Propagation Finie dans les Réseaux et Modèles Entraînés avec la Perte IS :
Une Borne Unifiée et un Exemple Détaillé

21 août 2025

1 Introduction

Nous présentons une borne unifiée, agnostique au domaine, sur la vitesse finie à laquelle l'influence, l'information ou les corrélations peuvent se propager dans les réseaux avec interactions locales (graphes, réseaux, circuits). Nous fournissons ensuite un exemple détaillé montrant comment la dynamique de prédiction neuronale entraînée avec la perte d'Itakura–Saito (IS) obéit à un « cône de lumière » discret exact sous linéarisation NTK. La perte IS améliore le conditionnement spectral (et donc la robustesse et la convergence *à l'intérieur* du cône) en normalisant les erreurs relatives, mais elle ne change pas la vitesse de propagation fondamentale définie par la localité architecturale.

2 Limite Généralisée de Vitesse dans les Réseaux

Considérons un réseau (graphe, réseau ou circuit computationnel) comme des nœuds V et des arêtes E, avec une distance métrique $d(x, y)$ entre les nœuds et des règles de mise à jour *locales*. Chaque nœud $x \in V$ a un état $\varphi_x(t)$ évoluant dans le temps t selon une dynamique locale. Soit $\mathcal{O}_x$ un observable localisé en x.

Théorème 1 (Borne de Propagation Unifiée dans les Réseaux)**.** *Supposons que le réseau satisfait :*

- *(i)* ***Localité :*** *les mises à jour de φ_x dépendent uniquement des nœuds dans un voisinage borné de x ;*
- *(ii)* ***Intensité d'interaction finie :*** *chaque mise à jour locale est bornée par une constante de Lipschitz g (par ex. échelle d'énergie, bande passante, facteur de Lipschitz) ;*
- *(iii)* ***Métrique bien définie :*** *le graphe admet une distance $d(\cdot, \cdot)$.*

Alors il existe une vitesse finie $v > 0$ (la vitesse de propagation du réseau*) et des constantes $C, \xi > 0$ telles que pour deux observables localisés quelconques $\mathcal{O}_x(t)$ et $\mathcal{O}_y(0)$,*

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

Interprétations.

- Dans les **systèmes de spins quantiques**, cela redonne la borne de Lieb–Robinson avec v la vitesse LR.
- Dans les **réseaux de communication**, v reflète les capacités des arêtes et d la longueur du chemin (diamètre).

— Dans le **consensus distribué**, les taux sont contraints par le trou spectral du Laplacien : des trous spectraux plus grands $\Rightarrow$ effectivement un v plus grand.
— Dans les **circuits / réseaux profonds**, v correspond aux couches par unité de temps (croissance du champ réceptif), donc $d(x, y)/v$ minore la profondeur minimale pour propager l'influence.

Conséquences. (a) Aucune propagation instantanée : il y a toujours un « cône de lumière » fini. (b) L'invariance d'échelle/conforme peut améliorer la robustesse et le conditionnement, mais ne supprime pas la limite de vitesse v.

3 Exemple Détaillé : Vitesse de Propagation Finie dans la Dynamique de Prédiction Neuronale sous Perte IS

Nous dérivons une borne explicite de type cône de lumière pour la dynamique de prédiction neuronale entraînée avec la perte IS via linéarisation NTK, et montrons comment la perte IS améliore le conditionnement sans changer la limite de vitesse.

3.1 Cadre : Dynamique NTK avec Perte IS

Soit $\{(x_i, y_i)\}_{i=1}^n$ des échantillons d'entraînement et $f_\theta(x) \in \mathbb{R}$ un modèle scalaire. Notons $f_\theta := (f_\theta(x_1), \ldots, f_\theta(x_n))^\top$ et $y \in \mathbb{R}^n$. Pour une donnée (y_i, f_i) la perte d'Itakura–Saito est

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

de sorte que

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

Près de $f \approx y$, la perte est localement quadratique avec une courbure $W := \mathrm{diag}(1/y_1^2, \ldots, 1/y_n^2)$ et un gradient $\nabla_f L \approx W(f - y)$.

Supposons une descente de gradient en lot complet et une linéarisation NTK avec le noyau NTK empirique fixe $K \in \mathbb{R}^{n \times n}$:

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

Avec l'erreur $e_t := f_t - y$ on obtient la dynamique linéaire

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

Hypothèse de localité. Supposons que les échantillons sont les nœuds d'un graphe (V, E) avec une distance $d(i, j)$. Supposons que K est *local de portée* R :

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

Ceci est valable pour les CNN/réseaux à champ réceptif fini, les GNN à R sauts, et les noyaux localisés. Puisque W est diagonal, $A{=}KW$ partage la même sparsité que K.

3.2 Un Cône de Lumière Discret pour l'Influence de Prédiction

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

Lemme 1 (Bande sous localité)**.** *Si* (6) *est vraie et W est diagonal, alors $A^t = (KW)^t$ est local de portée tR : $(A^t)_{ij} = 0$ si $d(i,j) > tR$.*

Démonstration. A a la même sparsité que K. Le produit de matrices locales de portée R_B et R_C est local de portée (R_B+R_C) par l'inégalité triangulaire sur les chemins. Itérer donne la localité de portée tR pour A^t. □

Théorème 2 (Cône de lumière discret sous entraînement IS)**.** *Sous* (6)*, pour tout* $t \geq 0$*,*

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

Équivalemment, une perturbation unitaire au nœud j à l'étape 0 ne peut affecter $f_t(i)$ si $d(i,j) > tR$.

Démonstration. Développons $(I-\eta A)^t = \sum_{k=0}^{t} \binom{t}{k} (-\eta)^k A^k$. Par le Lemme 1, A^k est local de portée kR. Si $d(i,j) > tR$, alors $d(i,j) > kR$ pour tout $k \leq t$, donc $(A^k)_{ij} = 0$ et l'entrée (i,j) de la somme s'annule. □

Ainsi la vitesse de propagation maximale est $v = R$ nœuds/étape — une propriété architecturale indépendante de la courbure ou de la taille du pas (sous réserve de stabilité spectrale).

3.3 Stabilité Spectrale, Taux de Convergence et Conditionnement du Hessien

Soit $\lambda_{\max}(A)$ le rayon spectral. La descente de gradient (5) est linéairement stable si

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

Sous stabilité, les erreurs se contractent à un taux gouverné par le spectre de A (et le conditionnement de la base propre). Près de $f \approx y$, le Hessien dans l'espace de prédiction est $H_f \approx W$, donc $A{=}KW$ agit comme un *préconditionneur implicite* : il pondère moins les grandes cibles et plus les petites via $W_{ii}{=}1/y_i^2$. Ceci aplatit le spectre relativement à MSE ($W{=}I$), améliorant le conditionnement et accélérant la convergence *à l'intérieur* du cône de lumière, sans altérer R (et donc sans augmenter v).

3.4 Queue Exponentielle au-delà de la Localité Stricte

Si K n'est pas strictement bandé mais décroît rapidement hors diagonale,

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

et W est borné ($0 < w_{\min} \leq W_{ii} \leq w_{\max} < \infty$), des estimations sous-multiplicatives donnent des constantes $C, \xi > 0$ telles que

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

c.-à-d. une borne de type Lieb–Robinson : suppression exponentielle en dehors d'un cône de lumière linéaire.

3.5 Exemple Concret en 1D (Noyau Tridiagonal)

Sur une chaîne 1D avec $d(i,j) = |i-j|$ et K tridiagonal (voisins immédiats ; R=1), le Lemme 1 implique que $(KW)^t$ est de bande $(2t+1)$, d'où

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

Une perturbation en j n'influence que les indices avec $|i-j| \leq t$ après t étapes : v=1 nœud/étape. Si $|y_i|$ varie, IS définit W_{ii}=$1/y_i^2$, ce qui redimensionne les magnitudes d'influence mais ne peut pas créer des entrées hors bande. La stabilité nécessite $\eta < 2/\lambda_{\max}(KW)$; IS réduit typiquement $\lambda_{\max}$ relativement à MSE, élargissant la plage de tailles de pas stables.

4 Conclusion

Sous localité, la dynamique de prédiction neuronale entraînée avec la perte IS obéit à une vitesse de propagation finie stricte v=R (nœuds/étape). La perte IS (invariante d'échelle/conforme dans l'espace de prédiction/cible) améliore le conditionnement spectral et la robustesse *à l'intérieur* du cône via la courbure diagonale W=diag$(1/y^2)$, mais elle ne modifie pas la limite de vitesse causale définie par la localité architecturale.