

Sujet de thèse

Titre de la thèse : Apprentissage profond basé sur la physique pour la modélisation de la dynamique complexe. Applications au climat

Directrice ou directeur de thèse : Patrick Gallinari, **Professeur à Sorbonne Université – ISIR**

Co-encadrement : M. Levy et S. Thiria du laboratoire LOCEAN

Laboratoire d'accueil : ISIR (*Institut des Systèmes Intelligents et de Robotique*), Campus Pierre et Marie Curie, 4 place Jussieu, 75005 Paris.

Début de la thèse : octobre / novembre 2022

Le sujet de recherche est ouvert et, selon le profil du candidat, il peut être orienté davantage vers la théorie ou vers l'application.

Personne à contacter

Prénom Nom : Patrick Gallinari

Email : patrick.gallinari(at)sorbonne-universite.fr

Envoyer votre candidature par mail, avec [*sujet de la thèse*] en objet, un CV, une lettre de motivation, les notes obtenues en master, et lettres de recommandation si possible.

Description du sujet (en français)

Contexte :

L'apprentissage profond commence à être développé pour le calcul scientifique dans des domaines traditionnellement dominés par les modèles physiques comme les sciences de la terre, les sciences du climat, les sciences biologiques, etc. Il est particulièrement prometteur pour les problèmes impliquant des processus qui ne sont pas complètement compris ou qui sont trop complexes pour être modélisés analytiquement. Des chercheurs de différentes communautés ont commencé à explorer (i) comment intégrer les connaissances physiques et les données pour la modélisation de phénomènes complexes, et (ii) comment repousser les limites des méthodes et de la théorie actuelles de l'apprentissage automatique pour ces problèmes de modélisation, deux directions stimulantes. Nous considérons ici les approches d'apprentissage profond pour la modélisation de systèmes dynamiques complexes caractérisant les phénomènes naturels, un sujet de recherche récent et en pleine expansion (Willard et al. 2020, Thuerey et al. 2021). Des problèmes et des applications motivants seront issus de la science du climat (de Bezenac et al. 2018, Ayed et al. 2020).

Objectifs scientifiques :

L'objectif global de la thèse est le développement de nouveaux modèles exploitant des données d'observation ou de simulation pour la modélisation de dynamiques spatio-temporelles complexes caractérisant des phénomènes physiques tels que ceux qui sous-tendent les observations en sciences de la terre et du climat. Les outils classiques de modélisation de ces dynamiques en physique et en mathématiques appliquées reposent sur les équations aux dérivées partielles (EDP). Malgré leurs succès dans différents domaines, les approches d'apprentissage actuelles sont nettement insuffisantes pour de tels problèmes. Utiliser

Sous la co-tutelle de :

L'apprentissage pour la physique soulève de nouveaux problèmes qui exigent de repenser les idées à la base de l'apprentissage.

Axes de recherche :

Systèmes hybrides - Intégrer la physique et l'apprentissage profond

Souvent il existe des connaissances physiques préalables décrites par des EDP pour caractériser le phénomène sous-jacent. Une question clé est alors de savoir comment combiner ces connaissances avec les informations extraites des données. L'apprentissage peut venir en complément des modèles numériques et nous permettre de prendre en compte des informations non présentes dans le modèle ou d'intégrer des données d'observation. Il peut également servir de modèle de prototypage rapide. Des tentatives initiales pour résoudre des problèmes similaires existent dans des travaux récents tels que (de Bezenac et al. 2018, Harlim et al. 2020, Yin et al. 2021, Dona et al. 2022). Ceci sera développé pour le projet de thèse avec l'objectif d'analyser et de développer différents cadres de systèmes hybrides.

Généralisation de domaine pour l'apprentissage de dynamiques

Les modèles physiques explicites sont accompagnés de garanties et peuvent être utilisés dans n'importe quel contexte (également appelé domaine ou environnement) où le modèle est valide. Ce n'est pas le cas des réseaux de neurones qui n'offrent aucune garantie de généralisation à de nouveaux environnements physiques. Nous proposons ici d'attaquer ce problème en nous inspirant de cadres d'apprentissage récents développés pour traiter ce nouveau sujet de recherche qu'est la généralisation à un domaine, tels que (Yin et al. 2021b, Wang et al. 2021).

Apprentissage à plusieurs échelles

La modélisation des processus physiques dynamiques nécessite souvent la prise en compte de plusieurs échelles spatio-temporelles. Par exemple, dans le domaine du climat, les phénomènes globaux sont influencés par des dynamiques opérant à une échelle plus petite. Des problèmes similaires se posent, par exemple, en dynamique des fluides. L'apprentissage à différentes échelles est une question ouverte. La plupart des déploiements actuels de réseaux de neurones pour l'apprentissage de dynamiques utilisent une discrétisation spatio-temporelle fixe. Des avancées récentes (Sitzman 2020, Lindel et al. 2021, Li 2021) s'appuyant sur des représentations implicites, permettent d'apprendre un espace de fonctions au lieu de flux discrets et ouvrent la possibilité de généraliser à différentes résolutions spatio-temporelles. Ceci sera utilisé comme point de départ pour l'apprentissage multi-échelles avec des réseaux de neurones.

Profil recherché : Master en informatique ou en mathématiques appliquées, ou école d'ingénieur·e·s. Formation et expérience en apprentissage automatique. Bonnes compétences techniques en programmation.

Environnement de travail :

Le contrat de thèse est pour trois ans débutant en octobre/novembre 2022. Il ne comprend pas d'obligation d'enseignement, mais il est possible d'en faire si on le souhaite. Le/la doctorant(e) travaillera à Sorbonne Université (S.U.), Campus Pierre et Marie, dans le centre de Paris. Il/elle intégrera l'équipe Machine Learning and Deep Learning for Information Access de S.U. au laboratoire ISIR. (Institut des Systèmes Intelligents et de Robotique). En ce qui concerne le climat, le candidat sera co-encadré par M. Levy et S. Thiria du laboratoire LOCEAN, <https://www.locean-ipsl.upmc.fr/>.

Sous la co-tutelle de :

Description du sujet (en anglais)

Physics Based Deep Learning for Modeling Complex Dynamics. Applications to Climate

Context:

Deep Learning is beginning to be explored for scientific computing in domains traditionally dominated by physics models (first principles) like earth science, climate science, biological science, etc. It is particularly promising in problems involving processes that are not completely understood, or computationally too complex to solve by running the physics inspired model. The direct use of pure machine learning approaches has however met limited successes for scientific computing. Hence, researchers from different communities have started to explore (i) how to integrate physics knowledge and data, and (ii) how to push the limits of current ML methods and theory; two challenging directions. We consider here deep learning approaches for the modeling of complex dynamical systems characterizing natural phenomena, a recent and fast growing research topic (Willard et al. 2020, Thuerey et al. 2021). Motivating problems and applications will come from climate science (de Bezenac et al. 2018, Ayed et al. 2020).

Scientific Objective:

The global objective is the development of new models leveraging observation or simulation data for the modeling of complex spatio-temporal dynamics characterizing physical phenomena such as those underlying earth-science and climate observations. The classical modeling tools for such dynamics in physics and applied mathematics rely on partial differential equations (PDE). Despite their successes in different areas, current ML based approaches are notably insufficient for such problems. Using ML for physics raises new challenging problems and requires rethinking fundamental ML ideas.

Research directions:

Hybrid systems - Integrating Physics and Deep Learning

In many situations, there is available some prior physical knowledge provided by PDEs for characterizing the underlying phenomenon. A key issue is then how to combine this prior knowledge with information extracted from the data. ML could come as a complement to numerical models and allow us to take into account information not present in the model or to integrate observation data. Alternatively, it could come as a surrogate model for fast prototyping. From a ML perspective, physical priors help guide and constrain the learning process. Initial attempts to solve similar problems can be found in recent work such as (de Bezenac et al. 2018, Harlim et al. 2020, Yin et al. 2021, Dona et al. 2022). This will be further developed for the PhD project with the objective of analyzing and developing different integration frameworks.

Domain generalization for deep learning as dynamical models

Explicit physical models come with guarantees and can be used in any context (also called domain or environment) where the model is valid. This is not the case for DNNs, and we have no guarantee that they can be extrapolated to new physical environments. We propose here to tackle the problem by drawing inspiration from recent ML frameworks developed for handling the new research topic of domain generalization, such as (Yin et al. 2021b, Wang et al. 2021).

Learning at Multiple Scales

Modeling dynamical physical processes often requires considering multiple spatio-temporal scales. For example in climate, global phenomena are influenced by dynamics operating at a smaller scale. Similar problems occur e.g. in computational fluid dynamics. Learning at different scales is an open issue in ML. Most current DNN deployments for learning dynamics operate at a

Sous la co-tutelle de :

fixed spatio-temporal discretization. Recent advances (Sitzman 2020, , Lindel et al. 2021, Li 2021) relying on implicit representations, allow us learning a function space instead of discrete flows and open the possibility for generalizing at different spatio-temporal resolutions. This will be used as starting point for learning at different scales with DNNs.

Required Profile: Master in computer science or applied mathematics, Engineering school. Background and experience in machine learning. Good technical skills in programming.

Position and Working Environment

The PhD studentship is a three years position starting in October/November 2022. It does not include teaching obligation, but it is possible to engage if desired. The PhD candidate will work at Sorbonne Université (S.U.), Pierre et Marie Campus in the center of Paris. He/She will integrate the Machine Learning and Deep Learning for Information Accesss team at S.U. at the ISIR lab. (Institut des Systèmes Intelligents et de Robotique). On the Climate side, the candidate will be co-advised by M. Levy and S. Thiria from LOCEAN laboratory, <https://www.locean-ipsl.upmc.fr/>

References :

Ayed, I., de Bezenac, E., Pajot, A. and Gallinari, P. 2020. Learning the Spatio-Temporal Dynamics of Physical Processes from Partial Observations. *ICASSP (2020)*.

de Bezenac, E., Pajot, A. and Gallinari, P. 2018. Deep Learning For Physical Processes: Incorporating Prior Scientific Knowledge. *ICLR (2018)*.

Dona, J., Dechelle, M., Jacques-Dumas, V., Gallinari, P. and Levy, M. 2022. Learning Inverse and Forward Models from Partially Known Dynamics. *ICLR (2022)*.

Harlim, J., Jiang, S.W., Liang, S. and Yang, H. 2021. Machine learning for prediction with missing dynamics. *Journal of Computational Physics*. 428, (2021), 109922.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A. and Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations. *ICLR (2021)*, 1–16.

Lindell, D.B., Van Veen, D., Park, J.J. and Wetzstein, G. 2021. BACON: Band-limited Coordinate Networks for Multiscale Scene Representation. <http://arxiv.org/abs/2112.04645> (2021).

Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G. and University, S. 2020. Implicit Neural Representations with Periodic Activation Functions. *Neurips (2020)*.

Wang, R., Walters, R. and Yu, R. 2021. Meta-Learning Dynamics Forecasting Using Task Inference. <http://arxiv.org/abs/2102.10271> (2021), 1–20.

Willard, J.D., Jia, X., Xu, S., Steinbach, M. and Kumar, V. 2020. Integrating physics-based modeling with machine learning: A survey. *arXiv (2020)*, 1–34.

(Yin et al. 2021a)Yin, Y., Le Guen, V., Dona, J., de Bezenac, E., Ayed, I., Thome, N. and Gallinari, P. 2021. Augmenting Physical Models with Deep Networks for Complex Dynamics Forecasting. *ICLR (2021)*.

(Yin et al. 2021b)Yin, Y., Ayed, I., de Bézenac, E., Baskiotis, N. and Gallinari, P. 2021. LEADS: Learning Dynamical Systems that Generalize Across Environments. *Neurips (2021)*.