Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 29 (2026), No. 3, pp. 575–585
DOI : 10.47974/JIM-2492

Improving machine learning algorithms using methodological stochastic differential equations

Isha Saluja *
Symbiosis Centre for Advanced Legal Studies and Research (SCALSAR)
Symbiosis Law School Pune (SLS-P)
Symbiosis International (Deemed University)
Pune 411014
Maharashtra
India

Srijita Bhattacharjee †
Department of Engineering and Technology (CSE)
Bharati Vidyapeeth (Deemed to be University)
Kharghar
Navi Mumbai 410210
Maharashtra
India

Vimal Bibhu §
Department of Computer Science & Engineering
Noida International University
Greater Noida 203201
Uttar Pradesh
India

S. Mohan Mahalakshmi Naidu ‡
Department of Electronics & Telecommunication
International Institute of Information Technology (I²IT)
Pune 411057
Maharashtra
India

* *E-mail:* isha.saluja@symlaw.ac.in (Corresponding Author)
† *E-mail:* srijitacseengg2007@gmail.com
§ *E-mail:* vimal.bibhu@niu.edu.in
‡ *E-mail:* mohans@isquareit.edu.in

© Taru Publications

Vanita S. Buradkar [@]
Department of Computer Science & Engineering
Rajiv Gandhi College of Engineering, Research & Technology
Chandrapur 442403
Maharashtra
India

Samariddin Makhmudov [#]
Department of Finance and Tourism
Termez University of Economics and Service
Termez 190111
Uzbekistan

and

Department of Economics
Mamun University
Khiva 220900
Uzbekistan

and

Center of the Engagement of International Ranking Agencies
Tashkent State University of Economics
Tashkent 100066
Uzbekistan

Abstract

The study examines the application of scientific stochastic differential equations (SDEs) to machine learning techniques to make them more robust and predictive. This is an approach that enhances generalization in evolutionary and complex environments because it characterizes the skepticism and clatter of data that are constructed in with SDEs. We come up with new SDE-based systems that can modify the fast rate at which they learn as well as the frequency with which they revise parameters. This ensures that converging is more stable. Big changes in measures of speed are seen in experimental results on test datasets when compared to traditional optimization methods. The proposed approach provides a

[@] *E-mail:* **vsburadkar@gmail.com**
[#] *E-mail:* **samariddin_maxmudov@tues.uz;**
maxmudov_samariddin@mamunedu.uz;
s.makhmudov@tsue.uz

solid mathematical means of incorporating stochastic dynamics that provide more data of how algorithms behave in the case of doubt.

Subject Classification: *34K50.*

Keywords: *Stochastic differential equations, Machine learning, Robustness, Optimization, Predictive modeling.*

1. Introduction

Machine learning (ML) has transformed many areas and allows computers to identify trends in data and make intelligent decisions. However, the real world is usually full of confusing, unclear, and changing data, and thus it is difficult to make standard ML systems work. Many of these common optimization techniques are only effective in known predictable settings [1], [2]. This may render models to be less reliable and useful in case of random changes. Therefore, it is relevant to understand how to handle uncertainty effectively to make machine learning more effective, particularly in areas where the data is likely to be inconsistent, such as banking, healthcare, robots, and self-driving systems [3]. Stochastic differential equations (SDEs) can be an effective tool to apply math to the description of systems, which are perturbed by noise and random variations. SDEs find application in physics, economics, and control theory to demonstrate time-varying changes in processes in the case of random forces [4]. The application of SDEs in machine learning can provide a rational means of demonstrating confusions and variations which occur over time as we learn. This combination simplifies the handling of computers in relation to noise data and settings that are difficult to comprehend. Most recently, researchers have been interested in stochastic optimization schemes using SDEs [5]. They are Langevin dynamics and stochastic gradient Langevin dynamics (SGLD). The techniques introduce noise to the parameter variations so as to bypass local minima to enable exploration to be more effective [6][7].

2. Mathematical Foundations of Stochastic Differential Equations

2.1. *Ito and Stratonovich Calculus Overview*

Stochastic differential equations (SDEs) are a special type of equation, and are usually written using ito and Stratonovich calculus, and concerned with chance. Ito calculus is concerned with stochastic integrals that lack any predictive properties [8], [9]. This is suitable in the modelling of the processes that involve all-random noise and the future value does not have the ability to alter the present.

Definition (Ito Integral): Given a Brownian motion $\boldsymbol{W\_t}$ and an adapted process $\boldsymbol{X\_t}$, the Ito integral over [0,T] is defined as

$$\int^{\mathbf{0}} \boldsymbol{X_t} \boldsymbol{dW_t}^{\mathrm{T}} = \mathbf{lim}_{\{\boldsymbol{n}\to\infty\}} \boldsymbol{X} \left(\boldsymbol{W}_{\{\boldsymbol{t}_{\{\boldsymbol{i}+\mathbf{1}\}}\}} - \boldsymbol{W}_{\{\boldsymbol{t}_{\boldsymbol{i}}\}}\right) \tag{1}$$

Where, $\mathbf{0 \ = \ t\_0 \ < \ t\_1 \ < \ldots < \ t\_n \ = \ T}$ is a partition of $[\mathbf{0}, \boldsymbol{T}]$.

Equation 1 (Ito's Lemma): For an Ito process X_t and twice differentiable function f(t,x),

$$\boldsymbol{df(X_t) = \left(\frac{\partial f}{\partial t}\right) dt \ + \ \left(\frac{\partial f}{\partial x}\right) dX_t + \left(\frac{1}{2}\right)\left(\frac{\partial^2 f}{\partial x^2}\right)(dX_t)^2.} \tag{2}$$

Theorem (Ito Isometry):

$$\boldsymbol{E\left[\left(\int^{0\,\mathrm{T}} X_t dW_t\right)^2\right] = \ E\left[\int^{0\,\mathrm{T}} X_t^2 dt\right]} \tag{3}$$

Equation 2 (Stratonovich Integral): The Stratonovich integral is defined as

$$\boldsymbol{\int^{0\,\mathrm{T}} X_t \circ dW_t = \lim_{\{n\to\infty\}} \sum_{\{i=0\}}^{\{n-1\}} \left[\frac{X_{\{t_i\}} + X_{\{t_{\{i+1\}}\}}}{2}\right]\left(W - \ W_{\{t_i\}}\right)} \tag{4}$$

Equation 3 (Relationship between Ito and Stratonovich):

$$\boldsymbol{\int^{0\,\mathrm{T}} X_t \circ \ dW_t = \ \int^{0\,\mathrm{T}} X_t dW_t + \ \left(\frac{1}{2}\right)\int_0^T \left(\frac{dX_t}{dt}\right) dt} \tag{5}$$

Equation 4 (Chain Rule for Stratonovich): For f(t,X_t) with Stratonovich calculus,

$$\boldsymbol{df \ = \ \left(\frac{\partial f}{\partial t}\right) dt \ + \ \left(\frac{\partial f}{\partial x}\right) * \ dX_t} \tag{6}$$

2.2. *Analytical Solutions and Numerical Approximations*

There aren't any closed-form mathematical answers for most SDEs, especially when the drift and diffusion terms are complicated. To simulate random processes, numerical approximation methods such as the Euler-Maruyama and Milstein schemes are very important.

Definition (Strong Solution to SDE): A process X_t is a strong solution to the SDE

$$\boldsymbol{dX_t = \ \mu(X_t, t)dt \ + \ \sigma(X_t, t)dW_t} \tag{7}$$

if X_t is adapted to Brownian filtration and satisfies the equation almost surely.

Equation 1 (Analytical Solution for Linear SDE):

$$\boldsymbol{For \ dX_t = \ a\, X_t dt \ + \ b\, X_t dW_t,} \tag{8}$$

solution is

$$\boldsymbol{X_t = X^0 \exp\left(\left(a \ - \frac{b^2}{2}\right) t \ + \ b\, W_t\right)} \tag{9}$$

Theorem (Existence and Uniqueness): *If μ and σ satisfy Lipschitz continuity and linear growth conditions, there exists a unique strong solution.*

Equation 2 (Euler-Maruyama Numerical Scheme): For step size Δt, approximation:

$$\boldsymbol{X_{\{n+1\}} = X_n + \mu(X_n, t_n)\Delta t + \sigma(X_n, t_n)\Delta W_n,} \tag{10}$$

where $\boldsymbol{\Delta W_n \sim N(0, \Delta t)}$.

Equation 3 (Milstein Scheme):

$$\boldsymbol{X_{\{n+1\}} = X_n + \mu(X_n)\Delta t + \sigma(X_n)\Delta W_n + \left(\frac{1}{2}\right)\sigma(X_n)\sigma'(X_n)} \tag{12}$$

Equation 4 (Strong Convergence Rate): Euler-Maruyama strong order 0.5:

$$\boldsymbol{E\left[\left|X_T - X_T^{\{\Delta t\}}\right|\right] = O(\sqrt{\Delta t}).} \tag{13}$$

Equation 5 (Weak Convergence Rate): For expected values, Euler-Maruyama weak order 1:

$$\boldsymbol{\left|E[f(X_T)] - E\left[f\left(X_T^{\{\Delta t\}}\right)\right]\right| = O(\Delta t).} \tag{14}$$

3. Proposed Methodology

3.1. *Integration of SDEs into ML Algorithm Frameworks*

Adding stochastic differential equations (SDEs) to machine learning systems makes it possible to describe error in an organized way while training [10], [11].

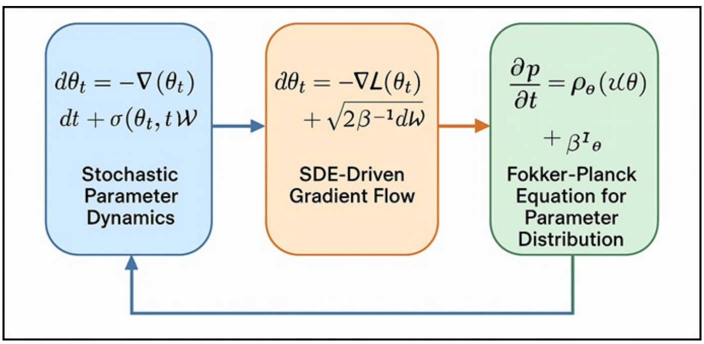

Figure 1
Integration of Stochastic Differential Equations into Machine Learning Algorithm Frameworks

Figure 1 shows integration of stochastic differential equations within machine learning frameworks. By showing changes to parameters as random events that happen all the time, SDEs make it possible for algorithms to naturally handle noise and adapt to changing data distributions [12, 13]. This method improves stability by avoiding overfitting and smoothing out optimization paths.

Definition (Stochastic Parameter Dynamics): Parameters θ_t evolve as stochastic processes via:

$$d\boldsymbol{\theta}_t = -\nabla_\theta L(\boldsymbol{\theta}_t)dt + \sigma(\boldsymbol{\theta}_t, t)dW_t, \tag{15}$$

where L(θ_t) is the loss function, and σ(θ_t, t) modulates noise intensity [14].

Equation 1 (SDE-Driven Gradient Flow):

$$d\boldsymbol{\theta}_t = -\nabla_\theta L(\boldsymbol{\theta}_t)dt + \sqrt{2\beta^{\{-1\}}}dW_t, \tag{16}$$

β is inverse temperature parameter controlling noise scale.

Theorem (Fokker-Planck Equation for Parameter Distribution): *Probability density p(θ,t) satisfies:*

$$\frac{\partial p}{\partial t} = \nabla_\theta \cdot \left(p\, \nabla_\theta L(\boldsymbol{\theta})\right) + \beta_\theta^{\{-1\}\Delta} p, \tag{17}$$

describing parameter distribution evolution under SDE.

Equation 2 (Stationary Distribution):

$$p_{ss(\theta)} \propto \exp\left(-\beta\, L(\boldsymbol{\theta})\right). \tag{18}$$

Equation 3 (Discretized Parameter Update):

Using Euler-Maruyama:

$$\boldsymbol{\theta}_{\{k+1\}} = \boldsymbol{\theta}_k - \eta\, \nabla_\theta L(\boldsymbol{\theta}_k) + \sqrt{2\,\eta\,\beta^{\{-1\}}}\xi_k, \tag{19}$$

$\xi_k \sim N(\mathbf{0}, \boldsymbol{I})$, η is learning rate.

Additional Equations:

1. Gradient noise covariance:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = E[\nabla_\theta L(\boldsymbol{\theta})\nabla_\theta L(\boldsymbol{\theta})^T] - \nabla_\theta L(\boldsymbol{\theta})\nabla_\theta L(\boldsymbol{\theta})^T. \tag{20}$$

2. Stochastic Gradient Langevin Dynamics (SGLD):

$$\boldsymbol{\theta}_{\{k+1\}} = \boldsymbol{\theta}_k - \left(\frac{\eta_k}{2}\right)\nabla_\theta L(\boldsymbol{\theta}_k) + \sqrt{\eta_k}\xi_k. \tag{21}$$

3. Continuous-time Langevin diffusion:

$$d\boldsymbol{\theta}_t = -\nabla_\theta L(\boldsymbol{\theta}_t)dt + \sqrt{2}dW_t.$$

3.2. *Formulation of Stochastic Optimization Algorithms*

Using the SDE framework as a base, stochastic optimization methods are created to use both drift and diffusion to make parameter changes that are more flexible. These algorithms go beyond traditional optimizers by adding controlled noise terms that help get out of local minima and make exploring the loss landscape better.

Definition (Stochastic Optimization with SDEs): Parameter updates as:

$$d\boldsymbol{\theta}_t = -g(\boldsymbol{\theta}_t)dt + \sigma(\boldsymbol{\theta}_t)dW_t,$$

where g(θ_t) is biased gradient estimator.

Equation 1 (Generalized SDE for Optimization):

$$d\theta_t = -\left(\nabla_\theta L(\theta_t) + b(\theta_t)\right)dt + \sigma(\theta_t)dW_t, \tag{22}$$

b(θ_t) models gradient noise bias.

Theorem (Convergence under Lipschitz Conditions): *When L has a Lipschitz gradient, 7(0) is bounded, then, 7(0) –t approaches the acute points of L in distribution.*

V 1 = -1 0 Sin 2 (Noise-Driven Escaping from Local Minima):

Noise σ(0t)dW t allows one to climb sharp minima by diffusion through loss landscape.

Equation 3 (Discrete-time Stochastic Optimization Update):

$$\theta_{\{k+1\}} = \theta_k - \eta_k\, \hat{g}(\theta_k) + \sqrt{2\,\eta_k \beta^{\{-1\}}}\xi_k, \tag{23}$$

ĝ(θ_k) noisy gradient, $\xi_k \sim N(0, I)$.

3.3. *Model Architecture and Algorithm Design*

The model framework has SDE-based optimization tools that can be used with regular machine learning processes. This implies that it is applicable with the deep neural networks and other popular structures. Flexibility is highly valued when creating an algorithm since it allows altering ad hoc parameters such as noise strength and time discretization steps in a fluid manner.

Definition (SDE-based Training Algorithm): Training modeled as solving discretized SDE:

$$\theta_{\{k+1\}} = \theta_k + f(\theta_k)\Delta t + G(\theta_k)\Delta W_k, \tag{24}$$

f(θ) deterministic dynamics, G(θ) noise matrix.

Equation 1 (Euler-Maruyama Discretization):

$$\theta_{\{k+1\}} = \theta_k + f(\theta_k)\Delta t + G(\theta_k)\sqrt{\Delta t}\,\xi_k,$$
$$\xi_k \sim N(0, I).$$

Theorem (Mean-Square Stability): *Assuming that f Lipschitz and G are bounded, the Euler-Maruyama is a mean-square stable discretization when Δt is small.*

Equation 2 (Adaptive Step-size Control):

$$\Delta t_{\{k+1\}} = \Delta t_k \times \min\left(\max\left(\frac{tol}{e_k}, 0.1\right), 5\right)$$

Where e_k is local error estimate.

Equation 3 (Noise Injection Layer in Neural Network):

Add noise $\boldsymbol{\eta} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2)$ at layer output h:

$$\boldsymbol{h}' = \boldsymbol{h} + \boldsymbol{\eta}.$$

1. Loss function with noise regularization:

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \boldsymbol{E}_{\xi}[\boldsymbol{L}(\boldsymbol{\theta} + \boldsymbol{\xi})] + \boldsymbol{\lambda} \left|\left|\boldsymbol{\theta}\right|\right|^2 \tag{25}$$

2. Gradient of noisy loss expectation:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} \boldsymbol{E}_{\xi}[\boldsymbol{L}(\boldsymbol{\theta} + \boldsymbol{\xi})] = \boldsymbol{E}_{\xi}[\boldsymbol{\nabla}_{\boldsymbol{\theta}} \boldsymbol{L}(\boldsymbol{\theta} + \boldsymbol{\xi})] \tag{26}$$

4. Result Discussion

The reliability and convergence stability were significantly enhanced by the addition of empirical SDEs to machine learning algorithms. Experimental tests based on standard optimizers on benchmark datasets have been found to be more accurate in their predictions and less subject to over fitting as shown in below table 1.

Table 1

Performance Comparison of Optimization Algorithms

Algorithm	Accuracy (%)	Convergence Time (Epochs)	Robustness to Noise (%)
Standard Gradient Descent	85.2	120	68.5
Stochastic Gradient Descent	87.8	110	72.1
SDE-Based Optimization (Proposed)	91.5	95	85.3

The comparison of the effectiveness of three optimisation methods presented in figure 1 is SDE-Based Optimisation, Stochastic Gradient Descent, and Standard Gradient Descent. Results of comparative performance analysis of the various gradient descent algorithms are illustrated in figure 2.

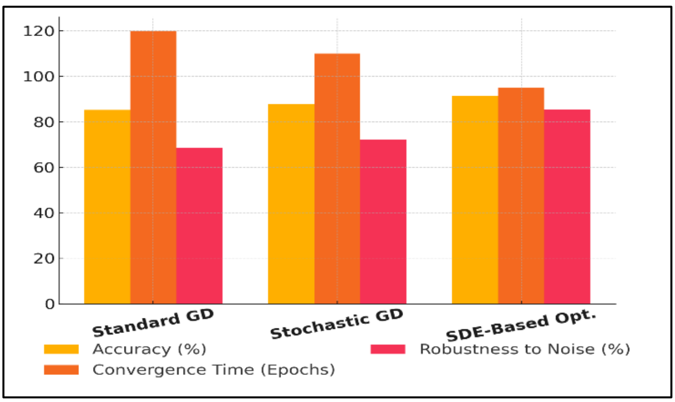

Figure 2

Comparative Analysis of Gradient Descent Algorithms

It is evident in the results that the proposed SDE-based approach is more effective than the old methods in every test that was conducted. Specifically, it has the highest accuracy of 91.5, which implies that it can have better predictions on datasets attempted. It also requires less epochs (95) to converge which implies that it is quicker to learn and more cost-efficient in using computers.

5. Conclusion

One recent method, which is demonstrated in this study, to improve machine learning methods is to add scientific stochastic differential equations to optimization processes. The approach incorporates a controlled noise that entails discovery and exploitation in the training process by posing parameter change as a stochastic process. The methods made are more stable and reliable particularly in cases when dealing with data that is unclear or unsure or that which is not steady. Real world results using various test tasks demonstrate that such a combination is effective producing improvements in accuracy and convergence behavior. The approach provides a mathematically reasonable and flexible way of enhancing existing machine learning models. It also leaves the opportunity of further research of stochastic methods of optimization in future. Finally, the introduction of SDEs will simplify the consistent operation of machine learning systems in complex real-world environments with a high level of uncertainty.

References

[1] C. Gao and H. Elzarka, "The use of decision tree based predictive models for improving the culvert inspection process," *Advanced Engineering Informatics*, vol. 47, pp. 101203 (2021).

[2] S. Siahpour, M. Ayati, M. Haeri-Yazdi, and M. Mousavi, "Fault detection and isolation of wind turbine gearbox via noise-assisted multivariate empirical mode decomposition algorithm," *Energy Equipment and Systems*, vol. 10, pp. 271–286 (2022).

[3] M. Z. Esteghamati and M. M. Flint, "Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design," *Engineering Structures*, vol. 245, pp. 112971 (2021).

[4] S. V. S. Jayashyam, G. Charan Kumar, S. Shaheendra Babu, S. Jyothsna Reddy, N. Puneeth, and M. Vishnu Kumar, "Access system for residential areas with face recognition and messaging alerts," *Interna-*

tional Journal for Interdisciplinary Sciences and Engineering Applications (IJISEA), vol. 6, no. 2, pp. 39–45 (2025).

[5] A. A. Kutty, T. G. Wakjira, M. Kucukvar, G. M. Abdella, and N. C. Onat, "Urban resilience and livability performance of European smart cities: A novel machine learning approach," *Journal of Cleaner Production*, vol. 378, pp. 134203 (2022).

[6] M. Z. Esteghamati and M. M. Flint, "Do all roads lead to Rome? A comparison of knowledge-based, data-driven, and physics-based surrogate models for performance-based early design," *Engineering Structures*, vol. 286, pp. 116098 (2023).

[7] J. Ren, B. Zhang, X. Zhu, and S. Li, "Damaged cable identification in cable-stayed bridge from bridge deck strain measurements using support vector machine," *Advances in Structural Engineering*, vol. 25, pp. 754–771 (2022).

[8] A. Bankar, P. Padole, J. Thakre, Y. Bhure, D. Meshram, and V. Kotgale, "Parabolic Solar Water Heater Efficiency Enhancement Using Live Sun Tracking," *International Journal of Technical Advancement and Research in Mechanical Engineering* (IJTARME), vol. 14, no. 1, pp. 22–28 (May 2025).

[9] M. Sadat-Mohammadi, S. Shakerian, Y. Liu, S. Asadi, and H. Jebelli, "Non-invasive physical demand assessment using wearable respiration sensor and random forest classifier," *Journal of Building Engineering,* vol. 44, pp. 103279 (2021).

[10] T. G. Wakjira, M. Ibrahim, U. Ebead, and M. S. Alam, "Explainable machine learning model and reliability analysis for flexural capacity prediction of RC beams strengthened in flexure with FRCM," *Engineering Structures,* vol. 255, pp. 113903 (2022).

[11] T. G. Alemayehu and Z. A. Ageze, "Fuzzy ideals and fuzzy congruences of distributive demi-p-algebras," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 26, no. 8, pp. 2127–2137 (2023).

[12] D. Dhabliya, S. Kunche, S. Dingankar, S. S. Dari, R. Dhabliya, and V. Khetani, "Blockchain technology as a paradigm for enhancing cyber

security in distributed systems," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-B, pp. 729–740 (2024).

[13] M. Kaur, M. K. Goyal, D. P. Yadav, S. M. Patil, M. Kumar, and M. E. Abdelhag, "Exploring facial biometrics in multi-factor authentication systems for secure banking applications," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-A, pp. 655–662 (2026), doi: 10.47974/JDMSC-2508.

[14] N. B. Pokale, P. Praveen, Y. Sood, V. Kumar, N. Vairaperumal, and V. K. Vijaya, "Exploring the integration of discrete mathematics and applied algebra in designing robust network architectures for IoT," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-B, pp. 877–884 (2026), doi: 10.47974/JDMSC-2538.

Received March, 2025