Map of *How Transformers Attend — Part I*

Aprendizaje automático · análisis matemático de medidas de atención

22 results · **0** entirely in Lean · **9** partly formal · **3** in Isabelle · **0** checked by a program · written proof only · (a result can carry several tags)

Dependencies in the written proof: an arrow A → B means that this proof of B uses A. Every result rests on its written proof; the tags record what was checked besides, and each statement's "Evidence & scope" block says how far. The colour of a box is its strongest tag, in the order above; the tags in the box give the rest. Boxes with a thick border are the main results; click a box to read the statement.

Background and Notation

- Definition 4.1 — Planos rotatorios con y sin ciclo completo
- Definition 4.2 — Distancias de referencia de la geometría RoPE

Modelling γ: Geometry, Approximations and Empirical Tests

- Proposition 5.1 — La suma de cosenos no implica el candidato Padé

Why Power Law: The Maximum Entropy Foundation

- Theorem 6.1 — MaxEnt selecciona una familia de potencias bajo un momento fijado — ◐ part in Lean
- Theorem 6.2 — Existencia y unicidad del exponente MaxEnt en soporte finito — ◐ part in Lean

Finite Power-Law Families and Infinite-Support Limits

- Theorem 7.1 — Varianza continua de la logdistancia en uno — ◐ part in Lean
- Theorem 7.2 — Entropía de la familia finita y sus regímenes asintóticos — ◐ part in Lean
- Definition 7.3 — Regímenes de sumabilidad del soporte infinito

Layer Heterogeneity and the Proposed Critical Layer

- Definition 10.1 — Coordenada histórica de capa crítica

KV-Cache Compression: Mass Budgets and Recorded Tests

- Theorem 11.1 — Cuantil continuo de masa y cuantil discreto de tokens

Coordinate Ablations and the Limits of Band Interpretation

- Definition 13.1 — Agrupación de planos rotatorios

Corpus Contrasts and the Unvalidated Architecture Ratio

- Definition 16.1 — Contraste entre corpus con protocolo emparejado

Effective Coordinates and Dynamic Control Proposals

- Proposition 19.1 — Causalidad por prefijos del controlador corregido

Rational Coordinates and the Limits of the RG Analogy

- Conjecture 20.1 — Hipótesis empírica sobre alucinación, aún no validada

Eight Formally Verified Result Groups

- Theorem K.1 — Nulo uniforme y calibración de la pendiente — ◐ part in Isabelle
- Theorem K.2 — Sesgo de exposición afín — ◐ part in Isabelle
- Theorem K.3 — Curvatura de mezclas de potencias — ✅ Isabelle
- Theorem K.4 — Orden de cuantiles en soporte finito — ◐ part in Isabelle
- Theorem K.5 — Identidades de la familia exponencial finita — ✅ Isabelle
- Theorem K.6 — Cotas de truncación para una fila congelada — ◐ part in Isabelle
- Theorem K.7 — Estabilidad del softmax ante cambios de puntaje — ✅ Isabelle
- Theorem K.8 — Integral coseno y residuo RoPE explícito — ◐ part in Isabelle

> La Parte I reúne cinco teoremas y dos proposiciones del cuerpo, ocho teoremas de calibración y cotas en el apéndice K, seis definiciones y una conjetura empírica. Las pruebas Isabelle cubren ocho grupos del apéndice; Lean verifica partes de cuatro resultados del cuerpo. Ninguna de esas pruebas establece cómo se entrena un modelo.

🚀 **Start here:** **Theorem K.1**: Nulo uniforme y calibración de la pendiente.

Main results:

- **Proposition 5.1**: La suma de cosenos no implica el candidato Padé
- **Theorem 6.1**: MaxEnt selecciona una familia de potencias bajo un momento fijado
- **Theorem 6.2**: Existencia y unicidad del exponente MaxEnt en soporte finito
- **Theorem 7.2**: Entropía de la familia finita y sus regímenes asintóticos
- **Theorem 11.1**: Cuantil continuo de masa y cuantil discreto de tokens
- **Theorem K.1**: Nulo uniforme y calibración de la pendiente
- **Theorem K.2**: Sesgo de exposición afín
- **Theorem K.3**: Curvatura de mezclas de potencias
- **Theorem K.4**: Orden de cuantiles en soporte finito
- **Theorem K.5**: Identidades de la familia exponencial finita
- **Theorem K.6**: Cotas de truncación para una fila congelada
- **Theorem K.7**: Estabilidad del softmax ante cambios de puntaje
- **Theorem K.8**: Integral coseno y residuo RoPE explícito

🧭 Fields: where the paper lives and what it connects

Tools taken from other branches

- **Familias exponenciales y teoría de la información** → Desigualdad de Gibbs, unicidad MaxEnt, entropía finita y cuantiles en soportes declarados. — Theorem 6.1, Theorem 6.2, Theorem 7.2, Theorem 11.1, Theorem K.5, Theorem K.4
- **Análisis de regresión y de softmax** → Covarianza en una malla ponderada, sesgo del nulo, mezclas, truncación y cotas de estabilidad. — Theorem K.1, Theorem K.2, Theorem K.3, Theorem K.6, Theorem K.7
- **Geometría RoPE y sumas oscilatorias** → Frecuencias geométricas, ciclos completos e integral coseno con residuo; la hipótesis sobre scores es adicional. — 4.1, 4.2, Proposition 5.1, Theorem K.8

What the results serve

- **Medición de atención y evaluación de intervenciones** ← El nulo y las cotas ayudan a diseñar sondas y controles; los resultados de modelos entrenados se

> evalúan con datos separados. — Theorem K.1, Theorem K.6, Theorem 11.1, Proposition 19.1