# The Schwarzian Bridge in a Single Sigmoid Neuron

## Spherical Inputs, All-Power Laws, and Empirical Resolution


Lluis Eriksson
Independent Researcher
lluiseriksson@gmail.com


August 2026


**Abstract**

A sigmoid is increasing on scalars, yet its bend controls two different geometries. The classical Schwarzian criterion implies that a negative-Schwarzian activation is not matrix-monotone of order two. We give a closed logistic formula showing every distinct two-point Löwner determinant to be negative, a dimension-minimal $2 \times 2$ positive-semidefinite perturbation witness certified with interval arithmetic, and the corresponding classical serial-chain consequence.

Our central link is that the same local invariant determines the leading quadratic onset coefficient of saturation anisotropy. At a centered inflection of an increasing activation $g$, for a sensitivity $h_p = (g')^p$ and isotropic spherically symmetric input $X$,

$$\kappa_{p,X}(r) = 1 - q_X p\, Sg(0) r^2 + O(r^4), \qquad q_X = \frac{\mathbb{E}\|X\|^4}{d(d+2)}.$$

The Gaussian coefficient is the special case $q_X = 1$. The adjacent-point Löwner determinant is $g'(0)^2 Sg(0)\delta^2/6 + O(\delta^3)$. Negative Schwarzian therefore simultaneously gives local noncommutative order failure and positive initial radial–tangential anisotropy. We turn the equality of limits into a finite-scale inequality with explicit constants determined by derivative suprema and input moments.

For isotropic Gaussian inputs in dimension $d \geq 2$, population curvature at a teacher of norm $r$ has one radial eigenvalue and $d-1$ equal tangential eigenvalues. For any nonnegative sensitivity profile $h$ with finite defining expectations, these eigenvalues are exactly

$$\alpha_h(r) = \mathbb{E}h(rZ), \qquad \beta_h(r) = \mathbb{E}[Z^2 h(rZ)], \qquad Z \sim N(0,1),$$

and, if its zeroth and second Lebesgue moments are finite and positive,

$$\alpha_h(r) \sim \frac{\varphi(0)m_0}{r}, \qquad \beta_h(r) \sim \frac{\varphi(0)m_2}{r^3}.$$

Thus saturation suppresses magnitude information by two additional powers: the radial–tangential ratio obeys $\alpha_h(r)/\beta_h(r) \sim (m_0/m_2)r^2$. For every real $p > 0$ and $h_p = \sigma'^p$, we prove strict monotonicity and the unified identities

$$m_0(p) = \frac{\Gamma(p)^2}{\Gamma(2p)}, \qquad \kappa_p(r) \sim \frac{r^2}{2\psi_1(p)}, \qquad \kappa_p(r) = 1 + \frac{p}{2}r^2 - \frac{p}{8}r^4 + \frac{p(p+1)}{16}r^6 + O(r^8).$$

For $n$ samples we prove an explicit two-sided relative Löwner bound for the weighted empirical Gram matrix, radial-eigenvalue and eigenspace-angle bounds, and the sufficient logistic scaling $n \geq C_p r[d + \log(1/\delta)]/\varepsilon^2$, with a computable constant depending only on $p$. An empty-transition-slab argument proves a necessary $c_p r \log(1/\delta)$ term, while a matching product lower bound remains open. Prior work contains a one-sided $p = 1$ empirical bound, bilateral constant-factor Hessian sandwiches, the Bernoulli population exponents, and the one-unit squared-output reduction. Our additions include an explicit pointwise, block-resolved $(1 \pm \varepsilon)$ theorem for every logistic power $p > 0$, with saturation scaling, angle control and a slab obstruction, together with the spherical and quantitative bridges and unified all-$p$ endpoint laws. A deterministic replay and immutable public source record accompany the proofs.

**Keywords.** single neuron; sigmoid; matrix monotonicity; Löwner order; Schwarzian derivative; Fisher information; Gaussian design; weighted sample covariance; finite-sample concentration.

# 1 Question, result, and claim boundary

The scalar logistic curve is

$$\sigma(s) = \frac{1}{1+e^{-s}}, \qquad \sigma'(s) = \frac{1}{4}\operatorname{sech}^2(s/2).$$

It is strictly increasing, so larger scalar scores produce larger outputs. We first ask whether this order survives spectral functional calculus on noncommuting matrices. It does not, at the smallest possible matrix size and on every interval. We then return to the ordinary teacher neuron $\sigma(w_\star^\top X)$ with $X \sim N(0, I_d)$. There, a perturbation perpendicular to $w_\star$ rotates the separating hyperplane, while a parallel perturbation changes its confidence scale and must be detected through samples near the transition layer. Our goal is to compute both failures exactly.

The main contributions are:

1. a closed logistic specialization of the classical negative-Schwarzian obstruction, proving directly that every nontrivial two-point Löwner matrix is indefinite;
2. a dimension-minimal $2 \times 2$ witness $A \preceq B$ but $\sigma(A) \not\preceq \sigma(B)$, with outward-rounded eigenvalue certification, and a Löwner-order reformulation of the known serial-composition consequence;
3. a spherical-input Schwarzian bridge equating the local order defect with the leading radial–tangential anisotropy, including its exact fourth-moment multiplier $q_X$;
4. an explicit finite-scale bridge bounding the discrepancy between the normalized Löwner determinant and anisotropy at nonzero $(\delta, r)$;
5. a profile-generic two-eigenvalue refinement under Gaussian design, with finite-$r$ brackets and, for every $p > 0$, closed moments, strict ratio monotonicity and exact endpoint laws for $h_p = \sigma'^p$;
6. a two-sided relative concentration theorem for the empirical sensitivity matrix, with an explicit radial-eigenvector angle, all-$p$ logistic threshold and necessary slab/rank obstructions; and
7. exact local consequences for stationary fixed-step optimization and estimation, plus a one-command replay of the identities, interval witness, inequalities, and figure.

The matrix-order theorem concerns spectral functional calculus $\sigma(A)$, not entrywise activation of a vector. The radial theorem is not a claim about deep networks, global optimization, implicit bias, or arbitrary inputs. Its local coefficient extends from Gaussian to isotropic spherical inputs, while its large-saturation constants remain Gaussian. The two questions meet at a precise boundary: scalar monotonicity controls neither noncommuting PSD perturbations nor the conditioning of parameter directions.

## 2 A sigmoid is not an order-preserving operator

For Hermitian matrices, $A \preceq B$ means that $B - A$ is positive semidefinite. A scalar function $f$ is *matrix-monotone of order two* on an interval $J$ if $f(A) \preceq f(B)$ for every pair of $2 \times 2$ Hermitian matrices with spectra in $J$ and $A \preceq B$, where $f(A)$ is defined by spectral functional calculus. If $A$ and $B$ commute, scalar monotonicity is sufficient because they are simultaneously diagonalizable. The question is what happens without commutativity.

For distinct $x, y \in J$, define the two-point Löwner matrix

$$L_f(x,y) = \begin{pmatrix} f'(x) & \dfrac{f(x)-f(y)}{x-y} \\ \dfrac{f(x)-f(y)}{x-y} & f'(y) \end{pmatrix}. \tag{1}$$

The fixed-order Löwner criterion says that matrix monotonicity of order two requires $L_f(x,y) \succeq 0$ for every pair [1, 2]. The equivalent local criterion $Sf \geq 0$ for increasing smooth functions is classical [3, Lemma 10]; the result below is an explicit logistic specialization, not a new qualitative matrix-monotonicity criterion.

Cook, Hammerlindl and Tucker define

$$\chi_f(x,y) = f'(x)f'(y)\left(\frac{x-y}{f(x)-f(y)}\right)^2$$

and study the class $\chi_f \leq 1$, including sigmoid activations and serial compositions [4]. If $D_f(x,y) = (f(x)-f(y))/(x-y)$, then

$$\det L_f(x,y) = D_f(x,y)^2\,[\chi_f(x,y) - 1]. \tag{2}$$

Thus their two-point inequality is algebraically the nonpositive Löwner-determinant inequality. Our contribution in this subsection is the strict closed logistic identity and the certified matrix witness, not the underlying Schwarzian/composition mechanism.

**Theorem 2.1** (closed logistic two-point determinant)**.** *Let*

$$s_{\alpha,b}(x) = \frac{1}{1+e^{-\alpha(x-b)}}, \qquad \alpha > 0.$$

*For every $x \neq y$, the Löwner matrix $L_{s_{\alpha,b}}(x,y)$ has negative determinant. Consequently the logistic activation is not matrix-monotone of order two on any nondegenerate interval.*

*Proof.* Put $u = e^{\alpha(x-b)}$, $v = e^{\alpha(y-b)}$, and $r = v/u = e^{\alpha(y-x)}$. Direct cancellation gives the exact identity

$$\det L_{s_{\alpha,b}}(x,y) = \frac{\alpha^2 u^2}{(1+u)^2(1+v)^2}\left[r - \frac{(r-1)^2}{(\log r)^2}\right]. \tag{3}$$

For $r \neq 1$, write $t = \log r$. Strict convexity of sinh on $(0,\infty)$ gives

$$|r-1| = 2e^{t/2}|\sinh(t/2)| > e^{t/2}|t| = \sqrt{r}\,|\log r|.$$

The bracket in (3) is therefore strictly negative. Every nondegenerate interval contains a distinct pair, so the Löwner criterion completes the proof. □

This is stronger than finding one bad scale: changing the slope $\alpha$, shifting the bias $b$, or restricting the score range never removes the obstruction. It also identifies the precise failure. The diagonal entries in (1) are positive, but the divided difference is too large for their geometric mean.

**Proposition 2.2** (dimension-minimal explicit PSD witness)**.** *For the standard logistic sigmoid, let*

$$A = \begin{pmatrix} -1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \qquad B = A + \frac{1}{100}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{4}$$

*Then $A \preceq B$ but $\sigma(A) \not\preceq \sigma(B)$. At 256-bit outward-rounded Arb precision,*

$$\operatorname{spec}(B) \subset [-0.4901000, -0.4900999] \cup [0.5100999, 0.5101000], \tag{5}$$

$$\operatorname{spec}(\sigma(B) - \sigma(A)) \subset [-9.9150, -9.9147]\,10^{-5} \;\cup\; [0.0047990, 0.0047992], \tag{6}$$

$$\det(\sigma(B) - \sigma(A)) \in [-4.7583, -4.7581]\,10^{-7}. \tag{7}$$

*Proof.* $B - A$ has eigenvalues 0 and $1/50$. If $a = 1/2$, $\varepsilon = 1/100$, and $d = (a^2 + \varepsilon^2)^{1/2}$, then $B$ has eigenvalues $\lambda_\pm = \varepsilon \pm d$ and

$$\sigma(B) = mI + n(B - \varepsilon I), \quad m = \frac{\sigma(\lambda_+) + \sigma(\lambda_-)}{2}, \quad n = \frac{\sigma(\lambda_+) - \sigma(\lambda_-)}{2d}.$$

Substitution reduces the three enclosures to scalar exponential and square-root balls. The replay evaluates them with directed Arb rounding; the negative eigenvalue in (6) proves the claim. □

The witness is illustrative, while Theorem 2.1 is the proof for every interval. The derivative form also explains why an indefinite Löwner matrix produces witnesses: for diagonal $A = \operatorname{diag}(x, y)$, the Fréchet derivative in the all-ones PSD direction is precisely the Schur product with $L_f(x, y)$.

## 2.1 Serial depth does not repair the order

For a $C^3$ function with nonzero derivative, its Schwarzian is

$$Sf = \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2. \tag{8}$$

The logistic sigmoid and a scaled hyperbolic tangent satisfy

$$Ss_{\alpha,b} = -\frac{\alpha^2}{2}, \qquad S\tanh(\alpha x + \beta) = -2\alpha^2. \tag{9}$$

**Corollary 2.3** (serial spectral reformulation)**.** *Let $F$ be a finite scalar composition of positive-slope affine maps and at least one increasing logistic or hyperbolic-tangent activation. Then $SF < 0$ everywhere, and $F$ is not matrix-monotone of order two on any nondegenerate interval.*

*Proof.* The composition identity

$$S(f \circ g) = (Sf \circ g)(g')^2 + Sg \tag{10}$$

and (9) give $SF < 0$, since positive affine maps have zero Schwarzian and all derivatives in the chain are positive. For $h \to 0$, Taylor expansion of the two-point determinant gives

$$\det L_F(x, x+h) = \frac{F'(x)^2}{6}\, SF(x)\, h^2 + O(h^3). \tag{11}$$

It is negative for all sufficiently small nonzero $h$. Every interval contains such a pair. □

The corollary restates, in spectral Löwner order, the serial closure proved in the nowhere-coexpanding framework [4]. It is deliberately about *serial scalar chains*. It does not cover sums of neurons, skip connections, multivariate architectures, or entrywise coordinate order. Its content is that composing more of the same S-shaped bend cannot turn the resulting scalar function into a PSD-order-preserving spectral activation.

## 3 Two models, one sensitivity matrix

Write $r = \|w_\star\|$ and $u = w_\star/r$ when $r > 0$. We use two standard observation models.

**Squared-output regression.** For the realizable population loss

$$\mathcal{L}_{\mathrm{sq}}(w) = \frac{1}{2}\mathbb{E}\big[\sigma(w^\top X) - \sigma(w_\star^\top X)\big]^2, \tag{12}$$

the residual vanishes at $w_\star$, so its Hessian is the Gauss–Newton matrix

$$H_{\mathrm{sq}}(w_\star) = \mathbb{E}\big[\sigma'(w_\star^\top X)^2 X X^\top\big]. \tag{13}$$

The same matrix, divided by a known output-noise variance $\tau^2$, is the Fisher information of $Y = \sigma(w_\star^\top X) + \varepsilon$, $\varepsilon \sim N(0, \tau^2)$.

**Bernoulli logistic neuron.** If $Y \mid X \sim \mathrm{Bernoulli}(\sigma(w_\star^\top X))$, the expected negative-log-likelihood Hessian and the Fisher information coincide:

$$H_{\mathrm{B}}(w_\star) = \mathbb{E}\big[\sigma'(w_\star^\top X) X X^\top\big]. \tag{14}$$

Both are instances of the sensitivity matrix

$$H_h(r, u) := \mathbb{E}\big[h(r u^\top X) X X^\top\big], \qquad h \geq 0. \tag{15}$$

The power of the reduction below is that no high-dimensional integration remains.

Amari, Karakida and Oizumi already derived the exact one-unit Gaussian Fisher decomposition for $h = \phi'^2$, including the radial–tangential split and the bias block [20]. We do not claim that decomposition or the squared-output branch as new. For the Bernoulli profile $h = \sigma'$, Chen and Mazumdar recently identified the same radial and orthogonal Hessian functions and proved their $r^{-3}$ and $r^{-1}$ orders in a finite-sample analysis of logistic regression [13]. Their follow-up proves the minimax norm-estimation rate $\Theta(\sqrt{r^3/n})$ [14]. We do not claim those exponents or the Bernoulli optimization penalty as new. The purpose of the next sections is to isolate the profile-generic mechanism, give finite-radius brackets and exact logistic constants, prove strict monotonicity of both eigenvalue ratios, and connect their small-$r$ coefficients to the Schwarzian order defect.

## 4 Universal Gaussian saturation theorem

Let $\varphi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ and define the even sensitivity moments

$$m_j(h) := \int_{\mathbb{R}} t^j h(t)\, dt \quad (j = 0, 2, 4),$$

whenever finite. Evenness of $h$ is not needed for the spectral reduction or leading asymptotics, but it holds in our sigmoid applications and makes the interpretation symmetric.

**Theorem 4.1** (exact spectrum and finite-$r$ brackets)**.** *Let $d \geq 2$, $X \sim N(0, I_d)$, $r > 0$, $u \in S^{d-1}$, and let $h : \mathbb{R} \to [0, \infty)$ be measurable with $\mathbb{E}[(1 + Z^2) h(rZ)] < \infty$. Then*

$$H_h(r, u) = \alpha_h(r)(I - uu^\top) + \beta_h(r) uu^\top, \tag{16}$$

*where*

$$\alpha_h(r) = \mathbb{E}h(rZ), \qquad \beta_h(r) = \mathbb{E}[Z^2 h(rZ)], \qquad Z \sim N(0, 1). \tag{17}$$

*If $m_0, m_2 < \infty$ and are positive, then*

$$\alpha_h(r) \sim \frac{\varphi(0)m_0}{r}, \qquad \beta_h(r) \sim \frac{\varphi(0)m_2}{r^3}, \qquad \kappa_h(r) := \frac{\alpha_h(r)}{\beta_h(r)} \sim \frac{m_0}{m_2}r^2. \tag{18}$$

*If $m_0, m_2, m_4 < \infty$, the following nonasymptotic brackets hold for every $r > 0$:*

$$\frac{\varphi(0)}{r}\left(m_0 - \frac{m_2}{2r^2}\right) \le \alpha_h(r) \le \frac{\varphi(0)m_0}{r}, \tag{19}$$

$$\frac{\varphi(0)}{r^3}\left(m_2 - \frac{m_4}{2r^2}\right) \le \beta_h(r) \le \frac{\varphi(0)m_2}{r^3}. \tag{20}$$

*Negative lower endpoints are interpreted as valid but uninformative bounds.*

*Proof.* Decompose $X = Zu + Y$, where $Z \sim N(0,1)$, $Y \sim N(0, I - uu^\top)$, and $Z, Y$ are independent. The mixed terms vanish and $\mathbb{E}YY^\top = I - uu^\top$, giving (16)–(17). Changing variables $t = rz$ gives the exact one-dimensional formulae

$$\alpha_h(r) = \frac{\varphi(0)}{r}\int_{\mathbb{R}} h(t)e^{-t^2/(2r^2)}dt, \quad \beta_h(r) = \frac{\varphi(0)}{r^3}\int_{\mathbb{R}} t^2 h(t)e^{-t^2/(2r^2)}dt. \tag{21}$$

Dominated convergence proves (18). Finally, $1 - x \le e^{-x} \le 1$ for $x \ge 0$ inserted into (21) proves (19)–(20). □

## 5 Spherical and finite-scale Schwarzian bridges

The matrix-order and saturation calculations meet exactly at an inflection point. Gaussian independence makes the full saturation curve one-dimensional, but it is not responsible for the local bridge: rotational symmetry and suitable moments suffice. The fourth moment identifies the coefficient below; the stated sixth moment controls its $O(r^4)$ remainder.

**Theorem 5.1** (spherical local order–anisotropy bridge)**.** *Let $d \ge 2$ and let $X \in \mathbb{R}^d$ be isotropic and spherically symmetric with $\mathbb{E}\|X\|^6 < \infty$. Let $g \in C^5(\mathbb{R})$ satisfy $g' > 0$ and $g''(0) = 0$. For $p > 0$, suppose $h_p = (g')^p$ has bounded fourth derivative. For $u \in S^{d-1}$ put $Z = u^\top X$ and*

$$H_{p,X}(r,u) = \mathbb{E}[h_p(rZ)XX^\top] = \alpha_{p,X}(r)(I - uu^\top) + \beta_{p,X}(r)uu^\top, \qquad \kappa_{p,X} = \frac{\alpha_{p,X}}{\beta_{p,X}}.$$

*Then, with*

$$q_X := \frac{\mathbb{E}\|X\|^4}{d(d+2)},$$

*one has*

$$\lim_{r\downarrow 0}\frac{\kappa_{p,X}(r) - 1}{r^2} = -q_X p\, Sg(0) = -\frac{6q_X p}{g'(0)^2}\lim_{\delta\to 0}\frac{\det L_g(0,\delta)}{\delta^2}. \tag{22}$$

*In particular, $Sg(0) < 0$ makes $L_g(0,\delta)$ indefinite and $\kappa_{p,X}(r) > 1$ at all sufficiently small nonzero scales. Standard Gaussian input has $q_X = 1$.*

*Proof.* Rotations fixing $u$ force the displayed two-eigenspace form, with

$$\beta_{p,X}(r) = \mathbb{E}[Z^2 h_p(rZ)], \qquad \alpha_{p,X}(r) = \frac{1}{d-1}\mathbb{E}[(\|X\|^2 - Z^2)h_p(rZ)].$$

Choose a unit $v \perp u$. Spherical symmetry and isotropy give

$$\mathbb{E}Z^4 = 3q_X, \qquad \mathbb{E}[Z^2(v^\top X)^2] = q_X.$$

Taylor's formula, cancellation of odd terms under $X \overset{d}{=} -X$, and the sixth-moment assumption therefore give

$$\alpha_{p,X}(r) = h_p(0) + \frac{q_X h_p''(0)}{2} r^2 + O(r^4), \qquad \beta_{p,X}(r) = h_p(0) + \frac{3 q_X h_p''(0)}{2} r^2 + O(r^4).$$

Taking the ratio yields $\kappa_{p,X}(r) = 1 - q_X h_p''(0) r^2 / h_p(0) + O(r^4)$. Since $\log h_p = p \log g'$ and $g''(0) = 0$,

$$\frac{h_p''(0)}{h_p(0)} = p \frac{g'''(0)}{g'(0)} = pSg(0).$$

The adjacent-point expansion (11) gives the remaining equality. $\square$

The equality of limits has a computable finite-scale version. The constants below are deliberately explicit rather than optimized.

**Theorem 5.2** (explicit finite-scale bridge)**.** *Under Gaussian input, retain the hypotheses of Theorem 5.1 and write*

$$a = h_p(0), \quad b = h_p''(0), \quad M_4 = \|h_p^{(4)}\|_\infty.$$

*Choose $r_0 > 0$ so that*

$$\frac{3|b|}{2} r_0^2 + \frac{5M_4}{8} r_0^4 \le \frac{a}{2}, \tag{23}$$

*and set*

$$C_h = \frac{3M_4}{2a} + 3\left(\frac{b}{a}\right)^2 + \frac{5|b|M_4}{4a^2} r_0^2.$$

*For a chosen $\delta_0 > 0$, put*

$$A = g'(0), \ c = g'''(0), \ d_4 = g^{(4)}(0), \ M_5 = \sup_{|x| \le \delta_0} |g^{(5)}(x)|,$$

$$K = \frac{|c|}{6} + \frac{|d_4|\delta_0}{24} + \frac{M_5 \delta_0^2}{120}, \qquad D_g = \frac{|A d_4|}{12} + \frac{7 A M_5 \delta_0}{120} + \delta_0 K^2.$$

*Then for $0 < r \le r_0$ and $0 < |\delta| \le \delta_0$,*

$$\left|\kappa_p(r) - 1 + pSg(0) r^2\right| \le C_h r^4, \tag{24}$$

$$\left|\det L_g(0,\delta) - \frac{A^2 Sg(0)}{6}\delta^2\right| \le D_g |\delta|^3, \tag{25}$$

$$\left|\frac{\kappa_p(r) - 1}{r^2} + \frac{6p}{A^2} \frac{\det L_g(0,\delta)}{\delta^2}\right| \le C_h r^2 + \frac{6pD_g}{A^2}|\delta|. \tag{26}$$

*Thus the order defect predicts anisotropy at finite, independently chosen scales with a certified error bar.*

*Proof.* Gaussian Taylor remainders give

$$\alpha_p = a + \frac{b}{2} r^2 + R_\alpha, \quad |R_\alpha| \le \frac{M_4}{8} r^4, \qquad \beta_p = a + \frac{3b}{2} r^2 + R_\beta, \quad |R_\beta| \le \frac{5M_4}{8} r^4.$$

Condition (23) makes $\beta_p \ge a/2$. Subtracting $1 - (b/a) r^2$ before dividing by $\beta_p$ yields

$$\left|\kappa_p - 1 + \frac{b}{a} r^2\right| \le \left[\frac{3M_4}{2a} + 3\left(\frac{b}{a}\right)^2 + \frac{5|b|M_4}{4a^2} r_0^2\right] r^4,$$

which is (24) because $b/a = pSg(0)$.

Taylor expansion at zero gives

$$g'(\delta) = A + \frac{c}{2}\delta^2 + \frac{d_4}{6}\delta^3 + R_1, \qquad \frac{g(\delta) - g(0)}{\delta} = A + \frac{c}{6}\delta^2 + \frac{d_4}{24}\delta^3 + R_2,$$

where $|R_1| \le M_5|\delta|^4/24$ and $|R_2| \le M_5|\delta|^4/120$. Writing the determinant as $A(g'(\delta) - A) - 2A(V - A) - (V - A)^2$, with $V = (g(\delta) - g(0))/\delta$, gives (25) with the displayed $D_g$. Finally, divide (24) by $r^2$, divide (25) by $\delta^2$, multiply the latter by $6p/A^2$, and use the triangle inequality. $\square$

For the standard logistic sigmoid, $S\sigma = -1/2$. Theorem 5.1 therefore predicts the exact initial coefficient $p/2$ for every $h_p = \sigma'^p$, recovered globally below. Equation (26) strengthens the conceptual coefficient match into a falsifiable finite-scale comparison.

For a general $h$, $\kappa_h = \alpha_h/\beta_h$ in (18) is an anisotropy ratio; it is the spectral condition number only after the eigenvalue ordering is known. The two powers are geometric. At large $r$, only a score slab of width $O(r^{-1})$ remains unsaturated, producing $\alpha \asymp r^{-1}$. A radial perturbation carries an additional factor $Z^2 = O(r^{-2})$ inside that slab, producing $\beta \asymp r^{-3}$. Both modes are supported by the near-boundary slab; the radial mode is weaker because it carries the additional $Z^2$ factor.

# 6 All-power logistic saturation laws

For every real $p > 0$ set $h_p(t) = \sigma'(t)^p$. The cases $p = 1$ and $p = 2$ are Bernoulli Fisher and squared-output curvature, respectively, but the analytic law is not restricted to those observation models. The Gamma transform below is classical generalized-logistic distribution theory [22, eqs. (2.3)–(2.5)]; our use of it is geometric.

**Lemma 6.1** (closed sensitivity moments for every $p > 0$)**.** *Let $\psi_j$ denote the polygamma function of order $j$. Then*

$$\widehat{h}_p(k) := \int_{\mathbb{R}} e^{ikt} h_p(t)\,dt = \frac{\Gamma(p + ik)\Gamma(p - ik)}{\Gamma(2p)}, \tag{27}$$

*and*

$$m_0(p) = \frac{\Gamma(p)^2}{\Gamma(2p)}, \qquad m_2(p) = 2\psi_1(p)m_0(p), \qquad m_4(p) = \big(12\psi_1(p)^2 + 2\psi_3(p)\big)m_0(p). \tag{28}$$

*In particular,*

| *profile* | $m_0$ | $m_2$ | $m_4$ |
|---|---|---|---|
| $h_1 = \sigma'$ | 1 | $\pi^2/3$ | $7\pi^4/15$ |
| $h_2 = \sigma'^2$ | $1/6$ | $(\pi^2 - 6)/18$ | $7\pi^4/90 - 2\pi^2/3$ |

*Proof.* With $x = e^t$,

$$h_p(t) = \frac{x^p}{(1 + x)^{2p}}, \qquad \widehat{h}_p(k) = \int_0^\infty \frac{x^{p+ik-1}}{(1 + x)^{2p}}\,dx = B(p + ik, p - ik),$$

which proves (27). Since $h_p(t) = O(e^{-p|t|})$, all polynomial moments are finite and differentiation under the Fourier integral is justified. Differentiating its logarithm at zero gives $(\log \widehat{h}_p)''(0) = -2\psi_1(p)$ and $(\log \widehat{h}_p)^{(4)}(0) = 2\psi_3(p)$. Since $m_2 = -\widehat{h}_p''(0)$ and $m_4 = \widehat{h}_p^{(4)}(0)$, (28) follows. The table uses the standard integer polygamma values. $\square$

**Theorem 6.2** (strict all-power logistic anisotropy)**.** *For every $p > 0$, $\kappa_p(0) = 1$, $\kappa_p(r) > 1$ for $r > 0$, and $\kappa_p$ is strictly increasing on $(0, \infty)$. At the two endpoints,*

$$\kappa_p(r) = 1 + \frac{p}{2}r^2 - \frac{p}{8}r^4 + \frac{p(p+1)}{16}r^6 + O(r^8) \quad (r \downarrow 0), \tag{29}$$

*and*

$$\kappa_p(r) = \frac{r^2}{2\psi_1(p)} + 1 + \frac{\psi_3(p)}{4\psi_1(p)^2} + O(r^{-2}) \quad (r \to \infty). \tag{30}$$

*Thus $\kappa_1(r) \sim 3r^2/\pi^2$ and $\kappa_2(r) \sim 3r^2/(\pi^2 - 6)$.*

*Proof.* Normalize the density $q_{p,r}(z) \propto h_p(rz)\varphi(z)$. Then $\beta_p(r)/\alpha_p(r) = \mathbb{E}_{q_{p,r}} Z^2$. If $r_2 > r_1$, the derivative on $z > 0$ of the log likelihood ratio is

$$\frac{d}{dz} \log \frac{h_p(r_2 z)}{h_p(r_1 z)} = -p\big[r_2 \tanh(r_2 z/2) - r_1 \tanh(r_1 z/2)\big] < 0.$$

Thus the distribution of $|Z|$ under $q_{p,r}$ decreases strictly in monotone-likelihood-ratio order as $r$ increases, and its strictly increasing statistic $Z^2$ has decreasing expectation. Hence $\alpha_p/\beta_p$ increases strictly. At $r = 0$ isotropy gives one. Lemma 6.1 inserted into Theorem 4.1, with one further term in $e^{-t^2/(2r^2)}$, gives the large-$r$ statement. More explicitly, $m_6(p) < \infty$ by the exponential tail, and Taylor remainder bounds for $e^{-x}$ give

$$\alpha_p(r) = \frac{\varphi(0)}{r}\left(m_0 - \frac{m_2}{2r^2} + \frac{m_4}{8r^4} + O(r^{-6})\right), \qquad \beta_p(r) = \frac{\varphi(0)}{r^3}\left(m_2 - \frac{m_4}{2r^2} + O(r^{-4})\right).$$

Their ratio yields (30). Finally,

$$h_p(t) = 4^{-p}\left[1 - \frac{p}{4}t^2 + \frac{p(1+3p)}{96}t^4 - \frac{p(15p^2+15p+4)}{5760}t^6 + O(t^8)\right].$$

For each fixed $p > 0$, every derivative of $h_p$ is bounded. Taylor's theorem therefore bounds the displayed remainder globally by $\|h_p^{(8)}\|_\infty |t|^8/8!$; after setting $t = rZ$, this is integrable in both $\alpha_p$ and $\beta_p$ because the Gaussian has a finite tenth moment. Inserting the Gaussian moments through $\mathbb{E}Z^8 = 105$ in the coefficients and taking the ratio gives (29). ☐

Figure 1 shows the two statistically canonical members of the all-power family. The solid and dashed curves are direct Gaussian integrals; the dotted curves are the asymptotic constants in (30).

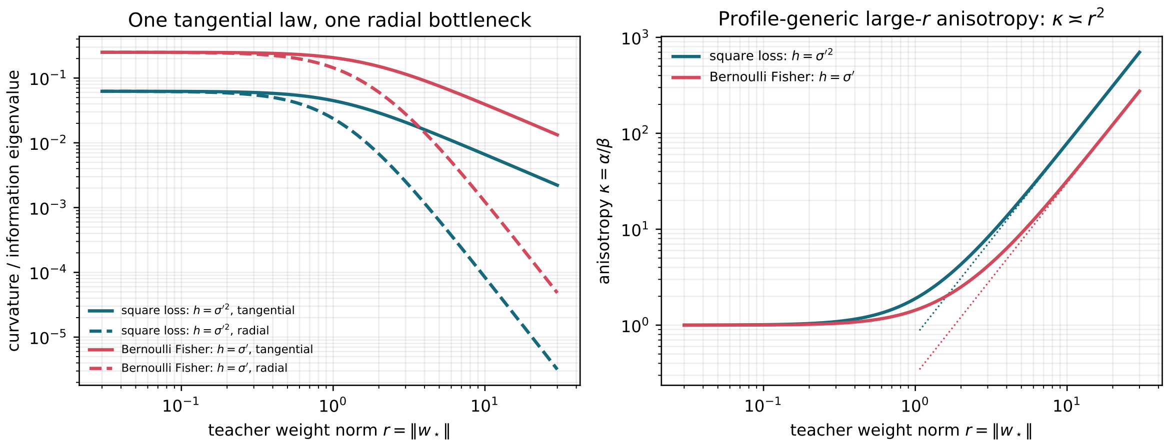


Figure 1: Exact curvature/information eigenvalues and their ratio. Saturation leaves $d-1$ tangential directions at order $r^{-1}$ but pushes the radial direction to order $r^{-3}$. The quadratic condition-number law is not a fitted exponent; it follows from Theorem 4.1.

The cancellation of every $p^2$ contribution in the fourth-order ratio coefficient is worth noting: the full all-power family has the correction $-pr^4/8$, followed by the first nonlinear-in-$p$ term $p(p+1)r^6/16$.

## 7 Finite-sample resolution of the weighted Gram matrix

The population eigenvalues alone do not say when a sample resolves the radial mode. We now answer that question for the fixed-oracle matrix

$$\widehat{H}_{h,n}(r,u) = \frac{1}{n}\sum_{i=1}^{n} h(ru^\top X_i)X_iX_i^\top, \qquad X_i \overset{\text{iid}}{\sim} N(0,I_d).$$

Prior self-concordant analyses already give bilateral constant-factor empirical-Hessian sandwiches, including a uniform result for Gaussian logistic regression [16, Eq. (92)]; fixed-point standardized-Hessian concentration is also available under generic matrix-Bernstein conditions [17]. Chardon, Lerasle and Mourtada prove a uniform one-sided lower bound for the empirical Bernoulli Hessian ($p=1$) under the sufficient condition $n \gtrsim r(d+t)$ [15, Theorem 6]. We do not claim the first bilateral concentration or the first finite-sample lower bound. The result below instead gives tunable two-sided relative error for a fixed teacher, applies to a general bounded profile and hence every $p>0$, separates radial, tangential and cross errors, and returns an eigenspace angle.

Put $m=d-1$, $Z=u^\top X$, $Y=X-Zu$, and $W=h(rZ)$. Define

$$\alpha = \mathbb{E}W, \quad \beta = \mathbb{E}[WZ^2], \quad \gamma_j = \mathbb{E}[W^2Z^{2j}] \quad (j=0,1,2),$$

and the envelopes

$$H = \|h\|_\infty, \qquad K_1 = \sup_z z^2 h(rz)^2, \qquad K_2 = \sup_z z^2 h(rz).$$

**Theorem 7.1** (explicit relative empirical spectrum)**.** *Assume $0 \le h \le H$, $K_1, K_2 < \infty$, and $\alpha,\beta>0$. Let $0<\delta<1$ and set*

$$t = \log(12/\delta), \qquad u_0 = m\log 9 + t,$$

$$S_0^+ = n\gamma_0 + \sqrt{2nH^2\gamma_0 t} + \frac{H^2t}{3}, \qquad S_1^+ = n\gamma_1 + \sqrt{2nK_1\gamma_1 t} + \frac{K_1t}{3},$$

$$e_T = \sqrt{\frac{2\gamma_0 t}{n}} + \frac{Ht}{3n} + \frac{4}{n}\big(\sqrt{S_0^+u_0} + Hu_0\big), \tag{31}$$

$$e_R = \sqrt{\frac{2\gamma_2 t}{n}} + \frac{K_2t}{3n}, \qquad q_n = \frac{\sqrt{S_1^+}}{n}\big(\sqrt{m}+\sqrt{2t}\big). \tag{32}$$

*Then, with probability at least $1-\delta$, in the radial–tangential basis*

$$\widehat{H}_{h,n} = \begin{pmatrix} C & b_n \\ b_n^\top & c_n \end{pmatrix}$$

*satisfies simultaneously*

$$\|C-\alpha I_m\|_{\mathrm{op}} \le e_T, \qquad |c_n-\beta| \le e_R, \qquad \|b_n\|_2 \le q_n. \tag{33}$$

*Consequently, for*

$$\varepsilon_n := \max\left\{\frac{e_T}{\alpha}, \frac{e_R}{\beta}\right\} + \frac{q_n}{\sqrt{\alpha\beta}}, \tag{34}$$

*one has the two-sided relative Löwner bound*

$$(1-\varepsilon_n)H_h(r,u) \preceq \widehat{H}_{h,n}(r,u) \preceq (1+\varepsilon_n)H_h(r,u). \tag{35}$$

*If* $G := \alpha-\beta-e_T-e_R > 0$, *the empirical matrix has a unique eigenvalue* $\widehat{\beta}$ *below the tangential block spectrum, and its unit eigenvector* $\widehat{u}$ *obeys*

$$\tan\angle(\widehat{u},u) \le \frac{q_n}{G}, \qquad |\widehat{\beta}-\beta| \le e_R + \frac{q_n^2}{G}. \tag{36}$$

*Every other ordered eigenvalue differs from* $\alpha$ *by at most* $\max\{e_T,e_R\}+q_n$.

*Proof.* Write

$$C = \frac{1}{n}\sum_i W_iY_iY_i^\top, \qquad b_n = \frac{1}{n}\sum_i W_iZ_iY_i, \qquad c_n = \frac{1}{n}\sum_i W_iZ_i^2.$$

Scalar Bernstein controls $\overline{W}-\alpha$ and, one-sidedly, $\sum_i W_i^2 \le S_0^+$. Conditional on the weights, for a fixed unit $v$,

$$v^\top \sum_i W_i(Y_iY_i^\top - I_m)v = \sum_i W_i(G_i^2-1), \qquad G_i \overset{\text{iid}}{\sim} N(0,1).$$

The exact Gaussian moment-generating function gives

$$\left|\sum_i W_i(G_i^2-1)\right| \le 2\sqrt{u_0\sum_i W_i^2} + 2Hu_0$$

outside conditional probability $2e^{-u_0}$. A 1/4-net of $S^{m-1}$ has at most $9^m$ points and $\|A\|_{\text{op}} \le 2\max_{v \text{ in net}}|v^\top Av|$, proving the first bound in (33). Bernstein applied to $W_iZ_i^2 \in [0,K_2]$ proves the radial bound.

Conditionally on the $Z_i$,

$$b_n \sim N\left(0, \frac{\sum_i W_i^2Z_i^2}{n^2}I_m\right).$$

A second one-sided Bernstein bound gives $\sum_i W_i^2Z_i^2 \le S_1^+$, and the Gaussian norm tail proves the cross bound. The listed failures total less than $9e^{-t} < \delta$.

Conjugating $\widehat{H}_{h,n} - H_h$ by $H_h^{-1/2} = \operatorname{diag}(\alpha^{-1/2}I_m, \beta^{-1/2})$ shows that its norm is at most (34); this is equivalent to (35). If $G>0$, then $c_n < \lambda_{\min}(C)$. Interlacing leaves one eigenvalue below $C$. The eigenvector equation and the Schur complement give respectively $\tan\angle(\widehat{u},u) \le q_n/G$ and $0 \le c_n - \widehat{\beta} \le q_n^2/G$. Weyl's inequality gives the remaining assertion. □

For $h_p = \sigma'^p$, all quantities in Theorem 7.1 are one-dimensional and computable. The elementary bound $\sigma'(s) \le e^{-|s|}$ gives, for every $p>0$,

$$H = 4^{-p}, \quad K_1 \le \frac{e^{-2}}{p^2r^2}, \quad K_2 \le \frac{4e^{-2}}{p^2r^2}, \tag{37}$$

and, with $\varphi(0) = (2\pi)^{-1/2}$,

$$\gamma_0 \le \frac{\varphi(0)}{pr}, \qquad \gamma_1 \le \frac{\varphi(0)}{2p^3r^3}, \qquad \gamma_2 \le \frac{3\varphi(0)}{2p^5r^5}. \tag{38}$$

Indeed, after $s = rZ$, use $h_p(s)^2 \le e^{-2p|s|}$ and $\int_{\mathbb{R}} |s|^j e^{-2p|s|}ds = 2j!/(2p)^{j+1}$ for $j = 0,2,4$.

**Corollary 7.2** (effective sample size for every logistic power)**.** *Let*

$$R_p^2 = \max\left\{\frac{3m_2(p)}{m_0(p)}, \frac{m_4(p)}{m_2(p)}\right\}.$$

*For $r \ge R_p$, $0 < \varepsilon \le 1/2$ and $0 < \delta < 1$, the explicit condition $\varepsilon_n \le \varepsilon$ in* (34)*, with* (37)*–*(38) *substituted, guarantees* (35)*. In particular, for each fixed $p > 0$ there are finite computable constants $C_p, C_p'$ such that*

$$n \ge C_p \frac{r\,[d + \log(12/\delta)]}{\varepsilon^2} \tag{39}$$

*is sufficient and, after increasing $C_p$ if necessary, $e_T + e_R \le (\alpha - \beta)/2$. Hence $G \ge (\alpha - \beta)/2$ and, on the same event,*

$$\sin\angle(\widehat{u}, u) \le C_p'\left(\sqrt{\frac{d + \log(1/\delta)}{nr}} + \frac{\sqrt{[d + \log(1/\delta)]\log(1/\delta)}}{n}\right).$$

*Thus the effective number of transition-layer observations is $n/r$: relative resolution of a population eigenvalue of order $r^{-3}$ does not itself require $r^3$ samples.*

*Proof.* The brackets in Theorem 4.1 and the definition of $R_p$ give

$$\alpha \ge \frac{\varphi(0)m_0(p)}{2r}, \qquad \beta \ge \frac{\varphi(0)m_2(p)}{2r^3}, \qquad \alpha - \beta \ge \frac{\varphi(0)m_0(p)}{2r}.$$

Insert these and (37)–(38) in (31)–(34). Each relative block term is bounded by a $p$-dependent constant times $\sqrt{r[d + \log(12/\delta)]/n} + r[d + \log(12/\delta)]/n$. The cross term has the same first order. Increasing a finite constant $C_p$ makes their sum at most $\varepsilon$ and also makes $e_T + e_R \le (\alpha - \beta)/2$. Thus $G \ge (\alpha - \beta)/2 \ge \varphi(0)m_0(p)/(4r)$, and substitution of $q_n$ into (36) gives the displayed bound with a finite computable $C_p'$. □

The sufficient product scaling in (39) is not claimed minimax-sharp. Two necessary obstructions can nevertheless be proved. First, positive weights imply $\operatorname{rank}(\widehat{H}_{p,n}) \le n$, so $n \ge d$ is necessary for a positive smallest eigenvalue. Second, choose any finite $a_p > 0$ satisfying

$$\int_{|s|>a_p} s^2 h_p(s)\,ds \le \frac{m_2(p)}{16}.$$

For $r \ge \max\{\sqrt{m_4(p)/m_2(p)},\, 4\varphi(0)a_p\}$, the empty-transition-slab event and Markov's inequality give

$$\Pr\left\{\lambda_{\min}(\widehat{H}_{p,n}) < \frac{\beta_p(r)}{2}\right\} \ge \frac{1}{2}\exp\left(-\frac{4\varphi(0)a_p n}{r}\right). \tag{40}$$

Hence failure probability at most $\delta < 1/2$ requires

$$n \ge \frac{r}{4\varphi(0)a_p}\log\frac{1}{2\delta}. \tag{41}$$

To verify (40), condition on no $|Z_i| \le a_p/r$. This event has probability at least the exponential shown; the conditional mean radial entry is at most $\beta_p/4$, so with conditional probability at least $1/2$ that entry, and therefore the smallest eigenvalue, is below $\beta_p/2$.

The lower bounds (41) and $n \ge d$ do not combine into a matching $\Omega(rd)$ theorem. Establishing that product lower bound requires a smallest-singular-value result for the heteroskedastic weighted Gaussian design and remains open. Nor is (41) a minimax parameter-estimation theorem: the teacher, profile and input law are fixed in this oracle Gram problem.

# 8 What the law costs

## 8.1 Stationary fixed-step gradient descent

For either logistic profile, assume $d \geq 2$ and that one constant scalar step size $\eta$ is reused at every iteration. At the teacher, the Jacobian of one gradient-descent step is $I - \eta H_h$. Its radial eigenvalue is $1 - \eta\beta$ and its tangential eigenvalue is $1 - \eta\alpha$. The best scalar step for this local quadratic model is

$$\eta_\star = \frac{2}{\alpha + \beta}, \qquad \rho_\star = \frac{\alpha - \beta}{\alpha + \beta} = \frac{\kappa - 1}{\kappa + 1}. \tag{42}$$

Consequently $-\log \rho_\star \sim 2/\kappa$. For worst-case local error with components allowed in both eigenspaces, the e-folding iteration count $T_\mathrm{e} := (-\log \rho_\star)^{-1}$ therefore has the exact leading scaling

$$T_{\mathrm{e,sq}} \sim \frac{3}{2(\pi^2 - 6)} r^2, \qquad T_{\mathrm{e,B}} \sim \frac{3}{2\pi^2} r^2 \tag{43}$$

for the optimally tuned stationary scalar-step linearization. This is not a global iteration bound or an algorithm-independent obstruction: away from the teacher the Hessian contains residual terms, and nonstationary polynomial methods can exploit the two-point spectrum. Indeed, the scalar schedule $\eta_1 = 1/\alpha$, $\eta_2 = 1/\beta$ makes $(I - \eta_2 H)(I - \eta_1 H) = 0$ in this exact local quadratic. Natural-gradient or matrix-preconditioned updates can also remove the local condition number by acting differently on the two eigenspaces.

## 8.2 Estimation

For $n$ independent Gaussian-output observations with known variance $\tau^2$, the Fisher information is $nH_{h_2}/\tau^2$. The Cramér–Rao inequality gives, along any unit tangential vector $v \perp u$ and the radial vector $u$,

$$\mathrm{Var}(v^\top \widehat{w}) \geq \frac{\tau^2}{n\alpha_2(r)}, \qquad \mathrm{Var}(u^\top \widehat{w}) \geq \frac{\tau^2}{n\beta_2(r)}. \tag{44}$$

The ratio of the displayed radial lower bound to the tangential lower bound is $\kappa_2(r)$. Hence tangential variance is bounded below at order $r/n$, while radial variance is bounded below at order $r^3/n$. The analogous Bernoulli bounds replace $h_2$ by $h_1$. The same quadratic ratio therefore appears in these coordinatewise information lower bounds and in stationary fixed-step descent, although the latter can be removed by a nonstationary schedule.

# 9 Bias and universality

A conventional neuron includes a bias $b$. At $b = 0$ the augmented parameter $(w, b)$ has Fisher/Hessian block

$$\mathbb{E}\left[ h(rZ) \begin{pmatrix} XX^\top & X \\ X^\top & 1 \end{pmatrix} \right].$$

For even $h$, the radial–bias cross term $\mathbb{E}[Zh(rZ)]$ vanishes. The bias eigenvalue equals $\alpha_h(r)$, so the spectrum contains one radial eigenvalue $\beta_h$ and $d$ tangential/bias eigenvalues $\alpha_h$. A nonzero bias preserves the $(d-1)$ tangential eigenspace but replaces the radial and bias coordinates by the explicit $2 \times 2$ block

$$\begin{pmatrix} \mathbb{E}[Z^2 h(rZ + b)] & \mathbb{E}[Zh(rZ + b)] \\ \mathbb{E}[Zh(rZ + b)] & \mathbb{E}[h(rZ + b)] \end{pmatrix}. \tag{45}$$

The centered theorem is therefore not an artifact of omitting the bias; it is the diagonal point of a fully explicit two-coordinate extension.

The exponent two is broader than the logistic curve. Theorem 4.1 shows that any localized sensitivity profile with finite positive $m_0$ and $m_2$ has the same $r^2$ anisotropy, with only the constant $m_0/m_2$ depending on the activation and observation model. Profiles with heavy sensitivity tails or vanishing second moment fall outside this universality class and can have different laws.

## 10 Relation to prior work

Table 1 states the claim boundary before the detailed discussion.

| Prior result | Result here | Exact difference |
|---|---|---|
| Fixed-order Löwner/Schwarzian criterion and serial closure [3, 4] | closed logistic determinant and certified $2 \times 2$ witness | explicit specialization and certificate; not a new criterion or closure law |
| One-unit Gaussian Fisher split for $\phi'^2$ [20] | spherical local bridge and Gaussian all-$p$ laws | new coefficient cross-identification and extensions; not a new one-unit decomposition |
| Bernoulli exponents and finite-sample logistic Hessian bounds [13, 16, 15, 17] | explicit pointwise block-resolved relative Gram bound for every $p > 0$ | tunable precision, all powers, angle and slab obstruction; not the first bilateral or one-sided Hessian bound |
| Generalized-logistic transforms and cumulants [22, 21] | Gamma/polygamma moments inside neuronal anisotropy | transform is classical; the unified anisotropy, endpoint laws and monotonicity are the contribution |
| Lam's historical v2 Fisher–Schwarzian construction [23] | Löwner–anisotropy coefficient identity | different objects and coefficient; no claim to the first general Fisher–Schwarzian connection |

Table 1: Result-by-result priority boundary.

Löwner matrices characterize fixed-order matrix monotonicity; modern treatments include Hiai–Sano and Heinävaara [1, 2]. Kozlovski and Sands prove the relevant classical equivalence between $Sf \geq 0$ and order-two matrix monotonicity and record the Schwarzian composition law [3]. Hence the qualitative logistic obstruction and its serial-chain extension are classical consequences; negative Schwarzian also implies the all-pair determinant sign through strict convexity of $(f')^{-1/2}$. Cook, Hammerlindl and Tucker develop the algebraically equivalent nowhere-coexpanding inequality, prove serial closure, and explicitly discuss sigmoid activations and one-dimensional neural compositions [4]. Neural papers have proposed analytic quantum activations and trainable matrix activations [5, 6], but their objectives are circuit implementation or matrix-valued parameterization, not preservation of PSD order by the spectral logistic function. In the primary sources we located, we did not find the closed logistic identity (3), the explicit certified witness, or the coefficient identity (22) linking the Löwner defect to Gaussian radial–tangential anisotropy. This is a bounded novelty statement, not a claim that no earlier source can exist.

Learning one neuron is already a nontrivial optimization problem. Yehudai and Shamir study when gradient methods learn a single neuron under broad activations and input distributions, including positive and negative results [7]. Diakonikolas et al. treat monotone neurons with adversarial label noise and obtain polynomial learners for logistic activations under log-concave inputs [8]. Wu extends learnability results to nonmonotone activations [9], and Vardi et al. show that adding a bias qualitatively changes the ReLU landscape [10]. Recent single-index work analyzes SGD under anisotropic Gaussian inputs [11]. Under Gaussian logistic regression, Hsu and Mazumdar study the signal-dependent difficulty of direction and temperature estimation [12]. Chen and Mazumdar identify the Bernoulli radial and tangential Hessian functions, prove their $r^{-3}$ and $r^{-1}$ orders and show tangential curvature exceeds radial curvature for every nonzero signal [13]; they subsequently prove a minimax norm-estimation rate of order $\sqrt{r^3/n}$ [14]. Ostrovskii and Bach prove a uniform bilateral constant-factor empirical-Hessian sandwich for Gaussian logistic regression [16, Eq. (92)], while Fisher et al. give a generic fixed-point standardized-Hessian concentration framework [17]. Chardon, Lerasle and Mourtada prove, for $p = 1$, a uniform one-sided empirical Hessian lower bound under the sufficient condition $n \gtrsim r(d + t)$ [15, Theorem 6]. We therefore do not claim the Bernoulli exponents, the $p = 1$ ordering, the effective transition-slab scale, the first bilateral Hessian concentration, or the first finite-sample lower bound. Our empirical contribution is the explicit pointwise $(1 \pm \varepsilon)$ approximation (35) for all $p > 0$, with separate block errors, saturation scaling, radial-eigenspace angle and slab obstruction.

Fisher geometry and natural gradient are classical [18]. Karakida et al. study broad Fisher spectral statistics for random deep networks [19]; Amari, Karakida and Oizumi also give the direct one-unit Gaussian $\phi'^2$ decomposition and bias coupling used here [20]. Our radial contribution is therefore not the two-eigenspace reduction. It is the profile-generic moment theorem, nonasymptotic brackets, closed all-power logistic constants, strict global monotonicity, and their connection through (22) to the Löwner order defect. The generalized-logistic Gamma transform and its polygamma cumulants in Lemma 6.1 are classical [22, 21]; the claim is their neural-geometric use, ratio monotonicity and endpoint package, not discovery of the distributional identity. Lam's v2 preprint connected Schwarzian curvature with Fisher–Rao geometry on manifolds of densities [23]. The current v3 explicitly corrects and supersedes v1–v2 and removes that architecture [24]. Neither version contains the one-neuron radial–tangential coefficient or its equality with the two-point Löwner defect. We cite v2 only as a dated historical adjacency; its $L^p$ parameter is unrelated to the sensitivity exponent in $h_p = \sigma'^p$.

## 11 Reproducibility and hostile checks

The supplied script `repro/verify_saturation_law.py` performs the following independent checks:

1. it spot-checks (3) at separated scales and constructs Proposition 2.2 with 256-bit directed Arb balls;
2. it checks the spherical fourth-moment coefficient for Gaussian and fixed-radius spherical inputs;
3. it compares numerical quadrature of $m_0, m_2, m_4$ with the Gamma/polygamma formula for several noninteger and integer $p$;
4. it checks the small-$r$ jet through $r^6$ and the refined large-$r$ expansion;
5. it evaluates $\alpha, \beta$ directly over a logarithmic radius grid and checks positivity and grid monotonicity for $p = 1, 2$;
6. it checks (24)–(26) at declared finite scales;
7. for three declared fixed random seeds it evaluates the empirical radial, tangential and cross blocks and verifies that those realized errors lie inside (33);
8. it checks both sides of (19)–(20) at declared finite radii; and
9. it regenerates Figure 1 and a JSON certificate with all evaluated values.

The numerical monotonicity and finite-radius checks are performed on declared grids; the universal statements are proved analytically above. The replay uses Python, python-flint, NumPy, SciPy, and Matplotlib. It is diagnostic rather than a substitute for the analytic proofs, which reduce to the Löwner determinant identity, the Schwarzian chain rule, Gaussian orthogonal decomposition, Bernstein/net bounds, an elementary exponential bound, and closed Fourier moments. The stable record, versioned source archives, audit reports and manifests are public at `https://arr-research.github.io/papers/ARR-2026-53CTRKDSP685PT51/`.

## 12 Limitations and next theorem

The Löwner theorem concerns spectral matrix functions, not entrywise vector activations; it does not say that a standard feedforward neuron reverses coordinatewise order. Full global saturation and empirical results assume Gaussian input. The spherical theorem controls only the local quadratic jet: covariance isotropy without orthogonal invariance is insufficient, and $q_X$ alone determines neither global monotonicity nor the large-$r$ constant. Elliptical Gaussians can be whitened, but radial/tangential directions then use the covariance metric.

The empirical theorem is pointwise for a fixed teacher and raw weighted Gram matrix. It is not a uniform landscape theorem, an estimator guarantee using labels, or a minimax lower bound. Its explicit net constants are conservative. The necessary $\Omega(r\log(1/\delta))$ slab factor and $n \geq d$ rank obstruction do not prove a matching $\Omega(rd)$ product lower bound. That requires a smallest-singular-value theorem for a heteroskedastic logistic-weighted Gaussian design and is the strongest proof-complete successor problem. The optimization result remains local, and the Cramér–Rao statement concerns regular unbiased estimation.

## AI assistance disclosure

AI tools assisted with public-literature and repository triage, proof stress testing, numerical replay, drafting, and PDF production. The author remains responsible for the mathematical claims and any remaining errors.

## References


[1] F. Hiai and T. Sano, *Löwner matrices of matrix convex and monotone functions*, Journal of the Mathematical Society of Japan 64(2) (2012), 343–364. `https://arxiv.org/abs/1007.2478`

[2] O. Heinävaara, *Characterizing matrix monotonicity of fixed order on general sets*, preprint (2019). `https://arxiv.org/abs/1906.06155`

[3] O. Kozlovski and D. Sands, *Higher order Schwarzian derivatives in interval dynamics*, Fundamenta Mathematicae 206 (2009), 217–239. `https://arxiv.org/abs/0812.2646`

[4] A. Cook, A. Hammerlindl and W. Tucker, *Nowhere coexpanding functions*, Chaos 33 (2023), 123105. `https://arxiv.org/abs/2303.12814`

[5] M. Maronese, C. Destri and E. Prati, *Quantum activation functions for quantum neural networks*, Quantum Information Processing 21 (2022), 128. `https://arxiv.org/abs/2201.03700`

[6] Z. Liu, S. Cao, Y. Li and L. Zikatanov, *Neural networks with trainable matrix activation functions*, Journal of Machine Learning for Modeling and Computing 6(2) (2025), 1–11. `https://arxiv.org/abs/2109.09948`

[7] G. Yehudai and O. Shamir, *Learning a Single Neuron with Gradient Methods*, Proceedings of Machine Learning Research 125 (2020), 3756–3786. `https://proceedings.mlr.press/v125/yehudai20a.html`

[8] I. Diakonikolas, V. Kontonis, C. Tzamos and N. Zarifis, *Learning a Single Neuron with Adversarial Label Noise via Gradient Descent*, Proceedings of Machine Learning Research 178 (2022), 4313–4361. https://proceedings.mlr.press/v178/diakonikolas22c.html

[9] L. Wu, *Learning a Single Neuron for Non-monotonic Activation Functions*, Proceedings of Machine Learning Research 151 (2022), 4178–4197. https://proceedings.mlr.press/v151/wu22c.html

[10] G. Vardi, G. Yehudai and O. Shamir, *Learning a Single Neuron with Bias Using Gradient Descent*, Advances in Neural Information Processing Systems 34 (2021), 28690–28700. https://arxiv.org/abs/2106.01101

[11] G. Braun, M. H. Quang and M. Imaizumi, *Learning a Single Index Model from Anisotropic Data with Vanilla Stochastic Gradient Descent*, Proceedings of Machine Learning Research 258 (2025), 1216–1224. https://proceedings.mlr.press/v258/braun25a.html

[12] D. Hsu and A. Mazumdar, *On the Sample Complexity of Parameter Estimation in Logistic Regression with Normal Design*, Proceedings of Machine Learning Research 247 (2024), 2418–2437. https://proceedings.mlr.press/v247/hsu24a.html

[13] J. Chen and A. Mazumdar, *Finite-Sample Performance of Gradient Descent in Logistic Regression with Gaussian Design*, arXiv:2606.21683 (2026). https://arxiv.org/abs/2606.21683

[14] J. Chen and A. Mazumdar, *Minimax Optimal Estimator and Improved Error Rate for the MLE in Logistic Regression with Gaussian Design*, arXiv:2608.17260 (2026). https://arxiv.org/abs/2608.17260

[15] H. Chardon, M. Lerasle and J. Mourtada, *Finite-sample performance of the maximum likelihood estimator in logistic regression*, arXiv:2411.02137v3 (first posted 2024; v3 2026). https://arxiv.org/abs/2411.02137v3

[16] D. M. Ostrovskii and F. Bach, *Finite-sample analysis of M-estimators using self-concordance*, Electronic Journal of Statistics 15(1) (2021), 326–391. https://doi.org/10.1214/20-EJS1780

[17] J. Fisher, L. Liu, K. Pillutla, Y. Choi and Z. Harchaoui, *Influence Diagnostics under Self-concordance*, Proceedings of Machine Learning Research 206 (2023), 10028–10076. https://proceedings.mlr.press/v206/fisher23a.html

[18] S.-i. Amari, *Natural Gradient Works Efficiently in Learning*, Neural Computation 10 (1998), 251–276. https://doi.org/10.1162/089976698300017746

[19] R. Karakida, S. Akaho and S.-i. Amari, *Universal Statistics of Fisher Information in Deep Neural Networks: Mean Field Approach*, Proceedings of Machine Learning Research 89 (2019), 1032–1041. https://proceedings.mlr.press/v89/karakida19a.html

[20] S.-i. Amari, R. Karakida and M. Oizumi, *Fisher Information and Natural Gradient Learning in Random Deep Networks*, Proceedings of Machine Learning Research 89 (2019), 694–702. https://proceedings.mlr.press/v89/amari19a.html

[21] C. J. Lee, A. Zito, H. Sang and D. B. Dunson, *Logistic-beta processes for dependent random probabilities with beta marginals*, Bayesian Analysis 20(4) (2025), 1345–1369. https://doi.org/10.1214/25-BA1541

[22] M. O. Ojo and A. K. Olapade, *On a Six-Parameter Generalized Logistic Distribution*, Kragujevac Journal of Mathematics 26 (2004), 31–38. https://imi.pmf.kg.ac.rs/kjm/pub/12616736649184_5.pdf

[23] H. P. G. Lam, *Real Bers embedding on the line: Fisher–Rao linearization, Schwarzian curvature, and scattering coordinates*, arXiv:2602.07373v2 (2026). https://arxiv.org/abs/2602.07373v2

[24] H. P. G. Lam, *Zero-energy scattering and the real Bers image on the line*, arXiv:2602.07373v3 (2026), correcting and superseding v1–v2. https://arxiv.org/abs/2602.07373v3