CRITICAL POINTS AND MODES OF THE KABATA–MATSUMOTO–OKUNO SEVEN-MODE GAUSSIAN MIXTURE

KOREA SUPERINTELLIGENCE LABS

Abstract. Kabata, Matsumoto and Okuno (arXiv:2608.01776) constructed an equally weighted three-component bivariate Gaussian mixture $p_{r,M}$ with at least seven modes for $r \in [97/100, 99/100]$ and $M \geq 10^4$, refuting the conjectured maximum $\binom{d+k-1}{d} = 6$ recorded by Améndola, Engström and Haase, and explicitly declined to say whether the seven modes are all of them. We settle two parts of that question. First, no mode lies on the three long ridges of the mixture: for $r \in [97/100, 99/100]$ and $1000 \leq M \leq 10^6$, no point x with $|n_j \cdot x - r| \leq 1/2$ and $16 \leq t_j \cdot x \leq M - 1$ is a local maximum. Second, for $r = 97/100$ and *every* $M \geq 10^4$, the mixture has exactly seven critical points in the rhombus $\Omega = \{|n_1 \cdot x| \leq 2,\ |n_2 \cdot x| \leq 2\}$, which contains the closed disc of radius 2 and hence the central triangle bounded by the three major axes; exactly four of them are modes (the origin and one near each pairwise intersection of the major axes), and the other three are saddle points. The count is uniform in M because a change of coordinates makes the gradient system rational in $(n_1 \cdot x,\ n_2 \cdot x,\ 1/(\sqrt{3}M))$; the certificate is an exclusion covering by 1451 boxes under a preconditioned (Krawczyk) test, together with 7×4 Krawczyk existence-and-uniqueness certificates. Every statement, including the seven-mode existence theorem itself (re-proved by a derivative-free argument valid from $M \geq 1000$), is verified in Lean 4 with Mathlib. The mode count on the whole plane, and the count at other values of r, remain open; a numerical Newton census, reported but not proved, finds thirteen critical points in the plane—seven maxima and six saddles—for every tested $r \in [0.9, 0.99]$.

1. Introduction

1.1. **The maximal mode problem.** A *mode* of a density is a local maximum point. Let $m(d, k)$ denote the maximum number of modes of a mixture of k Gaussian densities on $\mathbb{R}^d$. Améndola, Engström and Haase [1] record the following as Conjecture 5 (in the numbering of the arXiv version v3, the final version; attributed there to Sturmfels, AIM 2011):

> For all $d, k \geq 1$, $m(d, k) = \binom{d+k-1}{d}$.

For $(d, k) = (2, 3)$ the conjectured value is 6. The exact values known are $m(1, k) = k$, $m(d, 1) = 1$ and $m(d, 2) = d + 1$ (as recalled by Améndola and Rodriguez [2] after their Problem 6.7); the first unknown value is $m(2, 3)$.

1.2. **The seven-mode mixture.** Kabata, Matsumoto and Okuno [4] refuted the conjecture at $(2, 3)$ with an explicit family. Let R_θ be the counterclockwise rotation through θ, and

$$n_1 = \left(\tfrac{\sqrt{3}}{2}, \tfrac{1}{2}\right), \quad n_2 = \left(-\tfrac{\sqrt{3}}{2}, \tfrac{1}{2}\right), \quad n_3 = (0, -1), \qquad t_i = R_{\pi/2} n_i, \tag{1}$$

so that $t_1 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$, $t_2 = (-\frac{1}{2}, -\frac{\sqrt{3}}{2})$, $t_3 = (1, 0)$. For parameters $r \in [97/100, 99/100]$ and $M \geq 10^4$ the i-th component has mean $\mu_i = rn_i + Mt_i$ and covariance $\Sigma_i = n_i n_i^\mathsf{T} + M^2 t_i t_i^\mathsf{T}$; since $\det \Sigma_i = M^2$ for all i, the equally weighted mixture density is

$$p_{r,M}(x) = \frac{f(x)}{2\pi M}, \qquad f(x) = \frac{1}{3}\sum_{i=1}^{3} \exp\Big(-\tfrac{1}{2} q_i(x)\Big), \qquad q_i(x) = (n_i \cdot x - r)^2 + \frac{(t_i \cdot x - M)^2}{M^2}. \tag{2}$$

The line $\ell_i :\ n_i \cdot x = r$ contains the major axis of the i-th component; the three lines bound an equilateral triangle with inscribed radius r, and $p^* := (0, 2r) = \ell_1 \cap \ell_2$. The configuration is invariant under $R_{2\pi/3}$, which permutes the components cyclically, and has no reflection symmetry. Theorem 1 of [4] (v1, p. 3) reads, verbatim:

> For every $r \in \left[\frac{97}{100}, \frac{99}{100}\right]$, $M \geq 10^4$, the density $p_{r,M}$ has at least seven distinct nondegenerate local maximum points. More precisely, the origin is a local maximum point, and each of

the closed balls $\bar{B}(\mu_i, \frac{1}{2})$ $(i = 1, 2, 3)$, $\bar{B}(R_{2\pi k/3}p^*, \frac{2}{5})$ $(k = 0, 1, 2)$ contains a nondegenerate local maximum point in its interior.

Whether there are further modes is left open in [4]; its Remark 1 (p. 4) says, verbatim, "We do not investigate whether $p_{r,M}$ has modes other than the seven modes identified above, nor do we claim that its set of modes is finite", and its conclusion (§4, p. 9) ends with "Determining the exact maximum number of modes in this setting, or even whether it is finite, remains an open problem."

The finiteness half of that remark has since been settled in general: Améndola and Rodriguez [2] prove that every finite Gaussian mixture density has finitely many modes, and Okuno [7] proves unconditional finiteness of the critical set of every finite heteroscedastic Gaussian mixture; Wang [9] proves finiteness of the critical set for isotropic mixtures (equivalently, after a linear change of variables, homoscedastic ones), which does not cover the mixture (2). Upper bounds are far from 7: Nguyen [5] (Corollary 3.13 with Table 2) bounds the number of nondegenerate critical points of a heteroscedastic three-component bivariate mixture by 392, hence the number of modes by $\lfloor(392 + 1)/2\rfloor = 196$ when the modal set is finite (numbering of v1), and [2] records the state of the art, in the sentence following its Problem 6.7, as $7 \le m(2, 3) \le 196$. For *homoscedastic* three-component mixtures Okuno and Kabata [6] prove at most 15 critical points (their Theorem 1, which states that no non-degeneracy assumption is imposed) and at most 8 modes (their Theorem 2); the mixture (2) is heteroscedastic ([4], Remark 2), so that bound does not apply to it. No published result counts the modes of $p_{r,M}$ itself.

1.3. **Results.** Write $d_i(x) = n_i \cdot x - r$ for the signed distance from x to ℓ_i and $s_i(x) = t_i \cdot x$ for the coordinate along the i-th major axis. Each component's *ridge*—the strip $|d_i| \le \frac{1}{2}$ along ℓ_i between the central triangle and the mean μ_i—has length about M, and it is exactly where an eighth mode could hide: the ridgeline localisation of Ray and Lindsay [8] (restated as Lemma 2.5 of [5]) confines the critical points to a 2-dimensional set when $(d, k) = (2, 3)$ and therefore excludes nothing, and the finiteness results just cited are not quantitative.

- **The seven modes, derivative-free (Theorem 3.1).** For $r \in [97/100, 99/100]$ and $M \ge 1000$, f (equivalently $p_{r,M}$) has at least seven distinct local maximum points. This is Theorem 1 of [4] with the window widened from $M \ge 10^4$ to $M \ge 1000$; it is proved by comparing values on the boundaries of explicit compact sets, so no derivative of f is formed, at the price of not giving nondegeneracy. It is not claimed as new mathematics.
- **No mode on the mid-ridge (Theorem 4.1).** For $r \in [97/100, 99/100]$ and $1000 \le M \le 10^6$, no point with $|d_j(x)| \le \frac{1}{2}$ and $16 \le s_j(x) \le M - 1$ is a local maximum of f, for any j; along the first ridge, stepping towards the mean strictly increases f.
- **Exactly seven critical points and exactly four modes near the centre (Theorems 5.1 and 5.2).** Let $r = 97/100$ and

$$\Omega := \{x \in \mathbb{R}^2 : \ |n_1 \cdot x| \le 2 \text{ and } |n_2 \cdot x| \le 2\}, \tag{3}$$

  a rhombus containing the closed disc of radius 2 and hence the closed triangle $\bar{T} = \{x : n_i \cdot x \le r \ \forall i\}$ (whose vertices $R_{2\pi k/3}p^*$ are at distance $2r = 1.94$ from the origin). For every $M \ge 10^4$, f has exactly seven critical points in Ω, exactly four of which are local maxima (for f and for $p_{r,M}$ alike); the other three are not local maxima, and the Hessian determinant is negative throughout a box around each of them, so they are saddle points.
- **Negative controls (Proposition 6.1).** Among them: for $r \ge 21/20$ the origin is not a mode, so the hypothesis $r < 1$ is load-bearing.

The two counts are about different regions, and neither contradicts Theorem 1 of [4]: of the seven modes located there, the origin and the one in $\bar{B}(p^*, \frac{2}{5})$ lie in Ω, the two in $\bar{B}(R_{2\pi k/3}p^*, \frac{2}{5})$ $(k = 1, 2)$ are the rotations of the latter and also lie in Ω, and the three in $\bar{B}(\mu_i, \frac{1}{2})$ lie far outside (Remark 5.4). What remains open is the count on the rest of the plane—the part of each ridge with $s_j < 16$ and the neighbourhoods of the means—and the whole r-window for the exact count, whose price is discussed in §5.5.

1.4. **Method.** Theorems 3.1, 4.1 and 6.1 are elementary and derivative-free. Theorems 5.1 and 5.2 rest on a certificate: in the coordinates $\theta_i = n_i \cdot x$ and $w = 1/(\sqrt{3}M)$ the gradient system becomes rational in (θ_1, θ_2, w), with w ranging over the single interval $[0, 2^{-14}]$ for all $M \ge 10^4$; an adaptive bisection of $[-2, 2]^2 \times [0, 2^{-14}]$ into 1451 boxes excludes zeros from every box not contained in one of seven small balls, and a Krawczyk operator proves that each ball contains exactly one zero and classifies it by the sign of a Hessian enclosure. Exact critical-point censuses of explicit Gaussian mixtures are not new as a genre—Edelsbrunner, Fasy and Rote [3] carried one out, with indices and an Euler-characteristic check, for an isotropic family—and interval Newton methods are classical [10]; what is new is the count for this family, uniformly in M. All statements are machine-checked (§8).

2. Preliminaries

Throughout, $f = f_{r,M}$ is the function of (2), "mode" means local maximum point (as in [4]), and $x \cdot y$ is the Euclidean inner product on $\mathbb{R}^2$. Since $p_{r,M} = f/(2\pi M)$ with $M > 0$, f and $p_{r,M}$ have the same modes and the same critical points. The unit vectors satisfy $n_1 + n_2 + n_3 = 0$, $n_i \cdot n_j = -\frac{1}{2}$ for $i \neq j$, and (n_i, t_i) is an orthonormal frame, so

$$\sum_i d_i(x) = -3r, \qquad \sum_i d_i(x)^2 = \tfrac{3}{2}\|x\|^2 + 3r^2, \qquad \sum_i s_i(x)^2 = \tfrac{3}{2}\|x\|^2. \tag{4}$$

Expanding $q_i = d_i^2 + 1 - 2\varepsilon_i$ with $\varepsilon_i = s_i/M - s_i^2/(2M^2)$ gives

$$f(x) = \frac{e^{-1/2}}{3} \sum_i \exp\Big(-\tfrac{1}{2} d_i(x)^2 + \varepsilon_i(x)\Big), \qquad |\varepsilon_i(x)| = O(\|x\|/M), \tag{5}$$

which is the form in which all value comparisons below are made. The rotation $R := R_{2\pi/3}$ satisfies $Rn_1 = n_2$, $Rn_2 = n_3$, $Rn_3 = n_1$ and $f \circ R = f$, so local maxima are permuted by R.

The formal development indexes the components by $0, 1, 2$; its n_0, n_1, n_2 are the n_1, n_2, n_3 of (1). In particular the region (3) is written there with n_0 and n_1. We use the indexing of [4] in the prose and the formal indexing only in the listing of §8.

3. Seven modes by value comparison

Theorem 3.1. *Let $r \in [97/100, 99/100]$ and $M \geq 1000$. There are seven pairwise distinct points $z_1, \ldots, z_7 \in \mathbb{R}^2$, each of which is a local maximum point of $f_{r,M}$. The same holds with $p_{r,M}$ in place of $f_{r,M}$.*

For $M \geq 10^4$ the existence statement is Theorem 1 of [4], whose proof localises each mode by a strong-concavity lemma on a ball (its Lemma 1) and thereby also obtains nondegeneracy. The proof here is different and gives less: modes, but not nondegeneracy. The gain is the window $M \geq 1000$ and the absence of any derivative.

Proof sketch. The device is the following. If $K \subset \mathbb{R}^2$ is compact with interior U, $p \in U$, and $f < f(p)$ on $K \setminus U$, then a maximiser z of f on K lies in U, and since U is a neighbourhood of z on which $f \leq f(z)$, z is a local maximum point. Seven such pairs (K, p) are exhibited.

The means. For $K = \{|d_1| \leq \frac{1}{2},\ |s_1 - M| \leq M/2\}$ and $p = \mu_1$: on the boundary either $|d_1| = \frac{1}{2}$ or $|s_1 - M| = M/2$, so $q_1 \geq \frac{1}{4}$ and the first component contributes at most $\frac{1}{3}e^{-1/8} \leq \frac{1}{3} \cdot 0.8825$; the other two components are at distance at least $\frac{\sqrt{3}}{4}M - \frac{7}{4} \geq 431$ from their own axis lines and contribute at most 10^{-4} each. Hence $f < \frac{1}{3} \leq f(\mu_1)$ on the boundary. The other two means follow by the symmetry $f \circ R = f$.

The central points. For the rhombus $K = \{|d_1| \leq \frac{1}{2},\ |d_2| \leq \frac{1}{2}\}$ around p^*: on K one has $\|x\|^2 \leq 10$, so $|\varepsilon_i| \leq 1/250$ for $M \geq 1000$ by (5), and the first two terms of the sum in (5) at p^* already give $\Phi(p^*) \geq 2 \cdot \frac{249}{250} = 1.992$, where $\Phi = \sum_i \exp(-d_i^2/2 + \varepsilon_i)$. On the boundary one of $|d_1|, |d_2|$ equals $\frac{1}{2}$ and contributes at most $e^{-1/8} \leq 0.8825$; since $\sum_i d_i = -3r$ forces $|d_3| \geq 2.41 + d_2$, a five-way split of $d_2 \in [-\frac{1}{2}, \frac{1}{2}]$ bounds the other two terms by 1.09535, and $\Phi \leq \frac{250}{249} \cdot 1.97785 = 1.98580 < 1.992$. Nine rational bounds on e^{-a} (for $a = \frac{1}{8}, \frac{9}{128}, \frac{1}{32}, \frac{1}{128}, 1.824, 2.071, 2.333, 2.611, 2.904$) are obtained from truncated Taylor series with 16–20 terms.

The origin. With $u_i = -\theta_i^2/2 + r\theta_i + \varepsilon_i$ and $\theta_i = n_i \cdot x$ one has $f(x) = \frac{1}{3}e^{-(1+r^2)/2} \sum_i e^{u_i}$, $f(0) = e^{-(1+r^2)/2}$, and, by the identities (4), $\sum_i u_i = -\frac{3}{4}\|x\|^2(1 + M^{-2})$ exactly. The bound $e^u \leq 1 + u + u^2/2 + \frac{2}{9}|u|^3$ (Mathlib's `Real.exp_bound` at order 3) together with $\sum_i u_i^2 \leq 1.479496\|x\|^2$ gives $\sum_i e^{u_i} - 3 \leq -0.00943\|x\|^2 < 0$ for $0 < \|x\| \leq 1/400$, so the origin is a strict local maximum. The ball must be this small: the margin is $\frac{3}{4}(1 - r^2) - O(\|x\|)$, and the nearest saddle point is at distance 0.126 (§5).

Distinctness. The seven points are separated by their *profiles*—which of $|d_1|, |d_2|, |d_3|$ are $< \frac{1}{2}$: 000 for the origin, $110, 011, 101$ for the three central points, and $100, 010, 001$ for the three near the means. Proposition 6.1(ii) shows that no point has profile 111.

Range. The binding constraint is $|\varepsilon_i| \le 1/250$ on the rhombus, which needs $\|x\|/M \le 3.05 \cdot 10^{-3}$; at $M = 500$ the central inequality fails. □

4. No mode on the mid-ridge

Theorem 4.1. *Let $r \in [97/100, 99/100]$ and $1000 \le M \le 10^6$.*

(a) *If $x \in \mathbb{R}^2$ satisfies $|d_1(x)| \le \frac{1}{2}$ and $16 \le s_1(x) \le M-1$, then $f(x+\tau t_1) > f(x)$ for every $\tau \in (0,1]$.*

(b) *Consequently, for each $j \in \{1,2,3\}$, no x with $|d_j(x)| \le \frac{1}{2}$ and $16 \le s_j(x) \le M-1$ is a local maximum point of f.*

The strip in (b) is the whole of the j-th ridge except its first 16 units and its last unit before the mean. Only the initial piece of the ridge with $s_j \le 3.16$ meets the region Ω of §5: on the axis line ℓ_j one has $n_k \cdot x = -\frac{r}{2} \pm \frac{\sqrt{3}}{2} s_j$ for $k \ne j$, so ℓ_j leaves Ω at $s_j = (2+\frac{r}{2})\frac{2}{\sqrt{3}} = 2.87$ for $r = 97/100$, and the edges $|d_j| = \frac{1}{2}$ of the strip at $s_j = 2.58$ and 3.16. Nothing in [4] speaks about this region.

Proof sketch. Step by τ along t_1. Writing $c_j = n_j \cdot t_1$ and $e_j = t_j \cdot t_1$, the exact change of the j-th exponent is

$$q_j(x) - q_j(x+\tau t_1) = -\big(2\tau d_j c_j + \tau^2 c_j^2\big) - \frac{2\tau(s_j - M)e_j + \tau^2 e_j^2}{M^2}.$$

For the own component $c_1 = 0$, $e_1 = 1$, so the change is $[2\tau(M-s_1) - \tau^2]/M^2 \ge \tau/M^2$; since $q_1 \le \frac{5}{4}$ on the strip, $e^{-q_1/2} \ge \frac{3}{8}$, and $1 + u \le e^u$ turns this into a gain of at least $\frac{3}{16}\tau/M^2 \ge 1.875 \cdot 10^{-13}\tau$. The two foreign components sit at distance at least 12 from their own axis lines (from $n_j \cdot x = -\frac{1}{2} n_1 \cdot x \pm \frac{\sqrt{3}}{2} s_1$ and $s_1 \ge 16$), and $|d| e^{-d^2/2} \le \frac{8}{3} d^{-2} e^{-36} \le 10^{-17}$ there, using $e^{-36} = (e^{-9/2})^8 \le (0.0112)^8$. Four monomial bounds then bound the total foreign loss by $2 \cdot 10^{-16}\tau$, which the gain beats by a factor of about 900 even at $M = 10^6$. The point of the argument is that the loss bound is $O(\tau)$: it comes from applying $1 + u \le e^u$ to $u = (q_j(x) - q_j(x+\tau t_1))/2$, which is itself $O(\tau)$, whereas the naive bound "a foreign term loses at most its own value" is $O(1)$ and says nothing as $\tau \to 0$, the only regime that refutes a local maximum. Part (b) for $j = 1$ is immediate from (a) (the points $x + \tau t_1$ approach x); for $j = 2, 3$ apply R. The cap $M \le 10^6$ is only because the gain $\sim M^{-2}$ must beat the fixed loss 10^{-16}; raising the threshold 16 to 20 would move it past 10^{12}. □

5. The exact count near the centre

5.1. **Rational coordinates.** The only obstacle to exact rational arithmetic in (2) is the $\sqrt{3}$ in the vectors n_1, n_2, t_1, t_2. Put $\theta_i = n_i \cdot x$ and take $z = (\theta_1, \theta_2)$ as coordinates ($\theta_3 = -\theta_1 - \theta_2$); the inverse map is $\psi(z) = \big((z_1 - z_2)/\sqrt{3},\ z_1 + z_2\big)$. Then $d_j = \theta_j - r$ is linear with rational coefficients, and

$$\sqrt{3}\,(t_j \cdot x) = S_j(z), \quad S_1 = \theta_1 + 2\theta_2, \quad S_2 = -2\theta_1 - \theta_2, \quad S_3 = \theta_1 - \theta_2.$$

With the single scalar

$$w := \frac{1}{\sqrt{3}\,M}, \qquad M \ge 10^4 \iff w \le \frac{1}{\sqrt{3} \cdot 10^4} = 5.7735 \cdot 10^{-5} < 2^{-14}, \tag{6}$$

one has $(t_j \cdot x)/M = S_j w$ and $q_j = d_j^2 + (S_j w - 1)^2$, entirely rational in (θ_1, θ_2, w). The parameter M enters only through w, so the interval $W := [0, 2^{-14}]$ covers every $M \ge 2^{14}/\sqrt{3} = 9459.3\ldots$ at once; the theorems are stated for $M \ge 10^4$ to match [4]. Let $G_w(z) = \frac{1}{3}\sum_j e^{u_j}$ with $u_j = -q_j/2$, so that $G_w \circ \theta = f$, and let

$$F_i = \sum_j e^{u_j} P_{ij}, \qquad P_{ij} = d_j U_{ij} + (S_j w - 1)\, w\, V_{ij} = -\partial u_j / \partial z_i,$$

$$H_{ik} = \sum_j e^{u_j}\big(-P_{ij} P_{kj} + U_{ij} U_{kj} + w^2 V_{ij} V_{kj}\big),$$

where $U_{ij} = \partial d_j/\partial z_i$ and $V_{ij} = \partial S_j/\partial z_i$ are the integer matrices $\left(\begin{smallmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & -2 & 1 \\ 2 & -1 & -1 \end{smallmatrix}\right)$. Thus $F = \nabla_z(-3G_w)$ and H is its Jacobian, the Hessian of $-3G_w$. Both are expressions in the constants, the variables θ_1, θ_2, w, and $+, -, \times, (\cdot)^2, \exp$.

5.2. **Statements.** In the coordinates z, the region (3) is the square $[-2, 2]^2$. The certificate names seven boxes $B_k = \{z : \|z - c_k\|_\infty \le \rho_k\}$, $k = 0, \dots, 6$, with rational centres of denominator 2^{24}; Table 1 lists them, with $\rho_k = 1/256$ for the four central points and $\rho_k = 1/64$ for the three outer ones.

k	$c_k = (\theta_1, \theta_2)$ (exact)	c_k (decimal)	$\psi(c_k) \approx x$	type
0	$(0,\ 0)$	$(0,\ 0)$	$(0,\ 0)$	maximum
1	$(1056621,\ 1056627)/2^{24}$	$(0.06298,\ 0.06298)$	$(0.0000,\ 0.1260)$	saddle
2	$(-66039/2^{19},\ 1056621/2^{24})$	$(-0.12596,\ 0.06298)$	$(-0.1091,\ -0.0630)$	saddle
3	$(1056627/2^{24},\ -66039/2^{19})$	$(0.06298,\ -0.12596)$	$(0.1091,\ -0.0630)$	saddle
4	$(15335929/2^{24},\ 7668389/2^{23})$	$(0.91409,\ 0.91414)$	$(0.0000,\ 1.8282)$	maximum
5	$(-30672707/2^{24},\ 15335929/2^{24})$	$(-1.82824,\ 0.91409)$	$(-1.5833,\ -0.9141)$	maximum
6	$(7668389/2^{23},\ -30672707/2^{24})$	$(0.91414,\ -1.82824)$	$(1.5833,\ -0.9141)$	maximum

TABLE 1. The seven certified boxes at $r = 97/100$. Each B_k contains exactly one critical point of G_w for every $w \in [0, 2^{-14}]$; the type is that of the critical point.

Theorem 5.1. *Let $r = 97/100$. For every $M \ge 10^4$ the set $\{x \in \Omega : \nabla f_{r,M}(x) = 0\}$ has exactly seven elements. Equivalently, for every $w \in [0, 2^{-14}]$ the set $\{z \in [-2, 2]^2 : F(z) = 0\}$ has exactly seven elements, one in each box B_k of Table 1.*

Theorem 5.2. *Let $r = 97/100$. For every $M \ge 10^4$ the set $\{x \in \Omega :\ x$ is a local maximum point of $f_{r,M}\}$ has exactly four elements, and the same holds with $p_{r,M}$ in place of $f_{r,M}$. In the coordinates z, these four are the critical points in B_0, B_4, B_5, B_6; the critical points in B_1, B_2, B_3 are not local maximum points, and on each of the boxes $\|z - c_k\|_\infty \le \frac{17}{16}\rho_k$, $k = 1, 2, 3$, the interval enclosure of $\det H$ is negative.*

Remark 5.3 (Nondegeneracy and the Morse count). The classification certificate bounds the Hessian enclosure of $-3G_w$ on the enlarged box $\frac{17}{16}B_k$: for $k \in \{0, 4, 5, 6\}$ it is positive definite, and for $k \in \{1, 2, 3\}$ its determinant is negative. Since the box contains the critical point, Hess f is negative definite at the four modes and has negative determinant at the other three; so all seven are nondegenerate, the four modes in the sense of Theorem 1 of [4], and the other three genuine saddle points (index 1). The formal statements are those of Theorem 5.2: the count, “local maximum” for four, “not a local maximum” for three, and the sign of the determinant enclosure; the nondegeneracy phrased pointwise is the immediate consequence just described (the soundness of the enclosure on the enlarged box, applied at the critical point itself, which the box contains by construction) and is not a separately stated formal theorem. The Morse count in Ω is $4 - 3 = 1$, the Euler characteristic of the plane, and this is also what the numerical census of §7 finds for the whole plane $(7 - 6 + 0 = 1)$. Before Theorem 5.2, “all seven critical points in Ω are modes” would have been consistent with everything previously known; Theorem 5.2 refutes it rather than leaving it unproved.

Remark 5.4 (Relation to Theorem 1 of [4]). The seven modes of Theorem 1 of [4] and the four of Theorem 5.2 are counted on different sets; the following elementary arithmetic, which is not part of the formal development, relates them. The closed ball $\bar{B}(p^*, \frac{2}{5})$ lies in Ω (on it $|n_1 \cdot x|, |n_2 \cdot x| \le r + \frac{2}{5} \le 1.39$), so its modes are among the four of Theorem 5.2; of the four certified boxes only B_4 meets it—$\psi(B_4)$ lies within distance 0.15 of p^* (its centre corresponds to $x \approx (0, 1.8282)$, at distance 0.112), while $\psi(B_0)$, $\psi(B_5)$, $\psi(B_6)$ stay at distance more than 1.9 from p^*—so $\bar{B}(p^*, \frac{2}{5})$ contains exactly one mode, the one certified in B_4. The balls $\bar{B}(R_{2\pi k/3}p^*, \frac{2}{5})$ for $k = 1, 2$ are the images of $\bar{B}(p^*, \frac{2}{5})$ under R and R^2, so by $f \circ R = f$ each contains exactly one mode, the corresponding rotation of the mode in B_4. In the coordinates z the rotation R acts as $(\theta_1, \theta_2) \mapsto (-\theta_1 - \theta_2, \theta_1)$; the certified centres satisfy $c_5 = Rc_4$ and $c_6 = R^2 c_4$ exactly, the two rotated modes have $\|z - c_5\|_\infty$, resp. $\|z - c_6\|_\infty$, at most $\frac{1}{32}$, so they lie in Ω and, being neither the origin nor the mode in B_4 nor each other, are the modes certified in B_5 and B_6. The three means satisfy $|n_j \cdot \mu_i| = |-\frac{r}{2} \pm \frac{\sqrt{3}}{2}M| \ge \frac{\sqrt{3}}{2}M - \frac{1}{2} > 8659$ for $j \ne i$, so $\bar{B}(\mu_i, \frac{1}{2}) \cap \Omega = \emptyset$: the three modes near the means are outside Ω. Hence the four modes of Theorem 5.2 are the origin and the three modes of [4] near the pairwise intersections of the major axes, and Theorem 1 of [4] adds the three near the means; Theorem 4.1 excludes the ridges with $16 \le s_j \le M-1$. Whether f has modes elsewhere—on the initial segments $2 < \|x\|$, $s_j < 16$ of the ridges, or within distance $\frac{1}{2}$ of the means other than the one already known—is not settled here (§5.5).

5.3. **The certificate layer.** Two shared verified components of the development, `Cert/Interval` and `Cert/IntervalDeriv`, are used, exactly as follows.

Rational interval arithmetic and an enclosure of exp. An interval is a pair $I = [\mathrm{lo}, \mathrm{hi}]$ of rationals; membership $x \in I$ is for real x. The operations $+, -, \times$, a sharp square, and outward dyadic rounding to denominators 2^p come with soundness lemmas of the form $x \in I,\ y \in J \Rightarrow x \odot y \in I \odot J$. The enclosure of exp is total and hypothesis-free: at a rational point s, halve s to $|s/2^k| \le 1$ with $k = \lceil |s| \rceil$, apply the Taylor bound $|\exp t - \sum_{m<n} t^m/m!| \le |t|^n (n+1)/(n!\,n)$ for $|t| \le 1$ (Mathlib's `Real.exp_bound`), and square back k times with outward rounding between squarings; on an interval, evaluate at the two endpoints and use monotonicity. Expressions in the constants, variables, $+, -, \times, (\cdot)^2, \exp$ are evaluated on a box by structural recursion, with the single soundness theorem "the real value lies in the computed interval". A box on which the computed interval of some component of F excludes 0 contains no zero of F.

Krawczyk existence and uniqueness from a Jacobian enclosure. For $F : \mathbb{R}^n \to \mathbb{R}^n$ with an interval enclosure J of its Jacobian on a sup-norm box B of centre c and radius ρ, a rational matrix Y with an explicit rational left inverse Z, a rational K bounding the maximal absolute row sum of the interval matrix $\mathrm{Id} - YJ$, and a rational $\beta \ge \|YF(c)\|_\infty$: if $K < 1$ and $\beta + K\rho \le \rho$, then B contains *exactly one* zero of F. The proof is the classical one: $x \mapsto x - YF(x)$ is a K-contraction of B into itself (mean value inequality on the convex box, with the ℓ^1 row-sum norm as the dual of the sup norm), Banach's theorem gives a fixed point, and $K < 1$ gives uniqueness; the contraction estimate is the lemma called `kraw_dist_le` below, and the packaged statement `exists_unique_zero_in_box`.

The preconditioned exclusion test. A naive exclusion covering does not close: near the origin the three terms of F cancel to $\lambda_{\min}|z|$ with $\lambda_{\min} = 0.0224$ (the Jacobian eigenvalues at the origin are 0.0224 and 0.0672), while the width of the interval enclosure of F on a box of side h is about $2h$, so every box with $|z| \lesssim 10^2 h$ is undecided at every level. Instead, for a box with centre c (in the (θ_1, θ_2) coordinates; w is kept as an interval), let J be the enclosure of H on the box, Y the exact rational inverse of the midpoint of J, and K the row-sum constant of $\mathrm{Id} - YJ$ as above. If $F(x) = 0$ for some x in the box, then by the contraction estimate $\|x - c + YF(c)\|_\infty \le K\,\mathrm{rad}$, hence $\|YF(c)\|_\infty \le (1+K)\,\mathrm{rad}$; so

$$\mathrm{mig}\big(YF(c)\big) > (1+K)\,\mathrm{rad}$$

excludes the box, where mig is the mignitude (a lower bound on the absolute value) of the computed interval. Because $YF(c) \approx c - z^*$, this says "the centre is further from the zero than the box radius", and the conditioning drops out: the undecided region around a zero is one or two cells wide instead of $(\lambda_{\max}/\lambda_{\min})h$.

The covering. The driver is an adaptive bisection of the box $P = [-2,2]^2 \times [0, 2^{-14}]$: a box is accepted if it lies inside one of the seven boxes $B_k \times W$, or if it is excluded by the naive test on either component of F or by the preconditioned test; otherwise it is bisected along its widest coordinate (w counted at weight 64, so that W is refined once the (θ_1, θ_2) cells fall below about $4 \cdot 10^{-3}$) and both halves must be accepted, up to depth 24. The soundness of the driver is an induction on the depth whose only geometric input is that the two halves of a bisected interval cover it; no separate tiling lemma is needed. The covering closes with 1451 boxes (the deepest at depth 22; the formal check runs the driver with the bound 24); at depth 14 it leaves 58 boxes undecided, and the single box P is not excludable—both facts are recorded as computed controls.

The seven certificates. For each k the Krawczyk theorem is applied on $B_k \times W'$ where W' ranges over the four quarters of W: the residual bound β is dominated by the width of the w-interval, so W is split into four pieces, each ball receiving four certificates that share one preconditioner Y and its exact inverse Z. The $7 \times 4 = 28$ side conditions $0 \le K < 1$, $0 \le \beta$, $\beta + K\rho_k \le \rho_k$ are checked by one computation; the worst constants are $K = 0.7459$, $\beta = 5.88 \cdot 10^{-4}$ and step $\beta + K\rho = 3.50 \cdot 10^{-3} \le \rho = 3.91 \cdot 10^{-3}$. The seven boxes are pairwise disjoint and contained in $[-2,2]^2$ (two further computed checks), so the critical set in the square is the injective image of $\{0, \ldots, 6\}$.

Classification. On the enlarged box $\frac{17}{16}B_k \times W$ the interval enclosure $\mathcal{J}$ of H is either positive definite—$\mathcal{J}_{11}^{\mathrm{lo}} > 0$ and $(\mathcal{J}_{11}\mathcal{J}_{22} - \mathcal{J}_{12}^2)^{\mathrm{lo}} > 0$, which gives $v^\mathsf{T} Hv > 0$ for all $v \neq 0$ through $H_{11}\, v^\mathsf{T} Hv = (H_{11}v_1 + H_{12}v_2)^2 + \det H \cdot v_2^2$—or negative in an explicit rational direction $v \in \{(1,1), (-3,1), (-1,3)\}$, i.e. $(v^\mathsf{T} \mathcal{J} v)^{\mathrm{hi}} < 0$. Since $\operatorname{Hess} G_w = -H/3$, a one-dimensional argument converts the first case into a strict local maximum and the second into the negation of a local maximum: for $\varphi(\tau) = G_w(z + \tau v)$ and $\chi(\tau) = \langle F(z + \tau v), v\rangle$ one has $\chi(0) = 0$, $\chi' = v^\mathsf{T} Hv$ and $\varphi' = -\chi/3$, and two applications of strict monotonicity from a positive derivative give $G_w(z + v) < G_w(z)$ for all small v in the first case and $G_w(z + \tau v) > G_w(z)$ for all small $\tau > 0$ in the second. For the three saddles the determinant enclosure is additionally checked to be negative.

Back to the coordinates of [4]. $\theta : x \mapsto (n_1 \cdot x, n_2 \cdot x)$ is a linear homeomorphism with inverse ψ; $G_w \circ \theta = f$; local maxima transfer along θ in both directions; and, as a continuous linear map, θ gives $\nabla f(x) = 0 \iff F(\theta x) = 0$. Cardinalities transfer because ψ is injective, and $M \geq 10^4$ gives $w \in W$ by (6).

5.4. **Proof sketches.**

Proof sketch of Theorem 5.1. Fix $w \in W$. By the covering, every $z \in [-2, 2]^2$ with $F(z) = 0$ lies in some B_k; by the 28 Krawczyk certificates each B_k contains exactly one such z (the quarter of W containing w selects the certificate); the boxes are disjoint and inside the square. Hence the zero set of F in the square has exactly seven elements. Transfer to $\{x \in \Omega : \nabla f(x) = 0\}$ as described above. The claims that rest on computation are exactly: the covering (1451 boxes), the 28 side conditions, the seven left-inverse identities $ZY = \mathrm{Id}$, and the disjointness and containment of the seven boxes. □

Proof sketch of Theorem 5.2. The local maxima of G_w in the square are critical points, hence among the seven; the classification puts $k \in \{0, 4, 5, 6\}$ in and $k \in \{1, 2, 3\}$ out, so the set of local maxima is the injective image of $\{0, 4, 5, 6\}$. The additional computation is the classification check on the seven enlarged boxes (and the negativity of the three determinant enclosures). Transfer to f as above, and to $p_{r,M} = f/(2\pi M)$ because a positive constant multiple has the same local maxima. □

5.5. **What the certificate does not cover.**

The r-window. The certificates are at the single rational $r = 97/100$. A fourth interval variable for r does not work with one box per zero: the central saddle moves from $\theta_1 = 0.06298$ at $r = 0.97$ to $\theta_1 = 0.02031$ at $r = 0.99$, so one box covering the window needs $\rho \geq 0.0214$, while the contraction constant of that zero grows linearly with ρ ($K = 5.81, 2.96, 1.49, 0.746$ at $\rho = 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}$), giving $K \approx 4$ at $\rho = 0.0214$. Subdividing r into about 20 pieces of width 10^{-3} and repeating the whole certificate would settle the window at an estimated 10–15 minutes of computation; it was not attempted. *The rest of the plane.* The annulus $2 \leq \|x\| \leq 16$ contains, numerically, three further saddle points on the ridges (at $\|x\| = 7.06$ for $M = 10^4$); covering it needs a finer exp enclosure (some $|d_j| \approx 8$ there). The localisation of all critical points to $\bar{T}$, three thin strips along the axis lines with $|s_j| \leq 16$, and small neighbourhoods of the means—a specialisation of the ridgeline identity of [8]: at a critical point the three numbers $e^{-q_i/2} d_i$ agree pairwise up to $O(1/M)$, so either all d_i share a sign, which by $\sum d_i = -3r < 0$ puts x in the open triangle, or the dominant $|d_i|$ is $O(1/M)$—is elementary but has not been formalised. Together with Theorem 4.1 these are what separate the regional count from the global one.

6. Negative controls

Proposition 6.1.
(i) *For $r \in [21/20, 3/2]$ and $M \geq 1000$, the origin is not a local maximum point of $f_{r,M}$.*
(ii) *For $r \geq 97/100$ and every $x \in \mathbb{R}^2$, not all three of $|d_1(x)|, |d_2(x)|, |d_3(x)|$ are $< \frac{1}{2}$.*
(iii) *For every r and $M \geq 1000$, $f_{r,M}\big(\mu_1 + \frac{1}{100} t_1\big) > \frac{1}{3} \cdot \frac{8825}{10000}$.*
(iv) *For $r \in [97/100, 99/100]$ and $M \geq 1000$, $f_{r,M}(\mu_1) < f_{r,M}(0)$.*

(i) shows that the hypothesis $r < 1$ in Theorem 3.1 is load-bearing: the same third-order expansion as for the origin, run in reverse, gives $\sum_i e^{u_i} - 3 \geq 0.0693\|x\|^2 > 0$ on $0 < \|x\| \leq 1/400$, so the origin is a strict local *minimum.* This also follows in one line from Lemma 2 of [4], which computes $\nabla^2 f(0) = \frac{1}{2}(r^2 - 1)e^{-(1+r^2)/2} I$. (ii) is $\sum_i d_i = -3r \leq -2.91$ against $|\sum_i d_i| < \frac{3}{2}$; it is what makes the three central rhombi of §3 disjoint, and it fails for $r < 1/2$. (iii) says that a box of half-length $1/100$ along the major axis would not separate the mode near μ_1 from its boundary (the boundary bound used there is $\frac{1}{3}e^{-1/8} \leq \frac{1}{3} \cdot 0.8825$), so the half-length $M/2$ used in the proof is not slack. (iv) shows the seven modes do not all have the same height despite the threefold symmetry: $f(\mu_1) \leq \frac{1}{3}(1+2\cdot 10^{-4}) < (7/8)^8 \leq e^{-1} \leq f(0)$.

7. A numerical census, not proved

The following is computation outside the formal development and is stated as such. A multi-start Newton iteration on $\log(3f)$ (whose gradient is $O(1)$ everywhere, unlike that of f, which decays to 10^{-140} on the ridges), seeded on a 101^2 grid over $[-5,5]^2$ and on $3 \times 3 \times 4001$ points along the ridges out to $|s_j| = 2.2M$, and implemented twice independently, finds for each of the eight tested pairs $(r, M) = (0.97, 10^4)$, $(0.98, 10^4)$, $(0.985, 10^4)$, $(0.99, 10^4)$, $(0.975, 3 \cdot 10^4)$, $(0.97, 10^5)$, $(0.97, 10^6)$, $(0.9, 10^4)$ exactly 13 critical points: 7 nondegenerate local maxima, 6 saddle points, no local minimum (Euler count $7 - 6 + 0 = 1$). The maxima are the origin, three at distance $2(r + \alpha)$ from it, and three within 10^{-6} of the means; the saddles are three at distance $2(r + \alpha')$ and three on the ridges (at $\|x\| = 7.06$ for $M = 10^4$). Here α, α' are the two roots of $h(\alpha) = h(-3r - 2\alpha)$ with $h(d) = de^{-d^2/2}$: in the limit $M \to \infty$ the critical-point equation becomes $h(d_1) = h(d_2) = h(d_3)$ subject to $d_1 + d_2 + d_3 = -3r$, whose solutions are $(-r, -r, -r)$ (the origin) and the two (α, α, β) orbits, seven points, four of them maxima—in agreement with Theorems 5.1 and 5.2. The census also finds the saddle-node threshold $r^* \approx 0.853275$ below which only the origin survives in the limit, and, at $r = 1.05$, six maxima and one minimum (the origin), consistent with Proposition 6.1(i). So, numerically, the mixture (2) has exactly seven modes and does not give $m(2, 3) \geq 8$; what is proved is Theorems 4.1, 5.1 and 5.2.

8. Verification

Every theorem and proposition above is proved in Lean 4 [11] (v4.32.0) against Mathlib [12] (tag v4.32.0), with the statements in the listing below; the repository reference is to be added at submission. Theorems 3.1, 4.1 and Proposition 6.1 depend only on the three standard axioms `propext`, `Classical.choice`, `Quot.sound`, as reported by `#print axioms`; no derivative of f appears in their proofs. Theorems 5.1 and 5.2 depend in addition on six evaluations by `native_decide`, each a rational interval computation on an explicit finite object: the covering (`cover_ok`, 1451 boxes), the 28 Krawczyk side conditions (`certChk_ok`), the left-inverse identities (`ZY_ok`), the disjointness and containment of the seven boxes (`disj_ok`, `inside_ok`), and the Hessian-sign classification (`class_ok`); `exact_seven_critical_region` uses the first five and `exact_four_modes`, `exact_four_modes_pdens` all six, while `saddle_det_neg` carries its own evaluation. Every analytic step—the derivatives of F and H, the soundness of the box tests and of the covering driver, the Krawczyk theorem, the classification argument and the coordinate transfer—is checked by the kernel. The whole certificate module (all six evaluations and the controls) elaborates in 27.7 s with peak memory 3.0 GB; the three analytic modules take 10–20 s each on a warm cache; an exact-rational Python mirror of the interval layer produced the same 1451-box tree in 52 s before the formal run, and the whole computation for Theorems 5.1–5.2, numerics included, was about 0.8 CPU-hours.

Formal statements. The statements are reproduced verbatim from the sources (proof bodies omitted; the formal indexing $0, 1, 2$ of the components is as explained in §2).

```
-- Defs.lean
noncomputable def nvec : Fin 3 → ℝ × ℝ
  | 0 => (s3 / 2, 1 / 2)
  | 1 => (-(s3 / 2), 1 / 2)
  | 2 => (0, -1)
noncomputable def tvec (i : Fin 3) : ℝ × ℝ := (-(nvec i).2, (nvec i).1)
noncomputable def dd (r : ℝ) (i : Fin 3) (x : ℝ × ℝ) : ℝ := dotp (nvec i) x - r
noncomputable def ss (i : Fin 3) (x : ℝ × ℝ) : ℝ := dotp (tvec i) x
noncomputable def qq (r M : ℝ) (i : Fin 3) (x : ℝ × ℝ) : ℝ :=
  (dd r i x) ^ 2 + (ss i x - M) ^ 2 / M ^ 2
noncomputable def f (r M : ℝ) (x : ℝ × ℝ) : ℝ :=
  (1 / 3) * ∑ i : Fin 3, Real.exp (-(qq r M i x) / 2)
noncomputable def pdens (r M : ℝ) (x : ℝ × ℝ) : ℝ := f r M x / (2 * Real.pi * M)

-- Seven.lean  (Theorem 3.1)
theorem seven_modes {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100) (hM : 1000 ≤ M) :
```

```
    ∃ z : Fin 7 → ℝ × ℝ,
      (∀ i j : Fin 7, i ≠ j → z i ≠ z j) ∧ ∀ i : Fin 7, IsLocalMax (f r M) (z i)
theorem seven_modes_pdens {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100) (hM : 1000 ≤ M) :
    ∃ z : Fin 7 → ℝ × ℝ,
      (∀ i j : Fin 7, i ≠ j → z i ≠ z j) ∧ ∀ i : Fin 7, IsLocalMax (pdens r M) (z i)

-- Ridge.lean  (Theorem 4.1);  step x τ = x + τ·t₀
theorem ridge_step_increases {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100)
    (hM : 1000 ≤ M) (hM' : M ≤ 1000000) {x : ℝ × ℝ}
    (hd0 : |dd r 0 x| ≤ 1/2) (hs0a : 16 ≤ ss 0 x) (hs0b : ss 0 x ≤ M - 1)
    {τ : ℝ} (hτ : 0 < τ) (hτ1 : τ ≤ 1) :
    f r M x < f r M (step x τ)
theorem ridge_no_mode {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100)
    (hM : 1000 ≤ M) (hM' : M ≤ 1000000) {x : ℝ × ℝ}
    (hd0 : |dd r 0 x| ≤ 1/2) (hs0a : 16 ≤ ss 0 x) (hs0b : ss 0 x ≤ M - 1) :
    ¬ IsLocalMax (f r M) x
theorem ridge_no_mode_one {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100)
    (hM : 1000 ≤ M) (hM' : M ≤ 1000000) {y : ℝ × ℝ}
    (hdj : |dd r 1 y| ≤ 1/2) (hsa : 16 ≤ ss 1 y) (hsb : ss 1 y ≤ M - 1) :
    ¬ IsLocalMax (f r M) y
theorem ridge_no_mode_two {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100)
    (hM : 1000 ≤ M) (hM' : M ≤ 1000000) {y : ℝ × ℝ}
    (hdj : |dd r 2 y| ≤ 1/2) (hsa : 16 ≤ ss 2 y) (hsb : ss 2 y ≤ M - 1) :
    ¬ IsLocalMax (f r M) y

-- Exact.lean / ExactData.lean / ExactCover.lean  (Theorems 5.1, 5.2)
noncomputable def rr : ℝ := 97 / 100
def Region : Set (ℝ × ℝ) := {x | |dotp (nvec 0) x| ≤ 2 ∧ |dotp (nvec 1) x| ≤ 2}
def WI : Itv := ⟨0, 1 / 16384, by norm_num⟩
def bigI : Itv := ⟨-2, 2, by norm_num⟩
def critSet (w : ℝ) : Set (Fin 2 → ℝ) := {z | z 0 ∈ᵢ bigI ∧ z 1 ∈ᵢ bigI ∧ FR w z = 0}
def isMax : Fin 7 → Bool := ![true, false, false, false, true, true, true]
theorem exact_seven_critical_region {M : ℝ} (hM : (10 : ℝ) ^ 4 ≤ M) :
    {x : ℝ × ℝ | x ∈ Region ∧ fderiv ℝ (f rr M) x = 0}.ncard = 7
theorem exact_seven_critical {M : ℝ} (hM : (10 : ℝ) ^ 4 ≤ M) :
    (critSet (1 / (s3 * M))).ncard = 7
theorem critSet_eq_fderiv_zero {w : ℝ} (z : Fin 2 → ℝ) :
    FR w z = 0 ↔ fderiv ℝ (GG w) z = 0
theorem cover_ok : cover 24 bigP = true
theorem certChk_ok :
    ∀ idx : Fin 7, ∀ t : Fin 4,
      0 ≤ Kc idx t ∧ Kc idx t < 1 ∧ 0 ≤ Bc idx t ∧
        Bc idx t + Kc idx t * rhoQ idx ≤ rhoQ idx
theorem exact_four_modes {M : ℝ} (hM : (10 : ℝ) ^ 4 ≤ M) :
    {x : ℝ × ℝ | x ∈ Region ∧ IsLocalMax (f rr M) x}.ncard = 4
theorem exact_four_modes_pdens {M : ℝ} (hM : (10 : ℝ) ^ 4 ≤ M) :
    {x : ℝ × ℝ | x ∈ Region ∧ IsLocalMax (pdens rr M) x}.ncard = 4
theorem modes_in_region {w : ℝ} (hw : w ∈ᵢ WI) :
    {z : Fin 2 → ℝ | (z 0 ∈ᵢ bigI ∧ z 1 ∈ᵢ bigI) ∧ IsLocalMax (GG w) z}
      = (zpt hw) '' {idx : Fin 7 | isMax idx = true}
theorem zpt_isLocalMax {w : ℝ} (hw : w ∈ᵢ WI) {idx : Fin 7} (h : isMax idx = true) :
    IsLocalMax (GG w) (zpt hw idx)
theorem zpt_not_isLocalMax {w : ℝ} (hw : w ∈ᵢ WI) {idx : Fin 7} (h : isMax idx = false) :
    ¬ IsLocalMax (GG w) (zpt hw idx)
theorem class_ok : ∀ idx : Fin 7,
    (isMax idx = true → 0 < (Jcl idx 0 0).lo ∧ 0 < (detItv idx).lo) ∧
    (isMax idx = false → (qform idx).hi < 0)
theorem saddle_det_neg : ∀ idx : Fin 7, isMax idx = false → (detItv idx).hi < 0

-- Controls.lean  (Proposition 6.1)
theorem origin_not_isLocalMax_of_r_large {r M : ℝ} (hr : 21/20 ≤ r) (hr' : r ≤ 3/2)
    (hM : 1000 ≤ M) : ¬ IsLocalMax (f r M) ((0:ℝ), (0:ℝ))
theorem not_all_three_small {r : ℝ} (hr : 97/100 ≤ r) (x : ℝ × ℝ) :
    ¬ (|dd r 0 x| < 1/2 ∧ |dd r 1 x| < 1/2 ∧ |dd r 2 x| < 1/2)
```

```
theorem mean_box_needs_length {r M : ℝ} (hM : 1000 ≤ M) :
    (1/3) * (8825/10000 : ℝ)
      < f r M ((mu r M 0).1 + (1/100 : ℝ) * (tvec 0).1,
                (mu r M 0).2 + (1/100 : ℝ) * (tvec 0).2)
theorem modes_not_equal_height {r M : ℝ} (hr : 97/100 ≤ r) (hr' : r ≤ 99/100)
    (hM : 1000 ≤ M) : f r M (mu r M 0) < f r M ((0:ℝ), (0:ℝ))
```

Here `zpt hw idx` is the unique zero of `FR w` in the box B_{idx} (from the Krawczyk theorem), `GG w` is G_w, `FR w` is F, `Jcl`, `detItv` and `qform` are the interval enclosures of H, $\det H$ and $v^\mathsf{T} Hv$ on the enlarged boxes, `Kc`, `Bc` are the 28 constants K, β, `rhoQ` the radii, `mu r M i` the mean μ_{i+1}, and `s3` is $\sqrt{3}$.

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 (published online 2019); arXiv:1702.05066 (v3).

[2] C. Améndola, J. I. Rodriguez, A ridgeline correspondence criterion: the number of modes of a Gaussian mixture is finite, arXiv:2608.28558v1 [math.ST], 2026.

[3] H. Edelsbrunner, B. T. Fasy, G. Rote, Add isotropic Gaussian kernels at own risk: more and more resilient modes in higher dimensions, Discrete Comput. Geom. 49 (2013), no. 4, 797–822.

[4] Y. Kabata, H. Matsumoto, A. Okuno, At least seven modes in a heteroscedastic three-component bivariate Gaussian mixture, arXiv:2608.01776v1 [math.ST], 2026.

[5] H. D. Nguyen, Bounds on the number of modes of a Gaussian mixture density, arXiv:2605.15531v1 [math.ST], 2026.

[6] A. Okuno, Y. Kabata, On mixtures of three homoscedastic Gaussian densities: an unconditional sharper bound on the number of modes, arXiv:2607.17506v1 [math.ST], 2026.

[7] A. Okuno, Unconditional finiteness of the number of critical points of Gaussian mixture densities, arXiv:2608.30334v1 [math.ST], 2026.

[8] S. Ray, B. G. Lindsay, The topography of multivariate normal mixtures, Ann. Statist. 33 (2005), no. 5, 2042–2065; arXiv:math/0602238.

[9] H. Wang, On finite Gaussian mixtures: finiteness of the number of modes and an application to NPMLE, arXiv:2608.16675v1 [math.ST], 2026.

[10] R. Krawczyk, Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerschranken, Computing 4 (1969), no. 3, 187–201.

[11] L. de Moura, S. Ullrich, The Lean 4 theorem prover and programming language, in: Automated Deduction – CADE 28, Lecture Notes in Computer Science 12699, Springer, 2021, pp. 625–635.

[12] The mathlib Community, The Lean mathematical library, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381.