## Todo notes

**Entropy and moment details.** Differentiating the Dirichlet beta integral gives $\mathbb{E}[\log \Lambda_i] = \psi(\alpha_i) - \psi(A)$, where $A = \sum_i \alpha_i$. Substitution into $-\mathbb{E}[\log f_Y(Y)]$ yields the displayed entropy; its log moments are finite for every $\alpha_i > 0$. The same beta integral gives $\mathbb{E}[\Lambda_i] = \alpha_i/A$ and $\mathrm{Cov}(\Lambda_i, \Lambda_j) = \alpha_i \mathbf{1}_{i=j}/(A(A+1)) - \alpha_i\alpha_j/(A^2(A+1))$. With $Y = (\Lambda_1, \dots, \Lambda_d)^\top$ and $\Lambda_0 = 1 - \sum_{i=1}^d \Lambda_i$, we have $X_T = \sum_i v_i \Lambda_i = v_0 + E_d Y$, giving the stated mean and covariance. The full barycentric covariance is singular because the coordinates sum to one; the free-coordinate covariance is positive definite. . . . . 6

**Affine product coordinates.** The tangent and normal directions are orthogonal, while the simplex lies in the affine space $v_0+T$. Translate by $v_0$ before representing the law as a product on $T \times N$. Translation does not change covariance or relative entropy. . . . . 6

**Uniqueness of the Gaussian minimizer.** For fixed $\Sigma \succ 0$, set $B = \Sigma^{1/2}\widetilde{\Sigma}^{-1}\Sigma^{1/2}$. Then $\log\det\widetilde{\Sigma} + \mathrm{tr}(\widetilde{\Sigma}^{-1}\Sigma) = \log\det\Sigma + \mathrm{tr}(B) - \log\det B$. If $\lambda_i$ are the eigenvalues of $B$, the varying part is $\sum_i(\lambda_i - \log\lambda_i)$, uniquely minimized at $\lambda_i = 1$. Thus $B = I$ and $\widetilde{\Sigma} = \Sigma$. The mean term is uniquely minimized at $m = m_T$. Positive definiteness of $\Sigma$ follows from positive tangent covariance and positive-definite normal noise on the orthogonal space. . . . . 7

**Reference measure and Jacobian.** $H(\alpha_\sharp)$ is entropy relative to Lebesgue measure in the first $d$ barycentric coordinates. The affine map to the simplex multiplies $d$-dimensional Hausdorff measure by $J = d!V_d$, so entropy changes by $\log J$. KL is invariant because both densities transform by the same Jacobian; the $\log J$ term here converts the intrinsic entropy only. . . . . 7

**Determinant calculation.** With $A = \|\alpha_\sharp\|_1$ and $a = \alpha_\natural$, the matrix determinant lemma gives $\det(A\,\mathrm{diag}(a) - aa^\top) = A^d \prod_{i=1}^d \alpha_i\big(1 - a^\top(A\,\mathrm{diag}(a))^{-1}a\big) = A^{d-1}\prod_{i=0}^d \alpha_i$. Dividing by $[A^2(A+1)]^d$ gives $\det\mathrm{Cov}(Y) = \prod_i \alpha_i/[A^{d+1}(A+1)^d]$. The affine covariance map multiplies this determinant by $\det(E)^2 = (d!V_d)^2$. . . 7

**Uniform-case calculation.** With $k = d+1$ and $\alpha_i = 1$, $A = k$ and $\mathrm{Cov}(Y) = (kI_d - 1_d 1_d^\top)/(k^2(k+1))$. The coordinate simplex has volume $1/d!$, hence $H(1_{d+1}) = -\log(d!)$. Substitution into $C_{d,\alpha}$ gives $\frac{d}{2}\log(2\pi e) + \log(d!) - \frac{k}{2}\log k - \frac{d}{2}\log(k+1)$, which is the logarithm displayed below. This records the entropy's coordinate measure explicitly. . . . . 8

# GAUSSIAN MIXTURE APPROXIMATION FOR EMBEDDED SIMPLICES: DIMENSION- AND PARAMETER-DEPENDENT KULLBACK–LEIBLER BOUNDS

AUTHORSHIP PENDING

Abstract. We derive a geometry-independent Gaussian approximation for distributions near embedded simplices, combining arbitrary-parameter Dirichlet distributions with independent Gaussian noise in orthogonal normal spaces. The optimal ambient Gaussian matches the mean and covariance, yielding an explicit minimum forward KL divergence depending only on intrinsic dimension and Dirichlet parameters. For finite mixtures with shared weights, we prove a global upper bound given by the weighted sum of component-wise divergences. In the uniform case, this bound depends only on intrinsic dimension, independently of ambient dimension, simplex geometry, and normal covariances.

DRAFT

## 1. Introduction

Geometry and probability theory jointly describe physical systems and their chance of being is specific states, appearing naturally in geometric mechanics, computer graphics, and stochastic control [1, 2, 3, 4, 5]. Simplicial complexes [6] provide standard representations for quite general manifolds [6, 7], while Gaussian mixtures can represent probability measures of enough regularity and are commonly fitted using *Expectation-Maximization* (EM) [8, 9, 10, 11, 12]. When geometry is known, it can be exploited to arrive at a more efficient and tighter approximation bounds [13].

For embedded simplices, intrinsic and ambient dimensions differ, requiring full-dimensional Gaussian approximations to account for both tangential geometry and normal uncertainty. Related approaches include the Gaussian Mesh Renderer and geometry-aware mixtures for plane curves [14, 15], with applications in rendering [14], tracking, obstacle modeling, and trajectory planning [16]. However, rigorous measure theoretic approximation guarantees remain essential.

Our previous work derived closed-form Gaussian mixtures for uniform distributions on triangulated full-dimensional domains, with a dimension-dependent KL bound [13]. Here, we address the complementary lower-dimensional setting: a Dirichlet-distributed measure on an embedded simplex. We derive the optimal ambient Gaussian and its minimum forward KL divergence for arbitrary positive Dirichlet parameters, showing independence from the simplex embedding, geometry, scale, and normal covariance. We further establish a global KL bound for finite mixtures with shared component weights.

*Date*: November 4, 2026.

2020 *Mathematics Subject Classification.* To be determined.

*Key words and phrases.* To be determined.

Funding information to be confirmed.

**How to read the result.** The construction and the closed-form Gaussian mixture parameters require only finite-dimensional linear algebra [17, 18], numerical linear algebra [19, 20], calculus [21], and basic probability [22]. For analysis and calculus background, see Tao [23, 24]. In particular, one can use the formulas in Theorem 3.1 and Theorem 3.2 without studying measure theory. The proofs use density transformations, product probability laws, and convexity of relative entropy. Readers who want the measure-theoretic details can consult Tao [25] or Pollard [26]; the information-geometric interpretation of KL minimization is developed by Ay et al. [27]. The added normal Gaussian noise makes each distribution considered in the KL calculations have a density in the ambient space.

## 2. Ambient degeneracy and normal-noise lifting

Let $\sigma^d = \operatorname{conv}\{v_0, \dots, v_d\} \subset \mathbb{R}^D$ be a nondegenerate embedded simplex with $d < D$. Write

$$E_d = [v_1 - v_0, \dots, v_d - v_0] \in \mathbb{R}^{D\times d}, \qquad Y = (\Lambda_1, \dots, \Lambda_d)^\top,$$

where $\Lambda \sim \operatorname{Dirichlet}(\alpha_\sharp)$ for some $\alpha_\sharp \in \mathbb{R}^{d+1}_{++}$ and $\Lambda_0 = 1 - \sum_{i=1}^d \Lambda_i$. Then

$$X_T = \mathbf{V}\Lambda = v_0 + E_d Y.$$

The free-coordinate covariance $C_Y = \operatorname{Cov}(Y)$ is positive definite: for every nonzero $a \in \mathbb{R}^d$, $a^\top C_Y a = \operatorname{Var}(a^\top Y) > 0$, since the Dirichlet density is positive on the interior of the coordinate simplex and $a^\top y$ is not constant there. Consequently

$$\Sigma_T := \operatorname{Cov}(X_T) = E_d C_Y E_d^\top$$

has rank $d$. Indeed, $E_d$ has full column rank, so $\operatorname{rank}(E_d C_Y E_d^\top) = d$. In particular, $\Sigma_T$ is singular as a $D \times D$ ambient covariance when $d < D$; its nullspace is the normal space $N = \operatorname{range}(E_d)^\perp$.

The Gaussian with mean $\mathbb{E}[X_T]$ and covariance $\Sigma_T$ is therefore a degenerate Gaussian measure supported on the affine hull $v_0 + T$. It has no density with respect to $D$-dimensional Lebesgue measure. More strongly, the simplex-supported law $P_T$ is singular with respect to every nondegenerate ambient Gaussian $Q$: the affine hull has $P_T$-probability one and $Q$-probability zero. Hence $P_T$ is not absolutely continuous with respect to $Q$, and $D_{\mathrm{KL}}(P_T : Q) = +\infty$. The degenerate Gaussian can instead be considered on the affine hull with its intrinsic $d$-dimensional reference measure, but that is a different KL comparison from the ambient one used here.

To model uncertainty away from the simplex and obtain a full-dimensional ambient density, add an independent normal Gaussian $X_N \in N$ with covariance $\Sigma_N$ positive definite on $N$. The resulting covariance is

$$\operatorname{Cov}(X_T + X_N) = \Sigma_T + \Sigma_N \succ 0.$$

To see this, decompose any $z \in \mathbb{R}^D$ uniquely as $z = z_T + z_N$ with $z_T \in T$ and $z_N \in N$. Then

$$z^\top(\Sigma_T + \Sigma_N)z = z_T^\top \Sigma_T z_T + z_N^\top \Sigma_N z_N,$$

which is positive for every nonzero $z$, since $\Sigma_T$ is positive definite on $T$ and $\Sigma_N$ is positive definite on $N$. The added noise leaves the tangential mean and covariance unchanged while supplying positive variance in every missing normal direction. It also makes the target law itself full-dimensional; the subsequent KL bound is for this noise-augmented model, not for the original simplex-supported singular law.

DRAFT

## 3. Component and Mixture Construction

We first fix the geometric and probabilistic notation. Vectors are column vectors, $I_D$ denotes the identity matrix on $\mathbb{R}^D$, and $A^\top$ denotes the transpose of a matrix $A$. For a symmetric positive-definite matrix $S$, $\mathcal{N}_D(m, S)$ is the Gaussian distribution on $\mathbb{R}^D$ with mean $m$ and covariance $S$.

**Definition 3.1** (Simplex and barycentric coordinates)**.** *Given points $v_0, \dots, v_d \in \mathbb{R}^D$, let $\mathbf{V} = [v_0, \dots, v_d]$. Their convex hull is*

$$\operatorname{conv}\{v_0, \dots, v_d\} = \{\mathbf{V}\lambda : \lambda \in \mathbb{R}^{d+1}_+,\ \|\lambda\|_1 = 1\}.$$

*If the $d$ vectors $v_1 - v_0, \dots, v_d - v_0$ are linearly independent, this convex hull is a nondegenerate $d$-simplex. The coefficients $\lambda_i$ are its barycentric coordinates. Its intrinsic dimension is $d$, even when it lies in an ambient space $\mathbb{R}^D$ with $D > d$.*

**Definition 3.2** (Triangulation and intrinsic volume)**.** *A finite triangulation of a region or surface is a finite collection of simplices whose union is that set and whose interiors do not overlap; neighboring simplices may share faces such as edges or vertices. The $d$-dimensional volume $\mathrm{Vol}_d$ of a $d$-simplex is its ordinary length, area, or volume within its own $d$-dimensional affine hull. Thus the weight assigned to a simplex in a uniform distribution over a triangulated object is its $d$-volume divided by the sum of the $d$-volumes.*

**Definition 3.3** (Barycentric subdivision)**.** *For each nonempty face $F$ of a simplex, its barycenter is the average of its vertices. The barycentric subdivision replaces the simplex by smaller simplices whose vertices are barycenters of a nested sequence of faces. More precisely, a $d$-simplex with vertex matrix $\mathbf{V} = [v_0, \dots, v_d]$ is subdivided into one $d$-simplex for each ordering $i_0, \dots, i_d$ of $0, \dots, d$, with vertices*

$$b_{\{i_0\}},\ b_{\{i_0,i_1\}},\ \dots,\ b_{\{i_0,\dots,i_d\}}, \qquad b_I := \frac{\mathbf{V}1_I}{|I|}.$$

*Here $1_I \in \{0,1\}^{d+1}$ is the indicator vector of $I$. These smaller simplices cover the original simplex and have disjoint interiors. Applying this construction to every simplex in a triangulation gives a finer triangulation and preserves the original geometric domain.*

To refine only a selected part of a mesh, one may subdivide the selected simplices and leave the others unchanged, provided the subdivision agrees on every shared face. In practice, this means also refining neighboring simplices as needed so that both sides induce the same subdivision of their common face; otherwise the result may fail to be a triangulation. Repeating the barycentric subdivision produces progressively finer meshes. The Gaussian-mixture construction can then be applied to the resulting simplices, using their own vertices and intrinsic volumes to compute component parameters and weights. This gives a direct way to allocate more components to selected regions, although the dimension-dependent bound proved here does not by itself guarantee that local refinement reduces the actual KL error.

**Definition 3.4** (Tangent and normal spaces)**.** *For a nondegenerate simplex $\sigma = \operatorname{conv}\{v_0, \dots, v_d\}$, its tangent space is the linear space of directions parallel to its affine hull,*

$$T = \operatorname{span}\{v_i - v_0 : i = 1, \dots, d\}.$$

DRAFT

*Its normal space is* $N = T^\perp = \{z \in \mathbb{R}^D : z^\top t = 0 \text{ for every } t \in T\}$. *Every vector in* $\mathbb{R}^D$ *has a unique orthogonal decomposition into a tangent and a normal vector. A covariance on* $N$ *describes random displacement perpendicular to the simplex; when represented as a* $D \times D$ *matrix, it is zero on* $T$.

**Definition 3.5** (Dirichlet distribution and uniform simplex sampling)**.** *Write* $\alpha_\sharp = (\alpha_0, \dots, \alpha_d)^\top \in \mathbb{R}^{d+1}_{++}$ *for the full parameter vector and* $\alpha_\natural = (\alpha_1, \dots, \alpha_d)^\top \in \mathbb{R}^d_{++}$ *for the subvector associated with the* $d$ *free barycentric coordinates. Let* $1_{d+1} := (1, \dots, 1)^\top \in \mathbb{R}^{d+1}$. *The Dirichlet distribution* $\mathrm{Dirichlet}(\alpha_\sharp)$ *is a probability distribution on vectors* $\lambda \in \mathbb{R}^{d+1}_+$ *with* $\|\lambda\|_1 = 1$. *When* $\alpha_\sharp = 1_{d+1}$, *it is uniform over this coordinate simplex. Thus, if* $\Lambda \sim \mathrm{Dirichlet}(1_{d+1})$, *then* $X_T = \mathbf{V}\Lambda$ *is uniform with respect to* $d$*-dimensional volume on* $\sigma$.

**Definition 3.6** (Forward Kullback–Leibler divergence)**.** *If probability distributions* $P$ *and* $Q$ *have densities* $p$ *and* $q$ *with respect to the same volume measure, their forward Kullback–Leibler (KL) divergence is the relative entropy of* $p$ *with respect to* $q$ *[26, 27]:*

$$D_{\mathrm{KL}}(p : q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x,$$

*with the convention that it is infinite if* $q = 0$ *on a set where* $p > 0$. *The order matters: throughout this paper,* $p$ *is the target density and* $q$ *its Gaussian approximation. The divergence is nonnegative and is zero only when the two densities agree (up to sets of volume zero).*

In this paper, an *information projection onto Gaussians* means a Gaussian density $q$ that minimizes the forward KL divergence from a specified target density $p$. This is the projection of a probability model onto a chosen family of models under an information-geometric divergence [27]. The *mean* of a random vector is its expectation; its *covariance* is $\mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)^\top]$. A Gaussian with the same mean and covariance is called the moment-matched Gaussian. We prove that it is the information projection for each component considered here.

For a reader who only wants to construct the approximation, the procedure is direct: for each simplex and its Dirichlet parameters, use the weighted vertex mean and covariance formulas below; add the chosen normal covariance; and retain the target component weight. For a uniform distribution over a triangulated surface or curve, set all Dirichlet parameters to one and choose weights proportional to the simplices' intrinsic volumes. No isomorphism or algebraic-topology background is needed.

**Definition 3.7** (Dirichlet–Gaussian Distribution on an Embedded Simplex)**.** *Let* $\sigma^d = \mathrm{conv}\{v_0, \dots, v_d\} \subset \mathbb{R}^D$ *be a nondegenerate simplex with* $d < D$. *Its tangent and normal spaces are*

$$T = \mathrm{span}\{v_i - v_0\}_{i=1}^d, \qquad N = T^\perp.$$

*Let* $\Lambda = (\Lambda_0, \dots, \Lambda_d) \sim \mathrm{Dirichlet}(\alpha_\sharp)$ *for* $\alpha_\sharp \in \mathbb{R}^{d+1}_{++}$, *and define the simplex-valued random variable*

$$X_T = \mathbf{V}\Lambda \in \sigma^d \subset v_0 + T.$$

*Independently, let* $X_N \sim \mathcal{N}_N(0, \Sigma_N)$ *take values in* $N$, *with* $\Sigma_N$ *positive definite on* $N$; $\mathcal{N}_N$ *denotes an ordinary Gaussian in orthonormal coordinates on* $N$. *The Dirichlet–Gaussian random variable is* $X = X_T + X_N$.



DRAFT

Write $\mathcal{G}_D$ for the family of Gaussian densities on $\mathbb{R}^D$ with positive-definite covariance matrices. A Gaussian mixture is a finite weighted sum of Gaussian densities, where the nonnegative weights sum to one; probabilistically, it represents selecting a component at random according to those weights and then drawing from that component.

**Theorem 3.1** (Gaussian information projection for arbitrary Dirichlet parameters)**.** *Let $p$ be the density of $X = X_T + X_N$ from a nondegenerate $d$-simplex $\sigma^d = \operatorname{conv}\{v_0, \dots, v_d\} \subset \mathbb{R}^D$, $d < D$, with parameter vector $\alpha_\sharp = (\alpha_0, \dots, \alpha_d)^\top \in \mathbb{R}^{d+1}_{++}$ and independent $X_N \sim \mathcal{N}_N(0, \Sigma_N)$, where $\Sigma_N \succ 0$ on $N$. Write $\alpha_\natural = (\alpha_1, \dots, \alpha_d)^\top$ and let $E_d = [v_1 - v_0, \dots, v_d - v_0] \in \mathbb{R}^{D\times d}$. The unique minimizer over nondegenerate Gaussian densities $q \in \mathcal{G}_D$ of $D_{\mathrm{KL}}(p : q)$ is the density $q^*$ of the Gaussian*

$$\mathcal{N}_D(m_T, \Sigma_T + \Sigma_N),$$

*where*

$$m_T = \mathbb{E}[X_T] = v_0 + E_d \frac{\alpha_\natural}{\|\alpha_\sharp\|_1}, \qquad \Sigma_T = \operatorname{Cov}(X_T) = E_d \frac{\|\alpha_\sharp\|_1 \operatorname{diag}(\alpha_\natural) - \alpha_\natural \alpha_\natural^\top}{\|\alpha_\sharp\|_1^2(\|\alpha_\sharp\|_1 + 1)} E_d^\top.$$

*Here $\Sigma_T$ vanishes on $N$, and $\Sigma_N$ is extended by zero on $T$. The minimum divergence is*

$$C_{d,\alpha_\sharp} := \frac{d}{2}\log(2\pi e) + \frac{1}{2}\left(\log \prod_{i=0}^{d} \alpha_i - (d+1)\log\|\alpha_\sharp\|_1 - d\log(\|\alpha_\sharp\|_1 + 1)\right) - H(\alpha_\sharp),$$

*where $H(\alpha_\sharp)$ is the differential entropy of the Dirichlet density expressed in the first $d$ barycentric coordinates (the last coordinate is one minus their sum):*

$$H(\alpha_\sharp) = \log\left(\frac{\prod_{i=0}^{d}\Gamma(\alpha_i)}{\Gamma(\|\alpha_\sharp\|_1)}\right) + (\|\alpha_\sharp\|_1 - d - 1)\psi(\|\alpha_\sharp\|_1) - (\alpha_\sharp - 1_{d+1})^\top \psi(\alpha_\sharp),$$

*where $\psi$ acts componentwise on vectors. Here $\Gamma$ is the gamma function and $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ is the digamma function; the Dirichlet entropy formula is standard [28]. The constant $C_{d,\alpha_\sharp}$ is independent of the embedding dimension, simplex geometry, and normal covariance.*

*Proof.*

**Entropy and moment details.** Differentiating the Dirichlet beta integral gives $\mathbb{E}[\log \Lambda_i] = \psi(\alpha_i) - \psi(A)$, where $A = \sum_i \alpha_i$. Substitution into $-\mathbb{E}[\log f_Y(Y)]$ yields the displayed entropy; its log moments are finite for every $\alpha_i > 0$. The same beta integral gives $\mathbb{E}[\Lambda_i] = \alpha_i/A$ and $\operatorname{Cov}(\Lambda_i, \Lambda_j) = \alpha_i \mathbf{1}_{i=j}/(A(A+1)) - \alpha_i\alpha_j/(A^2(A+1))$. With $Y = (\Lambda_1, \dots, \Lambda_d)^\top$ and $\Lambda_0 = 1 - \sum_{i=1}^d \Lambda_i$, we have $X_T = \sum_i v_i\Lambda_i = v_0 + E_d Y$, giving the stated mean and covariance. The full barycentric covariance is singular because the coordinates sum to one; the free-coordinate covariance is positive definite.

**Affine product coordinates.** The tangent and normal directions are orthogonal, while the simplex lies in the affine space $v_0 + T$. Translate by $v_0$ before representing the law as a product on $T \times N$. Translation does not change covariance or relative entropy.

Use orthonormal coordinates for the affine decomposition into the tangent and normal spaces. In these coordinates, $P = P_T \otimes P_N$, where $P_T$ is the affine image of the stated Dirichlet distribution and $P_N = \mathcal{N}_N(0, \Sigma_N)$. The mean and covariance of $P$ are $m_T$ and

$\Sigma = \Sigma_T + \Sigma_N$, which is positive definite on $\mathbb{R}^D$. For a candidate Gaussian $\mathcal{N}_D(m, \widetilde{\Sigma})$, the KL divergence is $-h(P)$ plus

$$\frac{D}{2}\log(2\pi) + \frac{1}{2}\log\det\widetilde{\Sigma} + \frac{1}{2}\operatorname{tr}(\widetilde{\Sigma}^{-1}\Sigma) + \frac{1}{2}(m_T - m)^\top\widetilde{\Sigma}^{-1}(m_T - m).$$

This expression is uniquely minimized at $m = m_T$ and $\widetilde{\Sigma} = \Sigma$, proving the stated information projection.

**Uniqueness of the Gaussian minimizer.** For fixed $\Sigma \succ 0$, set $B = \Sigma^{1/2}\widetilde{\Sigma}^{-1}\Sigma^{1/2}$. Then $\log\det\widetilde{\Sigma} + \operatorname{tr}(\widetilde{\Sigma}^{-1}\Sigma) = \log\det\Sigma + \operatorname{tr}(B) - \log\det B$. If $\lambda_i$ are the eigenvalues of $B$, the varying part is $\sum_i(\lambda_i - \log\lambda_i)$, uniquely minimized at $\lambda_i = 1$. Thus $B = I$ and $\widetilde{\Sigma} = \Sigma$. The mean term is uniquely minimized at $m = m_T$. Positive definiteness of $\Sigma$ follows from positive tangent covariance and positive-definite normal noise on the orthogonal space.

In the orthogonal coordinates, both $P$ and $Q^*$ share the same normal factor $P_N$. Invariance of relative entropy under the coordinate isometry and its additivity for product measures therefore give [26]:

$$D_{\mathrm{KL}}(p : q^*) = D_{\mathrm{KL}}(p_T : q_T^*),$$

where $Q_T^*$ is the intrinsic Gaussian on the affine hull with mean $m_T$ and covariance $\Sigma_T|_T$. Let $V_d$ be the $d$-volume of the simplex. The affine map from the coordinate simplex to this simplex has volume factor $d!V_d$, so the differential entropy of $X_T$ is $H(\alpha_\sharp) + \log(d!V_d)$. Moment matching implies

$$D_{\mathrm{KL}}(p_T : q_T^*) = \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log\det_T(\Sigma_T) - H(\alpha_\sharp) - \log(d!V_d).$$

**Reference measure and Jacobian.** $H(\alpha_\sharp)$ is entropy relative to Lebesgue measure in the first $d$ barycentric coordinates. The affine map to the simplex multiplies $d$-dimensional Hausdorff measure by $J = d!V_d$, so entropy changes by $\log J$. KL is invariant because both densities transform by the same Jacobian; the $\log J$ term here converts the intrinsic entropy only.

Let $E \in \mathbb{R}^{d\times d}$ be the edge matrix of the simplex in an orthonormal basis of $T$, and let $\Sigma_\lambda$ be the covariance matrix of the first $d$ barycentric coordinates. The affine map gives $\det_T(\Sigma_T) = \det(E)^2\det(\Sigma_\lambda)$. The Dirichlet covariance formula yields

$$\det(\Sigma_\lambda) = \frac{\prod_{i=0}^d \alpha_i}{\|\alpha_\sharp\|_1^{d+1}(\|\alpha_\sharp\|_1 + 1)^d}, \qquad \det_T(\Sigma_T) = (d!V_d)^2\frac{\prod_{i=0}^d \alpha_i}{\|\alpha_\sharp\|_1^{d+1}(\|\alpha_\sharp\|_1 + 1)^d}.$$

**Determinant calculation.** With $A = \|\alpha_\sharp\|_1$ and $a = \alpha_\natural$, the matrix determinant lemma gives $\det(A\operatorname{diag}(a) - aa^\top) = A^d\prod_{i=1}^d \alpha_i\big(1 - a^\top(A\operatorname{diag}(a))^{-1}a\big) = A^{d-1}\prod_{i=0}^d \alpha_i$. Dividing by $[A^2(A+1)]^d$ gives $\det\operatorname{Cov}(Y) = \prod_i \alpha_i/[A^{d+1}(A+1)^d]$. The affine covariance map multiplies this determinant by $\det(E)^2 = (d!V_d)^2$.

Substitution gives $C_{d,\alpha_\sharp}$. The dependence on $V_d$ cancels, and the normal covariance has already canceled through the common normal factor. □

**Uniform-case calculation.** With $k = d+1$ and $\alpha_i = 1$, $A = k$ and $\mathrm{Cov}(Y) = (kI_d - 1_d 1_d^\top)/(k^2(k+1))$. The coordinate simplex has volume $1/d!$, hence $H(1_{d+1}) = -\log(d!)$. Substitution into $C_{d,\alpha}$ gives $\frac{d}{2}\log(2\pi e) + \log(d!) - \frac{k}{2}\log k - \frac{d}{2}\log(k+1)$, which is the logarithm displayed below. This records the entropy's coordinate measure explicitly.

**Corollary 3.1** (Uniform distribution on a simplex)**.** *Under the assumptions of Theorem 3.1, if $\alpha_\sharp = 1_{d+1}$, then $X_T$ is uniform on the simplex. With $1_d = (1,\dots,1)^\top \in \mathbb{R}^d$, its mean and covariance are*

$$m_T = v_0 + E_d \frac{1_d}{d+1}, \qquad \Sigma_T = E_d \frac{(d+1)I_d - 1_d 1_d^\top}{(d+1)^2(d+2)} E_d^\top.$$

*The minimum divergence becomes*

$$C_d = C_{d,1_{d+1}} = \log\left(\frac{(2\pi e)^{d/2} d!}{(d+1)^{(d+1)/2}(d+2)^{d/2}}\right),$$

*which depends only on $d$.*

**Theorem 3.2** (Gaussian-mixture bound for arbitrary Dirichlet parameters)**.** *Let $\{\sigma_j^{d_j}\}_{j=1}^M$ be a finite family of nondegenerate simplices embedded in $\mathbb{R}^D$, with $d_j < D$. For component $j$, let $\mathbf{V}_j = [v_{j,0},\dots,v_{j,d_j}]$, $E_j = [v_{j,1} - v_{j,0},\dots,v_{j,d_j} - v_{j,0}]$, and $\alpha_{\sharp,j} = (\alpha_{j,0},\dots,\alpha_{j,d_j})^\top \in \mathbb{R}^{d_j+1}_{++}$; set $\alpha_{\natural,j} = (\alpha_{j,1},\dots,\alpha_{j,d_j})^\top$. Let $\Lambda_j \sim \mathrm{Dirichlet}(\alpha_{\sharp,j})$ and $X_{T,j} = \mathbf{V}_j \Lambda_j$. Independently, let $X_{N,j} \sim \mathcal{N}_{N_j}(0, \Sigma_{N,j})$, and write $p_j$ for the density of $X_j = X_{T,j} + X_{N,j}$. Define*

$$m_{T,j} = \mathbb{E}[X_{T,j}] = v_{j,0} + E_j \frac{\alpha_{\natural,j}}{\|\alpha_{\sharp,j}\|_1}, \qquad \Sigma_{T,j} = \mathrm{Cov}(X_{T,j}) = E_j \frac{\|\alpha_{\sharp,j}\|_1 \operatorname{diag}(\alpha_{\natural,j}) - \alpha_{\natural,j}\alpha_{\natural,j}^\top}{\|\alpha_{\sharp,j}\|_1^2(\|\alpha_{\sharp,j}\|_1 + 1)} E_j^\top,$$

*where $\Sigma_{T,j}$ vanishes on $N_j$. Let $\mathbf{w} = (w_1,\dots,w_M)^\top \in \mathbb{R}^M_{++}$ with $\|\mathbf{w}\|_1 = 1$. Define $\mathbf{p}(x) = (p_1(x),\dots,p_M(x))^\top$ and let $q_j$ be the Gaussian density with mean $m_{T,j}$ and covariance $\Sigma_{T,j} + \Sigma_{N,j}$, where $\Sigma_{N,j}$ is extended by zero on $T_j$; set $\mathbf{q}(x) = (q_1(x),\dots,q_M(x))^\top$ and $\mathbf{c} = (C_{d_1,\alpha_{\sharp,1}},\dots,C_{d_M,\alpha_{\sharp,M}})^\top$. Let*

$$p(x) = \mathbf{w}^\top \mathbf{p}(x), \qquad q(x) = \mathbf{w}^\top \mathbf{q}(x).$$

*Then $q$ is a Gaussian mixture and*

$$D_{\mathrm{KL}}(p : q) \le \mathbf{w}^\top \mathbf{c}.$$

*If all components have common dimension $d$ and uniform parameter vectors $\alpha_{\sharp,j} = 1_{d+1}$, this reduces to $D_{\mathrm{KL}}(p : q) \le C_d$. For a uniform distribution over a same-dimensional triangulation whose simplex interiors are disjoint up to $d$-dimensional measure zero, the weights are proportional to the component $d$-volumes.*

*Proof.* The underlying log-sum inequality is given, for example, in Cover and Thomas, Section 2.7, Theorem 2.7.1, p. 31 [29]. For each $x$, apply the log-sum inequality to $a_j = w_j p_j(x)$ and $b_j = w_j q_j(x)$:

$$\left(\sum_j w_j p_j(x)\right) \log \frac{\sum_j w_j p_j(x)}{\sum_j w_j q_j(x)} \le \sum_j w_j p_j(x) \log \frac{w_j p_j(x)}{w_j q_j(x)} = \sum_j w_j p_j(x) \log \frac{p_j(x)}{q_j(x)}.$$

DRAFT

Integrating and using $\int p_j(x)\,\mathrm{d}x = 1$ gives

$$D_{\mathrm{KL}}(p : q) \le \sum_j w_j D_{\mathrm{KL}}(p_j : q_j) \le \sum_j w_j C_{d_j,\alpha_{\sharp,j}} = \mathbf{w}^\top \mathbf{c},$$

where the second inequality is Theorem 3.1 and $w_j \ge 0$. The shared weights are essential for this weighted-average form: with target weights $w_j$ and approximation weights $v_j$, the same argument contributes the additional label divergence $D_{\mathrm{KL}}(\mathbf{w} : \mathbf{v})$. Equality in the log-sum step holds exactly when $p_j(x)/q_j(x)$ is independent of $j$ among positive-weight components, for almost every $x$ (with standard zero-density conventions). □

**Remark 3.1.** *For components of different intrinsic dimensions, the bound is the weighted inner product* $\mathbf{w}^\top \mathbf{c}$*, with each* $C_{d_j,\alpha_{\sharp,j}}$ *computed at that component's dimension and parameter vector.*

**Example 3.1** (A mixture with different intrinsic dimensions)**.** *In ambient dimension* $D = 3$*, consider a mixture with weight* 0.7 *on an embedded line segment and weight* 0.3 *on an embedded triangle. Each component is perturbed by an independent Gaussian with positive-definite covariance on its own normal space. The componentwise Gaussian projections therefore define a two-component Gaussian mixture with the same weights. By the different-dimension bound,*

$$D_{\mathrm{KL}}(p : q) \le 0.7C_1 + 0.3C_2 \approx 0.2726 \text{ nats},$$

*regardless of the segment and triangle locations, shapes, orientations, or normal covariances. This is a certified upper bound; the exact mixture divergence may be smaller because the component densities can overlap.*

## 4. Conclusion

We developed a geometry-derived Gaussian mixture model for finite collections of embedded simplices. Each component represents a Dirichlet-distributed point on a simplex, with arbitrary positive parameter vector $\alpha_\sharp$, combined with independent Gaussian uncertainty in its normal space. The componentwise information projection is the ambient Gaussian matching the first two moments, and its forward KL divergence is the explicit constant $C_{d,\alpha_\sharp}$, depending only on intrinsic dimension and the Dirichlet parameter vector. For uniform points, $\alpha_\sharp = 1_{d+1}$ and this reduces to $C_d$.

Using joint convexity of relative entropy, we proved that the Gaussian mixture formed from these component projections has forward KL divergence at most $\mathbf{w}^\top \mathbf{c}$, with $\mathbf{w}$ and $\mathbf{c}$ as in Theorem 3.2. When every component is uniform and has common dimension $d$, this gives the bound $C_d$, even when component locations, shapes, normal spaces, and normal covariances differ. For uniform distributions over triangulations, volume-proportional weights give the corresponding geometric mixture. The bound concerns the mixture before any component reduction; it does not guarantee that compression or propagation through controlled dynamics preserves the same error.

The results cover arbitrary Dirichlet parameters and finite mixtures of simplex components. Extensions to adaptive triangulations, mixture reduction, and error propagation under stochastic dynamics are natural next steps. These would connect the present static approximation guarantee more directly to tracking and stochastic-control pipelines.

DRAFT

## 5. Acknowledgment

To Be Determined

## References


[1] Ralph Abraham and Jerrold Eldon Marsden. *Foundations of Mechanics*. 2nd ed. American Mathematical Society, 1978.

[2] Jerrold Eldon Marsden and Tudor Stefan Rațiu. *Introduction to Mechanics and Symmetry*. 2nd ed. Springer, 1999.

[3] Vladimir Igorevich Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 1989.

[4] Theodore Frankel. *The Geometry of Physics: An Introduction*. 3rd ed. Cambridge University Press, 2012.

[5] Wendell Fleming and Raymond Rishel. *Deterministic and Stochastic Optimal Control*. Springer, 1975.

[6] James R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.

[7] Dmitry Kozlov. *Combinatorial Algebraic Topology*. Springer, 2008.

[8] Vladimir Gilelevich Maz'ya and Gunther Schmidt. *Approximate Approximations*. American Mathematical Society, 2007.

[9] Yulong Lu and Andrew Stuart and Hendrik Weber. "Gaussian Approximations for Probability Measures on $\mathbb{R}^d$". In: *SIAM/ASA Journal on Uncertainty Quantification* 5 (2017), pp. 1136–1165.

[10] Arthur Pentland Dempster and Nan McKenzie Laird and Donald Bruce Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39.1 (1977), pp. 1–22.

[11] Chien-Fu Jeff Wu. "On the Convergence Properties of the EM Algorithm". In: *Annals of Statistics* 11.1 (1983), pp. 95–103.

[12] Geoffrey John McLachlan and Thriyambakam Krishnan. *The EM Algorithm and Extensions*. 2nd ed. Wiley, 2007.

[13] Ali Darijani et al. "Closed-Form Gaussian Mixture Representations of $d$-Dimensional Triangulated Domains". In: *2026 IEEE Sensor Data Fusion: Trends, Solutions, Applications (SDF 2026)* (Nov. 23–25, 2026). Accepted; to appear. Bonn, Germany, 2026.

[14] Xinpeng Liu and Fumio Okura. "Gaussian Mesh Renderer for Lightweight Differentiable Rendering". In: *arXiv preprint arXiv:2602.14493* (2026).

[15] Ali Darijani, Benedikt Stratmann, and Jürgen Beyerer. "A Geometric Gaussian Mixture Representation of Plane Curves". In: *arXiv preprint arXiv:2606.06505* (2026).

[16] Isin M. Balci and Efstathios Bakolas. "Density Steering of Gaussian Mixture Models for Discrete-Time Linear Systems". In: *arXiv preprint arXiv:2311.08500* (2023).

[17] Paul R. Halmos. *Finite-Dimensional Vector Spaces*. 2nd ed. Reprint of the 1958 second edition. New York: Springer-Verlag, 1974.

[18] Sheldon Axler. *Linear Algebra Done Right*. 4th ed. Undergraduate Texts in Mathematics. Cham: Springer, 2024.

[19] Lloyd N. Trefethen and III Bau David. *Numerical Linear Algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 1997.

[20] James W. Demmel. *Applied Numerical Linear Algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 1997.


UNDER PEER REVIEW

[21] Peter D. Lax and Maria Shea Terrell. *Calculus with Applications*. 2nd ed. Undergraduate Texts in Mathematics. New York: Springer, 2014.
[22] Sheldon Ross. *A First Course in Probability*. 10th ed. Pearson, 2019.
[23] Terence Tao. *Analysis I*. 4th ed. Springer, 2022.
[24] Terence Tao. *Analysis II*. 4th ed. Springer, 2022.
[25] Terence Tao. *An Introduction to Measure Theory*. American Mathematical Society, 2011.
[26] David Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, 2011.
[27] Nihat Ay et al. *Information Geometry*. Springer, 2017.
[28] Kai Wang Ng, Guo-Liang Tian, and Man-Lai Tang. *Dirichlet and Related Distributions*. Wiley, 2011.
[29] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd ed. Wiley-Interscience, 2006.

Author list, affiliations, and author order to be finalized.

DRAFT