Gulf Journal of Mathematics ISSN: 2309-4966
Vol 23, Issue 2 (2026) 1-18 https://doi.org/10.56947/bs4wx079

HERMITE STRUCTURE AND SPECTRAL PROPERTIES OF SYMMETRIC GATED ACTIVATION FUNCTIONS

CHIHEB EDDINE OUALHA[1*] and MOHAMMED LAKHDAR HADJI[2]

Abstract. We study a family of self-gated functions where the gate satisfies a reflection symmetry, and this forces a rigid Hermite structure. We then specialize the abstract theory to the Swish and GELU families by proving strict monotonicity of the spectral ratio, and deriving both the small- and large-parameter asymptotic expansions. For GELU specifically, we also derive the exact closed form for the spectral ratio. We then enrich the symmetric family by adding odd perturbations and show that the leading Hermite mode becomes continuously controllable.

Keywords. Random Matrix Theory, Hermite coefficients, activation functions, Swish, GELU.
2020 Mathematics Subject Classification. Primary 60B20; Secondary 68T07, 42C10.

1. Introduction

Artificial neural networks have become the main tool in machine learning today because they are capable of approximating complex nonlinear functions and have been successfully applied in different fields. Activation functions, which are components of the network architecture, affect the network's optimization process, the way the signal travels through the network, and the stability at the start of training. Older activations like ReLU [10], sigmoid, and hyperbolic tangent have been the subject of deep research. More recently, self-gated activation functions like Swish [5, 14], Mish [9], and GELU [6] have shown excellent performance in a number of real-world experiments. Related architecture-level variants that learn the activation shape itself have also been explored, for example, in fractional-spline neural networks [4].

Besides performance on data, activation functions can also be understood from their structure and spectral features. In the last few years, with Gaussian–Hermite expansions [16], the methods from random matrix theory [8, 17, 18] have turned out to be powerful instruments for unveiling the behavior of nonlinear random features and neural network representations. More precisely, the

Date: Received: May 10, 2026; Revised: Jun 9, 2026; Accepted: Jun 14, 2026.
* Corresponding author

theory of nonlinear random matrices developed recently by Pennington–Worah [12, 11] shows how the spectral distribution of Gram matrices after applying activations can be connected to a few scalar quantities dependent on the activation function. This insight has led to several research papers [1, 2, 13] investigating how different activation functions impact spectral properties at initialization.

Our goal is to study the relationship between symmetry and the Hermite composition of self-gated activation functions and from there, explore the spectral parameters that arise in nonlinear random matrix theory. More preciesly, we deal with activations of the form:

$$f_\beta(x) = xh(\beta x)$$

where the gate function h is positive, bounded, and satisfies the reflection symmetry

$$h(t) + h(-t) = 2C$$

for some constant $C > 0$. This class includes several commonly used activations, such as Swish and GELU, as particular cases.

Reflection symmetry turns out to severely restrict the possible form of the function

$$f_\beta(x) = Cx + e_\beta(x),$$

where e_β is an even function. So, the odd Hermite component is fully determined by the constant C: the odd Hermite coefficients are independent of the gating parameter β and all higher odd modes disappear. In particular, the leading Hermite coefficient does not change when β is varied, so the corresponding Pennington–Worah parameter cannot be adjusted just by varying the symmetric family. This inflexibility results in a structural barrier for reaching certain spectral regimes in the nonlinear random matrix model.

Inspired by this effect, we propose a general odd perturbation scheme of the form

$$f_{\alpha,\beta}(x) = xh(\beta x) + \alpha\psi(x),$$

where ψ is an odd function having a non-zero first Hermite coefficient. Perturbations like these maintain the symmetric gate architecture for the even sector while modifying the odd Hermite coefficients. This leads to regaining the continuous tuning of the leading Hermite mode and as a result, one is able to observe spectral properties that are not accessible in the original symmetric class.

From a spectral perspective, the derived family shows a larger attainable region in the parameter space controlling the limiting empirical spectrum. In particular, the perturbed class can reach the degenerate case $\zeta = 0$ which is related to the Marchenko–Pastur limit within the Pennington–Worah structure. More broadly, we describe the impact of odd perturbations on the pair of spectral parameters (η, ζ) and determine the least symmetry-breaking required to surpass the critical point.

We then specialize the general theory to the Swish and GELU families. For Swish, we establish strict monotonicity of the relevant spectral ratio and derive asymptotic expansions in both the small- and large-parameter regimes. For GELU, we obtain an explicit formula for the same spectral ratio, together with strict monotonicity and asymptotic expansions. These examples illustrate how the abstract Hermite rigidity mechanism translates into concrete spectral statements for widely used activation functions.

The main contributions of this paper can be summarized as follows.

- We identify a broad class of symmetric self-gated activations whose odd Hermite sector is rigid under parameter variation.
- We show that this rigidity fixes the leading Pennington–Worah parameter and restricts the accessible spectral regime.
- We derive strict monotonicity results within this framework for GELU and Swish.
- We derive asymptotic expansions for both small- and large-parameter asymptotic regimes.
- For GELU specifically, we derive a closed form for the ratio.
- We introduce a general odd perturbation framework that restores control of the leading Hermite mode.
- We characterize the reachable region of the resulting spectral parameters and show that the degenerate regime $\zeta = 0$ becomes accessible.

2. Preliminaries and notation

Throughout the paper, $Z \sim \mathcal{N}(0,1)$ denotes a standard Gaussian random variable. The probabilists' Hermite polynomials $\{H_k\}_{k\geq 0}$ form an orthogonal basis of $L^2(\mathbb{R},\gamma)$, where

$$d\gamma(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Any $f \in L^2(\mathbb{R},\gamma)$ has the Hermite expansion

$$f(x) = \sum_{k=0}^{\infty} c_k H_k(x), \qquad c_k = \frac{1}{k!}\mathbb{E}[f(Z)H_k(Z)].$$

We will use the two scalar quantities

$$\eta(f) := \mathbb{E}[f(Z)^2], \qquad \zeta(f) := \big(\mathbb{E}[f(Z)Z]\big)^2.$$

The next remark connects them to the Hermite coefficients.

Remark 2.1. If

$$f(x) = \sum_{k=0}^{\infty} c_k H_k(x),$$

then, since $H_1(x) = x$,

$$c_1 = \mathbb{E}[f(Z)Z].$$

Hence

$$\zeta(f) = c_1^2.$$

Moreover, Parseval's identity yields

$$\eta(f) = \sum_{k=0}^{\infty} k!\, c_k^2.$$

Definition 2.2. A function $g : \mathbb{R} \to \mathbb{R}$ is called a *symmetric gate with constant* C if

$$g(t) + g(-t) = 2C \qquad \forall t \in \mathbb{R},$$

for some constant $C > 0$.

Definition 2.3. We denote by $\mathcal{H}$ the class of all functions $h : \mathbb{R} \to \mathbb{R}$ such that:

(i) h is a symmetric gate with constant $C > 0$;
(ii) $h(t) > 0$ for all $t \in \mathbb{R}$;
(iii) there exists $M > 0$ such that $h(t) < M$ for all $t \in \mathbb{R}$.

Example 2.4. The sigmoid function [3] $\sigma(x) = (1 + e^{-x})^{-1}$ and the Gaussian CDF

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2}\, ds$$

belong to $\mathcal{H}$ with $C = \frac{1}{2}$. These are the gates underlying the Swish and GELU activations.

3. Rigidity of symmetric gated activations

We begin with the basic decomposition induced by the reflection symmetry.

Lemma 3.1. *Let*

$$f_\beta(x) = xg(\beta x),$$

where $g : \mathbb{R} \to \mathbb{R}$ *is a symmetric gate with constant* $C > 0$*. Then*

$$f_\beta(x) = Cx + e_\beta(x),$$

where e_β *is even.*

Proof. By symmetry of g,

$$g(-\beta x) = 2C - g(\beta x).$$

Therefore

$$f_\beta(-x) = (-x)g(-\beta x) = -2Cx + xg(\beta x),$$

so

$$f_\beta(x) - f_\beta(-x) = 2Cx.$$

Defining $e_\beta(x) := f_\beta(x) - Cx$ gives

$$e_\beta(-x) = f_\beta(-x) + Cx = f_\beta(x) - Cx = e_\beta(x),$$

so e_β is even. $\square$

The next proposition shows that the odd Hermite sector is rigid.

Proposition 3.2. *Let*

$$f_\beta(x) = xg(\beta x),$$

where g is a symmetric gate with constant $C > 0$. If

$$f_\beta(x) = \sum_{k=0}^{\infty} c_k(\beta) H_k(x),$$

then

$$c_1(\beta) = C, \qquad c_{2m+1}(\beta) = 0 \quad (m \geq 1).$$

In particular, every odd Hermite coefficient is independent of β.

Proof. By Theorem 3.1,

$$f_\beta(x) = Cx + e_\beta(x),$$

with e_β even. Since H_k is odd when k is odd, the product $e_\beta H_k$ is odd for odd k, so

$$\mathbb{E}[e_\beta(Z) H_k(Z)] = 0 \qquad (k \text{ odd}).$$

Thus, for odd k,

$$c_k(\beta) = \frac{C}{k!} \mathbb{E}[Z H_k(Z)].$$

For $k = 1$, $H_1(x) = x$, so $c_1(\beta) = C$. For $k = 2m + 1 \geq 3$, orthogonality to H_1 gives $\mathbb{E}[Z H_k(Z)] = 0$, hence $c_{2m+1}(\beta) = 0$. □

Corollary 3.3. *For every symmetric gated family $f_\beta(x) = xg(\beta x)$ with constant $C > 0$,*

$$c_1(\beta) = C$$

for all β.

Proposition 3.4. *Let $f_\beta(x) = xg(\beta x)$ where g is a symmetric gate with constant $C > 0$. Then*

$$\zeta(\beta) = \big(\mathbb{E}[f_\beta(Z) Z]\big)^2 = C^2$$

for every β.

Proof. By Remark 2.1, $\zeta(\beta) = c_1(\beta)^2$. The result follows from Corollary 3.3. □

Proposition 3.5. *Let*

$$f_\beta(x) = xh(\beta x),$$

where $h \in \mathcal{H}$. Then

$$C^2 \leq \eta(\beta) \leq MC,$$

for every $\beta \in \mathbb{R}$. Consequently,

$$1 \leq \frac{\eta(\beta)}{\zeta(\beta)} \leq \frac{M}{C}.$$

Proof. Since $0 \leq h \leq M$, we have $h^2 \leq Mh$, hence

$$\eta(\beta) = \mathbb{E}[Z^2 h(\beta Z)^2] \leq M\,\mathbb{E}[Z^2 h(\beta Z)].$$

Using symmetry of h and the Gaussian law,

$$2\mathbb{E}[Z^2 h(\beta Z)] = \mathbb{E}[Z^2(h(\beta Z) + h(-\beta Z))] = 2C\mathbb{E}[Z^2] = 2C,$$

so $\mathbb{E}[Z^2 h(\beta Z)] = C$, and hence $\eta(\beta) \leq MC$.

For the lower bound, Cauchy–Schwarz gives

$$\eta(\beta) \geq \big(\mathbb{E}[f_\beta(Z)Z]\big)^2 = \zeta(\beta) = C^2.$$

The ratio bound follows immediately. $\square$

4. The Swish family

We now specialize to the Swish gate

$$\sigma(x) = \frac{1}{1+e^{-x}},$$

and the activation

$$f_\beta(x) = x\sigma(\beta x), \qquad \beta \geq 0.$$

Since $\sigma(t) + \sigma(-t) = 1$, this family is symmetric with $C = \frac{1}{2}$. Hence

$$\zeta(f_\beta) = \frac{1}{4}$$

for every $\beta \geq 0$, and the nontrivial parameter dependence lies entirely in

$$\eta(\beta) = \mathbb{E}[Z^2\sigma(\beta Z)^2].$$

We define the Swish spectral ratio

$$R_{\mathrm{Sw}}(\beta) := \frac{\eta(f_\beta)}{\zeta(f_\beta)} = 4\mathbb{E}[Z^2\sigma(\beta Z)^2].$$

Theorem 4.1 (Strict monotonicity of the Swish spectral ratio)**.** *The function R_{Sw} is strictly increasing on $(0,\infty)$ and satisfies*

$$R_{\mathrm{Sw}}(0) = 1, \qquad \lim_{\beta\to\infty} R_{\mathrm{Sw}}(\beta) = 2.$$

Proof. At $\beta = 0$, $f_0(x) = x/2$, so $\eta(0) = \mathbb{E}[Z^2/4] = 1/4$ and therefore $R_{\mathrm{Sw}}(0) = 1$.

As $\beta \to \infty$, $\sigma(\beta x) \to \mathbf{1}_{\{x>0\}}$ for every $x \neq 0$, and dominated convergence yields

$$\eta(\beta) \to \mathbb{E}[Z^2\mathbf{1}_{\{Z>0\}}] = \frac{1}{2}.$$

Hence $R_{\mathrm{Sw}}(\beta) \to 2$.

For monotonicity, differentiate under the integral sign:

$$\eta'(\beta) = 2\mathbb{E}\big[Z^3\sigma(\beta Z)\sigma'(\beta Z)\big].$$

Using the evenness of σ' and the identity $\sigma(t) - \sigma(-t) > 0$ for $t > 0$, one symmetrizes to obtain

$$\eta'(\beta) = 2\int_0^\infty z^3\sigma'(\beta z)\big(\sigma(\beta z) - \sigma(-\beta z)\big)\,d\gamma(z) > 0$$

for every $\beta > 0$. Therefore $R'_{\mathrm{Sw}}(\beta) = 4\eta'(\beta) > 0$ on $(0,\infty)$. $\square$

Theorem 4.2 (Small-β expansion for Swish)**.** *As $\beta \to 0$,*

$$R_{\mathrm{Sw}}(\beta) = 1 + \frac{3}{4}\beta^2 - \frac{5}{8}\beta^4 + \frac{119}{192}\beta^6 + O(\beta^8).$$

Proof. The Taylor series of σ at the origin is

$$\sigma(t) = \frac{1}{2} + \frac{t}{4} - \frac{t^3}{48} + \frac{t^5}{480} + O(t^7).$$

Squaring and substituting $t = \beta Z$ gives

$$\eta(\beta) = \frac{1}{4} + \frac{3}{16}\beta^2 - \frac{5}{32}\beta^4 + \frac{119}{768}\beta^6 + O(\beta^8),$$

because the odd Gaussian moments vanish and

$$\mathbb{E}[Z^2] = 1, \quad \mathbb{E}[Z^4] = 3, \quad \mathbb{E}[Z^6] = 15, \quad \mathbb{E}[Z^8] = 105.$$

Multiplying by 4 yields the stated expansion. □

Theorem 4.3 (Large-β expansion for Swish)**.** *As $\beta \to \infty$,*

$$R_{\mathrm{Sw}}(\beta) = 2 - \frac{4\pi^2}{3\sqrt{2\pi}}\beta^{-3} + \frac{14\pi^4}{15\sqrt{2\pi}}\beta^{-5} + O(\beta^{-7}).$$

Proof. Using $\sigma'(t) = \sigma(t)(1-\sigma(t))$, we have $\sigma(t)^2 = \sigma(t) - \sigma'(t)$. Since symmetry implies $\mathbb{E}[Z^2\sigma(\beta Z)] = 1/2$, it follows that

$$\frac{1}{2} - \eta(\beta) = \mathbb{E}[Z^2\sigma'(\beta Z)].$$

A change of variables and the moments of the logistic density σ' yield

$$\mathbb{E}[Z^2\sigma'(\beta Z)] = \frac{1}{\sqrt{2\pi}}\left(\frac{\pi^2}{3}\beta^{-3} - \frac{7\pi^4}{30}\beta^{-5} + O(\beta^{-7})\right).$$

Since $R_{\mathrm{Sw}} = 4\eta = 2 - 4\mathbb{E}[Z^2\sigma'(\beta Z)]$, the result follows. □

5. The GELU family

Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-s^2/2}\, ds$$

be the Gaussian distribution function, and define

$$f_\beta(x) = x\Phi(\beta x), \qquad \beta \geq 0.$$

Since $\Phi(t) + \Phi(-t) = 1$, this is again a symmetric gated family with $C = \frac{1}{2}$, so

$$\zeta(f_\beta) = \frac{1}{4}.$$

Set

$$\eta_{\mathrm{G}}(\beta) := \mathbb{E}[Z^2\Phi(\beta Z)^2], \qquad R_{\mathrm{G}}(\beta) := 4\eta_{\mathrm{G}}(\beta).$$

Theorem 5.1 (Closed formula for the GELU spectral ratio)**.** *Let*

$$t = \beta^2, \qquad D = 1 + 2t, \qquad \rho = \frac{t}{1+t},$$

and define

$$\Theta = \frac{\pi}{2} + \arcsin\rho.$$

Then

$$\eta_{\mathrm{G}}(\beta) = \frac{1}{\pi D}\left[\frac{\Theta}{2} + t\big(\Theta + \sin\Theta\big)\right].$$

Consequently,

$$R_{\mathrm{G}}(\beta) = \frac{4}{\pi(1+2\beta^2)}\left[\frac{\Theta}{2} + \beta^2\big(\Theta + \sin\Theta\big)\right].$$

Proof. Let U, V, Z be independent standard Gaussian variables. Then

$$\Phi(\beta Z)^2 = \mathbb{P}(U \le \beta Z,\ V \le \beta Z \mid Z),$$

so

$$\eta_{\mathrm{G}}(\beta) = \mathbb{E}\big[Z^2 \mathbf{1}_{\{U\le \beta Z,\ V \le \beta Z\}}\big].$$

Set $A = U - \beta Z$ and $B = V - \beta Z$. Then (A, B) is centered Gaussian with variances $1+\beta^2$ and covariance β^2, hence correlation $\rho = \beta^2/(1+\beta^2)$. A Gaussian regression calculation gives

$$\mathbb{E}[Z \mid A, B] = -\frac{\beta}{D}(A+B), \qquad \mathrm{Var}(Z \mid A, B) = \frac{1}{D},$$

with $D = 1 + 2\beta^2$. Therefore

$$\mathbb{E}[Z^2 \mid A, B] = \frac{1}{D} + \frac{\beta^2}{D^2}(A+B)^2.$$

The standard quadrant formula [15] gives:

$$\mathbb{P}(A \le 0, B \le 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\rho = \frac{\Theta}{2\pi}.$$

To compute the truncated moment, normalize

$$X = \frac{A}{\sqrt{1+\beta^2}}, \qquad Y = \frac{B}{\sqrt{1+\beta^2}}.$$

Then (X, Y) is a centered bivariate Gaussian vector with unit variances and correlation

$$\rho = \frac{\beta^2}{1+\beta^2}.$$

Set

$$S = \frac{X+Y}{\sqrt{2(1+\rho)}}, \qquad T = \frac{X-Y}{\sqrt{2(1-\rho)}}.$$

Then S and T are independent standard Gaussian variables. In the (S, T)-plane, the region $\{X \le 0,\ Y \le 0\}$ is a wedge with opening angle

$$\Theta = \frac{\pi}{2} + \arcsin\rho.$$

Moreover,

$$X + Y = \sqrt{2(1+\rho)}\, S.$$

Using polar coordinates and rotational invariance of the standard Gaussian measure on $\mathbb{R}^2$, we obtain

$$\mathbb{E}\left[S^2 \mathbf{1}_{\{X\le 0,\ Y\le 0\}}\right] = \frac{1}{\pi}\int_{-\Theta/2}^{\Theta/2} \cos^2\theta\, d\theta = \frac{\Theta + \sin\Theta}{2\pi}.$$

Therefore

$$\begin{aligned}\mathbb{E}\left[(X+Y)^2\mathbf{1}_{\{X\le 0,\,Y\le 0\}}\right] &= 2(1+\rho)\frac{\Theta+\sin\Theta}{2\pi}\\ &= \frac{(1+\rho)(\Theta+\sin\Theta)}{\pi}.\end{aligned}$$

Since

$$A+B=\sqrt{1+\beta^2}(X+Y),$$

we get

$$\mathbb{E}\left[(A+B)^2\mathbf{1}_{\{A\le 0,\,B\le 0\}}\right] = (1+\beta^2)(1+\rho)\frac{\Theta+\sin\Theta}{\pi}.$$

But

$$(1+\beta^2)(1+\rho)=1+2\beta^2=D.$$

Thus

$$\mathbb{E}\left[(A+B)^2\mathbf{1}_{\{A\le 0,\,B\le 0\}}\right] = \frac{D(\Theta+\sin\Theta)}{\pi}.$$

Substitution gives the claimed closed form. □

Theorem 5.2 (Strict monotonicity for GELU)**.** *The function R_{G} is strictly increasing on $(0,\infty)$ and satisfies*

$$R_{\mathrm{G}}(0)=1,\qquad \lim_{\beta\to\infty}R_{\mathrm{G}}(\beta)=2.$$

Proof. Let $t=\beta^2$. From Theorem 5.1,

$$\eta_{\mathrm{G}}(\beta)=\frac{\Theta}{2\pi}+\frac{t\sin\Theta}{\pi(1+2t)},\qquad \Theta=\frac{\pi}{2}+\arcsin\left(\frac{t}{1+t}\right).$$

Differentiating in t gives a strictly positive expression for all $t\ge 0$, hence η_{G} and therefore R_{G} are strictly increasing on $(0,\infty)$.

$$\frac{d}{dt}\eta_G(\sqrt{t})=\frac{5t+3}{2\pi(1+t)^2(1+2t)^{3/2}}.$$

This quantity is strictly positive for every $t\ge 0$. Since $t=\beta^2$, we have, for $\beta>0$,

$$\frac{d}{d\beta}\eta_G(\beta)=2\beta\frac{5\beta^2+3}{2\pi(1+\beta^2)^2(1+2\beta^2)^{3/2}}>0.$$

Therefore $R_G(\beta)=4\eta_G(\beta)$ is strictly increasing on $(0,\infty)$.

At $\beta=0$, $f_0(x)=x/2$, so $R_{\mathrm{G}}(0)=1$. As $\beta\to\infty$, $\Phi(\beta x)\to\mathbf{1}_{\{x>0\}}$ for $x\ne 0$, and dominated convergence yields $\eta_{\mathrm{G}}(\beta)\to 1/2$, so $R_{\mathrm{G}}(\beta)\to 2$. □

Theorem 5.3 (Asymptotic expansions for GELU)**.** *As $\beta\to 0$,*

$$R_{\mathrm{G}}(\beta)=1+\frac{6}{\pi}\beta^2-\frac{10}{\pi}\beta^4+\frac{49}{3\pi}\beta^6+O(\beta^8).$$

As $\beta\to\infty$,

$$R_{\mathrm{G}}(\beta)=2-\frac{5\sqrt{2}}{3\pi}\beta^{-3}+\frac{43\sqrt{2}}{20\pi}\beta^{-5}+O(\beta^{-7}).$$

Proof. Both expansions follow by Taylor expansion of the closed form in Theorem 5.1. For $\beta \to 0$, one uses

$$\frac{\beta^2}{1+\beta^2} = \beta^2 - \beta^4 + \beta^6 + O(\beta^8), \qquad \arcsin x = x + \frac{x^3}{6} + O(x^5).$$

For $\beta \to \infty$, one expands in powers of β^{-1}. $\square$

6. Odd perturbations and controllable linear modes

The preceding results show that the symmetric gated family has a rigid odd Hermite sector. We now add an odd perturbation to break this rigidity while preserving the overall shape of the activation.

Definition 6.1. Let $h \in \mathcal{H}$ and let $\psi \in L^2(\mathbb{R}, \gamma)$ be odd with

$$\mathbb{E}[\psi(Z)Z] \neq 0.$$

For $\alpha \in \mathbb{R}$ and $\beta \geq 0$, define

$$f_{\alpha,\beta}(x) := xh(\beta x) + \alpha\psi(x).$$

Let

$$\kappa := \mathbb{E}[\psi(Z)Z], \qquad \nu := \mathbb{E}[\psi(Z)^2], \qquad \eta_h(\beta) := \mathbb{E}[Z^2 h(\beta Z)^2].$$

Theorem 6.2. *Let*

$$f_{\alpha,\beta}(x) = xh(\beta x) + \alpha\psi(x),$$

where $h \in \mathcal{H}$ and ψ is odd. Then the parameter α affects only the odd Hermite coefficients of $f_{\alpha,\beta}$.

Proof. By Theorem 3.1,

$$xh(\beta x) = Cx + e_\beta(x),$$

with e_β even. Therefore

$$f_{\alpha,\beta}(x) = e_\beta(x) + \big(Cx + \alpha\psi(x)\big).$$

The first term is even and the second is odd, so by parity orthogonality the parameter α only changes the odd Hermite sector. $\square$

Proposition 6.3. *Under the hypotheses above,*

$$c_1(\alpha, \beta) = C + \alpha\kappa.$$

Proof. Since $H_1(x) = x$,

$$c_1(\alpha, \beta) = \mathbb{E}[f_{\alpha,\beta}(Z)Z] = \mathbb{E}[Z^2 h(\beta Z)] + \alpha\mathbb{E}[\psi(Z)Z].$$

By Proposition 3.4, $\mathbb{E}[Z^2 h(\beta Z)] = C$, and by definition $\mathbb{E}[\psi(Z)Z] = \kappa$. Thus $c_1(\alpha, \beta) = C + \alpha\kappa$. $\square$

Proposition 6.4. *With the same assumptions,*

$$\zeta(\alpha, \beta) = \big(C + \alpha\kappa\big)^2,$$

and

$$\eta(\alpha, \beta) = \eta_h(\beta) + 2\alpha C\kappa + \alpha^2\nu.$$

Proof. The formula for ζ follows from Proposition 6.3. For η, expand the square:

$$\eta(\alpha,\beta) = \mathbb{E}\big[(Zh(\beta Z) + \alpha\psi(Z))^2\big] = \eta_h(\beta) + 2\alpha\mathbb{E}[Zh(\beta Z)\psi(Z)] + \alpha^2\nu.$$

Using $xh(\beta x) = Cx + e_\beta(x)$ with e_β even and ψ odd, we get $\mathbb{E}[e_\beta(Z)\psi(Z)] = 0$, hence

$$\mathbb{E}[Zh(\beta Z)\psi(Z)] = C\kappa.$$

Substitution gives the claimed formula. □

Corollary 6.5. *If $\kappa \neq 0$, then*

$$\alpha^* = -\frac{C}{\kappa}$$

is the unique critical parameter such that

$$\zeta(\alpha^*,\beta) = 0.$$

Moreover, if $\eta(\alpha^*,\beta) > 0$, then

$$\frac{\eta(\alpha,\beta)}{\zeta(\alpha,\beta)} \to \infty \qquad \text{as } \alpha \to \alpha^*.$$

Proof. This follows immediately from $\zeta(\alpha,\beta) = (C + \alpha\kappa)^2$. □

7. Reachable spectral region under odd perturbations

The previous sections show that the symmetric gated family has a rigid odd Hermite sector. We now make this rigidity quantitative by characterizing exactly what can be reached after adding an arbitrary odd perturbation.

Fix $h \in \mathcal{H}$ and write

$$f_\beta(x) = xh(\beta x) = Cx + e_\beta(x),$$

with e_β even. Let $\psi \in L^2_{\text{odd}}(\gamma)$ and define

$$f_{\beta,\psi} = f_\beta + \psi.$$

Because $H_1(x) = x$ has norm one in $L^2(\gamma)$, every odd perturbation admits the orthogonal decomposition

$$\psi = aH_1 + q, \qquad q \in L^2_{\text{odd}}(\gamma), \quad q \perp H_1.$$

Then

$$f_{\beta,\psi} = e_\beta + (C + a)H_1 + q,$$

with pairwise orthogonal components.

Theorem 7.1 (Exact reachable region)**.** *For fixed $\beta \geq 0$, the set*

$$\mathcal{R}_\beta := \{(\eta(f_{\beta,\psi}), \zeta(f_{\beta,\psi})) : \psi \in L^2_{\text{odd}}(\gamma)\}$$

is preciesly

$$\mathcal{R}_\beta = \{(\eta,\zeta) \in [0,\infty)^2 : \eta \geq \eta_h(\beta) - C^2 + \zeta\}.$$

Proof. Since the three components are orthogonal,

$$\zeta(f_{\beta,\psi}) = (C+a)^2, \qquad \eta(f_{\beta,\psi}) = \eta_h(\beta) - C^2 + \zeta(f_{\beta,\psi}) + \|q\|^2_{L^2(\gamma)}.$$

Thus every reachable pair satisfies

$$\eta \geq \eta_h(\beta) - C^2 + \zeta.$$

Conversely, given any (η, ζ) with this inequality, choose a so that $(C+a)^2 = \zeta$ and choose q orthogonal to H_1 with prescribed norm. Since $L^2_{\text{odd}}(\gamma) \cap H_1^{\perp}$ is infinite-dimensional, such a q always exists. This realizes the pair. □

Corollary 7.2 (Minimal perturbation for a prescribed first Hermite mode)**.** *Fix $\beta \geq 0$ and let $\zeta_0 \geq 0$. Among all odd perturbations ψ satisfying $\zeta(f_\beta + \psi) = \zeta_0$, the minimal $L^2(\gamma)$ norm is attained by a pure first-mode perturbation and equals*

$$\min\Big\{ |\sqrt{\zeta_0} - C|, \ |-\sqrt{\zeta_0} - C| \Big\}.$$

Corollary 7.3 (Minimal perturbation to reach the critical threshold)**.** *The minimal odd perturbation forcing $\zeta(f_\beta + \psi) = 0$ is*

$$\psi^*(x) = -Cx.$$

Its norm is $\|\psi^*\|_{L^2(\gamma)} = C$, and at this threshold

$$f_\beta + \psi^* = e_\beta$$

is even.

8. Pennington–Worah consequence

We now explain the random-matrix consequence of the previous computations. The point is not to reprove the Pennington–Worah theorem, but to specialize it to the present symmetry-breaking setting. Under the hypotheses of that theorem, the limiting empirical spectral distribution depends only on the pair (η, ζ).

For the one-parameter perturbation

$$f_{\alpha,\beta}(x) = xh(\beta x) + \alpha\psi(x),$$

Proposition 6.4 gives

$$\zeta(f_{\alpha,\beta}) = (C + \alpha\kappa)^2, \qquad \eta(f_{\alpha,\beta}) = \eta_h(\beta) + 2\alpha C\kappa + \alpha^2\nu.$$

If $\kappa \neq 0$, then the critical parameter

$$\alpha^* = -\frac{C}{\kappa}$$

forces $\zeta(f_{\alpha^*,\beta}) = 0$, and hence the corresponding limiting spectral law enters the degenerate Marchenko–Pastur regime.

Corollary 8.1 (Formal Pennington–Worah consequence)**.** *Assume that the activation $f_{\alpha,\beta}$ satisfies the hypotheses of the Pennington–Worah nonlinear random matrix theorem, and assume that in the corresponding high-dimensional Gaussian random-feature model the limiting empirical spectral distribution depends on the activation only through the parameters*

$$\eta(f) = \mathbb{E}[f(Z)^2], \qquad \zeta(f) = \big(\mathbb{E}[f(Z)Z]\big)^2.$$

Let

$$f_{\alpha,\beta}(x) = xh(\beta x) + \alpha\psi(x),$$

where ψ is odd and

$$\kappa = \mathbb{E}[\psi(Z)Z] \neq 0.$$

Then the critical parameter

$$\alpha^* = -\frac{C}{\kappa}$$

satisfies

$$\zeta(f_{\alpha^*,\beta}) = 0.$$

Consequently, within the Pennington–Worah limiting spectral description, the family reaches the degenerate case associated with the Marchenko–Pastur-type limit.

Proof. By Proposition 6.4,

$$\zeta(f_{\alpha,\beta}) = (C + \alpha\kappa)^2.$$

Therefore

$$\zeta(f_{\alpha,\beta}) = 0 \quad \Longleftrightarrow \quad C + \alpha\kappa = 0,$$

which gives the unique value

$$\alpha^* = -\frac{C}{\kappa}.$$

The final statement follows by applying the Pennington–Worah limiting spectral description at the parameter value $\zeta = 0$. □

Proof. This is an immediate consequence of the formula $\zeta(f_{\alpha,\beta}) = (C + \alpha\kappa)^2$ and the dependence of the limiting law only on (η, ζ). □

Theorem 8.2 (Quadratic approach with the first Hermite mode)**.** *For the pure first-mode perturbation*

$$f_{a,\beta}(x) = xh(\beta x) + ax,$$

one has

$$\zeta(f_{a,\beta}) = (C + a)^2, \qquad \eta(f_{a,\beta}) = \eta_h(\beta) + 2aC + a^2.$$

The critical value is $a^* = -C$, and at $a = a^*$,

$$\zeta(f_{a^*,\beta}) = 0, \qquad \eta(f_{a^*,\beta}) = \eta_h(\beta) - C^2.$$

Moreover, both spectral parameters approach the threshold quadratically as $a \to a^*$.

Proof. This is the special case of Proposition 6.4 with $\psi = H_1$, so $\kappa = \nu = 1$. □

9. Numerical illustrations of the spectral transition

In this section, we illustrate the effect of the odd perturbation

$$\psi(x) = xe^{-x^2/2}$$

on the Swish activation

$$f_{\alpha,\beta}(x) = x\sigma(\beta x) + \alpha\,\psi(x).$$

We track how the Pennington–Worah parameters and the empirical spectral distribution of the post-activation Gram matrix change as α varies, and compare these results with the unperturbed Swish case $\alpha = 0$.

Recall that the Swish gate satisfies

$$\sigma(t) + \sigma(-t) = 1,$$

so it belongs to the symmetric gated class with

$$C = \frac{1}{2}.$$

Throughout this section, expectations are approximated by Monte Carlo sampling with

$$Z_i \sim \mathcal{N}(0,1),$$

and empirical spectral distributions are computed from finite random feature matrices of the form

$$Y = f_{\alpha,\beta}(WX),$$

where the entries of W and X are independent standard Gaussian random variables.

9.1. **Evolution of the First Hermite Coefficient.**

We first investigate the effect of the perturbation parameter α on the square of the first Hermite coefficient

$$\zeta(\alpha,\beta).$$

By Proposition 6.3, the parameter α directly modifies the leading Hermite mode and therefore changes the value of ζ. In contrast, Proposition 3.2 shows that, for the symmetric gated family, the parameter β does not alter the odd Hermite sector.

Figure 1 illustrates the evolution of $\zeta(\alpha,\beta)$ as α varies. The numerical results confirm the existence of a critical value near

$$\alpha = -\frac{C}{\kappa},$$

for which

$$\zeta(\alpha,\beta) = 0.$$

In the Swish example, this critical value reduces to

$$\alpha = -\sqrt{2}.$$

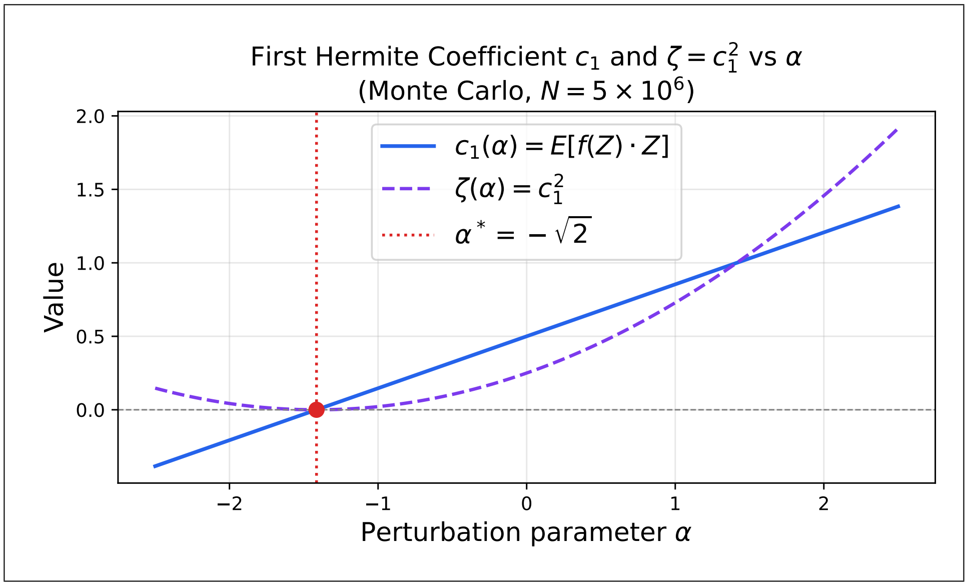

FIGURE 1. Evolution of the first Hermite coefficient $\zeta(\alpha, \beta)$ under variations of the perturbation parameter α.

9.2. **Comparison of Spectral Ratio Control.** We now compare the behavior of the Pennington–Worah ratio

$$\frac{\eta}{\zeta}$$

under variations of the Swish parameter β and the perturbation parameter α.

For the Swish family, Proposition 3.5 implies that the ratio remains confined to the bounded interval

$$1 \leq \frac{\eta(\beta)}{\zeta(\beta)} \leq 2.$$

Indeed, in this case one has $C = \frac{1}{2}$ and $M = 1$.

In contrast, Theorem 8.1 shows that the perturbed family can access arbitrarily large values of the ratio as α approaches the critical value

$$\alpha^* = -\frac{C}{\kappa},$$

which reduces to $\alpha^* = -\sqrt{2}$ in the Gaussian perturbation example.

Figure 2 compares the evolution of the ratio in the two settings. The Swish parameter β produces only bounded deformations, while the perturbation parameter α generates a substantially wider spectral regime.

9.3. **Empirical Spectral Distributions.** We finally examine the empirical spectral distribution of the post-activation Gram matrix

$$M = \frac{1}{m} Y^\top Y$$

for several representative values of the perturbation parameter α.

In particular, we compare:

$$\alpha = 0, \qquad \alpha = -1, \qquad \alpha \approx -\sqrt{2}, \qquad \alpha = 1.$$

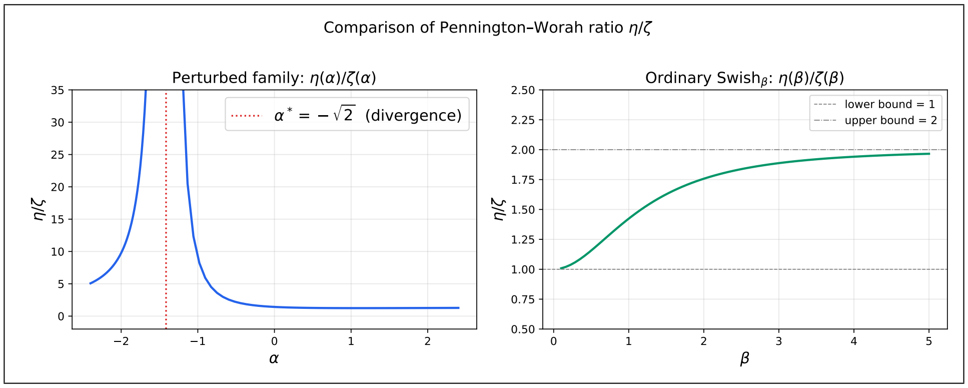

Figure 2. Comparison of the Pennington–Worah ratio η/ζ for the ordinary Swish parametrization and the perturbed family.

The numerical spectra illustrate the deformation predicted by the Pennington–Worah framework. As α approaches the critical regime, the spectrum progressively transitions toward a Marchenko–Pastur-type behavior associated with

$$\zeta \approx 0.$$

Away from the critical regime, the spectrum exhibits broader support and visible deviations from the classical Marchenko–Pastur profile.

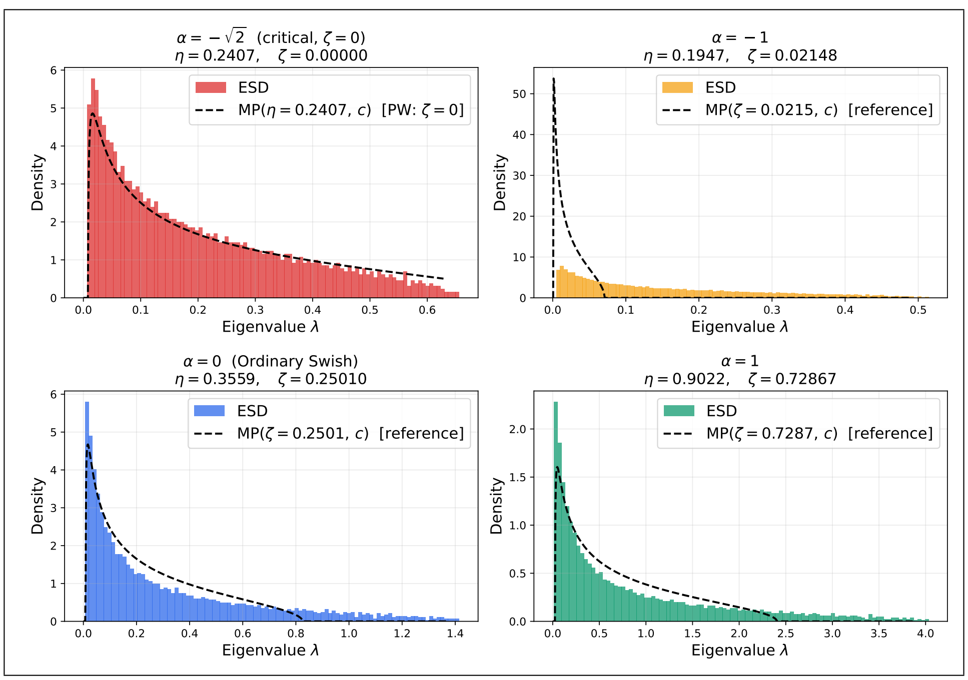

Figure 3. Empirical spectral distributions of the post-activation Gram matrix for different values of the perturbation parameter α.

The numerical experiments are consistent with the theoretical analysis developed in previous sections and illustrate how odd perturbations produce qualitatively different spectral regimes compared with the ordinary Swish parameterization.

10. Conclusion

We studied symmetric gated activations through their Hermite structure and the associated spectral quantities η and ζ. The reflection symmetry condition

$$h(t) + h(-t) = 2C$$

forces a rigid decomposition of the form

$$f_\beta(x) = Cx + e_\beta(x),$$

where e_β is even. As a consequence, the odd Hermite sector is completely determined by the constant C, and the leading Pennington–Worah parameter ζ remains fixed throughout the symmetric family. This rigidity is the basic structural feature underlying the results of the paper.

We then specialized this framework to two important examples, namely Swish and GELU. For Swish, we proved strict monotonicity of the spectral ratio and derived its small- and large-parameter asymptotic expansions. For GELU, we obtained an explicit closed formula for the spectral ratio, proved strict monotonicity, and derived corresponding asymptotic expansions. These computations show that, although the symmetric family has a rigid odd sector, the even part still produces a rich and tractable dependence on the gating parameter.

To overcome the rigidity of the symmetric class, we introduced odd perturbations of the form

$$f_{\alpha,\beta}(x) = xh(\beta x) + \alpha\psi(x).$$

This perturbation restores control of the first Hermite coefficient and yields an exact description of the resulting spectral parameters. In particular, we identified the reachable region in the (η, ζ)-plane, and showed how the threshold $\zeta = 0$ can be attained by a suitable choice of perturbation. This makes the symmetry-breaking mechanism quantitatively explicit within the Pennington–Worah framework.

Finally, the numerical illustrations confirm the theoretical picture: the unperturbed Swish family exhibits only bounded spectral variation, while the perturbed family can move into a much wider regime and approach the critical threshold predicted by the Hermite analysis. Overall, the results demonstrate that symmetry, Hermite structure, and random-matrix spectral behavior are tightly linked for self-gated activations, and that odd perturbations provide a natural way to break the rigidity of the symmetric setting.

Acknowledgement. The authors deeply thank the anonymous reviewers for their useful remarks and careful reading of the proofs presented in this paper.

References

[1] Adlam, B., Levinson, J., & Pennington, J. (2021). Random Matrix Perspective on Mixtures of Nonlinearities for Deep Learning. *arXiv preprint arXiv:1912.00827*. https://doi.org/10.48550/arXiv.1912.00827

[2] Benigni, L., & Péché, S. (2021). Eigenvalue distribution of some nonlinear models of random matrices. Electronic Journal of Probability, 26. https://doi.org/10.1214/21-EJP699

[3] Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association, 39*(227), 357–365. https://doi.org/10.1080/01621459.1944.10500699

[4] Bouaicha, A., Lamnii, M., & Ngadi, S. (2026). FRACCN: Fractional spline neural networks. *Gulf Journal of Mathematics, 22*(2). https://doi.org/10.56947/gjom.v22i2.4041

[5] Elfwing, S., Uchibe, E., & Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks, 107*, 3–11. https://doi.org/10.1016/j.neunet.2017.12.012

[6] Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*. https://arxiv.org/abs/1606.08415

[7] Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik, 1*(4), 457–483. https://doi.org/10.1070/SM1967v001n04ABEH001994

[8] Mehta, M. L. (2004). *Random matrices* (3rd ed.). Elsevier. https://doi.org/10.1016/C2009-0-22297-5

[9] Misra, D. (2019). Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*. https://arxiv.org/abs/1908.08681

[10] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814. https://icml.cc/Conferences/2010/papers/432.pdf

[11] Pennington, J., & Worah, P. (2017). Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems, 30*. https://papers.nips.cc/paper/6857-nonlinear-random-matrix-theory-for-deep-learning

[12] Pennington, J., & Worah, P. (2019). Nonlinear random matrix theory for deep learning. *Journal of Statistical Mechanics: Theory and Experiment, 2019*(12), 124005. https://doi.org/10.1088/1742-5468/ab3bc3

[13] Piccolo, V., & Schröder, D. (2021). Analysis of One-Hidden-Layer Neural Networks via the Resolvent Method *arXiv preprint arXiv:2105.05115*. https://doi.org/10.48550/arXiv.2105.05115

[14] Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*. https://arxiv.org/abs/1710.05941

[15] Sheppard, W.F. (1900), On the calculation of the double integral expressing normal correlation, Transactions of the Cambridge Philosophical Society 19, pp. 23–69.

[16] Szegő, G. (1975). *Orthogonal polynomials* (4th ed.). American Mathematical Society.

[17] Tao, T. (2012). *Topics in random matrix theory*. American Mathematical Society. https://doi.org/10.1090/gsm/132

[18] Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press. https://doi.org/10.1017/9781108231596

[1] Laboratory mathematical modeling and numerical simulation, Badji Mokhtar-Annaba University, 12, P.O.Box, 23000 Annaba, Algeria

Email address: chiheb-eddine.oualha@univ-annaba.dz

[2] Laboratory mathematical modeling and numerical simulation, Badji Mokhtar-Annaba University, 12, P.O.Box, 23000 Annaba, Algeria

Email address: mohammed-lakhdar.hadji@univ-annaba.dz