

Application of Langevin dynamics for optimization in machine learning tasks

Wednesday, 25 September 2024 15:30 (20 minutes)

A rapid growth of Machine Learning (ML) applications in different areas has been faced in recent years. Training of ML models is performed by finding such values of their parameters $x = \{x_1, x_2, \dots, x_N\}$ that optimize (minimize) the objective (loss) function $U(x)$. Usually, the number of parameters N is large and the training dataset is massive. Therefore, to reduce computational costs, the gradient $f = -dU(x)/dx$ of the objective function with respect to the model parameters is computed on relatively small subsets of the training data, called mini-batches. If these mini-batches are selected randomly from the training dataset, then the estimated values of the loss $\hat{U}(x)$ and its gradient $\hat{f} = -d\hat{U}(x)/dx$ are the stochastic approximations of their exact values. Therefore, it is natural to apply Langevin dynamics to treat this stochastic optimization problem. We consider the next discrete form of the Langevin equation:

$$\frac{\Delta x_{n+1} - \Delta x_n}{\Delta t^2} = \hat{f}_n - \gamma \frac{\Delta x_{n+1} + \Delta x_n}{2\Delta t}, \quad (1)$$

where n is an iteration number, $\Delta x_{n+1} = x_{n+1} - x_n$, Δt is a time step and $\gamma > 0$ is a viscous friction coefficient.

Now, it is straightforward to obtain the next parameter updating formula:

$$\Delta x_{n+1} = \rho \Delta x_n + \hat{f}_n \cdot \eta, \quad (2)$$

where $\rho = (1 - \gamma \Delta t / 2) / (1 + \gamma \Delta t / 2)$ is conventionally called a momentum coefficient and $\eta = \Delta t^2 (1 + \rho) / 2$ a learning rate constant.

Equation (2) was derived in our recent work where we have introduced Coolmomentum – a method for stochastic optimization by Langevin dynamics with simulated annealing [1]. To implement simulated annealing (or slow cooling, in physical terms), we apply a certain schedule for the gradual momentum coefficient decrease in the range

$$0 \leq \rho < 1 \quad (3)$$

In this talk we demonstrate that application of Langevin dynamics (2) with simulated annealing (3) to multidimension optimization tasks gives promising results in artificial intelligence [1], quantum computing [2] and optical engineering [3].

Acknowledgement: OB, MB and IO have received funding through the EURIZON project, which is funded by the European Union under grant agreement No.871072. AS acknowledges support by the National Research Foundation of Ukraine, project No.2023.03/0073.

[1] O. Borysenko, M. Byshkin, Sci Rep 11, 10705 (2021). <https://doi.org/10.1038/s41598-021-90144-3>.

[2] Daisuke Tsukayama et al. Jpn. J. Appl. Phys. 62, 088003 (2023). <https://dx.doi.org/10.35848/1347-4065/acea0a>.

[3] Z. Zhang et al. Photonics 10, 102 (2023). <https://doi.org/10.3390/photonics10020102>.

Primary author: BORYSENKO, Oleksandr (National Science Center "Kharkiv Institute of Physics and Technology")

Co-authors: Dr BRATCHENKO, Mykhailo (National Science Center "Kharkiv Institute of Physics and Technology"); Prof. LOMI, Alessandro (Università della Svizzera italiana); Mr OMELCHENKO, Ihor (V.N. Karazin Kharkiv National University); Prof. SOTNIKOV, Andrii (National Science Center "Kharkiv Institute of Physics and Technology")

Presenter: BORYSENKO, Oleksandr (National Science Center "Kharkiv Institute of Physics and Technology")

Session Classification: Afternoon Session 2

Track Classification: STATISTICAL PHYSICS AND KINETIC THEORY