The Finite Free Stam Inequality
An Introduction to Polynomial Convolution and Root Dynamics

Abstract

The classical Stam inequality formulates the superadditivity of entropy power, or equivalently, the decrease of Fisher information upon the addition of independent random variables. This paper proves the analogue of this result for real-rooted polynomials, where random variables are replaced by polynomials, addition by the "finite free additive convolution," and Fisher information by a quantity derived from the electrostatic repulsion of roots. We establish a key algebraic inequality—the Score-Gradient Inequality—via two applications of Cauchy–Schwarz, and combine it with a flow-based framework to obtain the full Finite Free Stam Inequality: $1/\Phi_n(p \boxplus_n q) \geq 1/\Phi_n(p) + 1/\Phi_n(q)$.

Contents

1 Introduction

1.1 Analogy with Probability Theory

In probability theory and information theory, the addition of independent random sources strictly increases uncertainty. The convolution of two distributions results in a distribution that is smoother than the originals.

Mathematically, this is captured by the **Stam Inequality**. Let X and Y be independent random variables with Fisher information $I(X)$ and $I(Y)$. Then:

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}.$$

Since Fisher information I measures the structural sharpness of a distribution, its reciprocal $1/I$ acts as a measure of smoothness or disorder. This inequality asserts that the total disorder of the sum is at least the sum of the individual disorders.

1.2 The Polynomial Framework

We replace random variables with **polynomials** according to the following correspondence:

Probability	$\longleftrightarrow$	**Polynomials**
Random Variable X	$\longleftrightarrow$	Polynomial $p(x)$
Distribution of X	$\longleftrightarrow$	Roots $\lambda_1, \ldots, \lambda_n$ of $p(x)$
Addition $X + Y$	$\longleftrightarrow$	Symmetric Additive Convolution $p \boxplus_n q$
Fisher Information $I(X)$	$\longleftrightarrow$	Finite Free Fisher Information $\Phi_n(p)$

Finite Free Probability provides a bridge between random matrix theory and the geometry of polynomials. In this framework, we investigate the validity of the Stam inequality for the finite free additive convolution.

This paper presents a self-contained proof of these phenomena. We treat the roots of a polynomial as charged particles on a line and define their "energy" (Fisher information) based on their mutual repulsion. We then show that convolution acts as a diffusion process, lowering the potential energy of the system.

2 Polynomials and Root Statistics

We consider the space of monic polynomials of degree n with real coefficients, denoted $\mathcal{P}_n$. We focus on the subset $\mathcal{P}_n^{\mathbb{R}}$ of polynomials with **all real roots**.

$$p(x) = \prod_{i=1}^{n}(x - \lambda_i), \quad \lambda_i \in \mathbb{R}.$$

We assume the roots are ordered $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Definition 2.1 (Root Statistics)**.** The first and second moments of the root distribution are defined as:

$$\mu(p) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i,$$

$$\sigma^2(p) = \frac{1}{n} \sum_{i=1}^{n} (\lambda_i - \mu)^2.$$

The variance can be computed directly from the coefficients without solving for the roots.

Lemma 2.1 (The Variance Formula)**.** *Let* $p(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots$. *Then:*

$$\sigma^2(p) = \frac{(n-1)a_1^2}{n^2} - \frac{2a_2}{n}.$$

Proof. Recall Vieta's formulas: $\sum \lambda_i = -a_1$ and $\sum_{i<j} \lambda_i \lambda_j = a_2$. The sum of squared roots is $\sum \lambda_i^2 = (\sum \lambda_i)^2 - 2 \sum_{i<j} \lambda_i \lambda_j = a_1^2 - 2a_2$. Substituting into the variance definition $\sigma^2(p) = \frac{1}{n} \sum \lambda_i^2 - \mu^2$ yields the result. $\qquad\square$

3 The Symmetric Additive Convolution

The addition of polynomials in this context is not pointwise addition but a convolution operation rooted in random matrix theory.

3.1 Matrix Models and Haar Integration

Let $p(x)$ be the characteristic polynomial of a symmetric matrix A, and $q(x)$ be that of B. The sum $A + B$ has a characteristic polynomial that depends on the relative eigenbasis of A and B. To obtain a basis-independent operation, we average over all possible relative orientations using the Haar measure.

Definition 3.1 (Haar Measure on $O(n)$)**.** The group of orthogonal matrices $O(n)$ admits a unique probability measure μ_{Haar} that is invariant under left and right multiplication: $\mu(S) = \mu(gS) = \mu(Sg)$ for any $g \in O(n)$ and measurable set S. This is the uniform distribution on the rotation group.

Definition 3.2 (Symmetric Additive Convolution)**.** The finite free additive convolution of p and q is defined as the expected characteristic polynomial of the sum of randomly rotated matrices:

$$p \boxplus_n q := \int_{O(n)} \det(xI - (A + QBQ^T)) \, d\mu_{\text{Haar}}(Q).$$

In this expression, QBQ^T represents the matrix B rotated by a random orthogonal matrix Q. The term $\det(xI - (A + QBQ^T))$ is a random polynomial. Its expectation is a deterministic polynomial whose roots effectively represent the "sum" of the root sets of p and q.

3.1.1   Relation between Matrix Eigenvalues and Differential Operators

A fundamental result by Marcus, Spielman, and Srivastava connects this matrix integral to a deterministic differential operator. This connection relies on the relationship between the eigenvalues of B and the coefficients of the operator.

Let the eigenvalues of B be the roots of $q(x)$, denoted $\beta_1, \ldots, \beta_n$. The characteristic polynomial is $q(x) = \prod_{i=1}^{n}(x - \beta_i) = \sum_{k=0}^{n} b_k x^{n-k}$. The coefficients b_k are the elementary symmetric polynomials of the eigenvalues:

$$b_k = (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le n} \beta_{i_1} \ldots \beta_{i_k}.$$

The expected characteristic polynomial expansion can be written in terms of these coefficients.

Theorem 3.1 (MSS Theorem). *The expectation over the orthogonal group factorizes into a differential operator acting on p:*

$$\mathbb{E}_Q[\det(xI - (A + QBQ^T))] = \sum_{k=0}^{n} b_k \frac{(n-k)!}{n!} \frac{d^k}{dx^k} p(x).$$

This theorem establishes that the eigenvalues of the matrix B (encapsulated in the coefficients b_k) determine the weights of the differentiation in the operator T_q.

3.2   The Differential Operator Representation

Based on the MSS theorem, we define the operator formally.

Definition 3.3 (Convolution Operator T_q). For a monic polynomial $q(x) = \sum_{k=0}^{n} b_k x^{n-k}$, define:

$$T_q := \sum_{k=0}^{n} \frac{(n-k)!}{n!} b_k \partial_x^k.$$

Then the convolution satisfies:

$$(p \boxplus_n q)(x) = T_q p(x).$$

This operator formalism reveals that convolution with q acts as a diffusion process on p, smoothing its features solely based on the root distribution of q.

Theorem 3.2 (Preservation of Real Roots). *If p and q have all real roots, then $p \boxplus_n q$ also has all real roots.*

Lemma 3.3 (Variance Additivity). *The variance is additive under convolution:*

$$\sigma^2(p \boxplus_n q) = \sigma^2(p) + \sigma^2(q).$$

Proof. By the convolution coefficient formula $c_k = \sum_{i+j=k} \frac{(n-i)!\,(n-j)!}{n!\,(n-k)!} a_i b_j$, we compute the first two coefficients of $p \boxplus_n q$.

For $k = 1$: only the terms $(i,j) = (1,0)$ and $(0,1)$ contribute (since $a_0 = b_0 = 1$):

$$c_1 = \frac{(n-1)!\,n!}{n!\,(n-1)!} a_1 + \frac{n!\,(n-1)!}{n!\,(n-1)!} b_1 = a_1 + b_1.$$

For $k = 2$: the contributing terms are $(i, j) \in \{(2, 0), (1, 1), (0, 2)\}$:

$$c_2 = a_2 + \frac{(n-1)!\,(n-1)!}{n!\,(n-2)!}\, a_1 b_1 + b_2 = a_2 + \tfrac{n-1}{n}\, a_1 b_1 + b_2.$$

Substituting into the variance formula (Lemma 2.1):

$$\sigma^2(p \boxplus_n q) = \frac{(n-1)c_1^2}{n^2} - \frac{2c_2}{n} = \frac{(n-1)(a_1 + b_1)^2}{n^2} - \frac{2}{n}\left(a_2 + \tfrac{n-1}{n} a_1 b_1 + b_2\right).$$

Expanding $(a_1 + b_1)^2 = a_1^2 + 2a_1 b_1 + b_1^2$ and collecting terms:

$$= \underbrace{\frac{(n-1)a_1^2}{n^2} - \frac{2a_2}{n}}_{\sigma^2(p)} + \underbrace{\frac{(n-1)b_1^2}{n^2} - \frac{2b_2}{n}}_{\sigma^2(q)} + \frac{2(n-1)a_1 b_1}{n^2} - \frac{2(n-1)a_1 b_1}{n^2}.$$

The cross-terms cancel exactly, yielding $\sigma^2(p \boxplus_n q) = \sigma^2(p) + \sigma^2(q)$. $\qquad\qquad\square$

4   Fisher Information: The Energy of Roots

We define the finite free Fisher information based on the electrostatic interaction of roots.

4.1   Scores and Forces

Roots repel each other with a force proportional to the inverse distance. The total force on a root λ_i is its **score**.

Definition 4.1 (The Score V_i)**.**

$$V_i = \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j}.$$

4.2   Finite Free Fisher Information

The Fisher information is defined as the sum of the squared scores, representing the total stress in the system.

Definition 4.2 (Fisher Information Φ_n)**.**

$$\Phi_n(p) = \sum_{i=1}^{n} V_i^2.$$

- **High Information:** Clustered roots produce large scores V_i, hence large Φ_n.

- **Low Information:** Well-separated roots produce small scores, hence small Φ_n.

Lemma 4.1 (Score via Derivatives)**.** *For any monic polynomial p of degree n with distinct roots $\lambda_1, \ldots, \lambda_n$,*

$$V_i = \frac{p''(\lambda_i)}{2\,p'(\lambda_i)}.$$

Proof. Since $p(x) = \prod_{j=1}^{n}(x - \lambda_j)$, its derivative at a root λ_i is $p'(\lambda_i) = \prod_{j \neq i}(\lambda_i - \lambda_j)$. To find $p''(\lambda_i)$, differentiate $p'(x) = \sum_{i=1}^{n} \prod_{j \neq i}(x - \lambda_j)$ once more. At $x = \lambda_i$, only the terms where x differentiates one factor of $\prod_{j \neq i}(x - \lambda_j)$ survive:

$$p''(\lambda_i) = 2 \sum_{k \neq i} \prod_{\substack{j \neq i \\ j \neq k}} (\lambda_i - \lambda_j) = 2\, p'(\lambda_i) \sum_{k \neq i} \frac{1}{\lambda_i - \lambda_k} = 2\, p'(\lambda_i)\, V_i.$$

Dividing both sides by $2\, p'(\lambda_i)$ gives $V_i = p''(\lambda_i)/(2\, p'(\lambda_i))$. □

Remark 4.1. This lemma says that the score at λ_i is simply the logarithmic derivative of p' evaluated at λ_i, divided by 2. It provides a coordinate-free way to compute scores directly from p and its derivatives.

4.3 Key Identities

Lemma 4.2 (Score-Root Identity).

$$\sum_{i=1}^{n}(\lambda_i - \mu)V_i = \frac{n(n-1)}{2}.$$

Proof. First, $\sum_{i=1}^{n} V_i = \sum_{i \neq j} \frac{1}{\lambda_i - \lambda_j} = 0$, since pairing (i,j) with (j,i) gives cancellation. Next,

$$\sum_{i=1}^{n} \lambda_i V_i = \sum_{i \neq j} \frac{\lambda_i}{\lambda_i - \lambda_j} = \sum_{i<j} \left(\frac{\lambda_i}{\lambda_i - \lambda_j} + \frac{\lambda_j}{\lambda_j - \lambda_i} \right) = \sum_{i<j} \frac{\lambda_i - \lambda_j}{\lambda_i - \lambda_j} = \binom{n}{2}.$$

Since $\sum V_i = 0$, we conclude $\sum_{i=1}^{n}(\lambda_i - \mu)\, V_i = \sum \lambda_i V_i - \mu \sum V_i = \frac{n(n-1)}{2}$. □

Lemma 4.3 (Fisher–Variance Inequality). $\Phi_n(p) \cdot \sigma^2(p) \geq \frac{n(n-1)^2}{4}$, *with equality if and only if there exists a constant c such that $V_i = c(\lambda_i - \mu)$ for all i.*

Proof. By Cauchy–Schwarz applied to the Score-Root Identity:

$$\frac{n^2(n-1)^2}{4} = \left(\sum_{i=1}^{n}(\lambda_i - \mu)\, V_i \right)^2 \leq \left(\sum_{i=1}^{n}(\lambda_i - \mu)^2 \right)\left(\sum_{i=1}^{n} V_i^2 \right) = n\, \sigma^2(p) \cdot \Phi_n(p).$$

Equality in Cauchy–Schwarz requires proportionality $V_i = c(\lambda_i - \mu)$ for a constant c. For $n = 2$ this holds automatically (see Lemma 4.4); for $n \geq 3$ it characterizes the Hermite-root configuration discussed after Theorem 5.4. □

Lemma 4.4 (The $n = 2$ Identity). *For quadratic polynomials ($n = 2$), information is inversely proportional to variance:*

$$\frac{1}{\Phi_2(p)} = 2\sigma^2(p).$$

Proof. Let roots be $-d/2$ and $d/2$. $V_1 = -1/d$, $V_2 = 1/d$. $\Phi_2 = 2/d^2$. $\sigma^2 = d^2/4$. Thus $1/\Phi_2 = d^2/2 = 2\sigma^2$. □

5 Local Analysis: Perturbation Theory

We examine the behavior of roots under convolution with a noise polynomial q_ϵ of small variance ϵ^2. For the perturbation expansion we assume q_ϵ is centered and symmetric, so its odd coefficients vanish and the first nontrivial correction is order ϵ^2.

Remark 5.1 (Why centering and symmetry are harmless)*.* The quantities $\sigma^2(p)$ and $\Phi_n(p)$ depend only on root differences, so they are invariant under translations $p(x) \mapsto p(x - c)$. The convolution satisfies $a_1(p \boxplus_n q) = a_1(p) + a_1(q)$, so shifting q by its mean simply shifts $p \boxplus_n q$ by the same amount without changing any variance or Fisher information. Hence we may assume q_ϵ is centered without loss of generality. Symmetry is a convenience: if the root set of q_ϵ is symmetric, then $b_1 = b_3 = \cdots = 0$, which makes the first correction term order ϵ^2. If q_ϵ is centered but not symmetric, odd coefficients produce only higher-order corrections (order ϵ^3 and above) and do not affect the leading ϵ^2 terms used in the dissipation identity.

5.1 Root Dynamics

The convolution operator induces a shift in the roots of p.

Lemma 5.1 (Shift of Roots)**.** *Let λ_i be the roots of p. The roots μ_i of $p \boxplus_n q_\epsilon$ satisfy:*

$$\mu_i = \lambda_i + \frac{\epsilon^2}{n-1} V_i + O(\epsilon^4).$$

Proof. Consider the operator representation $p \boxplus_n q_\epsilon = T_{q_\epsilon} p$. Since q_ϵ is centered with variance ϵ^2 and close to x^n, its coefficients satisfy $b_0 = 1$, $b_1 = 0$, and $b_2 = -n\epsilon^2/2$ (from the variance formula applied to q_ϵ). By the symmetry assumption, $b_3 = 0$, and higher coefficients are $O(\epsilon^4)$. The operator $T_{q_\epsilon} = \sum_k \frac{(n-k)!}{n!} b_k \partial_x^k$ therefore acts as:

$$T_{q_\epsilon} p(x) = p(x) + \frac{(n-2)!}{n!} b_2 p''(x) + O(\epsilon^4) = p(x) - \frac{\epsilon^2}{2(n-1)} p''(x) + O(\epsilon^4).$$

Here we used $(n-2)!/n! = 1/[n(n-1)]$ and $b_2 = -n\epsilon^2/2$.

Now let $\mu_i = \lambda_i + \delta_i$ with $\delta_i = O(\epsilon^2)$. Expanding $T_{q_\epsilon} p(\mu_i) = 0$:

$$0 = p(\lambda_i) + \delta_i\, p'(\lambda_i) + O(\delta_i^2) - \frac{\epsilon^2}{2(n-1)} p''(\lambda_i) + O(\epsilon^4).$$

Since $p(\lambda_i) = 0$ and $\delta_i^2 = O(\epsilon^4)$, we solve:

$$\delta_i = \frac{\epsilon^2}{2(n-1)} \cdot \frac{p''(\lambda_i)}{p'(\lambda_i)} + O(\epsilon^4).$$

By the Score via Derivatives lemma (Lemma 4.1), $p''(\lambda_i)/(2\, p'(\lambda_i)) = V_i$, so $\delta_i = \frac{\epsilon^2}{n-1} V_i + O(\epsilon^4)$. $\qquad\square$

Interpretation: Roots move in the direction of the repulsive force V_i. Isolated roots (small $|V_i|$) barely move, while roots in a cluster (large $|V_i|$) are pushed apart. The convolution acts like a physical relaxation.

5.2   Monotonicity of Fisher Information

This relaxation leads to a decrease in the total energy.

Lemma 5.2 (Change in Fisher Information)**.**

$$\Phi_n(p \boxplus_n q_\epsilon) = \Phi_n(p) - \frac{2\epsilon^2}{n-1}\mathcal{S}(p) + O(\epsilon^4),$$

*where $\mathcal{S}(p)$ is the **Score-Gradient Energy**:*

$$\mathcal{S}(p) = \sum_{1 \le i < j \le n} \frac{(V_i - V_j)^2}{(\lambda_i - \lambda_j)^2}.$$

Proof. From Lemma 5.1, the perturbed roots are $\mu_i = \lambda_i + \frac{\epsilon^2}{n-1}V_i + O(\epsilon^4)$. The new scores are:

$$V_i^{(\epsilon)} = \sum_{j \ne i} \frac{1}{\mu_i - \mu_j} = \sum_{j \ne i} \frac{1}{(\lambda_i - \lambda_j) + \frac{\epsilon^2}{n-1}(V_i - V_j) + O(\epsilon^4)}$$

$$= \sum_{j \ne i} \frac{1}{\lambda_i - \lambda_j} \cdot \frac{1}{1 + \frac{\epsilon^2}{n-1}\frac{V_i-V_j}{\lambda_i-\lambda_j} + O(\epsilon^4)}.$$

Using $\frac{1}{1+u} = 1 - u + O(u^2)$ for small u:

$$V_i^{(\epsilon)} = V_i - \frac{\epsilon^2}{n-1} \sum_{j \ne i} \frac{V_i - V_j}{(\lambda_i - \lambda_j)^2} + O(\epsilon^4).$$

Squaring and summing:

$$\Phi_n(p_\epsilon) = \sum_i \left(V_i^{(\epsilon)}\right)^2 = \sum_i V_i^2 - \frac{2\epsilon^2}{n-1} \sum_i V_i \sum_{j \ne i} \frac{V_i - V_j}{(\lambda_i - \lambda_j)^2} + O(\epsilon^4).$$

The key observation is that the cross-sum simplifies:

$$\sum_{i \ne j} \frac{V_i(V_i - V_j)}{(\lambda_i - \lambda_j)^2} = \sum_{i < j} \frac{V_i(V_i - V_j) + V_j(V_j - V_i)}{(\lambda_i - \lambda_j)^2} = \sum_{i < j} \frac{(V_i - V_j)^2}{(\lambda_i - \lambda_j)^2} = \mathcal{S}(p),$$

where in the second step we used $V_i(V_i - V_j) + V_j(V_j - V_i) = (V_i - V_j)^2$. This yields the stated formula. $\square$

Since $\mathcal{S}(p) \ge 0$, the Fisher information strictly decreases under convolution (it decreases strictly unless all scores are equal, i.e. $V_i = 0$ for all i, which only happens for $n = 1$).

5.3   The Score-Gradient Inequality

The score-gradient energy $\mathcal{S}(p)$ satisfies a fundamental lower bound relating it to the Fisher information and the variance. This inequality is the key algebraic input for the full Stam inequality.

Lemma 5.3 (Score Decomposition)**.**

$$\Phi_n(p) = \sum_{i=1}^n V_i^2 = \sum_{1 \le i < j \le n} \frac{V_i - V_j}{\lambda_i - \lambda_j}.$$

Proof. Expanding the left-hand side using the definition of V_i:

$$\sum_{i=1}^n V_i^2 = \sum_{i=1}^n V_i \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j} = \sum_{i \neq j} \frac{V_i}{\lambda_i - \lambda_j} = \sum_{i < j} \left(\frac{V_i}{\lambda_i - \lambda_j} + \frac{V_j}{\lambda_j - \lambda_i} \right) = \sum_{i < j} \frac{V_i - V_j}{\lambda_i - \lambda_j}. \quad \square$$

Theorem 5.4 (Score-Gradient Inequality)**.** *For any monic polynomial p of degree $n \geq 2$ with distinct real roots,*

$$\mathcal{S}(p) \cdot \sigma^2(p) \;\geq\; \frac{n-1}{2} \, \Phi_n(p),$$

that is,

$$\left(\sum_{1 \leq i < j \leq n} \frac{(V_i - V_j)^2}{(\lambda_i - \lambda_j)^2} \right) \cdot \left(\frac{1}{n} \sum_{i=1}^n (\lambda_i - \mu)^2 \right) \;\geq\; \frac{n-1}{2} \sum_{i=1}^n V_i^2.$$

Equality holds if and only if there exists a constant c such that $V_i = c(\lambda_i - \mu)$ for all i.

Proof. Write $T = \sum_{i=1}^n (\lambda_i - \mu)^2 = n\,\sigma^2$, $U = \Phi_n(p) = \sum V_i^2$, and $S = \mathcal{S}(p)$. The inequality to establish is

$$S\,T \;\geq\; \tfrac{n(n-1)}{2} \, U. \tag{1}$$

Step 1 (Cauchy–Schwarz on the Score-Root Identity). By Lemma 4.2 and the Cauchy–Schwarz inequality,

$$\frac{n^2(n-1)^2}{4} = \left(\sum_{i=1}^n (\lambda_i - \mu)\, V_i \right)^2 \leq T\,U. \tag{2}$$

Step 2 (Cauchy–Schwarz on the Score Decomposition). By Lemma 5.3 and the Cauchy–Schwarz inequality,

$$U^2 = \left(\sum_{i < j} \frac{V_i - V_j}{\lambda_i - \lambda_j} \right)^2 \leq \left(\sum_{i < j} \frac{(V_i - V_j)^2}{(\lambda_i - \lambda_j)^2} \right) \left(\sum_{i < j} 1 \right) = S \cdot \frac{n(n-1)}{2}. \tag{3}$$

Step 3 (Combination). From (3), $S \geq \frac{2U^2}{n(n-1)}$. Multiplying by T:

$$S\,T \;\geq\; \frac{2\,U^2\,T}{n(n-1)} = \frac{2\,U}{n(n-1)} \cdot T\,U \;\geq\; \frac{2\,U}{n(n-1)} \cdot \frac{n^2(n-1)^2}{4} = \frac{n(n-1)}{2}\,U,$$

where the second inequality uses (2). This establishes (1).

Equality analysis. Equality requires both Cauchy–Schwarz applications to be tight.

- Inequality (2) is an equality if and only if $V_i = c(\lambda_i - \mu)$ for some constant c.
- Inequality (3) is an equality if and only if $\frac{V_i - V_j}{\lambda_i - \lambda_j}$ is constant over all pairs $i < j$.

If $V_i = c(\lambda_i - \mu)$, then $\frac{V_i - V_j}{\lambda_i - \lambda_j} = c$, so both conditions hold simultaneously. Conversely, if $\frac{V_i - V_j}{\lambda_i - \lambda_j} = k$ for all $i < j$, then $V_i - k\lambda_i$ is constant across i; since $\sum V_i = 0$, this forces $V_i = k(\lambda_i - \mu)$. Therefore, equality holds if and only if $V_i = c(\lambda_i - \mu)$ for all i.

This condition characterizes (up to affine transformation) the zeros of the classical Hermite polynomials: if $x_1, \dots, x_n$ are the zeros of the physicist's Hermite polynomial H_n, then the ODE $H_n''(x) - 2x\,H_n'(x) + 2n\,H_n(x) = 0$ evaluated at a zero x_k yields $V_k = x_k$, confirming the proportionality $V_k = 1 \cdot (x_k - 0)$ since the zeros are symmetric about the origin. $\quad \square$

6   Global Analysis: The Flow Approach

We construct a continuous flow interpolating between p and $p \boxplus_n q$.

Definition 6.1 (Fractional Convolution Flow)**.** Write $b = \sigma^2(q) > 0$. We construct a one-parameter family $\{q_t\}_{t \in [0,1]}$ of polynomials such that $q_0 = x^n$ (the identity for $\boxplus_n$), $q_1 = q$, and the variance scales linearly: $\sigma^2(q_t) = t\, b$.

To see the semigroup structure, introduce the *normalized coefficients*

$$\kappa_k(p) := \frac{(n-k)!}{n!}\, a_k, \qquad \kappa_k(q) := \frac{(n-k)!}{n!}\, b_k.$$

Multiplying the convolution formula by $(n-k)!/n!$ gives the Cauchy product:

$$\kappa_k(p \boxplus_n q) = \sum_{i+j=k} \kappa_i(p)\, \kappa_j(q).$$

Equivalently, for the generating polynomial $K_p(z) = \sum_{k=0}^{n} \kappa_k(p) z^k$, we have

$$K_{p \boxplus_n q}(z) = K_p(z)\, K_q(z). \tag{4}$$

This identity is the precise algebraic content of the convolution.

We therefore define the fractional family by the power identity

$$K_{q_t}(z) := K_q(z)^{\,t}, \qquad t \in [0,1], \tag{5}$$

where $K_q(z)^{\,t}$ is expanded as a power series and truncated at degree n. Then $K_{q_0}(z) = 1$ (so $q_0 = x^n$), $K_{q_1}(z) = K_q(z)$, and by construction $K_{q_s}(z) K_{q_t}(z) = K_{q_{s+t}}(z)$, which yields the semigroup property $q_s \boxplus_n q_t = q_{s+t}$.

Define the *flow polynomial*:

$$p_t := p \boxplus_n q_t.$$

By variance additivity (Lemma 3.3),

$$\sigma^2(p_t) = \sigma^2(p) + t\, b. \tag{6}$$

Remark 6.1. Although $K_q(z)^{\,t}$ is a power series when t is non-integer, only the first n coefficients matter when acting on degree-n polynomials, because ∂_x^{n+1} annihilates them. Thus the definition of q_t via (5) is unambiguous at the operator level, and the semigroup identity follows from (4). In what follows we assume p_t has distinct real roots for $t \in [0,1]$ so that $\mathcal{S}(p_t)$ is well-defined.

Remark 6.2 (Why p_t has distinct real roots (proof sketch))**.** Real-rootedness is preserved by $\boxplus_n$, so each p_t is real-rooted. Distinctness follows by continuity and a discriminant argument. First, since p has simple roots, there is a $\delta > 0$ such that all root gaps at $t = 0$ are at least δ. The coefficients of p_t vary continuously in t, so the roots vary continuously as well; for sufficiently small t the gaps stay $> \delta/2$, so p_t has simple roots. If a collision occurred at some t_*, then the discriminant $\Delta(p_t)$ would vanish. But $t \mapsto \Delta(p_t)$ is a real-analytic function (it is a polynomial in the coefficients, which are analytic in t through K_{q_t}). A nonzero analytic function can only vanish at isolated points, so multiple roots can occur only for isolated t. This is enough for the flow argument: $\mathcal{S}(p_t)$ is defined for all t except possibly a discrete set, and the integral identity still holds by continuity (or by approximating q with a nearby polynomial that avoids collisions). In particular, for generic p, q the roots are simple for all $t \in [0,1]$.

6.1 The Dissipation Identity

Lemma 6.1 (Dissipation)**.**

$$\frac{d}{dt}\Phi_n(p_t) = -\frac{2b}{n-1}\,\mathcal{S}(p_t).$$

Proof. By the semigroup property, $p_{t+h} = p_t \boxplus_n q_h$ where $\sigma^2(q_h) = h\,b$. Applying Lemma 5.2 to the pair (p_t, q_h) gives the second-order expansion

$$\Phi_n(p_{t+h}) = \Phi_n(p_t) - \frac{2\,h\,b}{n-1}\,\mathcal{S}(p_t) + O(h^2).$$

Subtract $\Phi_n(p_t)$ and divide by h:

$$\frac{\Phi_n(p_{t+h}) - \Phi_n(p_t)}{h} = -\frac{2b}{n-1}\,\mathcal{S}(p_t) + O(h).$$

Taking the limit $h \to 0$ yields $\dot{\Phi}_n(p_t) = -(2b/(n-1))\,\mathcal{S}(p_t)$. □

6.2 The Integral Formula

We now derive a formula for the change in the *reciprocal* Fisher information.

Theorem 6.2 (Integral Identity)**.**

$$\frac{1}{\Phi_n(p \boxplus_n q)} - \frac{1}{\Phi_n(p)} = \frac{2b}{n-1}\int_0^1 \frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2}\,dt.$$

Proof. We work with the function

$$f(t) := \frac{1}{\Phi_n(p_t)},$$

and our goal is to compute $f(1) - f(0)$ by integrating $f'(t)$.

Step 1: Differentiate $f(t)$ via the chain rule. Since $f(t) = [\Phi_n(p_t)]^{-1}$, the standard single-variable chain rule gives

$$f'(t) = -\frac{1}{[\Phi_n(p_t)]^2} \cdot \frac{d}{dt}\Phi_n(p_t) = -\frac{\dot{\Phi}_n(p_t)}{\Phi_n(p_t)^2}.$$

(Here $\dot{\Phi}_n(p_t)$ is shorthand for $\frac{d}{dt}\Phi_n(p_t)$.)

Step 2: Substitute the dissipation identity. By Lemma 6.1, we have $\dot{\Phi}_n(p_t) = -\frac{2b}{n-1}\,\mathcal{S}(p_t)$. Plugging this in:

$$f'(t) = -\frac{-\frac{2b}{n-1}\,\mathcal{S}(p_t)}{\Phi_n(p_t)^2} = \frac{2b}{n-1}\cdot\frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2}.$$

Since $b > 0$ and $\mathcal{S}(p_t) \geq 0$ (it is a sum of squares), we conclude $f'(t) \geq 0$: the reciprocal Fisher information is non-decreasing along the flow.

Step 3: Integrate from 0 to 1. By the Fundamental Theorem of Calculus,

$$f(1) - f(0) = \int_0^1 f'(t)\,dt = \frac{2b}{n-1}\int_0^1 \frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2}\,dt.$$

Step 4: Identify the endpoints. At $t = 0$: $p_0 = p \boxplus_n q_0 = p \boxplus_n x^n = p$, so $f(0) = 1/\Phi_n(p)$.

At $t = 1$: $p_1 = p \boxplus_n q_1 = p \boxplus_n q$, so $f(1) = 1/\Phi_n(p \boxplus_n q)$.

Therefore:
$$\frac{1}{\Phi_n(p \boxplus_n q)} - \frac{1}{\Phi_n(p)} = \frac{2b}{n-1} \int_0^1 \frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2} \, dt. \qquad \square$$

Remark 6.3 (Why this identity matters). This integral identity is the engine of every result that follows. Notice three things:

1. The right-hand side is manifestly ≥ 0 (each factor in the integrand is non-negative). This immediately gives the Weak Stam inequality.
2. The right-hand side depends on q only through $b = \sigma^2(q)$ and the particular path p_t. Symmetry ($p \boxplus_n q = q \boxplus_n p$) lets us write an analogous identity starting from q.
3. Inserting a *lower bound* on $\mathcal{S}(p_t)$ (the Score-Gradient Inequality) turns this exact identity into a differential inequality that can be solved explicitly.

7   Main Results

We now use the Integral Identity (Theorem 6.2) together with the Score-Gradient Inequality (Theorem 5.4) to prove the Stam inequality in three stages of increasing strength. Throughout, we write
$$a := \sigma^2(p) > 0, \qquad b := \sigma^2(q) > 0.$$

7.1   The Weak Stam Inequality

Proposition 7.1 (Weak Stam). $\dfrac{1}{\Phi_n(p \boxplus_n q)} \geq \dfrac{1}{\Phi_n(p)}.$

Proof. By Theorem 6.2,
$$\frac{1}{\Phi_n(p \boxplus_n q)} - \frac{1}{\Phi_n(p)} = \underbrace{\frac{2b}{n-1}}_{> 0} \int_0^1 \underbrace{\frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2}}_{\geq 0} \, dt \; \geq \; 0.$$

The integrand is non-negative because $\mathcal{S}(p_t)$ is a sum of squares (Definition in Section 5.2) and $\Phi_n(p_t)^2 > 0$. $\qquad \square$

Remark 7.1. In words: convolving with *any* real-rooted polynomial can only decrease the Fisher information (equivalently, increase its reciprocal). The inequality is strict whenever $\mathcal{S}(p_t) > 0$ at some $t \in (0,1)$, which happens for every $n \geq 2$.

7.2   The Half-Stam Inequality

Theorem 7.2 (Half-Stam Inequality). *For any real-rooted polynomials p, q with distinct roots:*
$$\frac{2}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)}.$$

Proof. The Weak Stam (Proposition 7.1) applied to the pair (p, q) says
$$\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)}. \tag{7}$$

The convolution is commutative: $p \boxplus_n q = q \boxplus_n p$ (both equal the same Haar-averaged characteristic polynomial). So we may also apply the Weak Stam to the pair (q, p), giving

$$\frac{1}{\Phi_n(\underbrace{q \boxplus_n p}_{= p \boxplus_n q})} \geq \frac{1}{\Phi_n(q)}. \tag{8}$$

Add (7) and (8):

$$\frac{1}{\Phi_n(p \boxplus_n q)} + \frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)},$$

which is exactly $\frac{2}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)}$. □

Remark 7.2. The Half-Stam is already a non-trivial result. It says the reciprocal Fisher information of the convolution is at least the *average* of the reciprocal Fisher informations of p and q. The proof uses nothing about $\mathcal{S}$ beyond non-negativity. To upgrade from "average" to "sum", we need the quantitative lower bound on $\mathcal{S}$ provided by the Score-Gradient Inequality.

7.3 The Full Stam Inequality

The idea is to use the Score-Gradient Inequality to convert the integral identity into a *differential inequality* of the form $f'(t) \geq g(t) f(t)$, which can be solved by the integrating-factor method (Grönwall's lemma).

Theorem 7.3 (Full Stam Inequality). *For any real-rooted polynomials $p, q \in \mathcal{P}_n^{\mathbb{R}}$ with positive variances,*

$$\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)}.$$

Proof. Write $f(t) = 1/\Phi_n(p_t)$ as before.

Step 1: From the integral identity to a differential inequality.

From the proof of Theorem 6.2 we know

$$f'(t) = \frac{2b}{n-1} \cdot \frac{\mathcal{S}(p_t)}{\Phi_n(p_t)^2}. \tag{9}$$

We now use the Score-Gradient Inequality (Theorem 5.4) applied to the polynomial p_t (which has distinct real roots):

$$\mathcal{S}(p_t) \cdot \sigma^2(p_t) \geq \frac{n-1}{2} \Phi_n(p_t),$$

i.e.,

$$\mathcal{S}(p_t) \geq \frac{(n-1)\Phi_n(p_t)}{2\sigma^2(p_t)}. \tag{10}$$

Substituting (10) into (9):

$$\begin{aligned}
f'(t) &\geq \frac{2b}{n-1} \cdot \frac{1}{\Phi_n(p_t)^2} \cdot \frac{(n-1)\Phi_n(p_t)}{2\sigma^2(p_t)} \\
&= \frac{2b}{n-1} \cdot \frac{n-1}{2\sigma^2(p_t)} \cdot \frac{1}{\Phi_n(p_t)} \\
&= \frac{b}{\sigma^2(p_t)} \cdot f(t).
\end{aligned}$$

Since $\sigma^2(p_t) = a + tb$ (equation (6)), this becomes

$$f'(t) \geq \frac{b}{a + tb} f(t). \tag{11}$$

This is a first-order linear differential inequality of the form $f' \geq g(t)\, f$ with $g(t) = b/(a + tb)$.

Step 2: Solve the differential inequality (Grönwall).

Divide both sides of (11) by $f(t) > 0$:

$$\frac{f'(t)}{f(t)} \geq \frac{b}{a + tb}.$$

The left-hand side is $\frac{d}{dt} \ln f(t)$. The right-hand side is $\frac{d}{dt} \ln(a + tb)$ (check: $\frac{d}{dt} \ln(a + tb) = \frac{b}{a+tb}$). So:

$$\frac{d}{dt} \ln f(t) \geq \frac{d}{dt} \ln(a + tb).$$

Integrate both sides from $t = 0$ to $t = 1$:

$$\ln f(1) - \ln f(0) \geq \ln(a + b) - \ln a,$$

i.e.,

$$\ln \frac{f(1)}{f(0)} \geq \ln \frac{a + b}{a}.$$

Exponentiate (the exponential is increasing):

$$\frac{f(1)}{f(0)} \geq \frac{a + b}{a}.$$

Recalling $f(0) = 1/\Phi_n(p)$ and $f(1) = 1/\Phi_n(p \boxplus_n q)$:

$$\boxed{\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{a + b}{a} \cdot \frac{1}{\Phi_n(p)}.} \qquad \text{(Forward bound)} \tag{12}$$

Step 3: The reverse bound by symmetry.

Since $p \boxplus_n q = q \boxplus_n p$, we can run the identical argument starting from q and flowing toward $p \boxplus_n q$. Concretely, define $\hat{p}_s = q \boxplus_n r_s$ where r_s is the fractional semigroup of p with $\sigma^2(r_s) = s\, a$, so that $\hat{p}_0 = q$ and $\hat{p}_1 = q \boxplus_n p = p \boxplus_n q$.

The same calculation as Steps 1–2, but with the roles of a and b swapped, yields

$$\boxed{\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{a + b}{b} \cdot \frac{1}{\Phi_n(q)}.} \qquad \text{(Reverse bound)} \tag{13}$$

Step 4: Combine the two bounds.

We have two inequalities:

$$\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{a + b}{a} \cdot \frac{1}{\Phi_n(p)}, \qquad \frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{a + b}{b} \cdot \frac{1}{\Phi_n(q)}.$$

Define weights that sum to 1:

$$w_1 = \frac{a}{a+b}, \qquad w_2 = \frac{b}{a+b}, \qquad w_1 + w_2 = 1.$$

Multiply the forward bound (12) by w_1 and the reverse bound (13) by w_2:

$$w_1 \cdot \frac{1}{\Phi_n(p \boxplus_n q)} \geq w_1 \cdot \frac{a+b}{a} \cdot \frac{1}{\Phi_n(p)} = \frac{a}{a+b} \cdot \frac{a+b}{a} \cdot \frac{1}{\Phi_n(p)} = \frac{1}{\Phi_n(p)}, \qquad (14)$$

$$w_2 \cdot \frac{1}{\Phi_n(p \boxplus_n q)} \geq w_2 \cdot \frac{a+b}{b} \cdot \frac{1}{\Phi_n(q)} = \frac{b}{a+b} \cdot \frac{a+b}{b} \cdot \frac{1}{\Phi_n(q)} = \frac{1}{\Phi_n(q)}. \qquad (15)$$

Observe how the factors cancel perfectly in each line. Now add (14) and (15):

$$\underbrace{(w_1 + w_2)}_{=1} \cdot \frac{1}{\Phi_n(p \boxplus_n q)} \;\geq\; \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)},$$

which is exactly the Stam inequality. $\qquad\square$

Remark 7.3 (Why the weights work). The choice of weights $w_1 = a/(a+b)$, $w_2 = b/(a+b)$ is not arbitrary—it is the unique choice that makes both factors $w_i \cdot (a+b)/\sigma^2$ simplify to 1. Any other weighting would leave residual factors that prevent a clean cancellation. This is the step where the variance additivity (the fact that $\sigma^2(p \boxplus_n q) = a+b$) is essential.

Remark 7.4 (Summary of what each ingredient contributes). Let us recap which tool is responsible for each part of the proof:

Ingredient	Where used	What it gives
$\mathcal{S} \geq 0$ (sum of squares)	Weak & Half Stam	Monotonicity: $1/\Phi_n$ increases
Score-Gradient Ineq.	Full Stam, Step 1	Quantitative ODE: $f' \geq g\,f$
Grönwall integration	Full Stam, Step 2	Forward/reverse bounds
Commutativity of $\boxplus_n$	Full Stam, Step 3	Reverse bound by symmetry
Variance additivity	Full Stam, Step 4	Weight cancellation

Remark 7.5 (Strength of the forward and reverse bounds). The forward bound (12) by itself says $1/\Phi_n(p \boxplus_n q) \geq (1 + b/a)/\Phi_n(p)$, which is *stronger* than $1/\Phi_n(p)$ alone by the multiplicative factor $1 + b/a > 1$. In terms of the efficiency ratio $\eta(p) = 4\Phi_n(p)\sigma^2(p)/[n(n-1)^2]$, the forward bound is equivalent to $\eta(p \boxplus_n q) \leq \eta(p)$: convolution always regularizes (pushes η toward its minimum value 1).

8 Alternative Characterizations

The Stam inequality can be reformulated in several instructive ways that highlight different aspects of the convolution process.

8.1 The Efficiency Ratio and Regularization

A natural way to measure the "non-Gaussianity" of a polynomial's root distribution is by comparing its Fisher information to the lower bound established in Lemma 4.3.

Definition 8.1 (Efficiency Ratio)**.** For $p \in \mathcal{P}_n^{\mathbb{R}}$ with positive variance, the *efficiency ratio* $\eta(p)$ is defined as:

$$\eta(p) := \frac{4\Phi_n(p)\sigma^2(p)}{n(n-1)^2}.$$

By the Fisher–Variance inequality, $\eta(p) \geq 1$, with equality if and only if $V_i = c(\lambda_i - \mu)$ for all i (the Hermite-root configuration; $n = 2$ is the simplest case).

In this language, the Stam inequality is related to the following regularization property.

Theorem 8.1 (Regularization Property)**.** *For $p, q \in \mathcal{P}_n^{\mathbb{R}}$ with positive variances, the efficiency ratio of the convolution satisfies:*

$$\eta(p \boxplus_n q) \leq \frac{\sigma^2(p)\,\eta(p) + \sigma^2(q)\,\eta(q)}{\sigma^2(p) + \sigma^2(q)}.$$

Proof. Write $a = \sigma^2(p)$ and $b = \sigma^2(q)$. Recall the forward and reverse bounds from the proof of the Full Stam Inequality (Theorem 7.3):

$$\Phi_n(p \boxplus_n q) \leq \frac{a\,\Phi_n(p)}{a+b} \quad \text{and} \quad \Phi_n(p \boxplus_n q) \leq \frac{b\,\Phi_n(q)}{a+b}.$$

Multiply the first by $\frac{a}{a+b}$ and the second by $\frac{b}{a+b}$, then add:

$$\Phi_n(p \boxplus_n q) \leq \frac{a^2\,\Phi_n(p) + b^2\,\Phi_n(q)}{(a+b)^2}.$$

Multiplying both sides by $\frac{4(a+b)}{n(n-1)^2}$:

$$\underbrace{\frac{4\,\Phi_n(p \boxplus_n q)(a+b)}{n(n-1)^2}}_{=\,\eta(p\boxplus_n q)} \leq \frac{a \cdot \frac{4a\,\Phi_n(p)}{n(n-1)^2} + b \cdot \frac{4b\,\Phi_n(q)}{n(n-1)^2}}{a+b} = \frac{a\,\eta(p) + b\,\eta(q)}{a+b}.$$

Note that $\sigma^2(p \boxplus_n q) = a + b$ was used in identifying $\eta(p \boxplus_n q)$. $\qquad\square$

This theorem asserts that the efficiency ratio of the convolution is bounded by the variance-weighted average of the efficiency ratios of the components: convolution always pushes the root configuration toward a more regular (lower η) state.

8.2   Convex Mixing Interpretation

The symmetric additive convolution can be viewed as a form of convex mixing at the level of the matrix model. Let A and B be centered symmetric matrices with characteristic polynomials p and q. The functional $\Psi_n(M) = \sigma^2(M)\,\Phi_n(\chi_M)$, where χ_M denotes the characteristic polynomial of M, is scale-invariant.

Remark 8.1 (Convexity under Haar Averaging). The Regularization Property can be interpreted as a matrix-level Jensen's inequality: the Haar integration over $O(n)$ in the definition of $\boxplus_n$ acts as an averaging operation that always reduces the efficiency ratio η. Heuristically, rotating B by a random orthogonal matrix Q and adding it to A produces a sum whose eigenvalues are more "spread out" than either A or B alone. Since η measures deviation from the most spread-out configuration (the Hermite roots), this averaging cannot increase η.

This interpretation complements the flow-based proof: both perspectives confirm that convolution acts as a regularization process, driving root configurations toward the universally optimal Hermite arrangement.

9   Conclusion

We have established the full Finite Free Stam Inequality, linking the geometry of polynomial roots to information-theoretic inequalities.

1. We defined root interactions via electrostatic forces and their associated Fisher information $\Phi_n(p)$.

2. We proved the Score-Gradient Inequality (Theorem 5.4): $\mathcal{S}(p) \cdot \sigma^2(p) \geq \frac{n-1}{2}\Phi_n(p)$, via two applications of Cauchy–Schwarz to the Score-Root Identity and the Score Decomposition.

3. We derived a precise integral formula for the increase in the reciprocal Fisher information along the convolution flow.

4. We combined the Score-Gradient Inequality with the flow to obtain the full Stam inequality (Theorem 7.3):

$$\frac{1}{\Phi_n(p \boxplus_n q)} \geq \frac{1}{\Phi_n(p)} + \frac{1}{\Phi_n(q)},$$

with equality characterized by the Hermite polynomial condition $V_i = c(\lambda_i - \mu)$.

References

[1] A. Marcus, D. Spielman, N. Srivastava, *Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem*, Ann. Math. 182 (2015).

[2] A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control (1959).