le cnam

# Proposition d'un sujet de thèse :

# Observateurs ensemblistes pour réseaux de neurones : applications à leur supervision

## Contexte

Les réseaux de neurones sont des systèmes numériques permettant de modéliser et de simuler la relation entre les entrées et les sorties d'un système à l'aide de couches de neurones artificiels. Ils peuvent être utilisés de multiples façons pour la supervision de systèmes dynamiques, par exemple en tant que classificateurs afin de catégoriser les comportements normaux ou anormaux, ou encore en tant que modèle de comportement normal du système afin de fournir une estimation de la sortie future attendue. Ce projet se concentre sur ce dernier cas pour lequel des estimateurs à base de réseaux de neurones récurrents sont utilisés. Ils sont implantés en utilisant des approches itératives ou auto-régressives pour lesquelles des estimations ou des mesures passées sont utilisées pour prédire la sortie d'un système dynamique.

Les méthodes ensemblistes [1] y compris les méthodes par intervalles [2] sont des outils mathématiques basés sur des systèmes monotones [3] qui sont des systèmes dynamiques (à temps continu ou à temps discret) dont le flot préserve un ordre partiel. Elles prennent en compte des bornes sur les variables inconnues du système considéré, sous hypothèse que les bornes des conditions initiales et des quantités incertaines sont connues. Ce projet de thèse s'inscrit dans ce cadre méthodologique, et vise à développer des algorithmes ensemblistes systématiques et unifiés qui quantifient la robustesse et assurent la sûreté des réseaux de neurones, en déterminant la sur-approximation de bornes de leurs sorties sous les hypothèses que les incertitudes sont inconnues mais bornées et que toutes les entrées du réseau de neurones soient dans un ensemble borné.

Les résultats théoriques seront validés à l'aide du kit de développement Jetson Nano et d'une plate-forme expérimentale de servomoteur. Le projet est discuté avec des chercheurs de l'industrie, ce qui pourrait entrainer leur implication dans la définition des tests. En effet, quantifier la robustesse et assurer la sûreté des réseaux de neurones est un défi, par exemple pour la conduite autonome.

## Objectif

Bien que certaines solutions satisfaisantes soient connues pour les systèmes basés sur des modèles, les algorithmes ensemblistes manquent encore de généralité pour les réseaux de neurones [4]. En d'autres termes, il n'existe pas encore de méthode unifiée et systématique pour la conception de tels outils. Par conséquent, le but de ce projet est de combiner l'apprentissage automatique avec la théorie du contrôle pour répondre à cette demande. Le projet poursuit plus particulièrement trois objectifs :

- Objectif 1 : Etudier la bibliographie existante sur les méthodes ensemblistes pour les réseaux de neurones et développer un algorithme ensembliste calculant la sortie étroitement bornée des réseaux de neurones à entrée bornée. Dans le même esprit que ce qui est fait pour propager la covariance d'estimation d'un filtre de Kalman, l'évaluation de la robustesse d'un estimateur de réseau de neurones quantifie à quel point l'estimateur peut réduire l'influence de l'incertitude sur la prédiction.
- Objectif 2 : Développer des méthodes, pour la sûreté des réseaux de neurones, basées sur l'algorithme ensembliste de l'Objectif 1. Grâce à l'utilisation de la somme géométrique, des seuils serrés sont acquis et une alarme est générée lorsque des anomalies surviennent.
- Objectif 3 : Développer un démonstrateur pour valider expérimentalement les méthodes et les preuves d'assurance des Objectifs 1-2. Pour un scénario de validation, à définir durant la thèse, une plate-forme expérimentale de servomoteur comprenant le kit de développement NVIDIA Jetson Nano sera utilisée.

# Echéancier

La procédure pour arriver à nos objectifs est divisée en quatre étapes de travail.

- Etape 1 (4 mois) : Revue de littérature. Etude des propriétés structurelles des réseaux de neurones (réseaux de neurones à propagation avant, récurrents, transformeurs…) pour identifier les compatibilités avec les techniques ensemblistes en vue de répondre à l'Objectif 1.
- Etape 2 (21 mois) : Développer un algorithme ensembliste et un algorithme de détection de défauts pour des réseaux de neurones sélectionnés en première étape.
- Etape 3 (6 mois) : Développer un système de démonstration pour valider les contributions théoriques en deuxième étape.
- Etape 4 (5 mois) : Rédaction de la thèse.

# Bibliographie

[1] A. Chen, T.N. Dinh, T. Raïssi, and Yi Shen, "Outlier-robust set-membership estimation for discrete-time linear systems," International Journal of Robust and Nonlinear Control, Vol. 32, No. 4, pp. 2313-2329, 2022.

le cnam

[2] F. Zhu, Y. Fu, and T.N. Dinh, "Asymptotic Convergence Unknown Input Observer Design via Interval Observer," Automatica, Vol. 147, pp. 110744, 2023.

[3] W.M. Haddad, V. Chellaboina, and Q. Hui, "Nonnegative and compartmental dynamical systems," Princeton university press, 2009.

[4] A. Venzke and S. Chatzivasileiadis, "Verification of neural network behaviour: Formal guarantees for power system applications, " IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 383–397, 2021.

# Encadrement de thèse

Dr. Thach Ngoc Dinh
ngoc-thach.dinh@lecnam.net

Pr. Mathieu Moze
mathieu.moze@lecnam.net

Dr. Jérémy Van Gorp
jeremy.vangorp@lecnam.net