DIVERSITY WITH AXIOMATIC GUARANTEES IN POLYNOMIAL TIME: RESOLVING AN OPEN PROBLEM ON DIVERSITY MEASURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Mironov & Prokhorenkova (2025) formulated three axioms that a measure of the diversity of n objects, given only their pairwise dissimilarities (with no metric assumption), ought to satisfy: *monotonicity*, *uniqueness* and *continuity*. They showed that no measure in their survey satisfies all three, and constructed two measures that do but are NP-hard to evaluate. They asked whether *any* measure satisfying the axioms can be computed in practice, or whether all of them are hard. **We resolve this open problem in the affirmative.** We give an explicit measure Div_μ, which satisfies all three axioms and is computable to accuracy ε in time polynomial in n, $1/\varepsilon$ and the input encoding length. It is not an approximation scheme for some other, harder measure: Div_μ is itself the object that satisfies the axioms. The measure is built in two steps. *Maximum diversity* in the sense of Leinster & Meckes (2016), namely *Leinster-Cobbold diversity* maximised over abundance distributions, satisfies all three axioms once integrated over the scale of the exponential similarity matrices, but its single-scale evaluation is NP-hard in general. Replacing it by its doubly nonnegative relaxation keeps every axiom and yields a convex optimization problem that can be solved to prescribed accuracy in polynomial time. The relaxation is *exact* when the dissimilarity is of negative type, including Euclidean, ℓ_1 and tree distances. On these distances, Div_μ coincides with scale-integrated maximum diversity, and its evaluation reduces to convex quadratic programming with one-dimensional numerical integration. The measure is practical to evaluate: on synthetic Gaussian embeddings of 1000 objects, evaluation via the convex quadratic formulation takes a median of 1.67 seconds using 80 quadrature nodes.

1 INTRODUCTION

Language models, image generators and recommender systems often return collections of outputs. A useful collection should contain distinct alternatives, rather than repeated versions of a few objects. Given $\binom{n}{2}$ pairwise dissimilarities, a diversity measure aggregates them into a score for comparing collections or guiding their improvement.

Mironov & Prokhorenkova (2025) surveyed the measures in current use: average, minimum and maximum pairwise distance and their "sum" variants, the Vendi score (Friedman & Dieng, 2023), determinantal point process volumes (Wilhelm et al., 2018), Rényi kernel entropy (Jalali et al., 2023), Leinster–Cobbold diversity (Leinster & Cobbold, 2012), #Circles (Xie et al., 2023), Hamiltonian diversity (Hu et al., 2024) and the energy of Velikonivtsev et al. (2024). They showed that each of these measures behaves badly on some elementary configuration. From those failures they distilled three simple and natural axioms: *monotonicity* (pushing objects apart increases diversity), *uniqueness* (replacing a duplicate by any genuinely new object increases diversity) and *continuity*. None of the measures in their survey satisfies all three. They then constructed two that do and showed both are NP-hard, by reduction from the maximum clique problem. This led them to pose:

Open problem (Mironov & Prokhorenkova, 2025)**.** Construct a diversity measure that satisfies monotonicity, uniqueness and continuity and can be computed in practice, or prove that every such measure is NP-hard to compute.

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.

We resolve this open problem by constructing a diversity measure $\mathrm{Div}_\mu : \mathcal{D}_n \to \mathbb{R}$ with three guarantees:

(i) It satisfies monotonicity, uniqueness and continuity on the full domain of dissimilarity data, without requiring a triangle inequality (Theorem 5).

(ii) For our default choice of μ, measure Div_μ can be evaluated to additive accuracy ε in time polynomial in the number of objects n, $1/\varepsilon$ and the input encoding length (Theorem 6).

(iii) On dissimilarities of negative type, including Euclidean, ℓ_1 and tree distances, it agrees exactly with scale-integrated maximum diversity. Each scale can then be evaluated by convex quadratic programming (Corollary 8).

How the measure works The guiding intuition is that diversity should represent an effective number of distinct objects. Given a similarity matrix Z and probabilities p on the objects, the quantity $p^\top Zp$ is the expected similarity between two independent draws. Its reciprocal is large when the probability mass is spread across dissimilar objects. Maximizing this reciprocal over p gives *maximum diversity* (Leinster & Meckes, 2016). Optimized probabilities also handle duplicates naturally: identical copies share the weight that could have been assigned to a single representative.

Two obstacles remain. First, computing maximum diversity is NP-hard for general similarity matrices. We write $p^\top Zp = \langle Z, pp^\top \rangle$ and replace the rank-one matrix $pp^\top$ by a positive semidefinite, entrywise nonnegative matrix P whose entries sum to one. This doubly nonnegative relaxation yields a convex semidefinite program that can be solved to prescribed accuracy in polynomial time.

Second, a single similarity scale can overlook a strict improvement, thus violating the monotonicity axiom. We therefore average the relaxed score over the exponential similarity matrices $Z_\theta(D) = (e^{-\theta d_{ij}})$, using the probability density $2/(1+\theta)^3$ on $(0, \infty)$. Large values of θ resolve small differences between objects. Giving positive weight to arbitrarily fine scales supplies the strictness required by monotonicity and uniqueness, while the integrated measure remains continuous.

Practical evaluation and optimization On synthetic Gaussian embeddings of 1000 objects, evaluation via the convex quadratic formulation takes a median of 1.67 seconds using 80 quadrature nodes (Table 1). Geometric examples illustrate how Div_μ distinguishes configurations containing duplicates and how optimizing it produces different configurations from optimizing Average or Energy (Section 8).

Selecting a maximum-diversity subset remains NP-hard in general. We therefore develop tools for local improvement, including derivatives for continuously varying objects and bounds for screening one-object replacements using cached feasible solutions (Section 7).

Paper organization Section 2 reviews related work, and Section 3 states the setting and axioms. Sections 4 and 5 introduce optimized distributions and construct the scale-averaged measure Div_μ. Section 6 establishes polynomial-time evaluation and exactness of the quadratic formulation for negative-type dissimilarities. Section 7 studies global subset selection and develops tools for local improvement. Section 8 illustrates configuration rankings, point optimization, and evaluation cost. Section 9 discusses applications and limitations. Proofs and numerical implementation details appear in the appendices.

2 Related work

Axiomatic approaches to diversity have a long history in economics and ecology. Weitzman (1992) introduced a recursive diversity function, while Nehring & Puppe (2002) and Bossert et al. (2003) developed axiomatic treatments of diversity and the comparison of sets. Solow & Polasky (1994) proposed an inverse-similarity construction related to the magnitude of metric spaces (Leinster, 2013). Meckes (2015) further studied connections between magnitude, maximum diversity and capacities.

Our construction builds on the similarity-sensitive diversity measures of Leinster & Cobbold (2012). Leinster & Meckes (2016) showed that maximizing these measures over abundance distributions gives a common maximum across diversity orders. At order two, this is the reciprocal of a quadratic minimization over the probability simplex. Doubly nonnegative relaxations of such quadratic

problems are established optimization tools (Gökmen & Yıldırım, 2022). Our contribution is to combine this relaxation with integration over similarity scales and prove that the resulting measure satisfies all three axioms of Mironov & Prokhorenkova (2025) on the full dissimilarity domain.

The machine learning measures surveyed in Mironov & Prokhorenkova (2025) address related practical goals, including evaluating generated samples and selecting varied recommendations. In particular, Velikonivtsev et al. (2024) use inverse-distance energy to guide the generation of structurally diverse graphs. We compare with Average and Energy to illustrate behavior on configurations containing duplicates, where the proposed measure remains finite and satisfies the strict uniqueness requirement.

3 Setting

We follow Mironov & Prokhorenkova (2025). The construction below is defined for every integer $n \geq 1$. We state the axioms at each fixed size $n \geq 2$. We write $\mathbf{1} \in \mathbb{R}^n$ for the all-ones vector and set $J = \mathbf{1}\mathbf{1}^\top$. The space of real symmetric $n \times n$ matrices is denoted by $\mathbb{S}^n$, with inner product $\langle X, Y \rangle = \sum_{i,j} X_{ij} Y_{ij}$. Inequalities between matrices written $\geq$ are entrywise, $\succeq$ is the positive semidefinite order, and $\Delta_n := \{p \in \mathbb{R}^n : p \geq 0,\ \mathbf{1}^\top p = 1\}$.

Definition 1 (Dissimilarity data)**.** $\mathcal{D}_n$ is the set of $D = (d_{ij}) \in \mathbb{R}^{n \times n}$ satisfying

(D1) $d_{ij} \geq 0$ for all i, j, and $d_{ii} = 0$.

(D2) If $d_{ij} = 0$ then $d_{ik} = d_{jk}$ for all k.

(D3) $d_{ij} = d_{ji}$.

No triangle inequality is imposed. Condition (D2) says coinciding objects have the same distances to everything else. It makes $d_{ij} = 0$ an equivalence relation on $[n]$. We write $k(D)$ for the number of *distinct* objects. A *diversity measure* is a permutation-invariant function $\mathrm{Div} : \mathcal{D}_n \to \mathbb{R}$.

Axiom 1 (Monotonicity)**.** If $A, B \in \mathcal{D}_n$ with $A \geq B$ and $A \neq B$, then $\mathrm{Div}(A) > \mathrm{Div}(B)$.

Axiom 2 (Uniqueness)**.** Let $A, B \in \mathcal{D}_n$ satisfy $a_{ij} = b_{ij}$ for all $i, j > 2$, $b_{1j} = b_{2j}$ for all j, $a_{2j} = b_{2j}$ for all $j > 2$, and $a_{1j} > 0$ for all $j > 1$. Then $\mathrm{Div}(A) > \mathrm{Div}(B)$.

Axiom 3 (Continuity)**.** Div is continuous with respect to the topology $\mathcal{D}_n$ inherits from $\mathbb{R}^{n \times n}$.

Axiom 2 compares a configuration B in which object 1 duplicates object 2 with the configuration A obtained by moving object 1 somewhere distinct from all others, leaving objects $2, \ldots, n$ untouched. Crucially no relation between a_{1j} and b_{1j} is imposed: the new object may sit at distance 10^{-9} from all other objects in configuration.

Distances must use the same units across compared collections. Dividing each input matrix by its own diameter changes the function and can destroy monotonicity. A fixed reference scale chosen for an application is compatible with the construction.

4 Step 1: optimized distribution

The tension is between ignoring duplicates and responding continuously to new objects. At fixed collection size, uniqueness and continuity imply that redistributing duplicate copies among the same distinct objects cannot affect the score (Mironov & Prokhorenkova, 2025, Proposition 6.1). Counting distinct objects achieves duplicate invariance but is discontinuous. Averaging of pairwise distances is continuous but gives additional weight to an object when it is copied.

Optimized distributions resolve this weighting problem. Let $\mathcal{S}_n$ denote the set of symmetric $n \times n$ matrices with entries in $[0, 1]$ and unit diagonal, and let $Z \in \mathcal{S}_n$ be a similarity matrix. In the next section, we construct such a matrix from D using an exponential transformation. If objects I and I' are drawn independently with probabilities $p \in \Delta_n$, then their expected similarity is

$$\mathbb{E}[Z_{II'}] = p^\top Z p.$$

Small expected similarity means that the distribution p spreads its mass across dissimilar objects, so a maximally diverse distribution corresponds to $\min_{p\in\Delta_n} p^\top Zp$. Define *maximum diversity* as

$$D_{\max}(Z) = \left(\min_{p\in\Delta_n} p^\top Zp\right)^{-1}. \tag{1}$$

For k mutually dissimilar objects, represented by $Z = I_k$, uniform probabilities give $D_{\max}(I_k) = k$. Note that $D_{\max}$ ignores duplicates, since duplicate copies can only split the probability assigned to their shared location. This makes $D_{\max}$ a natural way to obtain an effective count of dissimilar objects.

The measure $D_{\max}$ has two problems: lack of strict monotonicity and computational complexity.

At a single similarity scale, an optimal distribution may assign zero weight to some objects. Increasing distances involving these objects can leave the optimal value unchanged, so $D_{\max}$ need not satisfy strict monotonicity. Appendix A gives an explicit three-object example. In the next section (Theorem 2) we restore the required strictness by integrating over all positive scales.

The second obstacle is computational. The following example shows that in general evaluating $D_{\max}$ is NP-hard (Leinster & Meckes, 2016, Proposition 9.1 and Remark 9.7). For a graph G with adjacency matrix A_G, taking $Z = I + A_G$ gives $\min_{p\in\Delta_n} p^\top Zp = 1/\alpha(G)$, where $\alpha(G)$ is its independence number. This is equivalent to maximum clique in the complement graph, thus evaluating $D_{\max}$ is NP-hard. This argument establishes NP-hardness for general similarity matrices. Our construction uses only the exponential similarity matrices introduced in the next section, but evaluation remains NP-hard even for this restricted family (Appendix B). Therefore in the next section we replace the quadratic minimization $\min_{p\in\Delta_n} p^\top Zp$ by a convex relaxation that can be solved to prescribed accuracy in polynomial time.

5 Step 2: the measure Div_μ

For a scale $\theta > 0$, define the exponential similarity matrix $Z_\theta(D) \in \mathcal{S}_n$ entrywise by

$$(Z_\theta(D))_{ij} := e^{-\theta d_{ij}} \in (0, 1]. \tag{2}$$

Larger θ resolves smaller positive dissimilarities: for every fixed $d_{ij} > 0$, the similarity $e^{-\theta d_{ij}}$ tends to zero as θ increases, while coincident objects retain similarity one.

Averaging maximum diversity over unbounded scales already gives a measure with the desired axioms.

Theorem 2. *Let μ be a probability measure on $(0,\infty)$ with $\mu((T,\infty)) > 0$ for every $T > 0$. Then $D \mapsto \int_0^\infty D_{\max}(Z_\theta(D))\,\mu(d\theta)$ satisfies monotonicity, uniqueness and continuity on $\mathcal{D}_n$.*

The proof is given in Appendix C.

Now let us resolve the computational obstacle, by relaxing the inner problem to a convex problem. Write $P = pp^\top$. Then $p^\top Zp = \langle Z, pp^\top\rangle$ becomes $\langle Z, P\rangle$. We drop the rank-one constraint on P and retain positive semidefiniteness, nonnegative entries and normalized total mass.

Definition 3 (Relaxed diversity at one scale)**.** For $Z \in \mathcal{S}_n$, define

$$\mathcal{F}_n = \{P \in \mathbb{S}^n : P \succeq 0,\ P \geq 0,\ \langle J, P\rangle = 1\}, \tag{3}$$

$$m(Z) = \min_{P\in\mathcal{F}_n} \langle Z, P\rangle, \qquad g(Z) = \frac{1}{m(Z)}. \tag{4}$$

Why this relaxation? The original variable $p \in \Delta_n$ is a probability vector. Its matrix $P = pp^\top$ is symmetric and has three properties:

$$\begin{aligned} x^\top Px &= (p^\top x)^2 \geq 0 && \text{for every } x \in \mathbb{R}^n, \\ P_{ij} &= p_i p_j \geq 0 && \text{for every } i, j, \\ \langle J, P\rangle &= \sum_{i,j} p_i p_j = (\mathbf{1}^\top p)^2 = 1. \end{aligned}$$

Thus $pp^\top$ is positive semidefinite, entrywise nonnegative, and has total mass one. It also has rank one. The relaxation retains the first three properties and drops the rank-one requirement, so every original matrix $pp^\top$ remains feasible. The objective is unchanged because $p^\top Z p = \langle Z, pp^\top \rangle$. The condition $\langle J, P \rangle = 1$ therefore preserves the total probability of the pair (I, I'). A relaxed feasible matrix need not factor as $pp^\top$, so it need not describe independent draws.

Definition 4 (Scale-averaged diversity)**.** Fix a probability measure μ on $(0, \infty)$, independently of the dissimilarity data. For $D \in \mathcal{D}_n$, define

$$\mathrm{Div}_\mu(D) = \int_0^\infty g(Z_\theta(D))\, \mu(d\theta). \tag{5}$$

Our default choice is $\mu(d\theta) = \rho(\theta)\, d\theta$ with $\rho(\theta) = 2/(1+\theta)^3$. For this choice,

$$\mathrm{Div}_\mu(D) = \int_0^\infty g(Z_\theta(D)) \frac{2\, d\theta}{(1+\theta)^3}. \tag{6}$$

We use this default for numerical evaluation and experiments. Results for general μ state their assumptions explicitly.

The three axioms hold for any probability measure that assigns positive mass beyond every finite scale (Theorem 5). We choose $\rho(\theta) = 2/(1+\theta)^3$ for its simplicity and quantitative properties. Its finite first moment gives a Lipschitz bound on sensitivity to input perturbations, while its polynomial tail gives a lower bound on the diversity gained by introducing a nearby new object. Its explicit quantile also simplifies numerical integration. Appendix D establishes these properties; the novelty bound is Proposition 15.

Theorem 5 (Axiomatic validity)**.** *Let μ be a probability measure on $(0, \infty)$ satisfying $\mu((T, \infty)) > 0$ for every $T > 0$. Then Div_μ satisfies monotonicity, uniqueness and continuity on $\mathcal{D}_n$. It is unchanged by adding or deleting a duplicate, and*

$$1 \le \mathrm{Div}_\mu(D) \le k(D).$$

The proof rests on three properties of the fixed feasible set. First, nonnegative entries of P make increasing dissimilarities nondecreasing for g. Second, merging duplicate rows and columns preserves feasibility and the objective. Third, after duplicates have been merged, $Z_\theta(D) \to I_{k(D)}$ as $\theta \to \infty$. At this limit the unique optimizer is $J_{k(D)}/k(D)^2$, which is positive on every pair. This gives strict monotonicity at sufficiently fine scales. Replacing a duplicate increases the limiting effective count by one, giving strict uniqueness. Full proofs are in Appendix C.

6 Computing the Measure

6.1 Polynomial-time evaluation

For the default density $\rho(\theta) = 2/(1+\theta)^3$, the quantile function is $\theta(v) = (1-v)^{-1/2} - 1$. Consequently,

$$\mathrm{Div}_\mu(D) = \int_0^1 g(Z_\theta(D))\big|_{\theta=(1-v)^{-1/2}-1}\, dv. \tag{7}$$

The integrand is nondecreasing and lies in $[1, n]$. Using midpoint quadrature on M equal intervals of $[0, 1]$, we approximate the integral by the average of its values at $v_j = (j + 1/2)/M$, for $j = 0, \ldots, M-1$. The quadrature error is at most $(n-1)/M$. Evaluation at each v_j requires one convex solve.

Theorem 6 (Polynomial-time evaluation)**.** *Let $\mu(d\theta) = 2(1+\theta)^{-3}\, d\theta$. For rational $D \in \mathcal{D}_n$ of encoding length B and $0 < \varepsilon < 1$, the value $\mathrm{Div}_\mu(D)$ can be computed to additive accuracy ε in time polynomial in n, B and $1/\varepsilon$.*

Theorem 6 gives a conservative guaranteed quadrature rule. In the experiments we use Gauss-Legendre quadrature and report its observed convergence separately from inner-solver accuracy. Appendix E gives the proof of Theorem 6, including the approximation and precision bounds needed for the semidefinite solves.

The same argument extends to other fixed probability measures whose quantiles at rational arguments can be computed to absolute error 2^{-b} in time polynomial in b and the argument's encoding length. After the quantile change of variable, the integrand remains nondecreasing and bounded between 1 and n, so the midpoint error bound $(n-1)/M$ is unchanged. For example, the exponential density $\rho(\theta) = e^{-\theta}$ has quantile $-\log(1-v)$ and admits the same polynomial-time evaluation guarantee. Our default density is chosen for its polynomial tail and finite first moment, rather than being required for tractability.

6.2 Exactness and quadratic programming

For negative-type dissimilarities, including Euclidean distances, each scale can be evaluated by convex quadratic programming. This follows from exactness of the relaxation for positive semidefinite similarity matrices, which we establish next.

Theorem 7 (Exactness for positive semidefinite similarities)**.** *Let $Z \in \mathcal{S}_n$. If $Z \succeq 0$, then $g(Z) = D_{\max}(Z)$.*

Negative-type dissimilarities generate positive semidefinite exponential similarity matrices at every scale, so Theorem 7 applies throughout the integral.

Corollary 8 (Negative-type dissimilarities)**.** *Let μ be any probability measure on $(0, \infty)$, and let $D \in \mathcal{D}_n$ satisfy $x^\top Dx \le 0$ whenever $\mathbf{1}^\top x = 0$. Then*

$$\mathrm{Div}_\mu(D) = \int_0^\infty D_{\max}(Z_\theta(D))\, \mu(d\theta). \tag{8}$$

Thus the proposed measure agrees exactly with scale-averaged maximum diversity on negative-type dissimilarities. Examples include Euclidean distances, squared Euclidean dissimilarities, ℓ_1 distances and tree distances.

For these inputs, each scale requires only the convex quadratic program in (1). Positive semidefiniteness is a sufficient condition for exactness, not a necessary one. When negative type is not known from the construction of the distance, it can be checked by testing $HDH \preceq 0$, where $H = I - J/n$ (Proposition 16). Appendix F contains the proofs of Theorem 7 and Corollary 8 and gives an optional nonnegative least squares formulation for positive definite similarity matrices (Proposition 17).

7 Optimizing diversity

In practical applications, a diversity measure can be used to guide the construction of a diverse collection, rather than only assign a score to a fixed one. We consider improving a collection of fixed size within a prescribed space of objects. Two existing objectives illustrate why the choice of measure matters for optimization:

$$\mathrm{Average}(D) = \frac{2}{n(n-1)} \sum_{i<j} d_{ij}, \qquad \mathrm{Energy}(D) = -\frac{2}{n(n-1)} \sum_{i<j} \frac{1}{d_{ij}}. \tag{9}$$

Average pairwise distance is a widely used objective for optimizing diversity. However, maximizing it can favor repeated copies of extreme objects rather than a collection of distinct alternatives: for example, among configurations of 16 points in the unit square with Euclidean distances, Average is maximized by placing four copies at each corner (Mironov & Prokhorenkova, 2025). To address such behavior, Velikonivtsev et al. (2024) proposed the inverse-distance objective Energy. Energy penalizes close pairs, but equals $-\infty$ whenever duplicates are present, so it cannot distinguish improvements between configurations that both contain duplicates. Our measure Div_μ remains finite and strictly rewards replacing a duplicate by a new object, even if other duplicates remain. We now develop tools for using Div_μ as an optimization objective.

There are several ways to improve a collection. When objects have continuous parameters, we can adjust those parameters using derivatives. For discrete objects, we can propose replacing one object and check whether the score increases. Alternatively, we can seek the globally best subset of a prescribed size from a finite candidate pool. We first explain the difficulty of this global problem, then develop tools for the two local approaches.

Global subset selection We show that selecting a maximum-diversity subset of size k is NP-hard in general by reducing the clique decision problem to it. Given a graph G and an integer k, this problem asks whether G contains k pairwise adjacent vertices, and is NP-complete when k is part of the input.

Represent the vertices as objects, assigning distance 2 to adjacent pairs and distance 1 to distinct nonadjacent pairs. By strict monotonicity, every k-clique has greater diversity than every nonclique of the same size. Thus, whenever a k-clique exists, an exact diversity-maximizing subset must be a clique. Checking the returned subset would therefore solve the clique decision problem. Consequently, a general polynomial-time exact algorithm would imply $\mathsf{P} = \mathsf{NP}$. We therefore focus on local improvement without a guarantee of global optimality.

Gradients for continuously changing objects Suppose each object is described by continuous parameters and the pairwise dissimilarities depend differentiably on these parameters. Examples include moving points within a region or adjusting the latent inputs of a differentiable generator. A gradient indicates how to change the objects to increase diversity.

Fix a scale $\theta > 0$. If all off-diagonal similarities are strictly less than one, equivalently there are no duplicate objects, and the minimizing matrix $P_\theta^\star$ is unique, then for a symmetric coordinate $i < j$,

$$\frac{\partial g(Z_\theta(D))}{\partial d_{ij}} = 2\theta\, g(Z_\theta(D))^2 (P_\theta^\star)_{ij} e^{-\theta d_{ij}}. \tag{10}$$

The chain rule converts these derivatives into derivatives with respect to the object parameters. A weighted quadrature rule combines the derivatives across scales, giving a direction for gradient ascent subject to the constraints on the objects. Appendix G.1 states the conditions for differentiating the exact integral.

At configurations with duplicates, the minimizing matrix may be nonunique and the measure may fail to be differentiable. An ordinary gradient need not exist, but directional derivatives along admissible changes remain available. They are characterized by an additional linear optimization over the minimizing matrices, which is again a semidefinite program. Thus derivative information can still guide local changes, although numerical derivative accuracy requires separate control. Appendix G.1 gives the formula and explains this distinction.

Screening a one-object replacement Suppose we propose replacing object i by a new candidate, leaving all other objects unchanged. We give an inexpensive sufficient test for improvement. It can accept some beneficial replacements without reoptimizing, but it need not detect every improvement.

Let D be the dissimilarity matrix of the current collection, and let D' be the matrix after the proposed replacement. At each scale θ, let P_θ be a feasible matrix previously computed when evaluating $g(Z_\theta(D))$. We keep these matrices fixed and compute $1/\langle Z_\theta(D'), P_\theta\rangle$ at each scale. Because they remain feasible for D', these values are lower bounds on $g(Z_\theta(D'))$: reoptimizing could only increase the scale scores. If the resulting lower bound on the proposed diversity exceeds an upper bound on the current diversity, with optimization and integration errors accounted for, we accept the replacement without solving the optimization problems for D'. Otherwise, the test is inconclusive.

So for each scale we keep a feasible matrix P_θ from a solution for D and its current objective value $a_\theta = \langle Z_\theta(D), P_\theta\rangle$. The same matrix P_θ remains feasible for the proposed collection D', with objective

$$a'_\theta = a_\theta + 2\sum_{j\neq i}\big(e^{-\theta d'_{ij}} - e^{-\theta d_{ij}}\big)(P_\theta)_{ij}, \qquad g(Z_\theta(D')) \geq \frac{1}{a'_\theta}. \tag{11}$$

Once the proposed distances d'_{ij} are known, this lower bound costs $O(n)$ operations per scale.

After accepting a replacement, the cached matrices remain feasible but need not remain optimal. Their objective values must be updated before screening further proposals. Note that this procedure applies to collections with duplicates and does not require differentiability. Appendix G.2 proves the screening bound and explains the cache updates. Appendix G.3 gives the bounds needed to account for optimization and integration errors when comparing the current and proposed diversity.

8 Numerical Illustrations

We illustrate the measure's behavior on geometric configurations and its evaluation cost on synthetic embeddings. Average and Energy provide the two contrasts discussed in Section 7: concentration at extreme points and loss of discrimination when duplicates are present. All experiments use Euclidean distances, the default density, and the convex quadratic formulation as justified by Theorem 7.

Experiment 1: Ranking configurations with duplicates Figure 1 extends the configurations of Mironov & Prokhorenkova (2025) with an almost-grid: one interior grid point is moved onto its neighbor. The distinct supports are nested, so uniqueness and duplicate invariance imply the strict ordering obtained by Div_μ. Average ranks both corner configurations above the full grid. Energy assigns $-\infty$ even to the almost-grid, leaving it tied with both collapsed configurations.

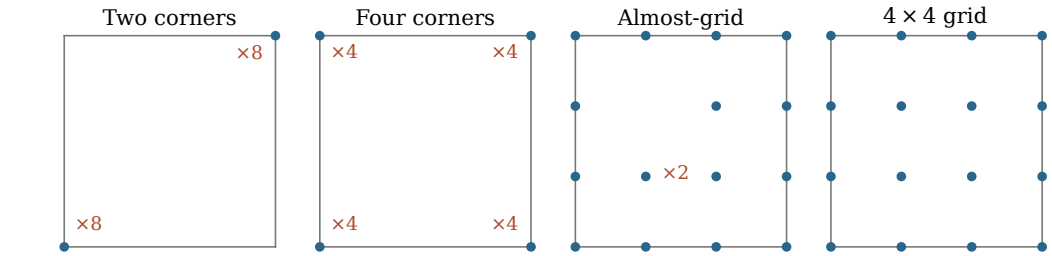


Figure 1: Four configurations of 16 points in $[0, 1]^2$, scored under all three measures. Labels indicate multiplicities. The almost-grid has one empty site and one duplicated point. Scores for Div_μ use 640 quadrature nodes.

Experiment 2: Optimizing points in a square We optimize 16 points in $[0, 1]^2$ from 25 shared uniform initializations (seeds 0–24), using the bound-constrained optimizer L-BFGS-B (Byrd et al., 1995) with common stopping settings. Energy is optimized through the equivalent harmonic mean $-1/\mathrm{Energy}$ to avoid infinite trial losses; reported scores retain the original definition. Across all 25 seeds, Average collapses to four corners, whereas Energy and Div_μ retain 16 distinct points. Energy places 14–15 points on the boundary. For Div_μ, different initializations produce both nearly regular grids and higher-scoring nongrid configurations. Figure 2 shows all three objectives from the same initialization (seed 0), alongside a higher-scoring Div_μ solution obtained from seed 7. Thus, the near-grid is a reproducible outcome of the optimization, but not the highest-scoring configuration found.

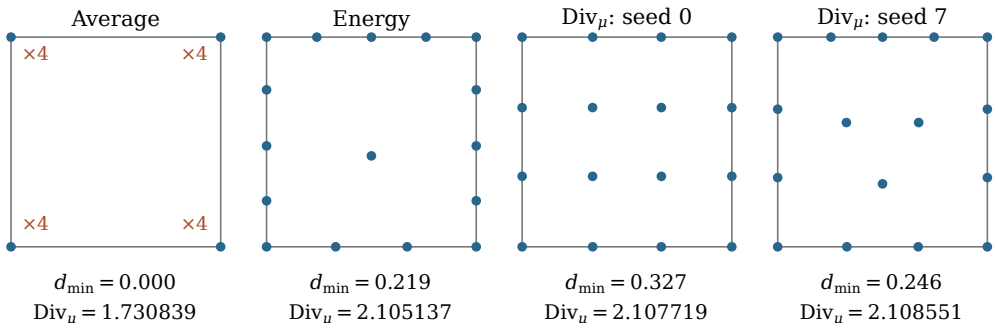


Figure 2: Local optimization: the first three panels share initialization seed 0; the fourth uses seed 7 and achieves higher Div_μ than the near-grid. Minimum separation and 640-node diversity scores are shown.

Experiment 3: Scaling to 1000 objects The two axiom-satisfying constructions of Mironov & Prokhorenkova (2025) are NP-hard to evaluate in general. To assess practical evaluation of our measure, we draw n independent standard Gaussian vectors in $\mathbb{R}^{32}$ and divide coordinates by the fixed constant $\sqrt{64}$, using ten seeds per size (seeds 0–9). Table 1 reports complete 80-node evaluations on an AMD EPYC 9V74 processor, with BLAS restricted to one thread. Times include distances, similarity matrices, and all inner solves, with quadrature nodes cached. At $n = 1000$, the median is 1.67 seconds. This study evaluates the Euclidean quadratic route.

n	Seconds, median [range]	$\lvert Q_{80} - Q_{160}\rvert$	Inner width
100	$0.127\ [0.115, 0.163]$	4.9×10^{-6}	3.7×10^{-6}
250	$0.269\ [0.248, 0.310]$	2.0×10^{-5}	1.8×10^{-6}
500	$0.580\ [0.546, 0.600]$	1.0×10^{-4}	3.2×10^{-6}
1000	$1.665\ [1.548, 1.754]$	1.5×10^{-4}	2.6×10^{-6}

Table 1: Evaluation cost. Times report the median and range of per-seed median runtimes (3 repetitions, 10 seeds). Q_m denotes diversity computed with m Gauss–Legendre nodes. Refinement changes and weighted inner interval widths are maxima over seeds.

Numerics and reproducibility Point optimization in Experiment 2 and timed evaluations in Experiment 3 use 80 Gauss–Legendre nodes in the quantile variable, with per-scale inner-solver interval tolerances of 10^{-8} and 10^{-4}, respectively. The reported configuration scores in Experiments 1 and 2 are reevaluated using 320 and 640 nodes with inner tolerances at most 10^{-9}. The largest score change across these configurations is 3.2×10^{-9}. Experiment 3 compares 80- and 160-node evaluations, with the largest changes reported in Table 1. Quadrature refinement is an empirical accuracy check. The floating-point inner intervals estimate error from the inner quadratic solves, not quadrature error. Appendix H provides input definitions, solver settings, additional optimization checks, and the timing protocol. Code, seeds, coordinates, raw results, and figure-generation scripts are available at `https://anonymous.4open.science/r/diversity-supplement`.

9 Discussion

Applications Diversity is used in image generation, recommender systems, and molecule generation, often as one component of a quality metric or optimization objective. Our measure could serve as this component alongside application-specific criteria such as fidelity, relevance, or molecular validity. Its axiomatic guarantees and practical evaluation on Euclidean inputs make it a promising candidate for assessing and encouraging diversity without the failures illustrated here. This paper establishes the theoretical foundation. Evaluating benefits in these applications remains future work.

Limitations Our scalability results concern Euclidean inputs. The general semidefinite formulation can be more expensive. Finite numerical accuracy can obscure small diversity gains. The choice of distances and scale distribution also determines which distinctions the measure rewards.

Different population sizes Our axiomatic comparisons and point-optimization experiments hold population size fixed. The construction nevertheless defines scores for every population size and is unchanged by adding duplicates. Studying comparisons between populations of different sizes, using common distance units and the same scale distribution, is another direction for future work.

10 Conclusion

We resolve the open problem posed by Mironov & Prokhorenkova (2025) by introducing a computationally tractable diversity measure that satisfies monotonicity, uniqueness, and continuity. To our knowledge, Div_μ is the first measure combining these three guarantees with polynomial-time evaluation to prescribed additive accuracy. It therefore avoids the undesirable behaviors captured by these axioms while remaining practical to evaluate on the Euclidean inputs studied here. This provides a foundation for using axiomatically justified diversity as a component of evaluation and optimization in applications.

AI USE STATEMENT

Claude (Anthropic) proposed the construction of the main measure Div_μ and generated the initial manuscript draft. ChatGPT (OpenAI) assisted with mathematical analysis and proof review, experimental design, implementation and execution of the synthetic experiments, interpretation of results, editing, and preparation of figures and reproducibility materials. The authors take responsibility for the final claims, proofs, code, experimental results, and disclosure, including all AI-assisted contributions.

REPRODUCIBILITY STATEMENT

The measure and evaluation procedures are specified in Sections 5 and 6. Proofs and numerical error bounds are provided in Appendices C–G. Appendix H specifies the experimental inputs, random seeds, solver settings, quadrature rules, and timing protocol. The supplementary archive contains the implementation, pinned dependencies, cached inputs and results, validation tests, and scripts for rerunning the experiments and regenerating all experimental figures and the timing table. Its README provides reproduction commands, and `RESULTS.md` records the provenance of the cached results. The archive is available at https://anonymous.4open.science/r/diversity-supplement.

REFERENCES

W. Bossert, P. K. Pattanaik, and Y. Xu. Similarity of options and the measurement of diversity. *Journal of Theoretical Politics*, 15(4):405–421, 2003.

R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

J. M. Danskin. *The Theory of Max-Min and its Application to Weapons Allocation Problems*. Springer-Verlag, Berlin, Heidelberg, 1967.

D. Friedman and A. B. Dieng. The Vendi score: a diversity evaluation metric for machine learning. *TMLR*, 2023.

Y. G. Gökmen and E. A. Yıldırım. On standard quadratic programs with exact and inexact doubly nonnegative relaxations. *Mathematical Programming*, 193(1):365–403, 2022. arXiv:2002.12659.

M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, Heidelberg, 1988.

X. Hu, G. Liu, Q. Yao, Y. Zhao, and H. Zhang. Hamiltonian diversity: effectively measuring molecular diversity by shortest Hamiltonian circuits. *Journal of Cheminformatics*, 16:94, 2024.

M. Jalali, C. T. Li, and F. Farnia. An information-theoretic evaluation of generative models in learning multi-modal distributions. In *NeurIPS*, 2023.

T. Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013.

T. Leinster and C. A. Cobbold. Measuring diversity: the importance of species similarity. *Ecology*, 93(3):477–489, 2012.

T. Leinster and M. W. Meckes. Maximizing diversity in biology and beyond. *Entropy*, 18(3):88, 2016. arXiv:1512.06314.

M. W. Meckes. Magnitude, diversity, capacities, and dimensions of metric spaces. *Potential Analysis*, 42:549–572, 2015. arXiv:1308.5407.

M. Mironov and L. Prokhorenkova. Measuring diversity: axioms and challenges. In *ICML*, 2025. arXiv:2410.14556.

K. Nehring and C. Puppe. A theory of diversity. *Econometrica*, 70(3):1155–1198, 2002.

I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, 1938.

A. R. Solow and S. Polasky. Measuring biological diversity. *Environmental and Ecological Statistics*, 1:95–103, 1994.

F. Velikonivtsev, M. Mironov, and L. Prokhorenkova. Challenges of generating structurally diverse graphs. In *NeurIPS*, 2024.

M. L. Weitzman. On diversity. *Quarterly Journal of Economics*, 107(2):363–405, 1992.

M. Wilhelm, A. Ramanathan, A. Bonomo, S. Jain, E. H. Chi, and J. Gillenwater. Practical diversified recommendations on YouTube with determinantal point processes. In *CIKM*, 2018.

Y. Xie, Z. Xu, J. Ma, and Q. Mei. How much space has been explored? Measuring the chemical space covered by databases and machine-generated molecules. In *ICLR*, 2023.

A FAILURE OF STRICT MONOTONICITY AT A FIXED SCALE

The measure Div_μ averages the scale score over all $\theta > 0$ rather than evaluating it at a single convenient scale. This appendix gives an explicit counterexample at $\theta = 1$, showing that maximum diversity evaluated at a fixed scale need not satisfy Axiom 1.

We demonstrate this with a counterexample of three objects: a pair of configurations $A, B \in \mathcal{D}_3$ with $A \geq B$ and $A \neq B$, but $D_{\max}(Z_1(A)) = D_{\max}(Z_1(B))$. The optimal distribution in Counterexample 9 gives one object zero weight, and pushing that object away from the other two leaves both the optimizer and the optimal value unchanged. Lemma 14 is what repairs this after integration: beyond some scale every optimizer has strictly positive entries, and μ gives every such tail positive mass.

Counterexample 9 (Single-scale monotonicity failure)**.** Consider the scale $\theta = 1$ and the similarity matrices

$$Z(s) = \begin{pmatrix} 1 & 1/2 & s \\ 1/2 & 1 & s \\ s & s & 1 \end{pmatrix}, \qquad s \in [3/4, 4/5].$$

Define $D(s)$ entrywise by $d_{ij}(s) = -\log Z(s)_{ij}$. Then $D(s) \in \mathcal{D}_3$ and $Z_1(D(s)) = Z(s)$. Because the scale is fixed at $\theta = 1$, it is convenient to specify the similarity matrix and recover the dissimilarities from it, rather than the other way round as elsewhere in the paper. Only the lower endpoint of the interval matters: $s \geq 3/4$ is what forces the bound below. The upper endpoint is arbitrary and merely keeps the arithmetic simple.

For any probability vector $p = (p_1, p_2, p_3)$, put $t = p_3$. Since $p_1 + p_2 = 1 - t$, completing the square in $p_1 - p_2$ gives

$$\begin{aligned} p^\top Z(s) p &= \frac{3}{4}(1-t)^2 + \frac{1}{4}(p_1 - p_2)^2 + 2st(1-t) + t^2 \\ &\geq \frac{3}{4}(1-t)^2 + \frac{3}{2}t(1-t) + t^2 \\ &= \frac{3}{4} + \frac{1}{4}t^2 \geq \frac{3}{4}. \end{aligned}$$

Equality is attained at $p = (1/2, 1/2, 0)$, so

$$D_{\max}(Z(s)) = \frac{4}{3} \qquad \text{for every } s \in [3/4, 4/5].$$

Now let $A = D(3/4)$ and $B = D(4/5)$; the smaller parameter gives the larger distances, so A is the more spread out configuration. The distance between objects 1 and 2 is unchanged, while both distances involving object 3 strictly increase:

$$a_{13} = a_{23} = \log(4/3) > \log(5/4) = b_{13} = b_{23}.$$

Thus $A \geq B$ and $A \neq B$, but

$$D_{\max}(Z_1(A)) = D_{\max}(Z_1(B)) = \frac{4}{3}.$$

This violates strict monotonicity at the fixed scale: the optimal distribution assigns zero weight to object 3 throughout the change.

B NP-HARDNESS FOR EXPONENTIAL SIMILARITY MATRICES

The graph construction in Section 4 uses similarities equal to zero or one. We show that the hardness persists when similarities arise from finite, strictly positive off-diagonal dissimilarities.

Proposition 10. *Computing $D_{\max}(Z_1(D))$ to additive accuracy $1/8$ is NP-hard, even for rational $D \in \mathcal{D}_n$ with strictly positive off-diagonal entries.*

Proof. Let G be a simple graph on $n \geq 3$ vertices with adjacency matrix A_G, and write $Z^{(0)} = I + A_G$. Its independence number $\alpha(G)$ is an integer in $[1, n]$, and

$$D_{\max}(Z^{(0)}) = \alpha(G)$$

by Leinster & Meckes (2016, Proposition 9.1).

Set $\delta = 1/(4n^2)$ and define

$$d_{ij} = \begin{cases} 0, & i = j, \\ \delta, & \{i,j\} \in E(G), \\ \delta^{-1}, & i \neq j,\ \{i,j\} \notin E(G). \end{cases}$$

The matrix D is symmetric with strictly positive off-diagonal entries, so $D \in \mathcal{D}_n$. Its entries are rational and have polynomial encoding length. Let $Z^{(1)} = Z_1(D) = (e^{-d_{ij}})$. On edges, $1 - e^{-\delta} \leq \delta$, while on nonedges, $e^{-1/\delta} \leq \delta$. The diagonal is unchanged, hence

$$\max_{i,j} |Z^{(1)}_{ij} - Z^{(0)}_{ij}| \leq \delta.$$

For every $p \in \Delta_n$, it follows that

$$\left| p^\top Z^{(1)} p - p^\top Z^{(0)} p \right| \leq \delta \sum_{i,j} p_i p_j = \delta.$$

Thus, writing

$$a = \min_{p \in \Delta_n} p^\top Z^{(0)} p, \qquad b = \min_{p \in \Delta_n} p^\top Z^{(1)} p,$$

we have $|a - b| \leq \delta$. Both matrices have unit diagonal and nonnegative entries, so each quadratic form is at least $\sum_i p_i^2 \geq 1/n$. Consequently,

$$|D_{\max}(Z_1(D)) - \alpha(G)| = \left| \frac{1}{b} - \frac{1}{a} \right| = \frac{|a-b|}{ab} \leq n^2 \delta = \frac{1}{4}.$$

If $\widehat{d}$ approximates $D_{\max}(Z_1(D))$ to additive accuracy $1/8$, then

$$|\widehat{d} - \alpha(G)| \leq \frac{1}{8} + \frac{1}{4} = \frac{3}{8} < \frac{1}{2}.$$

Rounding $\widehat{d}$ to the nearest integer therefore recovers $\alpha(G)$ exactly. Since the construction takes polynomial time and computing the independence number is NP-hard, obtaining such an approximation is also NP-hard. □

C Proof that the measures satisfy the axioms

We first establish properties of the optimization problem that hold at every scale. For matrices, let $\|\cdot\|_{\max}$ denote the largest absolute entry. Lemmas 11, 13 and 14 are the three properties previewed in Section 5: monotonicity of g in Z, invariance under merging duplicates, and the large-scale limit. Lemma 12 is an auxiliary fact about feasible matrices used in Appendices D and F.

Lemma 11 (Bounds, continuity, and monotonicity at one scale)**.** *The feasible set $\mathcal{F}_n$ is nonempty and compact, and for all $Z, W \in \mathcal{S}_n$:*

(i) for every $P \in \mathcal{F}_n$,

$$\frac{1}{n} \leq \operatorname{tr}(P) \leq \langle Z, P \rangle, \qquad \frac{1}{n} \leq m(Z) \leq 1, \qquad 1 \leq g(Z) \leq n; \tag{12}$$

(ii) $|m(Z) - m(W)| \leq \left(1 - \frac{1}{n}\right) \|Z - W\|_{\max}$, so m and $g = 1/m$ are continuous;

(iii) if $Z \leq W$ entrywise, then $m(Z) \leq m(W)$ and $g(Z) \geq g(W)$.

Proof. The matrix $e_1 e_1^\top$ is positive semidefinite, entrywise nonnegative and satisfies $\langle J, e_1 e_1^\top \rangle = 1$, so $\mathcal{F}_n$ is nonempty. The entries of every $P \in \mathcal{F}_n$ are nonnegative and sum to one, so $\mathcal{F}_n$ is bounded. Each of its three defining conditions is also closed, so $\mathcal{F}_n$ is compact.

For (i), positive semidefiniteness of P gives

$$1 = \mathbf{1}^\top P \mathbf{1} \le n \operatorname{tr}(P),$$

which is the lower bound on $\operatorname{tr}(P)$. For the second inequality, every $Z \in \mathcal{S}_n$ has nonnegative entries and unit diagonal, and $P \succeq 0$, so discarding the off-diagonal terms gives

$$\langle Z, P \rangle = \sum_{i,j} Z_{ij} P_{ij} \ge \sum_i Z_{ii} P_{ii} = \operatorname{tr}(P).$$

The upper bound on m follows by choosing $P = e_i e_i^\top$, and the bounds on g follow by taking reciprocals.

For (ii), let P attain $m(W)$ and use it as a feasible point for Z, so that $m(Z) \le \langle Z, P \rangle = m(W) + \langle Z - W, P \rangle$. Every matrix in $\mathcal{S}_n$ has unit diagonal, so $(Z - W)_{ii} = 0$ and only off-diagonal terms contribute:

$$\begin{aligned} \langle Z - W, P \rangle &= \sum_{i \ne j} (Z - W)_{ij} P_{ij} \le \|Z - W\|_{\max} \sum_{i \ne j} P_{ij} \\ &= \|Z - W\|_{\max} \big(1 - \operatorname{tr}(P)\big) \le \left(1 - \frac{1}{n}\right) \|Z - W\|_{\max}, \end{aligned}$$

using $P \succeq 0$, $\langle J, P \rangle = 1$ and the bound $\operatorname{tr}(P) \ge 1/n$ from (i). Exchanging Z and W gives the inequality in both directions.

Part (iii) is immediate from $P \succeq 0$ for every $P \in \mathcal{F}_n$. □

Lemma 12 (Feasible matrices dominate their rank-one part)**.** *Let $P \in \mathcal{F}_k$ and set $p = P\mathbf{1}$. Then $p \in \Delta_k$ and $P - pp^\top \succeq 0$.*

Proof. Nonnegativity of p follows from $P \succeq 0$, and $\mathbf{1}^\top p = \langle J, P \rangle = 1$, so $p \in \Delta_k$. For every $x \in \mathbb{R}^k$, the Cauchy Schwarz inequality for the positive semidefinite form P gives

$$(x^\top p)^2 = (x^\top P \mathbf{1})^2 \le (x^\top P x)(\mathbf{1}^\top P \mathbf{1}) = x^\top P x,$$

using $\mathbf{1}^\top P \mathbf{1} = \langle J, P \rangle = 1$. Since $x^\top p p^\top x = (x^\top p)^2$, this says exactly that $P - pp^\top \succeq 0$. □

Lemma 13 (Merging duplicates)**.** *Let $D \in \mathcal{D}_n$ have $k = k(D)$ distinct objects and let $\bar{D}$ be the matrix on those k objects. Then for every $\theta > 0$,*

$$g(Z_\theta(D)) = g(Z_\theta(\bar{D})). \tag{13}$$

In particular, adding or removing a copy of an existing object changes no scale score.

Proof. Let $M \in \{0,1\}^{k \times n}$ have one row per coincidence class, with $M_{ai} = 1$ when object i belongs to class a. Then

$$Z_\theta(D) = M^\top Z_\theta(\bar{D}) M.$$

For $P \in \mathcal{F}_n$, the matrix $MPM^\top$ belongs to $\mathcal{F}_k$ and has the same objective value. Conversely, choose one representative of each class and let $R \in \{0,1\}^{n \times k}$ put each coordinate on its representative. Then $MR = I_k$, and $RQR^\top \in \mathcal{F}_n$ for every $Q \in \mathcal{F}_k$, again with the same objective value. The two optimal values therefore coincide. □

Lemma 14 (The large-scale limit)**.** *Let $D \in \mathcal{D}_n$ with $k = k(D)$ distinct objects. Then*

$$\lim_{\theta \to \infty} g(Z_\theta(D)) = k(D), \tag{14}$$

and there is $\Theta > 0$ such that for every $\theta \ge \Theta$ each optimizer for $Z_\theta(\bar{D})$ has all entries strictly positive.

Proof. On k distinct objects, $Z_\theta(\bar{D}) \to I_k$ as $\theta \to \infty$. We first show that the unique minimizer of $\operatorname{tr}(P)$ over $\mathcal{F}_k$ is J_k/k^2. By Lemma 11(i) applied in dimension k, every $P \in \mathcal{F}_k$ has $\operatorname{tr}(P) \ge \mathbf{1}^\top P \mathbf{1}/k = 1/k$, using $\mathbf{1}^\top P \mathbf{1} = \langle J_k, P \rangle = 1$. Suppose equality holds. Since $\operatorname{tr}\big((I_k - J_k/k)P\big) = \operatorname{tr}(P) - \mathbf{1}^\top P \mathbf{1}/k$, equality is the statement

$$\operatorname{tr}\big((I_k - J_k/k)P\big) = 0.$$

Here $I_k - J_k/k$ is the orthogonal projection onto $\mathbf{1}^\perp$, so both factors are positive semidefinite; and for positive semidefinite X, Y the identity $\mathrm{tr}(XY) = \mathrm{tr}\left(X^{1/2}YX^{1/2}\right)$ exhibits $\mathrm{tr}(XY)$ as the trace of a positive semidefinite matrix, so $\mathrm{tr}(XY) = 0$ forces $X^{1/2}YX^{1/2} = 0$ and hence $XY = 0$.

Therefore $(I_k - J_k/k)P = 0$, that is $P = J_kP/k$, so every column of P is a multiple of $\mathbf{1}$ and $P = cJ_k$ for some $c \geq 0$. The normalization $\langle J_k, P\rangle = ck^2 = 1$ then forces $P = J_k/k^2$.

Now suppose some sequence of optimizers for $Z_\theta(\bar{D})$ did not converge to J_k/k^2 as $\theta \to \infty$. By compactness of $\mathcal{F}_k$ it would have a subsequence converging to some $P^\dagger \neq J_k/k^2$, and $P^\dagger$ would be a minimizer of $\mathrm{tr}(P)$ over $\mathcal{F}_k$, contradicting uniqueness. Hence every optimizer converges to J_k/k^2, whose entries are all positive, which gives the positivity claim; and the optimal values converge to $\mathrm{tr}(J_k/k^2) = 1/k$, which with Lemma 13 gives (14). □

We restate the two theorems of Section 5 before proving them. Both carry their main-text numbering.

Theorem 5 (Axiomatic validity)**.** *Let μ be a probability measure on $(0, \infty)$ satisfying $\mu((T, \infty)) > 0$ for every $T > 0$. Then Div_μ satisfies monotonicity, uniqueness and continuity on $\mathcal{D}_n$. It is unchanged by adding or deleting a duplicate, and*

$$1 \leq \mathrm{Div}_\mu(D) \leq k(D).$$

Proof. Permutation invariance follows by conjugating feasible matrices with the corresponding permutation matrix.

For monotonicity, suppose $A \geq B$ and $A \neq B$. The exponential similarity matrices obey $Z_\theta(A) \leq Z_\theta(B)$, so by Lemma 11(iii) the scale score for A is at least that for B. The coincidence partition of A refines that of B: if $a_{ij} = 0$ then $0 \leq b_{ij} \leq a_{ij} = 0$, so $b_{ij} = 0$ and objects coinciding in A also coincide in B. If A has more distinct objects, Lemma 14 gives strict inequality at every sufficiently large scale. Otherwise the two partitions have the same number of classes, and a refinement with the same number of classes is an equality, so they coincide. Merge their common classes and use an optimizer P_B for $Z_\theta(\bar{B})$. For sufficiently large θ, Lemma 14 makes its entries positive, and at least one matrix entry strictly decreases. Therefore

$$m(Z_\theta(\bar{A})) \leq \langle Z_\theta(\bar{A}), P_B\rangle < \langle Z_\theta(\bar{B}), P_B\rangle = m(Z_\theta(\bar{B})).$$

Thus the scale score strictly increases throughout a tail of scales.

For uniqueness, removing the duplicate from B leaves the common configuration of objects $2, \ldots, n$. By Lemma 13, this has the same scale score as B. Embedding its optimizer into A by adding a zero row and column gives $g(Z_\theta(A)) \geq g(Z_\theta(B))$ at every scale. Also $k(A) = k(B) + 1$, so Lemma 14 makes the inequality strict at all sufficiently large scales.

The mixing measure assigns positive mass to every tail. Integrating proves both strict axioms. Finally, the scale score is continuous in D by Lemma 11(ii) and bounded by n. Dominated convergence proves continuity of the integral. Invariance under adding or deleting a duplicate is Lemma 13, and applying Lemma 11(i) after duplicate merging gives the stated range. □

Theorem 2. *Let μ be a probability measure on $(0, \infty)$ with $\mu((T, \infty)) > 0$ for every $T > 0$. Then $D \mapsto \int_0^\infty D_{\max}(Z_\theta(D))\, \mu(d\theta)$ satisfies monotonicity, uniqueness and continuity on $\mathcal{D}_n$.*

Proof. Write $a(Z) = \min_{p \in \Delta_n} p^\top Zp$, so that $D_{\max}(Z) = 1/a(Z)$. The argument is the one used for g above, carried out on the compact simplex instead of on $\mathcal{F}_n$, and it follows the same three steps.

Bounds, continuity and monotonicity. We have $1/n \leq a(Z) \leq 1$, and the value is continuous and entrywise monotone in Z.

Merging duplicates. Duplicate weights merge by $p \mapsto Mp$, with M the class-indicator matrix of Lemma 13.

The large-scale limit. At I_k the unique minimizer of $\|p\|_2^2$ is $\mathbf{1}/k$. Hence every optimizing probability vector has positive entries for all sufficiently large scales, and $D_{\max}(Z_\theta(D)) \to k(D)$.

With these in hand the axioms follow as before. For monotonicity, if two ordered configurations have the same coincidence classes, then a strict decrease in a similarity matrix entry strictly decreases the

quadratic objective at such a positive optimizer; if their numbers of classes differ, the limits supply strictness. For uniqueness, adding an object preserves all old probability vectors by assigning it weight zero, which gives the weak inequality, and the distinct-object limits again make it strict on a tail. Integrating against the stated finite measure proves the two strict axioms, and continuity follows by dominated convergence using the bound $D_{\max} \le n$. □

D PROPERTIES OF THE DEFAULT DENSITY

The default density has an explicit tail, quantile and finite first moment:

$$\mu(d\theta) = \frac{2\,d\theta}{(1+\theta)^3}, \qquad \mu((T,\infty)) = \frac{1}{(1+T)^2}, \qquad \theta(v) = \frac{1}{\sqrt{1-v}} - 1. \tag{15}$$

Setting $T = 0$ proves normalization, and direct integration gives $\int_0^\infty \theta\,\mu(d\theta) = 1$. Every tail has positive mass, as required by the strictness proofs.

These three facts are what Section 5 appeals to. The finite first moment gives the stability bound below; the polynomial tail gives the novelty bound of Proposition 15; and the explicit quantile $\theta(v)$ is what turns the integral over $(0,\infty)$ into one over $[0,1]$, which is where the evaluation of Appendix E begins.

Stability The finite first moment gives a global continuity bound. Combining Lemma 11(ii) with $m \ge 1/n$ and $|e^{-\theta a} - e^{-\theta b}| \le \theta|a-b|$ gives

$$|\,\mathrm{Div}_\mu(A) - \mathrm{Div}_\mu(B)| \le n(n-1)\|A-B\|_{\max}. \tag{16}$$

This bound holds on the full dissimilarity domain, including configurations with duplicates.

Polynomial tails and small novelty Fine scales are needed to distinguish a new object that is very close to an existing one. The polynomial tail in Equation (15) gives these scales a quantitative amount of mass.

Proposition 15 (Quantitative uniqueness)**.** *Let $A, B \in \mathcal{D}_n$ be as in Axiom 2, let $k = k(B)$ and $\delta = \min_{j>1} a_{1j} > 0$. Then*

$$\mathrm{Div}_\mu(A) - \mathrm{Div}_\mu(B) \ge \frac{\delta^2}{2\big(\delta + \log(4k)\big)^2}. \tag{17}$$

Proof. Fix a scale and let P optimize the common configuration on objects $2,\dots,n$. Put $p = P\mathbf{1}$ and $h = g(Z_\theta(B)) \in [1,k]$. By Lemma 12, $P - pp^\top \succeq 0$. The old configuration has effective count h, so the new one should have effective count $h+1$; accordingly we give the new object weight $t = 1/(h+1)$ and rescale the rest. With this t, the matrix

$$\widetilde{P} = \begin{pmatrix} t^2 & t(1-t)p^\top \\ t(1-t)p & (1-t)^2 P \end{pmatrix}$$

is feasible for the new configuration. Its positive semidefiniteness follows by writing it as the outer product of $(t,(1-t)p)$ plus the positive semidefinite block $(1-t)^2(P-pp^\top)$. Its entries are nonnegative and sum to one. Let $\eta = e^{-\theta\delta}$. The new object's similarities to the old objects are at most η, so

$$m(Z_\theta(A)) \le \frac{1}{h+1} + \frac{2\eta h}{(h+1)^2}.$$

When $\theta \ge T := \log(4k)/\delta$, we have $\eta \le 1/(4k)$ and hence

$$m(Z_\theta(A)) \le \frac{1}{h+1} + \frac{h}{2k(h+1)^2} \le \frac{1}{h+1/2}.$$

The last inequality follows from $h(h+1/2) \le k(h+1)$, which holds because $h \le k$. Taking reciprocals gives $g(Z_\theta(A)) \ge h + 1/2$, while $g(Z_\theta(B)) = h$, so $g(Z_\theta(A)) - g(Z_\theta(B)) \ge 1/2$ for every $\theta \ge T$. The gap is nonnegative at every scale, so integrating over the tail $[T,\infty)$, which carries mass $\mu((T,\infty)) = (1+T)^{-2}$ by Equation (15), gives (17). □

For fixed k, this lower bound is quadratic in a small separation δ. An exponentially decaying scale tail would give an exponentially small lower bound through the same tail argument. The chosen density combines a polynomial tail with the finite first moment used in Equation (16). It also gives the explicit truncation bound

$$0 \le \mathrm{Div}_\mu(D) - \int_0^T g(Z_\theta(D))\,\mu(d\theta) \le \frac{n}{(1+T)^2}. \tag{18}$$

E Polynomial-time evaluation

We restate the theorem of Section 6 before proving it, with its main-text numbering.

Theorem 6 (Polynomial-time evaluation)**.** *Let $\mu(d\theta) = 2(1+\theta)^{-3}\,d\theta$. For rational $D \in \mathcal{D}_n$ of encoding length B and $0 < \varepsilon < 1$, the value $\mathrm{Div}_\mu(D)$ can be computed to additive accuracy ε in time polynomial in n, B and $1/\varepsilon$.*

Proof. Let D be rational with encoding length B, and let $0 < \varepsilon < 1$. The argument has four parts: the quadrature rule and its error, the rounding of the similarity matrix entries, the bit complexity of forming those entries, and the cost of the semidefinite solves.

Quadrature Using Equation (15), write

$$\mathrm{Div}_\mu(D) = \int_0^1 H(v)\,dv, \qquad H(v) = g(Z_\theta(D))\big|_{\theta=(1-v)^{-1/2}-1}.$$

Extend H by $H(0) = 1$ and $H(1) = k(D)$. Since H is nondecreasing, the midpoint rule of Section 6 applies, and its error bound

$$\left| \int_0^1 H(v)\,dv - \frac{1}{M}\sum_{j=0}^{M-1} H(v_j) \right| \le \frac{H(1)-H(0)}{M} \le \frac{n-1}{M}$$

follows by placing the midpoint sum between the left and right sums, whose difference telescopes to $(H(1)-H(0))/M$. Take $M = \lceil 2(n-1)/\varepsilon \rceil$ and evaluate each scale score to additive error $\varepsilon/2$.

Rounding the similarity matrix We give the precision details to account for the irrational similarity matrix entries. Approximate each entry of $Z_\theta(D)$ by a rational number with error at most $\tau = \varepsilon/(4n^2)$. Optimize this rational objective over $\mathcal{F}_n$ to value error at most τ. The true optimal value is then approximated to error at most 2τ. Clip this approximation to $[1/n, 1]$ and take its reciprocal. Clipping does not increase the error, and the reciprocal is n^2-Lipschitz on that interval. Thus the resulting scale score has error at most $\varepsilon/2$.

Bit complexity of the similarity matrix entries The largest node is $v_{M-1} = (M-1/2)/M$, so $1 - v_j \ge 1/(2M)$ for every j, and the quantile of Equation (15) gives

$$\theta(v_j) = \frac{1}{\sqrt{1-v_j}} - 1 \le \sqrt{2M} - 1.$$

Square roots, products and exponentials to the required absolute precision are computable with bit complexity polynomial in $B + \log M + \log(n/\varepsilon)$. For the scale computation, it suffices to approximate θ to error at most $\tau/(2\max(1, \max_{ij} d_{ij}))$. Large exponential arguments cause no difficulty for absolute error, since e^{-x} can be replaced by zero once a certified lower bound on x makes it smaller than the prescribed tolerance.

The semidefinite solves This semidefinite problem has a bounded feasible region and an explicit interior point relative to $\langle J, P\rangle = 1$:

$$P_0 = \frac{I_n + J}{n(n+1)}.$$

Its least eigenvalue and least entry are both $1/[n(n+1)]$. A Frobenius ball of radius $1/[2n(n+1)]$ around P_0 within the normalization hyperplane is feasible. Also every feasible P has $\|P\|_F \le 1$.

These are exactly the ingredients the ellipsoid method requires. A weak separation oracle for $\mathcal{F}_n$ is available in polynomial time: given P, test the entrywise condition directly, and for the semidefinite condition compute an approximate eigenvector v for the least eigenvalue, which yields the separating hyperplane $\langle vv^\top, \cdot \rangle \geq 0$ when $P \not\succeq 0$. Together with the inscribed ball around P_0 and the circumscribing radius 1, this places the problem in the standard framework of weak optimization over a convex body specified by such an oracle, which Grötschel et al. (1988) solve to objective error τ in time polynomial in n, the rational input length and $\log(1/\tau)$. There are $M = O(n/\varepsilon)$ solves, which proves the stated polynomial bound in n, B and $1/\varepsilon$. □

F Exactness for positive semidefinite similarities

We restate the theorem of Section 6 before proving it, with its main-text numbering. Corollary 8, whose proof follows, is not restated.

Theorem 7 (Exactness for positive semidefinite similarities)**.** *Let $Z \in \mathcal{S}_n$. If $Z \succeq 0$, then $g(Z) = D_{\max}(Z)$.*

Proof. Every $pp^\top$ with $p \in \Delta_n$ belongs to $\mathcal{F}_n$, so $m(Z) \leq \min_{p \in \Delta_n} p^\top Z p$. For the reverse inequality, take any $P \in \mathcal{F}_n$ and set $p = P\mathbf{1}$. By Lemma 12, $p \in \Delta_n$ and $P - pp^\top \succeq 0$. When $Z \succeq 0$,

$$\langle Z, P \rangle \geq \langle Z, pp^\top \rangle = p^\top Z p.$$

Minimizing and taking reciprocals proves $g(Z) = D_{\max}(Z)$. □

Proof of Corollary 8. Schoenberg's theorem (Schoenberg, 1938) states that a symmetric matrix D with zero diagonal satisfies $x^\top D x \leq 0$ whenever $\mathbf{1}^\top x = 0$ if and only if $(e^{-\theta d_{ij}})_{i,j}$ is positive semidefinite for every $\theta > 0$. Thus the hypothesis gives $Z_\theta(D) \succeq 0$ at every scale. Applying Theorem 7 and integrating against μ proves the claim. □

Proposition 16 (Testing negative type)**.** *Let $H = I - J/n$. Then D is of negative type, that is $x^\top D x \leq 0$ whenever $\mathbf{1}^\top x = 0$, if and only if $HDH \preceq 0$.*

Proof. H is the orthogonal projection onto $\mathbf{1}^\perp$, so as y ranges over $\mathbb{R}^n$ the vector $x = Hy$ ranges over $\mathbf{1}^\perp$, and $y^\top HDHy = x^\top Dx$. □

Practical caveats Testing one similarity matrix $Z_\theta(D)$ for positive semidefiniteness establishes convexity at that scale only. It does not establish negative type. When distances come from Euclidean embeddings, the condition is known analytically and no spectral test is needed. For numerical input without such a guarantee, eigenvalues near zero require an explicit tolerance or an error bound.

An alternative quadratic formulation. The implementation uses the nonnegative quadratic formulation below, normalizes its solution to the simplex, and refines it there when needed. This formulation remains valid for singular positive semidefinite matrices, including those with duplicates. We do not use the Cholesky-based least-squares reformulation, which requires positive definiteness.

Proposition 17 (Nonnegative least squares form)**.** *Let $Z \in \mathcal{S}_n$ with $Z \succeq 0$. Then*

$$D_{\max}(Z) = \max_{w \geq 0} \big(2\mathbf{1}^\top w - w^\top Z w \big).$$

If moreover Z is positive definite, with Cholesky factor L so that $Z = LL^\top$, this is equivalent to the nonnegative least squares problem $\min_{w \geq 0} \|L^\top w - L^{-1}\mathbf{1}\|_2^2$.

Proof. Write $w = sp$ with $p \in \Delta_n$ and $s \geq 0$. Optimizing the objective over s for fixed p gives $s = 1/(p^\top Z p)$ and value $1/(p^\top Z p)$, so maximizing over p returns $D_{\max}(Z)$. When $Z = LL^\top$ is positive definite, completing the square gives the stated least squares form. □

G Optimization formulas and numerical bounds

This appendix supports Section 7 in three parts. Appendix G.1 derives the scale-wise and integrated derivatives, and the directional derivatives that remain available at duplicate configurations. Appendix G.2 proves the screening bound obtained from a cached feasible matrix. Appendix G.3 collects the interval bounds used to certify a computed value. For brevity throughout, write $g_\theta(D) = g(Z_\theta(D))$.

G.1 Derivatives

Proposition 18 (Derivatives at a distinct configuration)**.** *Suppose all off-diagonal entries of D are positive, and let $i < j$.*

(i) Fix $\theta > 0$. If the minimizer $P_\theta^\star$ of $\langle Z_\theta(D), P\rangle$ over $\mathcal{F}_n$ is unique, then

$$\frac{\partial m(Z_\theta(D))}{\partial d_{ij}} = -2\theta e^{-\theta d_{ij}} (P_\theta^\star)_{ij},$$

and (10) follows by differentiating the reciprocal.

(ii) If in addition the minimizer is unique for μ-almost every θ and μ has finite first moment $\int_0^\infty \theta\, \mu(d\theta) < \infty$, then

$$\frac{\partial \operatorname{Div}_\mu(D)}{\partial d_{ij}} = \int_0^\infty 2\theta g_\theta(D)^2 (P_\theta^\star)_{ij} e^{-\theta d_{ij}}\, \mu(d\theta).$$

Proof. The feasible set $\mathcal{F}_n$ is compact and independent of D, and the objective is smooth in D, so Danskin's theorem (Danskin, 1967) gives (i); both symmetric matrix entries change with the coordinate d_{ij}. For (ii), the scale-wise difference quotients are bounded by an integrable multiple of $n^2\theta$, so dominated convergence permits differentiation under the integral. □

The first moment equals 1 for the default density. If the optimizer is not unique, the value has directional derivatives but an arbitrary optimizer need not yield an ordinary gradient. Approximate optimizers also require separate control of derivative error. A small value error alone does not supply it.

Directional derivatives at duplicate configurations Write

$$\mathcal{M}_\theta(D) = \arg\min_{P \in \mathcal{F}_n} \langle Z_\theta(D), P\rangle$$

for the set of minimizers at scale θ.

Proposition 19 (Directional derivative of the scale score)**.** *Let $D(t) = D + tH + o(t) \in \mathcal{D}_n$ as $t \downarrow 0$ be an admissible path, where H is symmetric with zero diagonal. Then*

$$g'_\theta(D; H) = 2\theta\, g_\theta(D)^2 \max_{P \in \mathcal{M}_\theta(D)} \sum_{i<j} e^{-\theta d_{ij}} P_{ij} H_{ij}.$$

Proof. Danskin's theorem applied to the family $\langle Z_\theta(D(t)), P\rangle$ over the fixed compact set $\mathcal{F}_n$, followed by differentiation of the reciprocal. □

Unlike Proposition 18, this requires neither uniqueness of the minimizing matrix nor absence of duplicates. The maximization is a linear semidefinite optimization over the optimal set, specified by

$$P \in \mathcal{F}_n, \qquad \langle Z_\theta(D), P\rangle = m(Z_\theta(D)).$$

Different directions can select different minimizing matrices, so these directional derivatives need not arise from a single ordinary gradient.

The bound $|g'_\theta(D; H)| \le n^2\theta \|H\|_{\max}$ and the finite first moment of μ also permit differentiation along the path under the integral. For numerical computation, the optimal-set constraint requires control of the optimal value. A small objective error from an approximate minimizing matrix alone does not guarantee a small derivative error. If the object-to-distance map is itself nondifferentiable at duplicates, its directional derivative must also be used.

G.2 Cached feasible matrices

Every feasible P_θ satisfies $m(Z_\theta(D')) \le \langle Z_\theta(D'), P_\theta \rangle$. Expanding the changed row and column gives (11). The denominator is positive because $\langle Z_\theta(D'), P_\theta \rangle \ge \mathrm{tr}(P_\theta) \ge 1/n$ by Lemma 11(i). Neither the inequality nor the linear update cost requires P_θ to be optimal. If it is optimal for D, the bound touches the current scale score at $D' = D$. After an accepted move, set $a_\theta \leftarrow a'_\theta$ before screening the next proposal with the same matrix.

For fixed positive quadrature weights w_r and scales θ_r, define $Q(D) = \sum_r w_r g_{\theta_r}(D)$. Then $\sum_r w_r / a'_r$ is a lower bound on $Q(D')$. If it exceeds an upper bound on $Q(D)$, the move improves Q. This does not on its own certify improvement of the integral. If $|Q(D) - \mathrm{Div}_\mu(D)| \le e_D$ and $|Q(D') - \mathrm{Div}_\mu(D')| \le e_{D'}$, a sufficient condition is

$$\sum_r \frac{w_r}{a'_r} - e_{D'} > U_Q(D) + e_D,$$

where $U_Q(D)$ is a valid upper bound on the current quadrature value.

G.3 Error bounds for numerical evaluation

For a positive semidefinite similarity matrix, let $p \in \Delta_n$ be any feasible probability vector, $u = p^\top Z p$ and $s = \min_i (Zp)_i$. Convexity gives, for every $q \in \Delta_n$,

$$q^\top Z q \ge u + 2(Zp)^\top (q - p) \ge 2s - u.$$

Hence, with $\ell = \max\{1/n, 2s - u\}$,

$$\ell \le m(Z) \le u, \qquad \frac{1}{u} \le g(Z) \le \frac{1}{\ell}. \tag{19}$$

For the general semidefinite problem, an upper bound on m comes from a feasible primal matrix. A lower bound ℓ comes from a decomposition $Z - \ell J = S + N$ with $S \succeq 0$ and $N \ge 0$, because $\langle Z, P \rangle = \ell + \langle S, P \rangle + \langle N, P \rangle \ge \ell$ for every $P \in \mathcal{F}_n$. The bound $1/n$ is always available.

Integration intervals can be obtained without a smoothness assumption. Let $F(v) = g_{\theta(v)}(D)$, extended by $F(0) = 1$ and $F(1) = k(D)$, and take a partition $0 = v_0 < \cdots < v_M = 1$. If $L_j \le F(v_j) \le U_j$, monotonicity implies

$$\sum_{j=0}^{M-1} (v_{j+1} - v_j) L_j \le \mathrm{Div}_\mu(D) \le \sum_{j=0}^{M-1} (v_{j+1} - v_j) U_{j+1}. \tag{20}$$

Refining this partition and the node intervals gives a valid comparison for the exact measure. Agreement between two Gauss rules is an empirical accuracy check and does not replace (20).

These are mathematical bounds in exact arithmetic. A rigorous floating-point certificate must also account for roundoff and primal or dual feasibility errors. The experiments evaluate (19) in standard double precision and therefore report numerical interval widths, rather than machine-verified enclosures.

H Experiment details

Shared conventions All three experiments use unsquared Euclidean distances and the default density $\rho(\theta) = 2/(1+\theta)^3$. The relaxation is exact for these inputs, so each scale is evaluated through the convex quadratic formulation. Distances are not normalized separately for each configuration.

H.1 The three experiments

Experiment 1: configurations Each configuration contains 16 points in $[0,1]^2$. The two-corner configuration places eight copies at each of $(0,0)$ and $(1,1)$. The four-corner configuration places four copies at each corner. The regular grid is $\{0, 1/3, 2/3, 1\} \times \{0, 1/3, 2/3, 1\}$. The almost-grid is obtained by moving $(1/3, 2/3)$ onto $(1/3, 1/3)$, leaving one empty site and one duplicated point. Exact duplicates are retained without perturbation.

Average and Energy are computed directly from (9), with Energy equal to $-\infty$ whenever a pair coincides. For Div_μ, we evaluate each configuration using 320 and 640 quadrature nodes and an inner score-interval tolerance of 10^{-9}. The figures report the 640-node scores. For the two-corner and four-corner configurations, symmetry gives the respective single-scale scores

$$\frac{2}{1+e^{-\sqrt{2}\theta}}, \qquad \frac{4}{1+2e^{-\theta}+e^{-\sqrt{2}\theta}}.$$

Independent adaptive integration of these expressions against ρ agrees with the reported scores to within 10^{-10}.

Experiment 2: initialization and optimization For each seed 0–24, NumPy's default random generator draws 16 independent uniform points in $[0,1]^2$. All three objectives use the same initial coordinates for each seed. We optimize the 32 coordinates with bound-constrained L-BFGS-B (Byrd et al., 1995), using SciPy's settings `maxiter=500`, `ftol=1e-11`, `gtol=1e-7`, and `maxls=40`. These are common stopping settings, not equal runtime budgets.

For Energy, the optimized objective is the harmonic mean of pairwise distances,

$$H(D) = -\frac{1}{\mathrm{Energy}(D)} = \left(\frac{2}{n(n-1)}\sum_{i<j}\frac{1}{d_{ij}}\right)^{-1}.$$

For configurations without duplicates, this is a strictly increasing transformation of Energy and has the same maximizers. At duplicate configurations, we set $H = 0$. This avoids infinite objective values at trial points. Reported Energy scores always use the original definition.

During optimization of Div_μ, we use 80 quadrature nodes and an inner score-interval tolerance of 10^{-8}. For distinct points, coordinate derivatives are computed from the scale-wise derivative formula and the chain rule. At exact duplicate trial points, the implementation uses bound-aware one-sided finite differences with step 10^{-5} and inner tolerance 10^{-9}. For Average, zero displacement contributes zero to the implemented derivative. For the harmonic-mean objective, a duplicate trial point is assigned a zero trial derivative. These conventions guide the numerical search and do not establish stationarity at nondifferentiable configurations.

The first three panels of Figure 2 use seed 0. The fourth shows the higher-scoring Div_μ configuration from seed 7. Displayed diversity scores are reevaluated with 320 and 640 nodes and inner tolerances at most 10^{-9}. The largest change across the displayed geometric configurations in Experiments 1 and 2 is 3.2×10^{-9}.

Experiment 2: additional checks The 25 random starts for Div_μ converge numerically to several different configurations. Ten runs reach nearly regular grids, with scores approximately 2.10771863 and 12 boundary points. Eleven reach higher-scoring nongrid configurations, with scores approximately 2.10855096 and 13 boundary points. The remaining four runs have 14 boundary points and lower scores than the best configuration found. Boundary counts here refer to coordinates lying exactly on the box constraints.

We also optimize five perturbed grids, obtained by adding independent Gaussian noise of standard deviation 0.03 to the grid coordinates and clipping to $[0,1]^2$. All five return to near-grid configurations. Selected configurations are further checked using finer quadrature, tighter optimization settings, and an alternative outer optimizer. Independent simplex quadratic solves using SLSQP agree with the production evaluator to within 1.1×10^{-13} at 640 nodes for the exact grid, a near-grid configuration, and a higher-scoring nongrid configuration.

For Energy, an independent implementation directly minimizes the mean reciprocal distance using SLSQP from the same 25 random initializations. Both this check and harmonic-mean L-BFGS-B produce configurations with 14–15 boundary points and 16 distinct points. These finite searches support the reported local behavior but provide no global-optimality guarantee.

Experiment 3: inputs For each $n \in \{100, 250, 500, 1000\}$ and seed 0–9, NumPy's default random generator draws n independent standard Gaussian vectors in $\mathbb{R}^{32}$. All coordinates are divided by the fixed constant $\sqrt{64}$. This sets a common reference scale across inputs, with expected squared pairwise distance equal to one.

H.2 Numerical protocol and environment

The inner solver and the diagnostics below are common to all three experiments. The quadrature settings and the timing protocol are stated for the experiment that uses them.

Inner solver and stopping rule At each quadrature node, we minimize

$$w^\top Z_\theta w - 2\mathbf{1}^\top w \quad \text{over } w \geq 0,$$

using L-BFGS-B and warm-starting from the preceding scale. The returned vector is normalized as $p = w/(\mathbf{1}^\top w)$. We then compute the numerical score interval from (19), using

$$u = p^\top Z_\theta p, \qquad \ell = \max\left\{\frac{1}{n},\, 2\min_i (Z_\theta p)_i - u\right\}.$$

Its lower and upper endpoints are $1/u$ and $1/\ell$.

If the interval is too wide, pairwise mass-transfer steps move probability from a positive coordinate with largest $(Z_\theta p)_j$ to a coordinate with smallest $(Z_\theta p)_i$. Each step uses exact line minimization along that direction. The solver stops when the score-interval width meets the requested tolerance and reports failure otherwise. Neither matrix inversion nor diagonal regularization is used.

Quadrature and numerical diagnostics Gauss–Legendre nodes on $(0, 1)$ are mapped to scales by $\theta(v) = (1 - v)^{-1/2} - 1$. The reported quadrature score is the weighted sum of the lower endpoints of the inner score intervals. The weighted inner width is the corresponding weighted sum of interval widths.

Experiment 3 uses 80 nodes for timing and 160 nodes for an untimed refinement check, both with inner tolerance 10^{-4}. Table 1 reports the largest absolute 80- versus 160-node score difference and the largest 80-node weighted inner width across the ten seeds.

Quadrature refinement measures observed numerical stability. It does not provide a rigorous bound on integration error. The inner intervals concern error from the quadratic solves and are evaluated in standard double precision. They are not machine-verified enclosures accounting for all rounding errors.

Timing protocol and environment Each input in Experiment 3 is evaluated three times. We take the median runtime for each seed, then report the median and range of these ten per-seed medians. Thus each table row summarizes 30 timed evaluations.

An untimed warm-up precedes the benchmark. Quadrature nodes are cached. Each timed evaluation constructs the pairwise distances, forms every exponential similarity matrix, and performs all inner optimization steps. Input generation, the 160-node refinement check, and writing result files are excluded. Inner solutions are warm-started across scales within an evaluation, but are not reused across timing repetitions.

The scaling benchmark uses Linux, an AMD EPYC 9V74 processor, Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0, and OpenBLAS 0.3.30. BLAS is restricted to one thread, and no GPU is used.

H.3 Reproduction

The accompanying code provides `diversity.py` for evaluation, `diagnostics.py` for Experiment 1, `optimize.py` for Experiment 2, and `scaling.py` for Experiment 3. The scripts `verify_diversity.py` and `verify_energy.py` implement the additional optimization checks. The script `report.py` generates figures and table summaries from the recorded results. Dependencies are specified in `requirements.txt`. Raw results include coordinates, seeds, solver settings, numerical diagnostics, and environment metadata.