# MODES OF A MIXTURE OF THREE HOMOSCEDASTIC GAUSSIANS: LOG-CONCAVITY AT DIAMETER TWO, THE GHOST-MODE THRESHOLD, AND A CERTIFIED ASYMMETRIC FOUR-MODE EXAMPLE

KOREA SUPERINTELLIGENCE LABS

Abstract. Let $f(x) = \sum_{j=1}^{3} w_j \exp(-\|x-\mu_j\|^2/2)$ be a mixture of three Gaussian densities with a common covariance (normalised to the identity), nonnegative weights of positive sum, and centres $\mu_j$ in a real inner product space. The largest possible number of modes (local maximisers) of $f$ is known to lie between 4, attained by the equilateral construction of Carreira-Perpiñán and Williams, and 8, the unconditional bound of Okuno and Kabata. We prove that if the centres have pairwise distance at most 2 then $f$ is log-concave, whatever the weights and the dimension, so every mode is a global maximiser and all modes take the same value; the constant 2 is exact, being the classical two-component threshold. For the equilateral equal-weight configuration of side $L$ the second derivative of $\log f$ at the barycentre is exactly $(L^2/3)(L^2/6-1)$, so the central "ghost" mode disappears at $L^2 = 6$, the threshold of Carreira-Perpiñán–Williams and of Edelsbrunner–Fasy–Rote in side-length form. Finally we give two explicit rational configurations in the plane, each with four rational balls of radius $10^{-3}$ that provably contain exactly one critical point of $f$, a maximiser of $f$ over the ball: Duistermaat's equilateral example, and a scalene one with three distinct weights and one pair of centres further apart than $\sqrt{6}$, so the four-mode region is not confined to the equilateral window; every published four-mode example we know of is equilateral with equal weights. All statements are machine-checked in Lean 4, the configuration-specific rational checks of the two examples by compiled code. A search over 9,288,663 configurations of the exact five-parameter reduction, reported as a computation, found no configuration with five modes or with more than seven critical points.

## 1. Introduction

A finite Gaussian mixture is *homoscedastic* when all of its components share one covariance matrix $\Sigma$. Since $x \mapsto \Sigma^{-1/2}x$ is an affine bijection that carries local maximisers to local maximisers, one may take $\Sigma = I$, and the normalising constant $(2\pi)^{-d/2}$ moves no maximiser either. Throughout, $E$ is a real inner product space, $\mu_0, \mu_1, \mu_2 \in E$, $w_0, w_1, w_2 \geq 0$ with $w_0 + w_1 + w_2 > 0$, and

$$f(x) = f_{w,\mu}(x) = \sum_{j=0}^{2} w_j \exp\Big(-\frac{\|x-\mu_j\|^2}{2}\Big), \qquad x \in E. \tag{1}$$

A *mode* of $f$ is a local maximiser. (We index the components by $0, 1, 2$ to match the formal development; the sources index them by $1, 2, 3$.)

**The question.** What is the largest number of modes of such a mixture? Carreira-Perpiñán and Williams conjectured that a homoscedastic mixture of $k$ Gaussians has at most $k$ modes [1, Section 2], and the conjecture is true on the line. It fails for $k = 3$ in the plane: three equally weighted components at the vertices of an equilateral triangle have, for a narrow window of side lengths, a fourth mode at the barycentre. The example is due to Duistermaat (reported in [1, Section 2 and Example 4]) and to Carreira-Perpiñán and Williams [3], who analysed the regular simplex with $k$ vertices; Edelsbrunner, Fasy and Rote [6] proved the simplex construction and studied its resilience in higher dimensions. On the other side, Améndola, Engström and Haase [1, Theorem 10] bound the number of modes of any mixture of $k$ Gaussians in $\mathbb{R}^d$ with finitely many modes by $2^{d+\binom{k}{2}}(5+3d)^k$, which is 42592 at $d = 2$, $k = 3$; Nguyen [8] sharpened the general bounds, and Okuno and Kabata [9, Section 2.2] report that his Corollary 3.20 gives 72 in the present setting, still under an a priori finiteness assumption. Okuno and Kabata [9] then proved, unconditionally, that a three-component homoscedastic mixture in any dimension has at most 15 critical points (their Theorem 1) and hence at most 8 modes (their Theorem 2). In their Section 2.2 they write:

> "By contrast, in the homoscedastic setting considered here, the largest number of modes currently known for three-component mixtures is 4, attained by the classical equilateral-triangle

> construction of Carreira-Perpiñán and Williams (2003a). To the best of our knowledge, no construction with 5 or more modes has been found in the homoscedastic three-component setting."

and they close with the conjecture that "the sharp maximum in the three-component homoscedastic setting is 4." The same gap is recorded, for bivariate mixtures, by Kabata, Matsumoto and Okuno [7, Remark 2]: the maximum "is currently known to lie between four and eight." For general (heteroscedastic) mixtures the relevant conjecture is Sturmfels' (AIM 2011), recorded as [1, Conjecture 5]: the maximum number of modes of a mixture of $k$ Gaussians in $\mathbb{R}^d$ is $\binom{d+k-1}{d}$, which is 6 at $d = 2$, $k = 3$.

**What this paper settles.** We do not close the gap [4, 8]. We settle three smaller questions around it, each with a machine-checked proof.

(a) *Log-concavity at diameter two* (Theorem 3.1). If the three centres have pairwise distance at most 2, then $\log f$ is concave on all of $E$, for every choice of nonnegative weights and in every dimension. Hence (Corollary 3.2) every mode is a global maximiser and any two modes take the same value. The constant 2 is exact (Proposition 3.4): it is the two-component threshold of Cule, Samworth and Stewart [5, 10], and the second derivative of $\log f$ at the midpoint of two equally weighted centres at distance $d$ is exactly $d^2(d^2/4-1)$. For two components the statement is classical; we did not find the three-component statement in print, and the survey of Carreira-Perpiñán and Williams [4] describes the known two-component unimodality conditions as "very particular" and not "easily generalisable". The proof is short once assembled, and a reader may reasonably regard it as folklore.

(b) *The exact ghost-mode threshold* (Propositions 4.1, 4.2). For an equilateral triangle of side $L$ with equal weights, the second derivative of $\log f$ at the barycentre towards a vertex is $(L^2/3)(L^2/6-1)$; if $L^2 > 6$ the barycentre is not a mode, in every ambient dimension. The threshold is known: it is $\sigma^2 = 1/(k-1)$ in the unit-circumradius normalisation of [3, Section 3] and $U_n = \sqrt{(n+1)/(2\pi)}$ in the unit-kernel normalisation of [6, eq. (16)]; both convert to $L^2 = 2k$. What is ours is the side-length form and the formal proof.

(c) *Certified four-mode configurations* (Theorems 5.2, 5.3). Two explicit configurations with rational centres and weights in $\mathbb{R}^2$, each with four explicit rational sup-norm balls of radius $10^{-3}$ such that every ball contains exactly one critical point of $f$ and that point maximises $f$ over the whole ball. The first is the rational form of Duistermaat's example (one side of squared length exactly 5.76, equal weights); the existence of its four modes is the published state of the art and is not claimed new. The second is *scalene with three distinct weights*: squared side lengths 5.4917, 5.7853, 6.1582 and weights 1 : 1.051434 : 1.107307, and it has at least four modes although its largest squared inter-centre distance exceeds 6, the equilateral threshold of (b). Every four-mode example we found in print [1, 3, 6, 8, 9] is equilateral with equal weights; what is and is not new in Theorem 5.3 is discussed after its proof.

Section 6 reports, as a computation and not as a theorem, a search over 9,288,663 configurations of the exact five-parameter reduction of the problem: no configuration with five modes and none with more than seven critical points was found, against the proved bound of 15. Section 7 describes the formal verification.

## 2. Preliminaries

2.1. **The line restriction.** Fix $x, v \in E$ and put $c_j = w_j e^{-\|x-\mu_j\|^2/2} \geq 0$ and $b_j = \langle \mu_j - x, v\rangle$. Expanding $\|x + sv - \mu_j\|^2$ gives, for every $s \in \mathbb{R}$,

$$\log f(x+sv) = \log\Big(\sum_{j=0}^{2} c_j e^{b_j s}\Big) - \frac{\|v\|^2}{2}\, s^2. \tag{2}$$

Write $p_j(s) = c_j e^{b_j s}$ and let $\rho_j = p_j / \sum_i p_i$ be the normalised masses (the posterior probabilities of the components at the point $x + sv$). Differentiating twice,

$$\frac{d^2}{ds^2}\log f(x+sv) = \mathrm{Var}_\rho(b) - \|v\|^2 = \frac{(\sum_j p_j b_j^2)(\sum_j p_j) - (\sum_j p_j b_j)^2}{(\sum_j p_j)^2} - \|v\|^2. \tag{3}$$

Both identities are proved in the formal development for the three-term sum $E_3(t) = \sum_j c_j e^{b_j t}$ and the function $t \mapsto \log E_3(t) - \kappa t^2/2$.

2.2. **Popoviciu's inequality, three-point form.**

**Lemma 2.1.** *Let $p_0, p_1, p_2 \geq 0$ and let $b_0, b_1, b_2 \in [m, m+B]$. Then*

$$\Big(\sum_j p_j b_j^2\Big)\Big(\sum_j p_j\Big) - \Big(\sum_j p_j b_j\Big)^2 \leq \frac{B^2}{4}\Big(\sum_j p_j\Big)^2.$$

*Proof.* Put $P = \sum p_j$, $Q = \sum p_j b_j$. Since each $b_j \in [m, m+B]$, $\sum_j p_j(b_j - m)(m + B - b_j) \geq 0$, which rearranges to $(\sum p_j b_j^2)P - Q^2 \leq (Q - mP)((m+B)P - Q)$, and the arithmetic–geometric mean inequality bounds the right-hand side by $B^2P^2/4$. □

2.3. **The mean map and the covariance identity.** Let $\rho_j(x) = w_j e^{-\|x-\mu_j\|^2/2}/f(x)$ (well defined since $f > 0$) and let

$$m(x) = \sum_j \rho_j(x)\,\mu_j$$

be the *mean map*, one step of mean-shift. Writing $f(x) = e^{-\|x\|^2/2}S(x)$ with $S(x) = \sum_j w_j e^{\langle \mu_j, x\rangle - \|\mu_j\|^2/2}$, the function $\log S$ is the cumulant generating function of the finite measure $\sum_j w_j e^{-\|\mu_j\|^2/2}\delta_{\mu_j}$, so

$$\nabla \log f(x) = m(x) - x, \qquad \nabla^2 \log f(x) = \mathrm{Cov}_{\rho(x)}(\mu) - I, \tag{4}$$

where $\mathrm{Cov}_{\rho(x)}(\mu)$ is the covariance of the centres under the probability vector $\rho(x)$. In particular the critical points of $f$ are the fixed points of $m$; since $m(x)$ is a convex combination of the centres, every critical point lies in their convex hull (cf. [9, Lemma 3], which places every critical point in the relative interior of the convex hull). In the formal development (4) is proved in coordinates on $\mathbb{R}^2$, in the forms $\nabla f(x) = f(x)\,(m(x) - x)$ and $Dm(x) = \mathrm{Cov}_{\rho(x)}(\mu)$; these are the two identities behind the certificate of Section 5. The one-dimensional identity (3) is the restriction of (4) to a line, and is what the proofs in Sections 3–4 actually use.

## 3. Log-concavity at diameter two

**Theorem 3.1.** *Let $E$ be a real inner product space, $w_0, w_1, w_2 \geq 0$ with $w_0 + w_1 + w_2 > 0$, and $\mu_0, \mu_1, \mu_2 \in E$ with $\|\mu_i - \mu_j\| \leq 2$ for all $i, j$. Then $x \mapsto \log f_{w,\mu}(x)$ is concave on $E$.*

*Proof.* It suffices to show that every line restriction (2) is concave on $\mathbb{R}$, and by (3) it suffices to show $\mathrm{Var}_\rho(b) \leq \|v\|^2$ for every $s$. The exponents satisfy $|b_i - b_j| = |\langle \mu_i - \mu_j, v\rangle| \leq \|\mu_i - \mu_j\|\,\|v\| \leq 2\|v\|$, so the three of them lie in an interval $[m, m + 2\|v\|]$ with $m = \min_j b_j$. Lemma 2.1 with $B = 2\|v\|$ gives $\mathrm{Var}_\rho(b) \leq B^2/4 = \|v\|^2$. The formal proof follows this route exactly: the one-variable statement (`concaveOn_lineLog`: if the $b_j$ lie in an interval of length $B$ and $B^2/4 \leq \kappa$ then $t \mapsto \log E_3(t) - \kappa t^2/2$ is concave on $\mathbb{R}$) is proved from the second-derivative criterion, and `logdens_concaveOn` transports it along the segment from $x$ to $y$ with $v = y - x$, $\kappa = \|v\|^2$. □

**Corollary 3.2.** *Under the hypotheses of Theorem 3.1, every mode $x$ of $f$ is a global maximiser, $f(y) \leq f(x)$ for all $y \in E$; and any two modes $x, y$ satisfy $f(x) = f(y)$.*

*Proof.* A local maximum of a concave function on a convex set is a global maximum; apply this to $\log f$, which has the same local and global maximisers as $f$ because $f > 0$ and log is increasing. The second statement follows from the first applied twice. □

*Remark* 3.3. Since the set of maximisers of a concave function is convex, the set of modes is convex under the diameter hypothesis (so it is a single point or a segment or a convex body of maximisers, never two isolated modes). This last sentence is an observation about concave functions and is not part of the formal development; the two statements of Corollary 3.2 are.

The constant 2 cannot be improved, and the obstruction is already present with two components: letting one weight vanish embeds the two-component family in ours.

**Proposition 3.4** (sharpness; the two-component threshold)**.** *Let $\mu_0, \mu_1, \mu_2 \in E$ and let $d = \|\mu_0 - \mu_1\|$. For the weights $(w_0, w_1, w_2) = (1, 1, 0)$, the restriction $\psi(s) = \log f\big(\frac{1}{2}(\mu_0 + \mu_1) + s(\mu_0 - \mu_1)\big)$ of the log-density to the line through the midpoint satisfies*

$$\psi'(0) = 0, \qquad \psi''(0) = d^2\Big(\frac{d^2}{4} - 1\Big).$$

*Consequently, if $d > 2$ then $\log f$ is not concave on $E$. Explicitly, on the real line the mixture with unit weights at 0 and 3 (and weight 0 on a third centre) has a non-concave log-density.*

*Proof.* With $x = \frac{1}{2}(\mu_0 + \mu_1)$ and $v = \mu_0 - \mu_1$ the data of (2) are $c_0 = c_1 = e^{-d^2/8}$, $c_2 = 0$, $b_0 = d^2/2$, $b_1 = -d^2/2$, $\|v\|^2 = d^2$; the two derivatives are direct computations from (3) at $s = 0$, where $\rho = (\frac{1}{2}, \frac{1}{2}, 0)$ and $\mathrm{Var}_\rho(b) = d^4/4$. If $\psi''(0) > 0$ and $\psi'(0) = 0$ then 0 is a strict local minimum of $\psi$; but concavity of $\log f$ on $E$ gives $\psi(s) + \psi(-s) \leq 2\psi(0)$, which together with the local minimum forces $\psi$ to be locally constant at 0, contradicting $\psi''(0) > 0$. This argument (`not_concaveOn_of_axis`) is the one used for all the non-concavity statements of this paper. $\square$

Proposition 3.4 is the "only if" half of the classical two-component result: Cule, Samworth and Stewart [5] state that "for $p \in (0, 1)$, the location mixture of standard univariate normal densities $f(x) = p\phi(x) + (1-p)\phi(x-\mu)$ is log-concave if and only if $\|\mu\| \leq 2$" (Section 2 of the arXiv version), restated in [10, Section 10.2] as: "$pN_d(-\mu, I) + (1-p)N_d(\mu, I)$ with $p \in (0, 1)$ has a log-concave density if and only if $\|\mu\| \leq 1$"; the equal-weight unimodality statement on the line, "if and only if $|\mu_2 - \mu_1| \leq 2\sigma$", is recorded in [1, Section 2], where it is attributed to Cohen (1953) and Eisenberger (1964) with Behboodian [2] as the reference given. What Proposition 3.4 adds is the exact midpoint second derivative and a formal proof inside the three-component family; only the equal-weight case $(1, 1, 0)$ is formalised. With Theorem 3.1, 2 is thus the exact threshold for log-concavity of a three-component homoscedastic mixture.

Two controls confirm that the hypothesis of Theorem 3.1 is neither vacuous nor secretly forcing the centres together.

**Proposition 3.5.** *(i) For $w > 0$ and any $\mu \in E$, the point $\mu$ is a mode of $f_{(w,w,w),(\mu,\mu,\mu)}$; three coincident centres satisfy the diameter hypothesis with diameter 0, and the density has a mode. (ii) For a unit vector $e \in E$ the centres $0, e, 2e$ have pairwise distances $1, 2, 1$; they satisfy the hypothesis, with equality for the outer pair.*

*Proof.* (i) $f(x) = 3we^{-\|x-\mu\|^2/2} \leq 3w = f(\mu)$. (ii) Norm computations. $\square$

## 4. The equilateral configuration and the ghost-mode threshold

Call $\mu_0, \mu_1, \mu_2 \in E$ a *centred equilateral triangle of side $L$* if $\mu_0 + \mu_1 + \mu_2 = 0$ and $\|\mu_i - \mu_j\|^2 = L^2$ for $i \neq j$. From these four relations one derives the Gram data $\|\mu_j\|^2 = L^2/3$ and $\langle \mu_i, \mu_j \rangle = -L^2/6$ ($i \neq j$). Take equal weights $w_0 = w_1 = w_2 = w > 0$ and restrict $\log f$ to the line through the barycentre 0 and the vertex $\mu_0$:

$$\phi(t) = \log f(t\mu_0) = \log\Big(we^{-L^2/6}\big(e^{(L^2/3)t} + 2e^{-(L^2/6)t}\big)\Big) - \frac{L^2}{6}\,t^2, \tag{5}$$

an instance of (2) with $x = 0$, $v = \mu_0$, $c_j = we^{-L^2/6}$ and exponents $(L^2/3, -L^2/6, -L^2/6)$.

**Proposition 4.1.** *For every $L \in \mathbb{R}$ and $w > 0$, the function $\phi$ of (5) satisfies*

$$\phi'(0) = 0, \qquad \phi''(0) = \frac{L^2}{3}\Big(\frac{L^2}{6} - 1\Big).$$

*In particular $\phi''(0) < 0$ for $0 < L^2 < 6$, $\phi''(0) = 0$ at $L^2 = 6$, and $\phi''(0) > 0$ for $L^2 > 6$.*

*Proof.* At $t = 0$ the masses are equal, $\rho = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the exponents sum to zero (because the centres do), so $\phi'(0) = \sum \rho_j b_j = 0$, and $\mathrm{Var}_\rho(b) = \frac{1}{3}\big((L^2/3)^2 + 2(L^2/6)^2\big) = L^4/18$; by (3), $\phi''(0) = L^4/18 - L^2/3$. $\square$

**Proposition 4.2.** *Let $\mu_0, \mu_1, \mu_2$ be a centred equilateral triangle of side $L$ in a real inner product space $E$, let $w > 0$, and suppose $L^2 > 6$. Then the barycentre 0 is not a mode of $f_{(w,w,w),\mu}$. Explicitly, for $\mu_0 = (2, -1, -1)$, $\mu_1 = (-1, 2, -1)$, $\mu_2 = (-1, -1, 2)$ in $\mathbb{R}^3$, a centred equilateral triangle of side $\sqrt{18}$, the origin is not a mode of the equal-weight mixture.*

*Proof.* If 0 were a mode of $f$, then 0 would be a local maximum of $\phi$. By Proposition 4.1, $\phi'(0) = 0$ and $\phi''(0) > 0$, so 0 is also a local minimum of $\phi$; hence $\phi$ is locally constant at 0 and $\phi''(0) = 0$, a contradiction. Note that the proof uses no strict-extremum machinery and no information about the ambient dimension. $\square$

**Proposition 4.3** (log-concavity fails above the threshold)**.** *Under the hypotheses of Proposition 4.2, $\log f_{(w,w,w),\mu}$ is not concave on $E$; in particular this holds for the explicit triangle of side $\sqrt{18}$ in $\mathbb{R}^3$. Consequently the constant 2 of Theorem 3.1 cannot be replaced by any constant exceeding $\sqrt{6}$.*

*Proof.* The argument of Proposition 3.4 applied to $\phi$: concavity gives $\phi(t)+\phi(-t) \le 2\phi(0)$, the local minimum from $\phi''(0) > 0$ then makes $\phi$ locally constant, contradicting $\phi''(0) > 0$. (Proposition 3.4 is sharper; this one is the natural control for the equilateral family.) □

**Sources for the threshold, and a caveat about the dimension.** Propositions 4.1 and 4.2 are not new. Carreira-Perpiñán and Williams [3, Section 3] place $M$ equally weighted isotropic components of variance $\sigma^2$ at the vertices of a regular simplex of unit circumradius in $\mathbb{R}^D$, $D = M - 1$, and state that the centre "is always a critical point, in particular a saddle, catastrophe and maximum if $\sigma^2 < 1/D$, $\sigma^2 = 1/D$ and $\sigma^2 > 1/D$, respectively". For $M = 3$ the side is $\sqrt{3}$ and $\sigma^2 = 1/2$ is $L^2/\sigma^2 = 6$. Edelsbrunner, Fasy and Rote [6] use the unit kernel $e^{-\pi\|x-z\|^2}$ (variance $\sigma_0^2 = 1/(2\pi)$) on the standard $n$-simplex scaled by $s$; they "find the transition by setting the distance equal to $2\sigma_0$, which gives $s$ equal to $U_n = \sqrt{(n+1)/(2\pi)}$" (their eq. (16)), and their Barycenter Lemma states that the barycentre "is a mode of $G_s$ for $s < U_n$ and it is a saddle of index 1 for $s > U_n$". The edge of the scaled simplex is $s\sqrt{2}$, so the transition is at $\text{edge}^2/\sigma_0^2 = 2(n+1)$, again 6 at $n = 2$. Neither source writes the threshold in side-length form.

Both sources are careful about the dimension, and so is Proposition 4.2: above the threshold the barycentre is a local *minimum* only inside the affine hull of the centres ($d = 2$); in ambient dimension at least 3 it is a saddle of index one, because in every direction orthogonal to that hull the second derivative of $\log f$ is $-1$ ("when the centre of the $M$-vertex simplex is a minimum in $M - 1$ dimensions it will be a saddle point in $M$ dimensions" [3, Section 3]). "Not a mode" is the dimension-free statement, and it is what is proved.

*Remark* 4.4 (the four-mode window, numerically). The lower end of the window is not computed here formally. Carreira-Perpiñán and Williams [3, Section 2] report four modes "for a narrow interval of scales $\sigma_1 < \sigma < \sigma_3$ (where $\sigma_1 = \sqrt{2}/2 \simeq 0.7071$ and $\sigma_3 \simeq 0.7391$)", i.e. $3/\sigma_3^2 \approx 5.492 < L^2/\sigma^2 < 6$. Our own numerical bisection of the one-dimensional reduced system (Section 6) puts the lower edge at $L_0^2 = 5.49128715\ldots$, where six critical points are born at once, in agreement with $\sigma_3$. Inside the window the equal-weight equilateral mixture has 7 critical points, 4 maxima and 3 saddles; above $L^2 = 6$ it has 3 maxima, 3 saddles and a minimum at the barycentre; below $L_0^2$ it has a single critical point. Duistermaat's example as drawn by Okuno and Kabata [9, Figure 1(b)], $\gamma_u = \gamma_v = \psi = 2.88$, is $L^2 = 5.76$; the example of [1, Example 4], circumradius 1 and $\sigma^2 = 0.53$, is $L^2/\sigma^2 = 3/0.53 \approx 5.660$. (The earlier arXiv versions v1 and v2 of [1] print the same example with "$\sigma = 0.72$" and covariance "$\sigma I_2$", i.e. $L^2/\sigma^2 = 3/0.72^2 \approx 5.787$; both points lie inside the window. We follow v3, the version of record.)

## 5. Certified four-mode configurations

Everything in this section is in the plane, $E = \mathbb{R}^2$, with rational centres and weights. The formal development works with the coordinate density $f(x) = \sum_j w_j \exp(-\sum_a (x_a - \mu_{j,a})^2/2)$ on $\mathbb{R}^2$ and proves that it agrees with (1) on $\mathbb{R}^2$ with the Euclidean norm, so that the results transfer to the notion of mode used in the rest of the paper.

5.1. **The certificate.** Let $c \in \mathbb{Q}^2$, $r \in \mathbb{Q}_{>0}$, and let $B = \{x : \|x - c\|_\infty \le r\}$ be the closed sup-norm ball. The following statement is the engine; it is proved in the formal development (`cert_mode`) from the interval-Krawczyk theorem of a shared certificate library, applied to the map $x \mapsto m(x) - x$ with the preconditioner $-I$, for which the Krawczyk map is exactly the mean map $m$ and the matrix $I - (-I)\, D(m - \mathrm{id})$ is exactly the covariance $\mathrm{Cov}_{\rho(x)}(\mu)$.

**Lemma 5.1** (the mode certificate)**.** *Suppose rational interval arithmetic on the box $B$ (with* exp *enclosed by a Taylor polynomial with rigorous remainder) establishes: (i) the enclosure of $S$ on $B$ does not contain* 0*; (ii) every row of the enclosure of* $\mathrm{Cov}_{\rho(x)}(\mu)$ *on $B$ has absolute row sum at most $K$, where $K < 1$; (iii) $\|m(c) - c\|_\infty \le \beta$; and (iv) $\beta + Kr < r$. Then there is exactly one point $p \in B$ with $\nabla f(p) = 0$; moreover $f(x) \le f(p)$ for every $x \in B$, and $p$ is a local maximiser of $f$.*

*Proof.* By (4), $Dm = \mathrm{Cov}_\rho(\mu)$, so (ii) makes $m$ a $K$-contraction of $B$ in the sup norm, and (iii)–(iv) make $m$ map $B$ into itself: $\|m(x) - c\|_\infty \le \|m(x) - m(c)\|_\infty + \|m(c) - c\|_\infty \le Kr + \beta < r$. Hence $m$ has a unique fixed point $p \in B$, and $\|p-c\|_\infty \le \beta/(1-K) < r$, so $p$ is interior. Since $\nabla f = f \cdot (m - \mathrm{id})$ and $f > 0$, the fixed points of $m$ are the critical points of $f$, which gives the first claim. For the second, no second-derivative argument is needed: *mean-shift ascent* says $f(x) \le f(m(x))$ for all $x$. Indeed $\log S$ is convex with gradient $m$, so Jensen's inequality for exp gives $e^{\langle m(x), y-x\rangle} S(x) \le S(y)$ for all $x, y$; with $y = m(x)$ and $f = e^{-\|\cdot\|^2/2} S$ this reads $f(m(x)) \ge e^{\|m(x)-x\|^2/2} f(x) \ge f(x)$. Iterating, $f(x) \le f(m^n(x)) \to f(p)$ by continuity, since $m^n(x) \to p$. Thus $p$ maximises $f$ over $B$, and being interior it is a local maximiser. ☐

The ascent property $f(x) \le f(m(x))$ is itself classical (mean-shift is an EM step); what we did not find in the literature is its combination with a contraction or Krawczyk certificate, or any interval-arithmetic or formally verified localisation of the modes of a Gaussian mixture (searched: a full-text arXiv corpus, the arXiv API, and the bibliographies of [1, 3, 6, 9]).

In both configurations below the same certificate data are used for all eight balls: $r = 10^{-3}$, $K = 39/40$, $\beta = 10^{-6}$ (so $\beta + Kr = 0.000976 < r$), and exp enclosed at dyadic precision $2^{-40}$ with 13 Taylor terms. The hypotheses (i)–(iv) are decidable statements about rational numbers and are evaluated by compiled code (Section 7).

5.2. **Duistermaat's configuration, certified.** An exactly equilateral triangle has no rational coordinates in $\mathbb{R}^2$, so we take the nearest rational isosceles one.

**Theorem 5.2.** *Let* $\mu_0 = (\frac{104}{75}, 0)$, $\mu_1 = (-\frac{52}{75}, \frac{6}{5})$, $\mu_2 = (-\frac{52}{75}, -\frac{6}{5})$ *and* $w_0 = w_1 = w_2 = 1$. *Then* $\mu_0 + \mu_1 + \mu_2 = 0$, $\|\mu_0 - \mu_1\|^2 = \|\mu_0 - \mu_2\|^2 = \frac{3604}{625} = 5.7664$ *and* $\|\mu_1 - \mu_2\|^2 = \frac{144}{25} = 5.76$. *For each of the four points*

$$c^{(1)} = (-0.4429251, -0.7655235), \qquad c^{(2)} = (-0.4429251, 0.7655235),$$
$$c^{(3)} = (-0.0115193, 0), \qquad c^{(4)} = (0.8857247, 0),$$

*the closed sup-norm ball of radius* $10^{-3}$ *about* $c^{(i)}$ *contains exactly one critical point* $p^{(i)}$ *of* $f$, *and* $p^{(i)}$ *maximises* $f$ *over that ball. The four balls are pairwise disjoint, so* $p^{(1)}, \dots, p^{(4)}$ *are four distinct modes: the mixture has at least four modes.*

The existence of four modes for the equilateral construction is the published state of the art ([9, Section 2.2], attributing it to [3]; [1, Example 4]; [6]) and is not claimed new. New here are the data: an explicit rational configuration carrying Duistermaat's $L^2 = 5.76$ as one side, the localisation of all four modes to within $10^{-3}$ with uniqueness of the critical point in each ball, and the formal proof. The third ball is the "ghost" mode near the barycentre (not exactly at the origin, because the rational triangle is only isosceles). The certified contraction constants (the largest row sums of the covariance enclosures) are 0.8896, 0.8896, 0.9736 and 0.8036; the ghost mode is the binding one, because at an exactly equilateral equal-weight triangle $\mathrm{Cov}_\rho(\mu) = (L^2/6)I$ at the barycentre, whose row sum 0.96 at $L^2 = 5.76$ leaves a margin of only 0.04 for the width of the enclosure. This is what forces $r = 10^{-3}$.

5.3. **A scalene configuration with three distinct weights.**

**Theorem 5.3.** *Let*

$$\mu_0 = (1.315877, -0.031276), \quad \mu_1 = (-0.667324, 1.217155), \quad \mu_2 = (-0.750037, -1.263048),$$
$$(w_0, w_1, w_2) = (1, 1.051434, 1.107307)$$

*(all coordinates exact decimals). Then*

$$\|\mu_0 - \mu_1\|^2 = \tfrac{2745833084081}{500000000000} = 5.4916661\dots,$$
$$\|\mu_0 - \mu_2\|^2 = \tfrac{289263145769}{50000000000} = 5.7852629\dots,$$
$$\|\mu_1 - \mu_2\|^2 = \tfrac{3079124180789}{500000000000} = 6.1582483\dots,$$

*the three squared side lengths are pairwise distinct, the three weights are pairwise distinct, and* $\|\mu_1 - \mu_2\|^2 > 6$. *For each of the four points*

$$c^{(1)} = (-0.5563378, -0.9554167), \quad c^{(2)} = (-0.4221259, 0.8058684),$$
$$c^{(3)} = (0.05283, 0.0070331), \quad c^{(4)} = (0.5809246, -0.00697),$$

*the closed sup-norm ball of radius* $10^{-3}$ *about* $c^{(i)}$ *contains exactly one critical point of* $f$*, which maximises* $f$ *over that ball; the four balls are pairwise disjoint. Hence this scalene, unequally weighted homoscedastic mixture of three Gaussians has at least four modes.*

*Proof of Theorems 5.2 and 5.3.* The side lengths and weight inequalities are rational arithmetic. For each of the eight balls the four hypotheses of Lemma 5.1 are evaluated in exact rational interval arithmetic with the data $r = 10^{-3}$, $K = 39/40$, $\beta = 10^{-6}$ (the certified row sums for Theorem 5.3 are 0.7393, 0.8671, 0.9692, 0.9696, all below $K$). Disjointness is the rational check that for each pair of centres some coordinate differs by more than $2r$. Lemma 5.1 then gives four distinct local maximisers, and the bridge lemma identifies them with modes of (1) on $\mathbb{R}^2$. □

Theorem 5.3 says that the four-mode region of the problem is not contained in $\{\max_{i<j}\|\mu_i - \mu_j\|^2 \le 6\}$: an equilateral equal-weight triangle with $L^2 > 6$ has no mode at its barycentre (Proposition 4.2), whereas this configuration has a pair of centres further apart than any four-mode equilateral configuration allows, and four modes anyway. Its squared side lengths spread by 12% and its weights by 11%. We are aware of no four-mode homoscedastic three-component example in print that is not equilateral with equal weights: the examples of [1, Example 4], [3], [6] and the seed construction of [8, Theorem 4.11] (regular simplex, equal weights, in its proof) are regular simplices with equal weights, and [9, Figure 1] keeps $w_1 = w_2 = w_3$ in all panels, its one asymmetric panel (c), $\gamma_u = 3.2$, $\gamma_v = 2.6$, $\psi = 2.5$, losing the central mode. Two near-misses must be recorded. Carreira-Perpiñán and Williams [3, Section 4] assert, without an example or numbers, that "adding a very small amount of noise to the centroids, mixing proportions or covariances in the triangle construction still preserves the appearance of a fourth mode at the centre. However, adding a larger (but still very small) amount of noise prevents it. Thus, the mode creation at the origin is generic but fragile." The same qualitative statement follows from the implicit function theorem, since all seven critical points of Duistermaat's example are (numerically) non-degenerate. And Edelsbrunner, Fasy and Rote [6, Section 4] pose the quantitative question as open: "How robust are they under moving individual kernels or changing their weights?" What neither supplies is an explicit example, and what no perturbation argument supplies is a four-mode configuration with two centres further apart than $\sqrt{6}$; a reader who takes the sentence of [3] as covering Theorem 5.3 should regard the theorem as a certified instance of a known phenomenon rather than as new.

*Remark* 5.4 (what is not certified). Both configurations have, numerically, exactly 7 critical points, 4 maxima and 3 saddles (Newton enumeration on a $70 \times 70$ grid over the convex hull, with the parity audit of Section 6 satisfied). "Exactly four modes" is not proved: that needs an exclusion covering of the convex hull outside the four balls and a certificate that each saddle is not a mode. An adaptive bisection with the same interval arithmetic does not close: the number of unresolved boxes doubles at every halving of the leaf size $(234, 532, 972, 1730, 3302$ at widths $2^{-3}, \ldots, 2^{-7})$, because near the ghost mode the derivative of $f \cdot (m - \mathrm{id})$ is a hundred times smaller than the interval overestimate.

5.4. **Controls for the certificate.** Three further rational evaluations, checked in Lean by `native_decide` in the same way as the certificates themselves, confirm that the certificate is neither vacuous nor indiscriminate. They are recorded as controls, not as results, and they are not bound to a row of the ledger. In the configuration of Theorem 5.2:

(i) (`wrong_sign_fails`) at the ghost-mode ball about $c^{(3)}$, replacing the preconditioner $-I$ by $+I$ gives a certified row sum $1156657022969/1099511627776 = 1.051973\ldots > 1$;

(ii) (`ball_too_large_fails`) at the same ball centre with the correct preconditioner, radius $1/50$ gives a row sum $1172921641085/1099511627776 = 1.066766\ldots > 1$; the value $1.731133\ldots$ at radius $1/10$ is a separate rational computation, not a Lean statement;

(iii) (`saddle_fails`) at the on-axis saddle $(0.1427732, 0)$ of the same density, with radius $10^{-3}$ and preconditioner $-I$, the row sum is $287660411605/274877906944 = 1.046502\ldots > 1$.

So the sign of the preconditioner is load-bearing (with $+I$ the Krawczyk map is $x \mapsto 2x - m(x)$), the radius cannot be enlarged freely, and the test discriminates: at a saddle it refuses to certify a mode. The Lean statements assert only that each row sum exceeds 1; the exact rationals and their decimal values were computed separately in exact rational arithmetic.

## 6. Computations outside the formal development

Nothing in this section is proved; it is reported as a computation, with its parameters.

**The five-parameter reduction.** Put $\mu_0 = 0$, $\mu_1 = u$, $\mu_2 = v$ with $u, v$ linearly independent, and

$$\gamma_u = \tfrac{1}{2}\|u\|^2, \quad \gamma_v = \tfrac{1}{2}\|v\|^2, \quad \psi = \langle u, v\rangle, \quad \kappa_u = \log(w_0/w_1), \quad \kappa_v = \log(w_0/w_2)$$

(this is the parametrisation of [9]). Every critical point is $x = \alpha u + \beta v$ with $(\alpha, \beta)$ in the open 2-simplex [9, Lemma 3], and in the coordinates $s = \log(\alpha/\rho_0)$, $t = \log(\beta/\rho_0)$, $\rho_0 = 1 - \alpha - \beta$, criticality reads $p(s,t) = q(s,t) = 0$ with

$$p = s + \kappa_u + \gamma_u - 2\gamma_u\alpha - \psi\beta, \qquad q = t + \kappa_v + \gamma_v - \psi\alpha - 2\gamma_v\beta,$$
$$\alpha = \frac{e^s}{1 + e^s + e^t}, \qquad \beta = \frac{e^t}{1 + e^s + e^t}.$$

By (4) a critical point is classified by the two nonzero eigenvalues of $\mathrm{Cov}_\rho(\mu)$, which are explicit in $(\alpha, \beta)$ and the five parameters (a mode iff both are below 1, a saddle iff exactly one exceeds 1), and the Jacobian of $(p, q)$ in $(s, t)$ is $I - \mathrm{Cov}_\rho(\mu)^\top$ in the appropriate basis. When all critical points are non-degenerate, $\#\max - \#\text{saddle} + \#\min = 1$ (the gradient field $m(x) - x$ of $\log f$ points inward on a large circle, so the indices sum to 1); in particular $\#\text{modes} \le (\#\text{critical} + 1)/2$, which is how [9] pass from 15 to 8. We use the relation as a *completeness audit*: an even critical-point count, or a count of zero, is a detection failure of the enumeration, never mathematics.

**The search.** Critical points were enumerated by vectorised Newton iteration from a $36 \times 36$ grid over the exact bounding rectangle $s \in (\min(0, \psi) - \gamma_u - \kappa_u,\ \max(2\gamma_u, \psi) - \gamma_u - \kappa_u)$ (and likewise for $t$), which follows from $p = q = 0$ and $0 < \alpha, \beta, \alpha + \beta < 1$. Sampling arms: broad random triangles and weights (log-uniform scale 0.8–12, weights $e^{-6}$–1); near-equilateral; near-collinear (aspect ratios down to $10^{-4}$); extreme weights ($|\kappa| \le 20$); wide scale; a $24 \times 24 \times 32 \times 26 \times 26$ grid over $(\|u\|, \|v\|, \angle(u, v), \kappa_u, \kappa_v)$; and a hill-climb on $(\#\text{modes}, \#\text{critical})$ with 300 restarts of 4000 steps. The implementation was cross-checked against a second, independent one (Newton on $\nabla \log f$ in $\mathbb{R}^2$ from a $45 \times 45$ grid plus 600 random starts, classified by the $2 \times 2$ Hessian) on 73 of 73 test configurations, and it reproduces the three panels of [9, Figure 1]. Over 9,288,663 configurations with a completed histogram:

| critical points | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | $\ge 8$ |
|---|---|---|---|---|---|---|---|---|---|
| configurations | 975 | 3,934,182 | 49,484 | 3,494,786 | 81,075 | 924,317 | 118,255 | 685,589 | 0 |

| modes | 0 | 1 | 2 | 3 | 4 | $\ge 5$ |
|---|---|---|---|---|---|---|
| configurations | 4,841 | 3,982,640 | 3,631,747 | 1,669,435 | 0 | 0 |

No configuration with five or more modes, and none with more than seven critical points, was found. The even and zero counts, 249,789 configurations or 2.69%, are audit failures; they are concentrated in the extreme-weight and extreme-aspect arms, where the bounding rectangle is hundreds of units wide, and in every one of them the found mode count is 1 or 2. So the enumeration is complete on 97.3% of the sampled configurations, and the incomplete 2.7% are the regimes in which one component is effectively switched off. The four-mode region is thin: zero hits in these arms, and one hit in a separate near-equilateral arm of three million samples, at $(\gamma_u, \gamma_v, \psi, \kappa_u, \kappa_v) = (2.7804, 3.0890, 2.9373, 0.0437, -0.0494)$, a triangle within about 6% of equilateral with weights within about 5% of equal.

**The four-mode region with equal weights.** Four further sweeps of 400,000 configurations each, at Newton grid 60: a near-equilateral arm ($\|u\|, \|v\| \in [2.0, 2.9]$, angle in $[0.85, 1.30]$, weights within $e^{\pm 0.6}$), a widened arm, an isosceles arm with exactly equal weights ($\|u\| \in [2.0, 3.0]$, $\|v\| = \|u\|(1 \pm 0.15)$, angle in $[0.9, 1.25]$), and a wide audit arm (scale in $[0.3, 30]$, $|\kappa| \le 12$). Audit failures: 1, 10, 0 and 73,050 of 400,000 respectively (18.3% in the wide arm). Only the equal-weight isosceles arm produced four-mode configurations, 44 of them, and all 44 have

$$\|u\| \in [2.3317, 2.3928], \qquad \|v\| \in [2.3270, 2.3929],$$
$$\text{apex angle} \in [59.32°, 60.57°], \qquad \text{diameter} \in [2.3527, 2.3942],$$

a small neighbourhood of the equilateral family, consistent with the exact equilateral window $L \in (2.34335, 2.44949)$ of Remark 4.4 (the scan's upper diameter falls short of 2.4495 only because the sampling rarely reaches the top of the window). Nobody has, to our knowledge, published the extent of the four-mode region; [1] say "for a small window of parameters" and [9] "across a wide range of configurations, we did not find any example with more than 4 modes". Theorem 5.3 shows that with unequal weights the region extends past the equilateral threshold.

**A sharper conjecture.** The observed maximum number of critical points is 7, against the proved bound of 15.

**Question 6.1.** Does every homoscedastic mixture of three Gaussians have at most 7 critical points?

A positive answer would imply the conjecture of [9] via $\#\text{modes} \leq (\#\text{critical} + 1)/2$.

## 7. Verification

Every theorem, proposition and corollary of Sections 3–5 is proved in Lean 4 on top of Mathlib, in the form recorded in Appendix A, as are Lemma 2.1, the identities (2)–(3) and (4) (on $\mathbb{R}^2$), and Lemma 5.1. For Theorem 3.1, Corollary 3.2, Propositions 3.4–4.3 and every lemma of Sections 2–4, `#print axioms` reports exactly `[propext, Classical.choice, Quot.sound]`. The same holds for the reasoning layer of Section 5 (the derivative identities, mean-shift ascent, the certificate Lemma 5.1, and the side-length and asymmetry statements of Theorems 5.2–5.3). The configuration-specific rational evaluations are decided by compiled code (`native_decide`): for each of the two configurations, one bundled check `checks` (the certificate hypotheses (i)–(iv) of all four balls, evaluated in a single call) and one disjointness check `sep`, and the three controls of Section 5.4 – seven `native_decide` calls in all, each contributing one axiom named `<theorem>._native.native_decide.ax_1_1`. Accordingly `#print axioms` reports, for `unique_critical` of either witness, the three standard axioms together with `checks._native.native_decide.ax_1_1`; for `four_modes` and `four_isMode`, those together with `sep._native.native_decide.ax_1_1`; and for each of the three controls, the three standard axioms and its own `_native.native_decide.ax_1_1`. Nothing uses `sorry`. Repository reference to be added at submission.

## Appendix A. The formal statements

Signatures quoted verbatim from the Lean sources, with the section variables in force; `E` is a real inner product space.

**Definitions.**

```
noncomputable def dens (w₀ w₁ w₂ : ℝ) (m₀ m₁ m₂ x : E) : ℝ :=
  w₀ * Real.exp (-‖x - m₀‖ ^ 2 / 2) + w₁ * Real.exp (-‖x - m₁‖ ^ 2 / 2)
    + w₂ * Real.exp (-‖x - m₂‖ ^ 2 / 2)

def IsMode (w₀ w₁ w₂ : ℝ) (m₀ m₁ m₂ x : E) : Prop :=
  IsLocalMax (dens w₀ w₁ w₂ m₀ m₁ m₂) x

noncomputable def E3 (c₀ c₁ c₂ b₀ b₁ b₂ t : ℝ) : ℝ :=
  c₀ * Real.exp (b₀ * t) + c₁ * Real.exp (b₁ * t) + c₂ * Real.exp (b₂ * t)

noncomputable def lineLog (c₀ c₁ c₂ b₀ b₁ b₂ κ t : ℝ) : ℝ :=
  Real.log (E3 c₀ c₁ c₂ b₀ b₁ b₂ t) - κ * t ^ 2 / 2

structure CentredEquilateral (L : ℝ) (m₀ m₁ m₂ : E) : Prop where
  sum : m₀ + m₁ + m₂ = 0
  d₀₁ : ‖m₀ - m₁‖ ^ 2 = L ^ 2
  d₀₂ : ‖m₀ - m₂‖ ^ 2 = L ^ 2
  d₁₂ : ‖m₁ - m₂‖ ^ 2 = L ^ 2

noncomputable def axisLog (L w : ℝ) : ℝ → ℝ :=
  lineLog (w * Real.exp (-(L ^ 2 / 3) / 2)) (w * Real.exp (-(L ^ 2 / 3) / 2))
    (w * Real.exp (-(L ^ 2 / 3) / 2)) (L ^ 2 / 3) (-(L ^ 2) / 6) (-(L ^ 2) / 6) (L ^ 2 / 3)

noncomputable def midLog (d : ℝ) : ℝ → ℝ :=
  lineLog (Real.exp (-(d ^ 2 / 4) / 2)) (Real.exp (-(d ^ 2 / 4) / 2)) 0
    (d ^ 2 / 2) (-(d ^ 2) / 2) 0 (d ^ 2)
```

The lemmas `logdens_axis` and `logdens_mid` identify `axisLog L w t` with `Real.log (dens w w w m₀ m₁ m₂ (t • m₀))` for a centred equilateral triangle, and `midLog` with the midpoint restriction of Proposition 3.4.

**Theorem 3.1.**

```
theorem popoviciu3 (p₀ p₁ p₂ b₀ b₁ b₂ m B : ℝ)
    (hp₀ : 0 ≤ p₀) (hp₁ : 0 ≤ p₁) (hp₂ : 0 ≤ p₂)
    (h₀ : m ≤ b₀) (h₀' : b₀ ≤ m + B) (h₁ : m ≤ b₁) (h₁' : b₁ ≤ m + B)
    (h₂ : m ≤ b₂) (h₂' : b₂ ≤ m + B) :
    (p₀ * b₀ ^ 2 + p₁ * b₁ ^ 2 + p₂ * b₂ ^ 2) * (p₀ + p₁ + p₂)
      - (p₀ * b₀ + p₁ * b₁ + p₂ * b₂) * (p₀ * b₀ + p₁ * b₁ + p₂ * b₂)
      ≤ B ^ 2 / 4 * (p₀ + p₁ + p₂) ^ 2

variable {c₀ c₁ c₂ : ℝ} (h₀ : 0 ≤ c₀) (h₁ : 0 ≤ c₁) (h₂ : 0 ≤ c₂) (hs : 0 < c₀ + c₁ + c₂)
variable (b₀ b₁ b₂ κ : ℝ)
theorem concaveOn_lineLog {m B : ℝ}
    (hb₀ : m ≤ b₀) (hb₀' : b₀ ≤ m + B) (hb₁ : m ≤ b₁) (hb₁' : b₁ ≤ m + B)
    (hb₂ : m ≤ b₂) (hb₂' : b₂ ≤ m + B) (hκ : B ^ 2 / 4 ≤ κ) :
    ConcaveOn ℝ Set.univ (lineLog c₀ c₁ c₂ b₀ b₁ b₂ κ)

variable {w₀ w₁ w₂ : ℝ} (h₀ : 0 ≤ w₀) (h₁ : 0 ≤ w₁) (h₂ : 0 ≤ w₂) (hs : 0 < w₀ + w₁ + w₂)
variable {m₀ m₁ m₂ : E}
variable (d₀₁ : ‖m₀ - m₁‖ ≤ 2) (d₀₂ : ‖m₀ - m₂‖ ≤ 2) (d₁₂ : ‖m₁ - m₂‖ ≤ 2)
theorem logdens_concaveOn :
    ConcaveOn ℝ Set.univ (fun x : E => Real.log (dens w₀ w₁ w₂ m₀ m₁ m₂ x))
```

**Corollary 3.2.** (same variables)

```
theorem isMaxOn_of_isMode {x : E} (hx : IsMode w₀ w₁ w₂ m₀ m₁ m₂ x) :
    ∀ y : E, dens w₀ w₁ w₂ m₀ m₁ m₂ y ≤ dens w₀ w₁ w₂ m₀ m₁ m₂ x

theorem modes_eq {x y : E} (hx : IsMode w₀ w₁ w₂ m₀ m₁ m₂ x) (hy : IsMode w₀ w₁ w₂ m₀ m₁ m₂ y) :
    dens w₀ w₁ w₂ m₀ m₁ m₂ x = dens w₀ w₁ w₂ m₀ m₁ m₂ y
```

**Proposition 3.4.**

```
theorem deriv2_midLog_zero (d : ℝ) :
    deriv (deriv (midLog d)) 0 = d ^ 2 * (d ^ 2 / 4 - 1)

variable {μ₀ μ₁ μ₂ : E}
theorem two_component_not_concaveOn (hd : 2 < ‖μ₀ - μ₁‖) :
    ¬ ConcaveOn ℝ Set.univ (fun x : E => Real.log (dens 1 1 0 μ₀ μ₁ μ₂ x))

theorem line_not_concaveOn :
    ¬ ConcaveOn ℝ Set.univ (fun x : ℝ => Real.log (dens 1 1 0 (0 : ℝ) 3 0 x))
```

**Proposition 3.5.**

```
theorem control_coincident_isMode (hw : 0 < w) (m : E) : IsMode w w w m m m m

theorem control_spread_two {e : E} (he : ‖e‖ = 1) :
    ‖(0 : E) - e‖ ≤ 2 ∧ ‖(0 : E) - (2 : ℝ) • e‖ ≤ 2 ∧ ‖e - (2 : ℝ) • e‖ ≤ 2 ∧
      ‖(0 : E) - (2 : ℝ) • e‖ = 2
```

**Propositions 4.1, 4.2, 4.3.** (variables `{L w : ℝ}` `{m₀ m₁ m₂ : E}`, and `(hw : 0 < w)` for the first two)

```
lemma deriv_axisLog_zero : deriv (axisLog L w) 0 = 0

theorem deriv2_axisLog_zero :
    deriv (deriv (axisLog L w)) 0 = L ^ 2 / 3 * (L ^ 2 / 6 - 1)

theorem barycentre_not_isMode (h : CentredEquilateral L m₀ m₁ m₂) (hw : 0 < w) (hL : 6 < L ^ 2) :
    ¬ IsMode w w w m₀ m₁ m₂ (0 : E)

theorem logdens_not_concaveOn (h : CentredEquilateral L m₀ m₁ m₂) (hw : 0 < w) (hL : 6 < L ^ 2) :
    ¬ ConcaveOn ℝ Set.univ (fun x : E => Real.log (dens w w w m₀ m₁ m₂ x))

namespace R3  -- v₀ v₁ v₂ : EuclideanSpace ℝ (Fin 3) are ![2,-1,-1], ![-1,2,-1], ![-1,-1,2]
lemma centred_equilateral : CentredEquilateral (3 * Real.sqrt 2) v₀ v₁ v₂

theorem not_concaveOn :
    ¬ ConcaveOn ℝ Set.univ
      (fun x : EuclideanSpace ℝ (Fin 3) => Real.log (dens 1 1 1 v₀ v₁ v₂ x))

theorem barycentre_not_isMode : ¬ IsMode 1 1 1 v₀ v₁ v₂ (0 : EuclideanSpace ℝ (Fin 3))
```

**Theorems 5.2 and 5.3.** In namespace `Homoscedastic3mix.Mode`, with `mu : Fin 3 → Fin 2 → ℚ` the centres and `w : Fin 3 → ℚ` the weights, the coordinate density and the centres as points of `EuclideanSpace ℝ (Fin 2)` are

```
noncomputable def dns (x : Fin 2 → ℝ) : ℝ :=
  ∑ j, (w j : ℝ) * Real.exp (-(∑ a, (x a - (mu j a : ℝ)) ^ 2) / 2)

noncomputable def ctrE (j : Fin 3) : EuclideanSpace ℝ (Fin 2) :=
  WithLp.toLp 2 (fun a => ((mu j a : ℚ) : ℝ))
```

Witness A (`WitA`):

```
def mu : Fin 3 → Fin 2 → ℚ := ![![104/75, 0], ![-52/75, 6/5], ![-52/75, -6/5]]
def wg : Fin 3 → ℚ := ![1, 1, 1]
def ctr : Fin 4 → Fin 2 → ℚ :=
  ![![-4429251/10000000, -1531047/2000000],
    ![-4429251/10000000,  1531047/2000000],
    ![-115193/10000000, 0],
    ![8857247/10000000, 0]]
def rad : ℚ := 1/1000
def Kc : ℚ := 39/40
def bc : ℚ := 1/1000000

theorem unique_critical (i : Fin 4) :
    ∃! q : Fin 2 → ℝ,
      q ∈ Metric.closedBall (fun a => ((ctr i a : ℚ) : ℝ)) ((rad : ℚ) : ℝ) ∧
        fderiv ℝ (dns mu wg) q = 0

theorem four_modes :
    ∃ p : Fin 4 → (Fin 2 → ℝ), Function.Injective p ∧
      (∀ i, IsLocalMax (dns mu wg) (p i)) ∧
      (∀ i a, |p i a - ((ctr i a : ℚ) : ℝ)| ≤ ((rad : ℚ) : ℝ))

theorem four_isMode :
    ∃ P : Fin 4 → EuclideanSpace ℝ (Fin 2), Function.Injective P ∧
      ∀ i, IsMode ((wg 0 : ℚ) : ℝ) ((wg 1 : ℚ) : ℝ) ((wg 2 : ℚ) : ℝ)
        (ctrE mu 0) (ctrE mu 1) (ctrE mu 2) (P i)
```

Witness B (`WitB`): the same three theorems with

```
def mu : Fin 3 → Fin 2 → ℚ :=
  ![![1315877/1000000, -31276/1000000],
    ![-667324/1000000, 1217155/1000000],
    ![-750037/1000000, -1263048/1000000]]
def wg : Fin 3 → ℚ := ![1, 1051434/1000000, 1107307/1000000]
def ctr : Fin 4 → Fin 2 → ℚ :=
  ![![-2781689/5000000, -9554167/10000000],
    ![-4221259/10000000, 2014671/2500000],
    ![5283/100000, 70331/10000000],
    ![2904623/5000000, -697/100000]]

theorem diameter_sq_gt_six : (6 : ℚ) < ∑ a, (mu 1 a - mu 2 a) ^ 2
```

together with `sides` (the three squared side lengths as exact rationals) and `asymmetric` (the three squared side lengths pairwise distinct and the three weights pairwise distinct). The certificate underlying both is the theorem `Homoscedastic3mix.Mode.cert_mode`, whose hypotheses are the four checks (i)–(iv) of Lemma 5.1 as decidable statements about rational intervals (`hS`, `hSpt`; `hKrow` with `hK1 : K < 1`; `hbeta`; `hstep : beta + K * r < r`) and whose conclusion is

```
    ∃ p : Fin 2 → ℝ,
      p ∈ Metric.closedBall (fun i => ((c i : ℚ) : ℝ)) (r : ℝ) ∧
      fmap mu w p = 0 ∧
      (∀ x ∈ Metric.closedBall (fun i => ((c i : ℚ) : ℝ)) (r : ℝ), fmap mu w x = 0 → x = p) ∧
      (∀ x ∈ Metric.closedBall (fun i => ((c i : ℚ) : ℝ)) (r : ℝ), dns mu w x ≤ dns mu w p) ∧
      IsLocalMax (dns mu w) p
```

where `fmap mu w x = fun a => mean mu w a x - x a` is $m(x) - x$.

**The controls of Section 5.4.** (namespace `Mode.Controls`, in the data of witness A; not bound to a ledger row)

```
theorem wrong_sign_fails :
    1 < ∑ j, (krawItv (fun i j => if i = j then (1 : ℚ) else 0)
        (fun a b => jacI mu wg prc mdc a b (ballBox (ctr 2) rad radNN)) 0 j).mag

theorem ball_too_large_fails :
    1 < ∑ j, (krawItv negId
        (fun a b => jacI mu wg prc mdc a b (ballBox (ctr 2) (1/50) (by norm_num))) 0 j).mag

def saddle : Fin 2 → ℚ := ![1427732/10000000, 0]
theorem saddle_fails :
    1 < ∑ j, (krawItv negId
        (fun a b => jacI mu wg prc mdc a b (ballBox saddle rad radNN)) 0 j).mag
```

(`prc = 40`, `mdc = 12`.) The Lean statements assert only that each row sum exceeds 1; the exact rationals and decimal values quoted in Section 5.4 were computed separately in exact rational arithmetic.

## References


[1] C. Améndola, A. Engström, C. Haase, Maximum number of modes of Gaussian mixtures, Information and Inference: A Journal of the IMA 9 (2020), no. 3, 587–600; arXiv:1702.05066v3. Example 4, Conjecture 5, Theorem 10 and Section 2 are numbered as in the arXiv v3 e-print.

[2] J. Behboodian, On the modes of a mixture of two normal distributions, Technometrics 12 (1970), no. 1, 131–139.

[3] M. Á. Carreira-Perpiñán, C. K. I. Williams, An isotropic Gaussian mixture can have more modes than components, Informatics Research Report EDI-INF-RR-0185, School of Informatics, University of Edinburgh, December 2003.

[4] M. Á. Carreira-Perpiñán, C. K. I. Williams, On the number of modes of a Gaussian mixture, Informatics Research Report EDI-INF-RR-0159, School of Informatics, University of Edinburgh, February 2003.

[5] M. Cule, R. Samworth, M. Stewart, Maximum likelihood estimation of a multi-dimensional log-concave density, Journal of the Royal Statistical Society, Series B 72 (2010), no. 5, 545–607 (with discussion). The sentence quoted in Section 3 is from the arXiv version, arXiv:0804.3989v1, Section 2.

[6] H. Edelsbrunner, B. T. Fasy, G. Rote, Add isotropic Gaussian kernels at own risk: more and more resilient modes in higher dimensions, Discrete & Computational Geometry 49 (2013), no. 4, 797–822. Equation (16), the Barycenter Lemma and Section 4 are cited by their numbering in the authors' manuscript version ("Discrete and Computational Geometry manuscript No."); the journal typesetting was not consulted.

[7] Y. Kabata, H. Matsumoto, A. Okuno, At least seven modes in a heteroscedastic three-component bivariate Gaussian mixture, arXiv:2608.01776v1 (2026). Remark 2 is numbered as in v1.

[8] H. D. Nguyen, Bounds on the number of modes of a Gaussian mixture density, arXiv:2605.15531v1 (2026). Corollary 3.20 and Theorem 4.11 are numbered as in v1.

[9] A. Okuno, Y. Kabata, On mixtures of three homoscedastic Gaussian densities: an unconditional sharper bound on the number of modes, arXiv:2607.17506v1 (2026). Section 2.2, Theorems 1 and 2, Lemma 3 and Figure 1 are numbered as in v1.

[10] R. J. Samworth, Recent progress in log-concave density estimation, Statistical Science 33 (2018), no. 4, 493–509; arXiv:1709.03154v1. Section 10.2 is numbered as in the arXiv version.