1 Challenges in deep learning and individual contributions to the field

The success of deep learning came with several issues: vulnerabilities against **adversarial attacks**, concern regarding **privacy leakages** of the train set, problematic behavior for **out-of-distribution** data, **instable training** in challenging tasks with high variance gradients, and **difficulties to reason** properly in LLM. The overarching objective of my research is to develop learning algorithms that are more robust to these uncertainties and defaults. In my researched I implemented Lipschitz and convexity constraints in neural networks, exploring theoretical guarantees and use cases. First, in Béthune et al. (2022) I characterized the expressiveness of these Lipschitz networks in classification tasks, showing an **accuracy/robustness tradeoff** controlled by entropic regularization of the loss, with **generalization guarantees**. After, in Béthune et al. (2023) I demonstrated that **signed distance functions** were a solution to a regularized optimal transport problem and that it could be used for **robust one class learning** and **neural implicit surfaces**. Finally, I showed in Béthune et al. (2023) that it was possible to perform **differentially private** training of Lipschitz networks without runtime or memory overhead. In this proposal, I will outline some concrete projects to overcome the shortcomings of existing work and apply these developed methodologies to new domains.

2 Research statement

Below I present a list of topics of personal interest, that I would like to pursue, in collaboration. I rated them with the associate risks.

- 🔵 **Software contribution.** No risk: this is a pure engineering project that takes time and may have a significant impact nonetheless.
- 🟢 **Low risk.** The literature around the required tools is plethoric. The project is likely to be realizable.
- 🟠 **Medium risk.** Existing literature suggests that the project is feasible, but some obstacles remain along the way.
- 🔴 **High risk.** The project is ambitious and the literature around the tools is scarce.

2.1 🔵 Improving existing toolbox in Jax

I believe that Jax (Bradbury et al., 2018) is a promising framework, thanks to its support for jit compilation on various hardware (including GPU), its vmap function transformation for easy and "bug-free" writing of code, and its extreme flexibility regarding automatic differentiation. Two open-source projects gathered my attention. The Jaxopt project (Blondel et al., 2021) on which I already contributed, a Jax library of optimizers with support of

implicit differentiation. The OTT-jax (Cuturi et al., 2022) project for computational optimal transport. Lipschitz neural networks are a central tool in neural optimal transport, and fruitful integration of these tools into the library are possible.

2.2 ● Conditional neural center-outward map for certified multivariate quantile regression

The center-outward map (del Barrio et al., 2022) is a generalization of median and quantiles in higher dimension, based on optimal transport (Peyré et al., 2017). Neural Monge map can be made statistically consistent with appropriate constraints and regularization (Korotin et al., 2021; Uscidda and Cuturi, 2023). By combining this with the recent conditional optimal transport (Bunne et al., 2022; Manupriya et al., 2023), it yields neural multivariate regression, with statistical guarantees. My goal is to combine these different works with Lipschitz constraints, to design the first neural multivariate quantile regressor with statistical consistency guarantees, and benchmark it on various tasks, including RL (see below).

2.3 ● Lipschitz constraints for stable control in reinforcement learning and robotic

Lipschitz constraint stabilizes control tasks (Song et al., 2023). I believe that entropic regularization in classification tasks with Lipschitz constraint (Béthune et al., 2022) is akin to Soft-Actor Critic's policy regularization (Haarnoja et al., 2018), suggesting that regularized Markov Decision Processes (Geist et al., 2019) can be solved with neural constraints on the policy. Mean regression in state-action value function Q can be replaced with quantile regression for stabilizing training with high-variance trajectories (Dabney et al., 2018), building upon the previous project on neural center-outward. My goal is to ensure robust and stable learning of policies in RL (robustness against noisy measurements, heavy-tailed rewards, or MDP shift), especially on the continuous control tasks used in robotic.

2.4 ● Lipschitz constraints for reliable neural computer graphics and physical simulations

The award-winning of Sharp and Jacobson (2022) shown that certified and robust queries were necessary for neural signed distance function (SDF) without aberrations. Lipschitz regularization for SDF was subsequently studied in Liu et al. (2022). I will attempt to scale their approach with Lipschitz constraint instead, using the formulation of Béthune et al. (2023). Furthermore, signed distance functions can be used in Monte-Carlo-based solving of Poisson equations, *without* discretization of the space, as proved in Sawhney and Crane (2020). My goal is to design robust solvers of PDE based on Lipschitz networks, leveraging their generalization capabilities to speed up solving on similar domains and giving formal approximation guarantees on the solution.

2.5 ● Theorize generalization in generative models with algorithmic information theory

Modern generative models, diffusion based (Yang et al., 2022) or LLM based (Vaswani et al., 2017), have the capabilities of performing variable-length computations, with next-token sampling or with iterative denoising based on Monte Carlo Langevin dynamics. The new framework of ν-information (Xu et al., 2019) offers tool to understand dataset complexity (Ethayarajh et al., 2022) through the lens of Kolmogorov complexity (Blier and Ollivier, 2018; Lee et al., 2022). My goal is to pose the basis of a framework to characterize the "difficulty" (e.g sample complexity) of reasoning tasks for deep learning, using these theories.

2.6 ● Toward implementing learn/forget operations in neural networks

Neural networks are known to suffer from catastrophic forgetting (Kirkpatrick et al., 2017). The field of continual learning offers tools to solve this problem (Parisi et al., 2019), and Lipschitz networks have shown some promises in this regard (Bonicelli et al., 2022). On the other hand, *machine unlearning* (Bourtoule et al., 2021) is emerging as a field that studies how to *unlearn* some datapoint, for example to fix licenses or privacy concerns, or discard noisy data, without retraining from scratch. Combining these two frameworks could allow the design cheap *learn/unlearn* operators in deep learning, for example to create a self-modifying memory in LLM, making them Turing-complete and capable of self-improvement, a necessary ingredient for adaptability.

References

L. Béthune, T. Boissin, M. Serrurier, F. Mamalet, C. Friedrich, and A. G. Sanz. Pay attention to your loss : understanding misconceptions about lipschitz neural networks. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

L. Béthune, T. Masséna, T. Boissin, Y. Prudent, C. Friedrich, F. Mamalet, A. Bellet, M. Serrurier, and D. Vigouroux. Dp-sgd without clipping: The lipschitz neural network way. *arXiv preprint arXiv:2305.16202*, 2023.

L. Béthune, P. Novello, G. Coiffier, T. Boissin, M. Serrurier, Q. Vincenot, and A. Troya-Galvis. Robust one-class classification with signed distance function using 1-Lipschitz neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 2245–2271. PMLR, 23–29 Jul 2023.

L. Blier and Y. Ollivier. The description length of deep learning models. *Advances in Neural Information Processing Systems*, 31, 2018.

M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert. Efficient and modular implicit differentiation. *arXiv preprint arXiv:2105.15183*, 2021.

L. Bonicelli, M. Boschini, A. Porrello, C. Spampinato, and S. Calderara. On the effectiveness of lipschitz-driven rehearsal in continual learning. *Advances in Neural Information Processing Systems*, 35:31886–31901, 2022.

L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot. Machine unlearning. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 141–159. IEEE, 2021.

J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL `http://github.com/google/jax`.

C. Bunne, A. Krause, and M. Cuturi. Supervised training of conditional monge maps. *Advances in Neural Information Processing Systems*, 35:6859–6872, 2022.

M. Cuturi, L. Meng-Papaxanthos, Y. Tian, C. Bunne, G. Davis, and O. Teboul. Optimal transport tools (ott): A jax toolbox for all things wasserstein. *arXiv preprint arXiv:2201.12324*, 2022.

W. Dabney, M. Rowland, M. Bellemare, and R. Munos. Distributional reinforcement learning with quantile regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

E. del Barrio, A. G. Sanz, and M. Hallin. Nonparametric multiple-output center-outward quantile regression. *arXiv preprint arXiv:2204.11756*, 2022.

K. Ethayarajh, Y. Choi, and S. Swayamdipta. Understanding dataset difficulty with v-usable information. In *International Conference on Machine Learning*, pages 5988–6008. PMLR, 2022.

M. Geist, B. Scherrer, and O. Pietquin. A theory of regularized markov decision processes. In *International Conference on Machine Learning*, pages 2160–2169. PMLR, 2019.

T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

A. Korotin, L. Li, A. Genevay, J. M. Solomon, A. Filippov, and E. Burnaev. Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark. *Advances in Neural Information Processing Systems*, 34:14593–14605, 2021.

Y. Lee, C. Finn, and S. Ermon. Relaxing the kolmogorov structure function for realistic computational constraints. In *NeurIPS 2022 Workshop on Information-Theoretic Principles in Cognitive Systems*, 2022.

H.-T. D. Liu, F. Williams, A. Jacobson, S. Fidler, and O. Litany. Learning smooth neural functions via lipschitz regularization. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–13, 2022.

P. Manupriya, R. K. Das, S. Biswas, S. Chandhok, and S. N. Jagarlapudi. Empirical optimal transport between conditional distributions. *arXiv preprint arXiv:2305.15901*, 2023.

G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019.

G. Peyré, M. Cuturi, et al. Computational optimal transport. *Center for Research in Economics and Statistics Working Papers*, 2017.

R. Sawhney and K. Crane. Monte carlo geometry processing: A grid-free approach to pde-based methods on volumetric domains. *ACM Transactions on Graphics*, 39(4), 2020.

N. Sharp and A. Jacobson. Spelunking the deep: Guaranteed queries on general neural implicit surfaces via range analysis. *ACM Transactions on Graphics*, 41(4):1–16, July 2022. ISSN 0730-0301, 1557-7368. doi: 10.1145/3528223.3530155.

X. Song, J. Duan, W. Wang, S. E. Li, C. Chen, B. Cheng, B. Zhang, J. Wei, and X. S. Wang. LipsNet: A smooth and robust neural network with adaptive Lipschitz constant for high accuracy optimal control. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*. PMLR, 23–29 Jul 2023.

T. Uscidda and M. Cuturi. The monge gap: A regularizer to learn all transport maps. In *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2023.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Y. Xu, S. Zhao, J. Song, R. Stewart, and S. Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2019.

L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 2022.