Finite Propagation in Networks & IS-Loss-Trained Models: A Unified Bound and a Worked Example

August 20, 2025

1 Introduction

We present a unified, domain-agnostic bound on the finite speed at which influence, information, or correlations can propagate in networks with local interactions (graphs, lattices, circuits). We then provide a worked example showing how neural prediction dynamics trained with the Itakura–Saito (IS) loss obey an exact discrete "light-cone" under NTK linearization. The IS loss improves spectral conditioning (and thus robustness/convergence *within* the cone) by normalizing relative errors, yet it does not change the fundamental propagation speed set by architectural locality.

2 Generalized Speed Limit in Networks

Consider a network (graph, lattice, or computational circuit) as nodes V and edges E, with a metric distance $d(x, y)$ between nodes and *local* update rules. Each node $x \in V$ has a state $\varphi_x(t)$ evolving in time t according to local dynamics. Let $\mathcal{O}_x$ denote an observable localized at x.

Theorem 1 (Unified Network Propagation Bound)**.** *Assume the network satisfies:*

(i) ***Locality:*** *updates of φ_x depend only on nodes within a bounded neighborhood of x;*

(ii) ***Finite interaction strength:*** *each local update is Lipschitz-bounded with constant g (e.g. energy scale, bandwidth, Lipschitz factor);*

(iii) ***Well-defined metric:*** *the graph admits a distance $d(\cdot, \cdot)$.*

Then there exists a finite velocity $v > 0$ (the network propagation speed*) and constants $C, \xi > 0$ such that for any two localized observables $\mathcal{O}_x(t)$ and $\mathcal{O}_y(0)$,*

$$\Big|\langle \mathcal{O}_x(t), \mathcal{O}_y(0)\rangle\Big| \;\leq\; C \exp\!\left(-\frac{d(x, y) - vt}{\xi}\right). \tag{1}$$

Interpretations.

- In **quantum spin systems**, this recovers the Lieb–Robinson bound with v the LR velocity.
- In **communication networks**, v reflects edge capacities and d the path length (diameter).
- In **distributed consensus**, rates are constrained by the Laplacian spectral gap: larger gaps $\Rightarrow$ effectively larger v.
- In **circuits / deep nets**, v corresponds to layers-per-unit-time (receptive-field growth), so $d(x, y)/v$ lower-bounds minimal depth to propagate influence.

Consequences. (a) No instantaneous propagation: there is always a finite "light-cone". (b) Scale/conformal invariance may improve robustness and conditioning, but does not remove the speed limit v.

3 Worked Example: Finite Propagation Speed in Neural Prediction Dynamics under IS Loss

We derive an explicit light-cone bound for neural prediction dynamics trained with IS loss via NTK linearization, and show how IS improves conditioning without changing the speed limit.

3.1 Setup: NTK Dynamics with IS Loss

Let $\{(x_i, y_i)\}_{i=1}^n$ be training samples and $f_\theta(x) \in \mathbb{R}$ a scalar model. Denote $f_\theta := (f_\theta(x_1), \ldots, f_\theta(x_n))^\top$ and $y \in \mathbb{R}^n$. For one datum (y_i, f_i) the Itakura–Saito loss is

$$\ell_{\mathrm{IS}}(y_i, f_i) \;=\; \frac{y_i}{f_i} - \log\left(\frac{y_i}{f_i}\right) - 1, \tag{2}$$

so that

$$\frac{\partial \ell_{\mathrm{IS}}}{\partial f_i} \;=\; \frac{f_i - y_i}{f_i^2}, \qquad \frac{\partial^2 \ell_{\mathrm{IS}}}{\partial f_i^2}\bigg|_{f_i = y_i} \;=\; \frac{1}{y_i^2}. \tag{3}$$

Near $f \approx y$, the loss is locally quadratic with curvature $W := \mathrm{diag}(1/y_1^2, \ldots, 1/y_n^2)$ and gradient $\nabla_f L \approx W(f - y)$.

Assume full-batch gradient descent and NTK linearization with fixed empirical NTK $K \in \mathbb{R}^{n \times n}$:

$$f_{t+1} \;=\; f_t \;-\; \eta\, K\, \nabla_f L(f_t) \;\approx\; f_t \;-\; \eta\, K\, W\, (f_t - y). \tag{4}$$

With error $e_t := f_t - y$ we get linear dynamics

$$e_{t+1} \;=\; \big(I - \eta A\big)\, e_t, \qquad A \;:=\; KW. \tag{5}$$

Locality assumption. Let samples be nodes of a graph (V, E) with distance $d(i, j)$. Suppose K is *range-R local*:

$$K_{ij} = 0 \quad \text{whenever } d(i, j) > R. \tag{6}$$

This holds for finite-receptive-field CNNs/grids, R-hop GNNs, and localized kernels. Since W is diagonal, $A{=}KW$ shares K's sparsity.

3.2 A Discrete Light-Cone for Prediction Influence

Let $J_t := \partial f_t / \partial f_0 = (I - \eta A)^t$.

Lemma 1 (Bandedness under locality)**.** *If* (6) *holds and W is diagonal, then $A^t = (KW)^t$ is range-tR local: $(A^t)_{ij} = 0$ whenever $d(i, j) > tR$.*

Proof. A has the same sparsity as K. The product of range-R_B and range-R_C local matrices is range-$(R_B{+}R_C)$ local by triangle inequality on paths. Iterate to get range-tR locality for A^t. □

Theorem 2 (Discrete light-cone under IS training)**.** *Under* (6)*, for any $t \geq 0$,*

$$\big(J_t\big)_{ij} \;=\; \big((I - \eta A)^t\big)_{ij} \;=\; 0 \qquad \textit{whenever } d(i, j) > tR. \tag{7}$$

Equivalently, a unit perturbation at node j at step 0 cannot affect $f_t(i)$ if $d(i, j) > tR$.

Proof. Expand $(I - \eta A)^t = \sum_{k=0}^{t} \binom{t}{k} (-\eta)^k A^k$. By Lemma 1, A^k is range-kR local. If $d(i,j) > tR$, then $d(i,j) > kR$ for all $k \le t$, so $(A^k)_{ij} = 0$ and the sum's (i,j) entry vanishes. $\square$

Thus the maximum propagation speed is $v = R$ nodes/step—an architectural property independent of curvature or step size (subject to spectral stability).

3.3 Spectral Stability, Convergence Rate, and Hessian Conditioning

Let $\lambda_{\max}(A)$ be the spectral radius. Gradient descent (5) is linearly stable if

$$0 \; < \; \eta \; < \; \frac{2}{\lambda_{\max}(A)}. \tag{8}$$

Under stability, errors contract at a rate governed by the spectrum of A (and eigenbasis conditioning). Near $f \approx y$, the prediction-space Hessian is $H_f \approx W$, so $A{=}KW$ acts as an *implicit preconditioner*: it down-weights large targets and up-weights small ones via $W_{ii}{=}1/y_i^2$. This flattens the spectrum relative to MSE ($W{=}I$), improving conditioning and accelerating convergence *inside* the light-cone, without altering R (hence not increasing v).

3.4 Exponential Tails beyond Strict Locality

If K is not strictly banded but decays rapidly off-diagonal,

$$|K_{ij}| \; \le \; C_0 \, e^{-d(i,j)/\xi_0}, \tag{9}$$

and W is bounded ($0 < w_{\min} \le W_{ii} \le w_{\max} < \infty$), submultiplicative estimates yield constants $C, \xi > 0$ such that

$$\left| \left(A^t\right)_{ij} \right| \; \le \; C \, e^{-\frac{d(i,j) - vt}{\xi}}, \qquad v := R, \tag{10}$$

i.e. a Lieb–Robinson type bound: exponential suppression outside a linear light-cone.

3.5 Concrete 1D Example (Tri-diagonal Kernel)

On a 1D chain with $d(i,j) = |i - j|$ and tri-diagonal K (nearest-neighbor; $R{=}1$), Lemma 1 implies $(KW)^t$ is $(2t{+}1)$-banded, hence

$$\left(J_t\right)_{ij} = 0 \quad \text{if } |i - j| > t. \tag{11}$$

A perturbation at j influences only indices with $|i - j| \le t$ after t steps: $v{=}1$ node/step. If $|y_i|$ varies, IS sets $W_{ii}{=}1/y_i^2$, which rescales influence magnitudes but cannot create out-of-band entries. Stability requires $\eta < 2/\lambda_{\max}(KW)$; IS typically reduces $\lambda_{\max}$ relative to MSE, widening the stable step-size range.

4 Conclusion

Under locality, neural prediction dynamics trained with IS loss obey a strict finite propagation speed $v{=}R$ (nodes/step). The IS loss (scale/conformal invariant in prediction/target space) improves spectral conditioning and robustness *within* the cone via the diagonal curvature $W{=}\mathrm{diag}(1/y^2)$, but it does not alter the causal speed limit set by architectural locality.