

The Schwarzian Bridge in a Single Sigmoid Neuron

Exact Löwner Determinants and Gaussian Saturation Anisotropy

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 2×2 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 $h_p = (g')^p$,

$$\frac{\mathbb{E}h_p(rZ)}{\mathbb{E}[Z^2h_p(rZ)]} = 1 - pSg(0)r^2 + O(r^4),$$

whereas the adjacent-point Löwner determinant is $g'(0)^2Sg(0)\delta^2/6 + O(\delta^3)$. Negative Schwarzian therefore simultaneously gives local noncommutative order failure and positive initial radial–tangential anisotropy.

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

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

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

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

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 the logistic sigmoid this ratio is the spectral condition number, and we evaluate its constants in closed form. Squared-output loss gives $\kappa(r) \sim 3r^2/(\pi^2 - 6)$, while Bernoulli likelihood gives $\kappa(r) \sim 3r^2/\pi^2$. Prior work contains the Bernoulli exponents and the one-unit squared-output spectral reduction; our additions are the Schwarzian bridge, profile-generic brackets, exact logistic constants, and strict ratio monotonicity. In both cases the anisotropy is strictly increasing from one. We prove explicit finite- r brackets, translate the law into optimally tuned stationary fixed-step local gradient-descent rates and Cramér–Rao lower bounds, and supply a deterministic replay. The bridge identifies a shared local source for two effects already present in one neuron, while the large- r theorem quantifies the later saturation regime.

Keywords. single neuron; sigmoid; matrix monotonicity; Löwner order; Schwarzian derivative; Fisher information; Gaussian design.

1 Question, result, and claim boundary

The scalar logistic curve is

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

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 2×2 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 Schwarzian bridge equating the local order defect with the leading radial–tangential anisotropy of the Gaussian sensitivity matrix at a centered inflection;
4. a profile-generic two-eigenvalue refinement under Gaussian design, with source-level finite- r brackets, closed logistic constants and strict ratio monotonicity for both $h = \sigma'$ and $h = \sigma'^2$; and
5. 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. 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 2×2 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

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

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

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

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*

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

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

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

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. $\square$

This is stronger than finding one bad scale: changing the slope α , 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*

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

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

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

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

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

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. $\square$

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 $A = \text{diag}(x, y)$, 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

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

The logistic sigmoid and a scaled hyperbolic tangent satisfy

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

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

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

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

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

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

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

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

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

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

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

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

$$H_B(w_\star) = \mathbb{E}[\sigma'(w_\star^\top X)XX^\top]. \quad (14)$$

Both are instances of the sensitivity matrix

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

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 [18]. 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 $h : \mathbb{R} \rightarrow [0, \infty)$ be measurable with $\mathbb{E}[(1 + Z^2)h(rZ)] < \infty$. Then*

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

where

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

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

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

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

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

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

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

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

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

5 The Schwarzian bridge

The matrix-order and saturation calculations meet exactly at an inflection point. The next statement applies beyond the logistic curve and turns that meeting into a coefficient identity.

Theorem 5.1 (local order–anisotropy bridge). *Let $g \in C^5(\mathbb{R})$ satisfy $g' > 0$ and $g''(0) = 0$. For $p > 0$, suppose $h_p = (g')^p$ is C^4 with bounded fourth derivative, and define*

$$\alpha_p(r) = \mathbb{E}h_p(rZ), \quad \beta_p(r) = \mathbb{E}[Z^2 h_p(rZ)], \quad \kappa_p(r) = \frac{\alpha_p(r)}{\beta_p(r)}.$$

Then

$$\lim_{r \downarrow 0} \frac{\kappa_p(r) - 1}{r^2} = -p Sg(0) = -\frac{6p}{g'(0)^2} \lim_{\delta \rightarrow 0} \frac{\det L_g(0, \delta)}{\delta^2}. \quad (22)$$

In particular, $Sg(0) < 0$ makes $L_g(0, \delta)$ indefinite for every sufficiently small $\delta \neq 0$ and makes $\kappa_p(r) > 1$ for all sufficiently small $r > 0$.

Proof. Taylor expansion under the Gaussian expectation, with the bounded fourth derivative controlling the remainder, gives

$$\alpha_p(r) = h_p(0) + \frac{h_p''(0)}{2} r^2 + O(r^4), \quad \beta_p(r) = h_p(0) + \frac{3h_p''(0)}{2} r^2 + O(r^4).$$

Consequently

$$\kappa_p(r) = 1 - \frac{h_p''(0)}{h_p(0)} r^2 + O(r^4).$$

Because $\log h_p = p \log g'$ and $g''(0) = 0$,

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

Finally, the adjacent-point expansion (11), applied to g at zero, yields the second equality in (22). $\square$

For the standard logistic sigmoid, $S\sigma = -1/2$. Theorem 5.1 therefore predicts the exact initial coefficients $1/2$ for $h_1 = \sigma'$ and 1 for $h_2 = \sigma'^2$, recovered globally below in (25)–(26). This equality of coefficients is the promised bridge: it does not merely place two sigmoid facts side by side, but identifies their common differential invariant.

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 The logistic constants and the exact saturation profile

For $p \in \{1, 2\}$ set $h_p(t) = \sigma'(t)^p$. The case $p = 1$ is Bernoulli Fisher curvature and $p = 2$ is squared-output curvature.

Lemma 6.1 (closed sensitivity moments). *The first three even moments are*

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. For h_1 , these are the normalization, variance, and fourth central moment of the standard logistic density. For $h_2(t) = \frac{1}{16} \text{sech}^4(t/2)$, its Fourier transform is

$$\widehat{h}_2(k) = \int_{\mathbb{R}} e^{ikt} h_2(t) dt = \frac{1}{6} \Gamma(2 + ik) \Gamma(2 - ik). \quad (23)$$

Differentiating at zero and using $\psi_1(2) = \pi^2/6 - 1$ and $\psi_3(2) = \pi^4/15 - 6$ gives the displayed values. $\square$

Theorem 6.2 (strict logistic anisotropy). *For $p = 1$ and $p = 2$, $\kappa_p(0) = 1$, $\kappa_p(r) > 1$ for $r > 0$, and κ_p is strictly increasing on $(0, \infty)$. Moreover,*

$$\kappa_1(r) \sim \frac{3}{\pi^2} r^2, \quad \kappa_2(r) \sim \frac{3}{\pi^2 - 6} r^2. \quad (24)$$

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

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

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 α_p/β_p increases strictly. At $r = 0$ isotropy gives one, and Lemma 6.1 inserted into Theorem 4.1 gives (24). $\square$

Figure 1 shows the entire saturation profile. The solid and dashed curves are direct Gaussian integrals; the dotted curves are the asymptotic constants in (24).

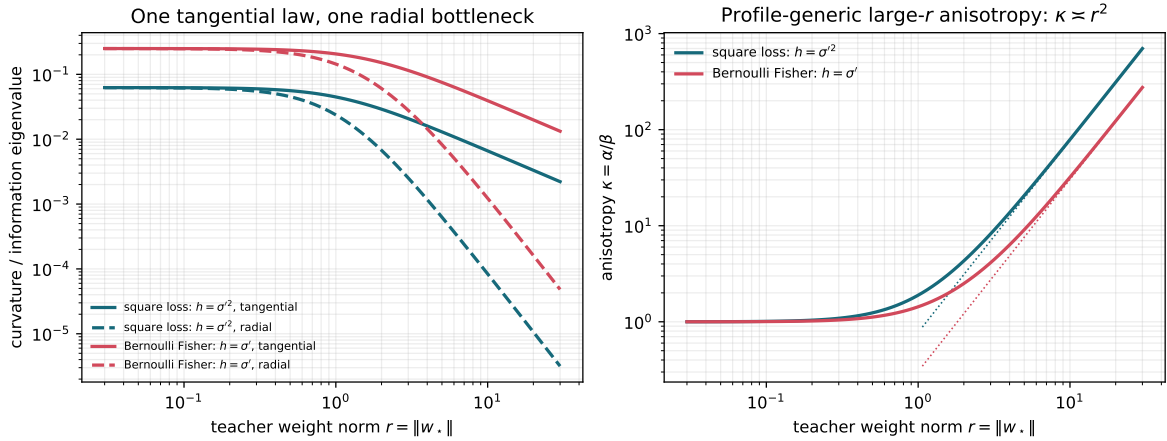


Figure 1: Exact curvature/information eigenvalues and their ratio. Saturation leaves $d - 1$ tangential directions at order r^{-1} but pushes the radial direction to order r^{-3} . The quadratic condition-number law is not a fitted exponent; it follows from Theorem 4.1.

At the nonsaturated endpoint the expansions are also explicit:

$$\kappa_1(r) = 1 + \frac{1}{2}r^2 - \frac{1}{8}r^4 + O(r^6), \quad (25)$$

$$\kappa_2(r) = 1 + r^2 - \frac{1}{4}r^4 + O(r^6). \quad (26)$$

They follow by expanding $\text{sech}^2(rZ/2)$ and $\text{sech}^4(rZ/2)$ and using Gaussian moments.

7 What the law costs

7.1 Stationary fixed-step gradient descent

For either logistic profile, assume $d \geq 2$ and that one constant scalar step size η 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

$$\eta_\star = \frac{2}{\alpha + \beta}, \quad \rho_\star = \frac{\alpha - \beta}{\alpha + \beta} = \frac{\kappa - 1}{\kappa + 1}. \quad (27)$$

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

$$T_{e,\text{sq}} \sim \frac{3}{2(\pi^2 - 6)} r^2, \quad T_{e,\text{B}} \sim \frac{3}{2\pi^2} r^2 \quad (28)$$

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.

7.2 Estimation

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

$$\text{Var}(v^\top \hat{w}) \geq \frac{\tau^2}{n\alpha_2(r)}, \quad \text{Var}(u^\top \hat{w}) \geq \frac{\tau^2}{n\beta_2(r)}. \quad (29)$$

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.

8 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 β_h and d tangential/bias eigenvalues α_h . A nonzero bias preserves the $(d - 1)$ tangential eigenspace but replaces the radial and bias coordinates by the explicit 2×2 block

$$\begin{pmatrix} \mathbb{E}[Z^2 h(rZ + b)] & \mathbb{E}[Zh(rZ + b)] \\ \mathbb{E}[Zh(rZ + b)] & \mathbb{E}[h(rZ + b)] \end{pmatrix}. \quad (30)$$

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.

9 Relation to prior work

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 in a finite-sample gradient-descent analysis [13], and subsequently prove a minimax norm-estimation rate of order $\sqrt{r^3/n}$ [14]. Chardon, Lerasle and Mourtada give complementary finite-sample MLE guarantees [15]. The radial contribution here is therefore an exact refinement, not discovery of those Bernoulli exponents: it gives profile-generic moment brackets and closed leading constants, including finite-radius bounds and strict ratio monotonicity for the squared-output profile $h = \sigma'^2$.

Fisher geometry and natural gradient are classical [16]. Karakida et al. study broad Fisher spectral statistics for random deep networks [17]; Amari, Karakida and Oizumi also give the direct one-unit Gaussian ϕ'^2 decomposition and bias coupling used here [18]. Our radial contribution is therefore not the two-eigenspace reduction. It is the profile-generic moment theorem, nonasymptotic brackets, closed logistic constants, strict global monotonicity, and their connection through (22) to the Löwner order defect. An earlier version of Lam’s preprint connects Schwarzian curvature with Fisher–Rao geometry on manifolds of densities [19]; it does not contain the Gaussian one-neuron radial–tangential coefficient or its equality with the two-point Löwner defect. The priority claimed here is only that coefficient-level cross-identification, not the first Schwarzian–Fisher connection in general.

10 Reproducibility and hostile checks

The supplied script `repro/verify_saturation_law.py` performs six 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 two logistic coefficients in the Schwarzian bridge at three small radii;
3. it compares numerical quadrature of m_0, m_2, m_4 with Lemma 6.1;
4. it evaluates α, β directly over a logarithmic radius grid and checks positivity and monotone anisotropy;
5. it checks both sides of (19)–(20) at declared finite radii; and
6. 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, an elementary exponential bound, and closed Fourier moments.

11 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. The radial theorem assumes isotropic Gaussian inputs. Elliptical Gaussians can be whitened, but the Euclidean notions of “radial” and “tangential” then inherit the covariance metric. Non-Gaussian inputs need not have a two-eigenvalue matrix. The optimization result is local, and the Cramér–Rao statement concerns regular unbiased estimation. Finite samples can also hide the transition slab entirely; quantifying the probability that an empirical Gram resolves the r^{-3} population mode is a separate concentration problem and is not implied by the Fisher parameter-estimation scale. A proof-complete successor should specify the estimator, norm, dimension dependence, and relative-versus-absolute precision, then give matching concentration and lower bounds rather than infer sample complexity from population curvature alone.

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:2411.02137 (2024). <https://arxiv.org/abs/2411.02137>
- [16] S.-i. Amari, *Natural Gradient Works Efficiently in Learning*, Neural Computation 10 (1998), 251–276. <https://doi.org/10.1162/089976698300017746>
- [17] 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>
- [18] 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>
- [19] 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>