Longtime behavior of mean-field Langevin dynamics for minimax problems in Machine Learning

Master 2 internship proposal

Supervisors: Loucas Pillaud-Vivien and Julien Reygner.
Duration: 6 months.
Location: CERMICS (École National des Ponts et Chaussées, Marne-la-Vallée).
Keywords: *Stochastic Differential Equation, Mean-field dynamics, Min-Max problems, Longtime behavior.*

Possibility to extend the internship to a PhD at the interface between probability and machine learning.

How to apply: Send your CV to loucas.pillaud-vivien@enpc.fr and julien.reygner@enpc.fr, together with a transcript of your Master 1 (or equivalent) and the name of a professor that can be contacted to recommend you.

Deadline for application: 29th of November 2024.

1. Context

On the one hand, the long-term behavior of the overdamped Langevin dynamics

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \tag{1}$$

has been the focus of an intense development [1] in recent decades; driven by its relationship with optimization, sampling, and functional inequalities. An interesting point of view comes from the fact that Eq.(1) can be seen as the Wasserstein Gradient flow of $F[\mu] := \int V d\mu + \beta^{-1}H[\mu]$, where H is the entropy functional and hence Eq.(1) can be interpreted as the descent of an entropic regularization of the minimization problem of V.

On the other hand, minimization-maximization problems, i.e. finding a pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that solves the program $\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} f(x, y)$ or equivalently finding a Nash equilibrium for two-players sum games, have been the focus of a lot of attention in Machine Learning in the last years with, e.g. the rise of Generative Adversarial Networks (GANs) or distributionally robust learning. A natural way to tackle such probems is to resort to local minimization-maximization algorithms, and their entropic-regularized version.

2. Question

This can be seen as the joint dynamics:

$$dX_t = -\nabla_x \left(\int_{y\in\mathcal{Y}} f(X_t, y) d\nu_t(y) \right) dt + \sqrt{2\beta^{-1}}dW_t \tag{2}$$

$$dY_t = \quad \nabla_y \left(\int_{x\in\mathcal{X}} f(x, Y_t) d\mu_t(x) \right) dt + \sqrt{2\beta^{-1}}dB_t\ , \tag{3}$$

where μ_t and ν_t are respectively the time-marginal laws of X_t and Y_t for all $t \geq 0$. This *mean-field* dynamics is in fact the Wassertein Gradient Descent-Ascent flow of the convex-concave functional

$$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \ni (\mu, \nu) \to F[\mu, \nu] := \int f d\mu d\nu + \beta^{-1} H[\mu] - \beta^{-1} H[\nu] \, .$$

Except for some particular cases [2, 3], the longtime behavior of such dynamics is still an open problem (see the related COLT issue [4]). The goal of this internship is to study the longtime behavior of the dynamics Eqs (2)-(3), including a quantitative description of the speed of convergence of $(\mu_t, \nu_t)_{t\geq 0}$ towards equilibrium. A possible approach to solve this problem could be either probabilistic methods such as couplings, or analytic methods such as functional inequalities.

References

[1] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[2] Carles Domingo-Enrich, Samy Jelassi, Arthur Mensch, Grant Rotskoff, and Joan Bruna. A mean-field analysis of two-player zero-sum games. *Advances in neural information processing systems*, 33:20215–20226, 2020.

[3] Yulong Lu. Two-scale gradient descent ascent dynamics finds mixed nash equilibria of continuous games: A mean-field perspective. In *International Conference on Machine Learning*, pages 22790–22811. PMLR, 2023.

[4] Guillaume Wang and Lénaïc Chizat. Open problem: Convergence of single-timescale mean-field langevin descent-ascent for two-player zero-sum games. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 5345–5350. PMLR, 30 Jun–03 Jul 2024.