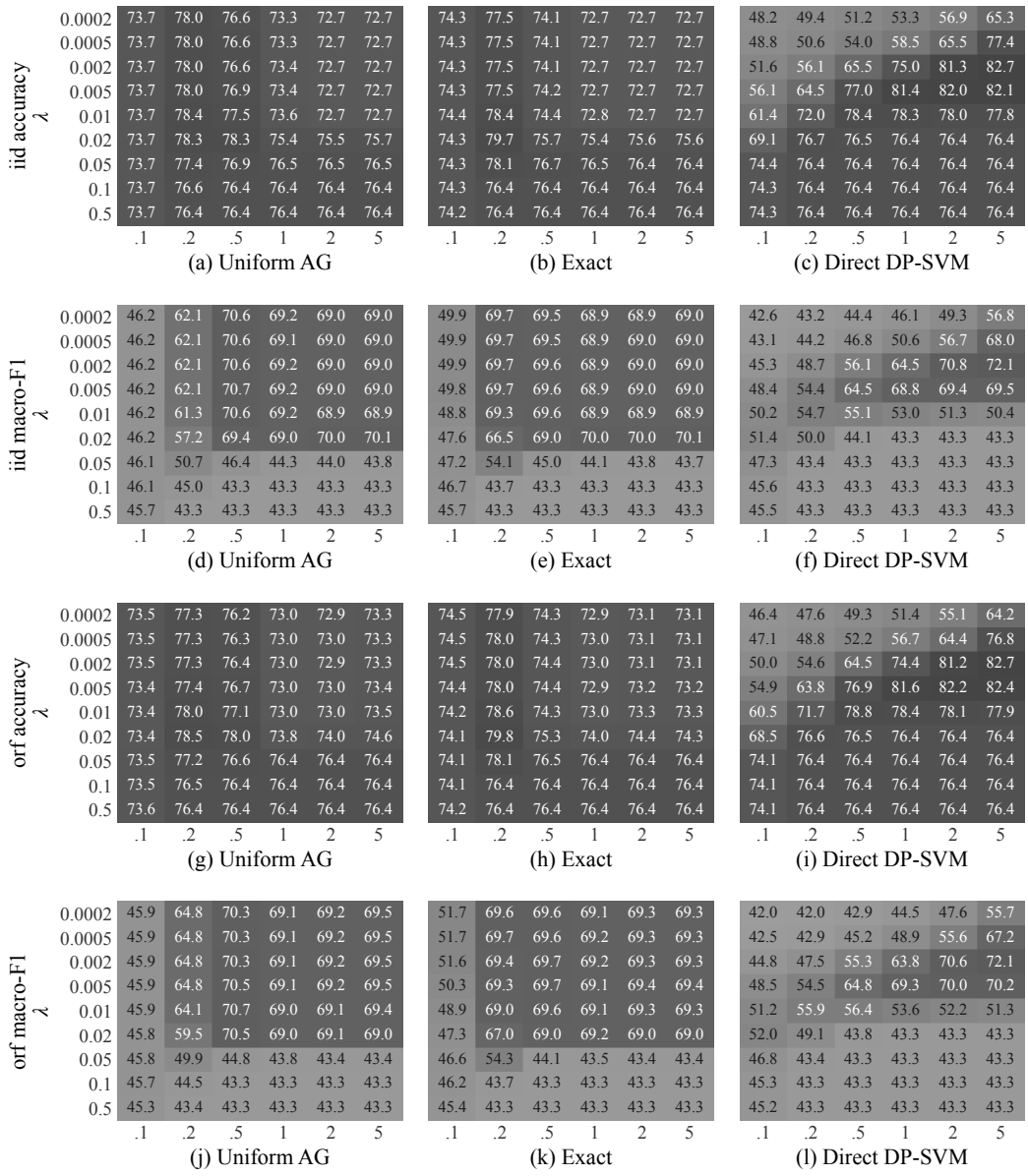


Figure 12: Adult: complete regularization grid with DP-scale. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

thogonal features (Yu et al., 2016) and couplings of their frequency norms (Reid et al., 2025) can reduce approximation variance. Our ensemble theorem compares the required isotropic noise for nonresonant designs with the same marginal spectrum on $\mathbb{R}^p$. A dependent design can improve approximation while its privacy profile exceeds the iid floor.

Gaussian calibration and covariance choice. Analytic Gaussian calibration supplies the conditional profile (Balle & Wang, 2018). General Gaussian mechanisms optimize the noise covariance for mean estimation on a bounded domain (Nikolov & Tang, 2024). Our main results keep the covariance isotropic and independent of the map, and obtain the exact joint profile by averaging the

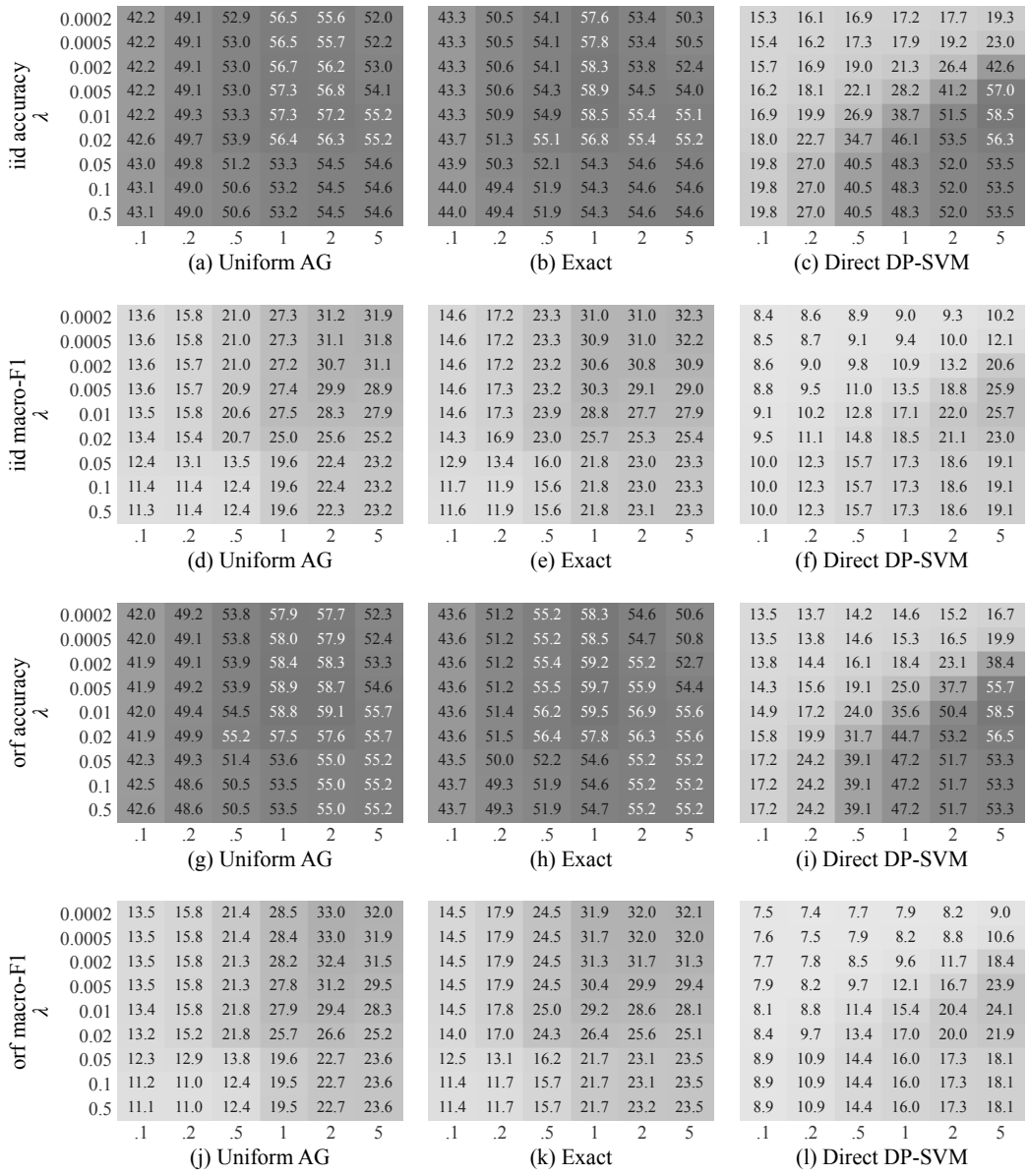


Figure 13: Covertype: complete regularization grid with DP-scale. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

conditional divergence over maps. Other covariance constructions exploit the geometry of counts and vector sums (Joseph & Yu, 2024; Lebeda, 2025).

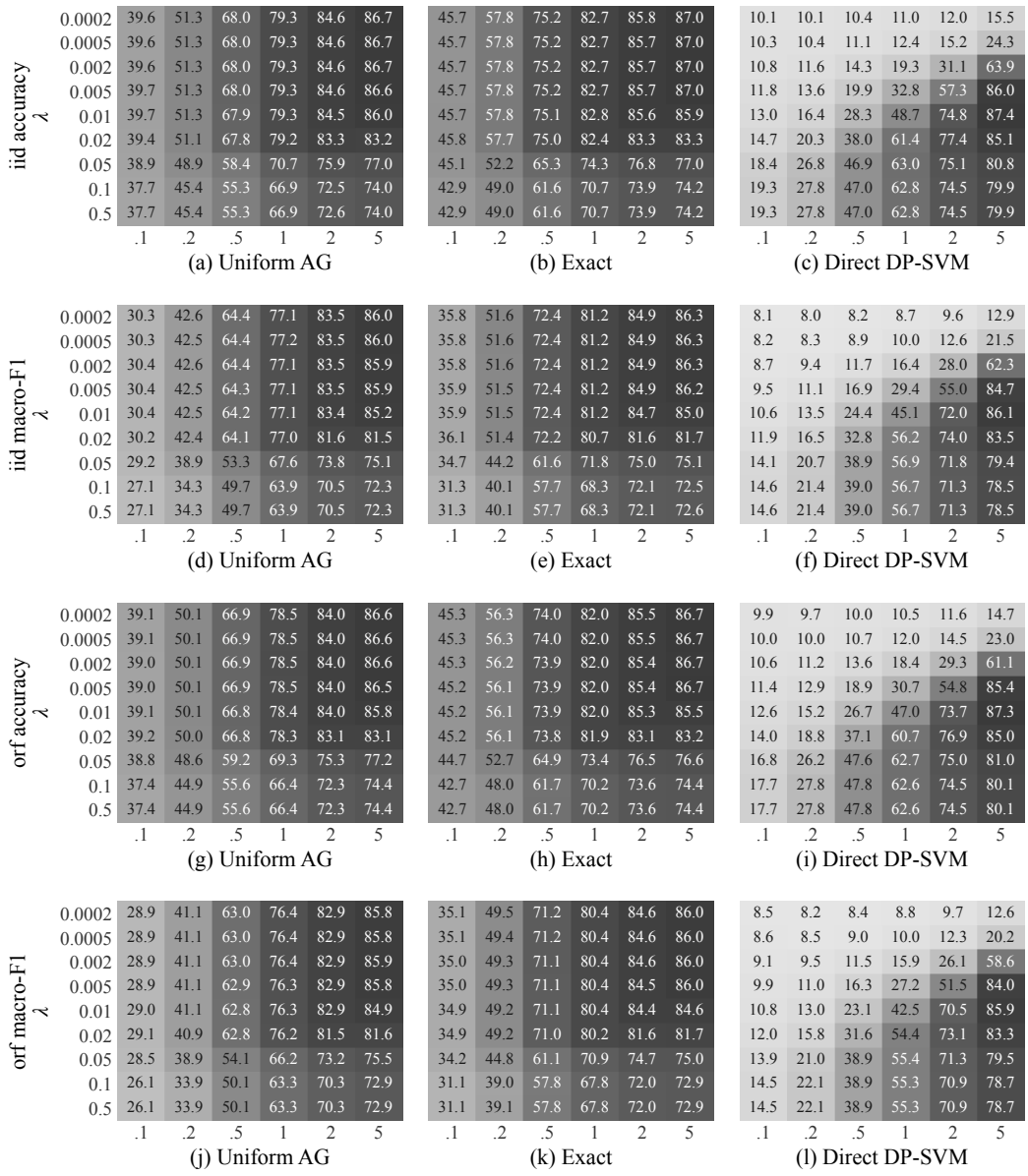


Figure 14: USPS: complete regularization grid with DP-scale. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

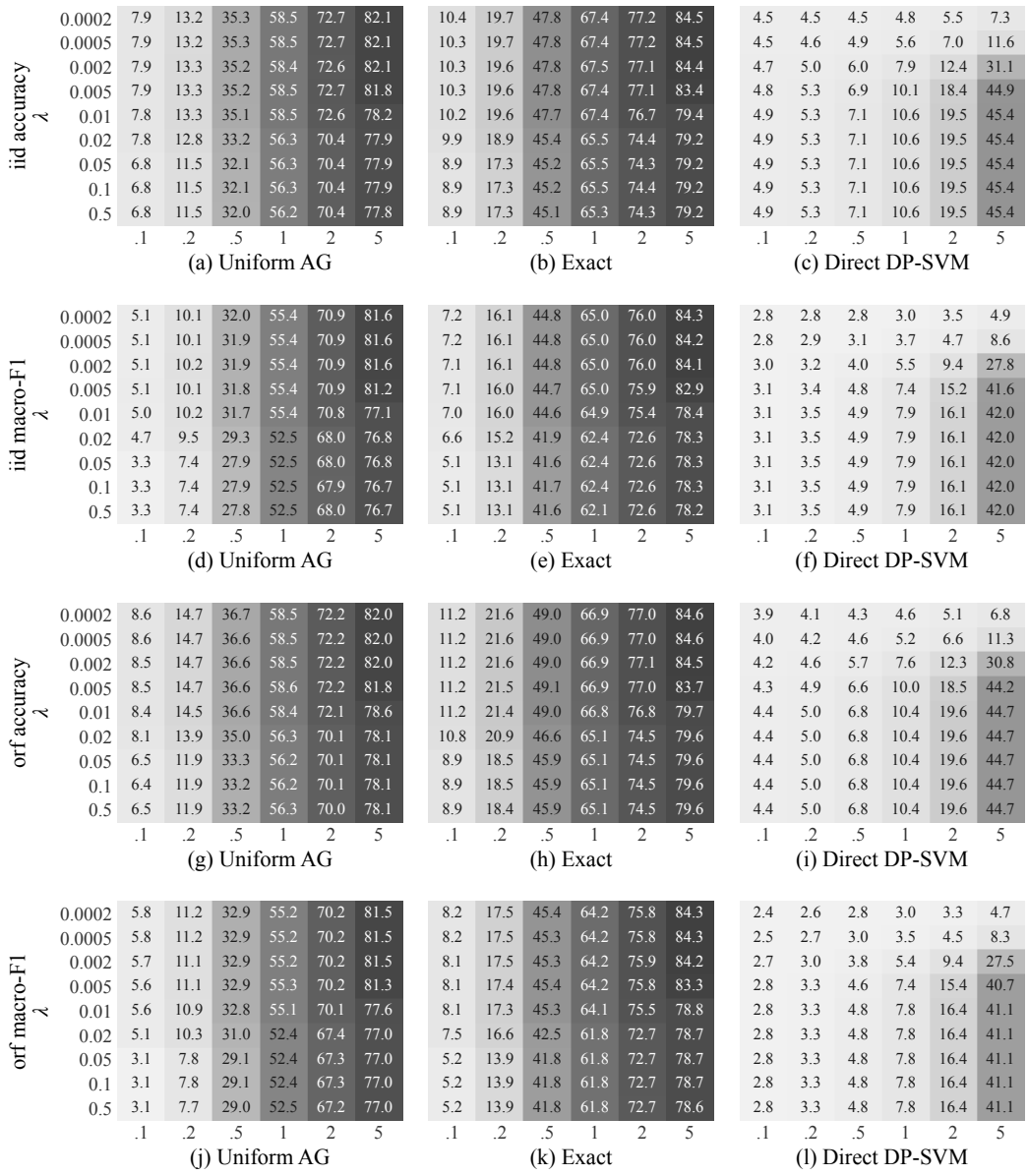


Figure 15: ISOLET: complete regularization grid with DP-scale. From top to bottom: iid accuracy, iid macro-F1, ORF accuracy, and ORF macro-F1. The three columns compare Uniform AG, Exact, and direct DP-SVM. Cells are means (%) over 30 trials. Within each panel, columns vary ε and rows vary λ.

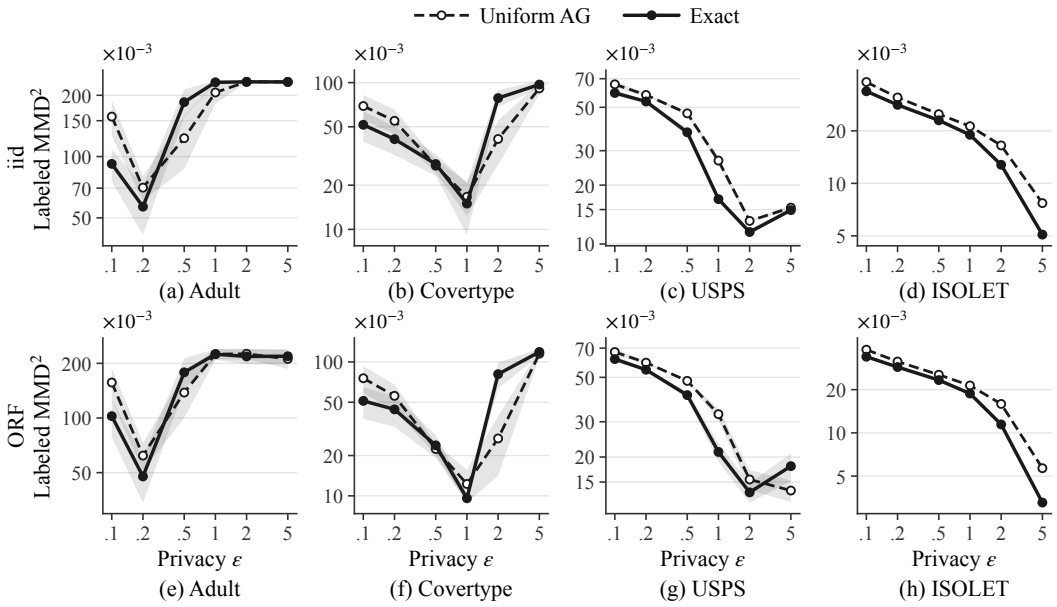


Figure 16: Labeled reconstruction MMD squared under DP-scale. Top: iid. Bottom: ORF. Calibrations share the privately estimated bandwidth within each trial. Bands are 95% intervals over 30 trials.