

# Lie Symmetry Net: Preserving Conservation Laws in Modelling Financial Market Dynamics via Differential Equations

Xuelian Jiang<sup>a</sup>, Tongtian Zhu<sup>b</sup>, Can Wang<sup>b</sup>, Yingxiang Xu<sup>a,\*</sup> and Fengxiang He<sup>c</sup>

<sup>a</sup>*School of Mathematics and Statistics, Northeast Normal University, Changchun, China*

<sup>b</sup>*College of Computer Science, Zhejiang University, Hangzhou, China*

<sup>c</sup>*School of Informatics, University of Edinburgh, Edinburgh, Scotland*

## ARTICLE INFO

### Keywords:

Physics-Informed Neural Networks

Lie Symmetry

Conservation Laws

Financial Market Dynamics

Black-Scholes Equation

Vařiček Equation

## ABSTRACT

This paper employs a novel Lie symmetry-based framework to model the intrinsic symmetries within financial market. Specifically, we introduce *Lie symmetry net* (LSN), which characterises the Lie symmetry of the differential equations (DE) estimating financial market dynamics, such as the Black-Scholes equation and the Vařiček equation. To simulate these differential equations in a symmetry-aware manner, LSN incorporates a Lie symmetry risk derived from the conservation laws associated with the Lie symmetry operators of the target differential equations. This risk measures how well the Lie symmetry is realised and guides the training of LSN under the structural risk minimisation framework. Extensive numerical experiments demonstrate that LSN effectively realises the Lie symmetry and achieves an error reduction of more than *one order of magnitude* compared to state-of-the-art methods. The code is available at this [URL](#).

## 1. Introduction

A classic approach for modelling financial market dynamics is via stochastic differential equations (SDEs). Through the application of the Feynman-Kac formula [1], these SDEs can be transformed into corresponding partial differential equations (PDEs), such as the Black-Scholes equation [2, 3, 4, 5] and the Vařiček equation [6]. such as the Black-Scholes equation [2, 3, 4, 5] and the Vařiček equation [6]. Traditionally, numerical methods such as finite volume methods [7] and B-spline interpolation methods [8, 9] are used to simulate these equations. In recent years, an emerging solution of solving these differential equations involves using AI-driven methods to fit their dynamics from sampled data, exemplified by Physics-Informed neural networks (PINNs) [10].

A defining characteristic of SDEs is their “symmetry”. A major family of mathematical tools to characterise the symmetry are Lie symmetry groups [11, 12, 13]. In traditional numerical methods, symmetry is crucial for solving these SDEs [14, 15]. The symmetry also facilitates solving both the Black-Scholes equation [16] and the Vařiček equation [17]. Our vision is that the Lie symmetry can represent some intrinsic symmetry in financial markets, though in an abstract manner. This abstract symmetry may shed light on discovering “new economics” that is not yet well understood.

However, this symmetry is largely untouched in existing AI-driven DE solvers. Without taking the symmetry into account, an AI-driven approach could learn an asymmetric solution that probably fits the training data, but unfortunately, mathematically wrong. This is caused by the imbalance or other limitations in the training data. When the learned solver is applied to unseen data, the performance is unsecured, as reported in a large volume of literature [18, 19].

This paper endeavours to answer the following fundamental question:


*Could Lie symmetry facilitate AI-driven DE solvers in simulating financial market dynamics, and how?*

Motivated by this question, we design Lie symmetry net (LSN), which enables the simulation of financial market dynamics while preserving Lie symmetry.

Similar to many symmetries in physics, the Lie symmetry can be transformed into conservation laws [11, 17, 20, 21]. Specifically, for the Black-Scholes and Vařiček equations, the conservation law derived from Lie symmetry is

$$D_t T^t + D_x T^x = 0,$$

\*Corresponding author: Yingxiang Xu

 [jiangxl133@nenu.edu.cn](mailto:jiangxl133@nenu.edu.cn) (X. Jiang); [raidon@zju.edu.cn](mailto:raidon@zju.edu.cn) (T. Zhu); [wcan@zju.edu.cn](mailto:wcan@zju.edu.cn) (C. Wang); [yxxu@nenu.edu.cn](mailto:yxxu@nenu.edu.cn) (Y. Xu); [F.He@ed.ac.uk](mailto:F.He@ed.ac.uk) (F. He)

where  $D_t$  represents the partial derivative with respect to time  $t$  or asset price  $x$ , and  $(T^t, T^x)$  represents the conservation vector subject to the symmetry condition (*i.e.*, Lie point symmetry operator)  $G$ , such that the action of  $G$  on the conservation vector satisfies  $G(T^t, T^x) = 0$  [22, 11].

In our LSN, we design a novel *Lie conservation residual* to quantify how well the Lie symmetry is realised on one specific point in the data space that comprises asset price and time. This Lie conservation residual then induces a *Lie symmetry risk* that aggregating the residual over the data space, and thus characterises how Lie symmetry is realised from a global view. It is worth noting that this Lie symmetry risk depends on the specific conservation law, and thus the specific Lie symmetry operator. This Lie symmetry risk is then integrated with risk functions measuring how well the LSN fits the sampled data [10, 23], and formulates the structural risk of LSN. We can optimise the LSN under the structural risk minimisation (SRM) framework [24] to learn an DE solver while preserving the Lie symmetries.

Extensive numerical experiments are conducted to verify the superiority of LSN. We compare LSN with state-of-the-art methods including IPINNs [25], sfPINNs [26], ffPINNs [26] and LPS [27]. The results demonstrate that LSN consistently outperforms these methods, achieving error reductions of more than an order of magnitude. Specifically, the error magnitude with single operator reaches  $10^{-3}$ , while with combined operators, it further decreases to  $10^{-4}$ .

The paper is structured as follows. Section 2 provides an overview of related work. Section 3 discusses the background of PINNs and SDEs. Section 4 introduces the methodology of LSN. Section 5 presents numerical experiments to validate the effectiveness of LSN. Finally, Section 6 draws conclusions and outlines directions for future research. Appendix A provides additional background and the theoretical analyses of LSN.

## 2. Related works

**Numerical equation solvers.** Numerical methods have long been essential for solving partial differential equations in various domains, including financial market modeling. Significant progress has been made in this area with models such as the Black-Scholes equation and the Vařiček equation. Traditional approaches, including finite volume methods [7] and B-spline interpolation methods [8, 9], have been widely applied to solve these equations [28, 16]. These grid-based techniques rely on discretizing the spatial and temporal domains, transforming the continuous equations into discrete problems suitable for simulation. However, these methods often come with high computational complexity, which may limit their applicability.

**Neural equation solvers.** In recent years, there has been a gradual increase in applying neural networks to solve differential equations. Two main approaches have emerged in this area. The first one, neural operator methods [29], focuses on learning the mapping between the input and output functions of the target equations. In contrast, the second approach, Physics-Informed Neural Networks (PINNs) [10], directly approximate the solution of the equations. PINNs and their variants, such as sfPINNs [26] and ffPINNs [26], have gained popularity for utilizing physical laws into the training process. Recent studies have successfully applied PINNs to solve financial equations, introducing efficient methods like IPINNs [25], which incorporates regularization terms for slope recovery. A more recent work by Akhound-Sadegh et al. [27] proposes to incorporate Lie symmetries into PINNs by minimizing the residual of the determining equations of Lie symmetries. While this approach offers an interesting direction, our LSN follows a different methodology. Specifically, their Lie point symmetry (LPS) method focuses on minimizing these symmetry residuals, whereas our LSN realises Lie symmetries by preserving the conservation laws derived from the Lie symmetry operators. These conservation laws are fundamental principles inherent to the system described by the differential equations. Additionally, LPS has been validated only on the Poisson and Burgers equations in their original paper and its effectiveness in leveraging inherent symmetries in financial markets remains unclear. In contrast, our comprehensive comparative experiments in financial domain, specifically on the Black-Scholes equation across various parameters, clearly demonstrate the superiority of LSN over LPS by reducing testing error by an order of magnitude.

## 3. Preliminaries

This section provides the essential background knowledge. We begin with an introduction to Physics-Informed Neural Networks. We then cover stochastic differential equations and explain how the Feynman-Kac formula allows for the transformation of a SDE into a corresponding PDE. To concretize this theoretical framework, we provide illustrative examples including the Black-Scholes equation and the Vařiček equation. For additional terminology related to finance and Lie symmetry, please refer to Appendix A.1.

### 3.1. Physics-Informed Neural Networks (PINNs)

PINNs solve partial differential equations by directly learning the solution. The use of PINNs to solve differential equations typically begins with generating a dataset  $S = (x^n, t^n)_{n=1}^N$  by randomly sampling points within the solution domain. To understand how PINNs operate, consider the following general form of partial differential equation,

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \mathcal{L}[u] & \text{for all } (x, t) \in \Omega \times [0, T], \\ u(x, 0) = \varphi(x) & \text{for all } x \in \Omega, \\ u(y, t) = \psi(y, t) & \text{for all } (y, t) \in \partial\Omega \times [0, T], \end{cases} \quad (1)$$

where  $\mathcal{L}[u]$  is a differential operator,  $\Omega$  is a bounded domain,  $\varphi(x)$  and  $\psi(y, t)$  are the initial and boundary conditions, respectively,  $T$  denotes the terminal time, and  $u(x, t)$  is the function to be solved. To solve Equation (1), PINNs model  $u$  as a neural network  $\hat{u}$  to approximate the exact solution by minimizing

$$\mathcal{L}_{PINNs} = \mathcal{L}_{PDE} + \mathcal{L}_{BC} + \mathcal{L}_{IC}, \quad (2)$$

where

$$\mathcal{L}_{PDE} = \left\| \frac{\partial \hat{u}(x, t)}{\partial t} - \mathcal{L}[\hat{u}](x, t) \right\|_{\Omega \times [0, T]}^2, \quad \mathcal{L}_{BC} = \|\hat{u}(y, t) - \psi(y, t)\|_{\partial\Omega \times [0, T]}^2, \quad \mathcal{L}_{IC} = \|\hat{u}(0, x) - \varphi(x)\|_{\Omega}^2. \quad (3)$$

The first term  $\mathcal{L}_{PDE}$  measures the residual of the PDE, while  $\mathcal{L}_{BC}$  and  $\mathcal{L}_{IC}$  quantify the errors in satisfying the boundary and initial conditions, respectively.

### 3.2. Stochastic Differential Equation (SDE)

SDE provide a mathematical framework for modeling systems influenced by random disturbances. To understand the dynamics of a stochastic process  $X_t$ , we consider the SDE of the following general form

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \quad (4)$$

where  $X_t$  represents the stochastic variable of interest and  $W_t$  is a standard Wiener process (as defined in Definition A.1). The functions  $\mu(X_t, t)$  and  $\sigma(X_t, t)$ , known as the drift and diffusion coefficients, respectively, are functions that characterise the deterministic and stochastic components of the dynamics.

**Feynman-Kac formula** [1]. The Feynman-Kac formula provides a critical theoretical framework to establish a connection between certain types of PDEs and SDEs. Given a payoff function  $f(x, t)$  and defined a discounting function  $r(x, t)$  to calculate the present value of future payoffs, if  $u(x, t)$  is a solution to the PDE:

$$\frac{\partial u}{\partial t} + \mu(x, t) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 u}{\partial x^2} - r(x, t)u = 0, \quad (5)$$

with the terminal condition  $u(x, T) = f(x)$ , then the solution  $u(x, T)$  to this PDE can be represented as:

$$E \left[ e^{-\int_t^T r(X_s, s)ds} f(X_T) \mid X_t = x \right], \quad (6)$$

where  $X_T$  denotes the value of Equation (4) at time  $T$ .

To illustrate the application of the Feynman-Kac formula, we present several canonical examples from finance.

**Example 1** (Black-Scholes equation [30]). Considering a frictionless and arbitrage-free financial market comprising a risk-free asset and a unit risky asset, the dynamics of the market can be modeled by the following SDE:

$$dx_t = rx_t dt + \sigma x_t dW_t, \quad (7)$$

where  $x$  denotes the price of a unit risky asset,  $t$  represent time,  $\sigma$  is the volatility,  $r$  is the risk-free interest rate and  $W_t$  is a standard Wiener process (refer to Definition A.1). Applying the feynman-Kac formula, the Black-Scholes equation for evaluating the price  $u(x, t)$  of a European call option (refer to Definition A.2) is derived as follows:

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 u}{\partial x^2} + rx \frac{\partial u}{\partial x} - ru = 0 & \Omega \times [0, T], \\ u(x, T) = \max(x - K, 0) & \Omega \times T, \\ u(0, t) = 0 & \partial\Omega \times [0, T], \end{cases} \quad (8)$$

where  $K$  is the strike price,  $T$  is the expiry time of the contract and  $\Omega$  is a bounded domain. The solution to this equation can be written as follows [25]:

$$u(x, t) = x\mathcal{N}(d_1) - K \exp^{-r(T-t)} \mathcal{N}(d_2), \quad d_1 = \frac{\ln(x/K) + (r + 0.5\sigma^2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2 = d_1 - \sigma\sqrt{T - t},$$

where  $\mathcal{N}$  denotes standard normal distribution.

**Example 2** (Vašíček Equation [6]). In a financial market characterised by short-term lending transactions between financial institutions, the evolution of short-term interest rates can be modeled by the following SDE:

$$dx_t = \lambda(\beta - x_t)dt + \sigma dW_t,$$

where  $\lambda, \beta > 0$ ,  $\sigma$  are constants and  $W_t$  is the Wiener process. Using the Feynman-Kac formula we can obtain the Vašíček pricing equation which is used to price risk-free bonds  $u(x, t)$ :

$$\begin{cases} \frac{\partial u}{\partial t} + \alpha \frac{\partial u}{\partial x^2} + \lambda(\beta - x) \frac{\partial u}{\partial x} + \gamma x u = 0 & \Omega \times [0, T], \\ u(x, T) = 1 & \Omega \times T, \\ u(x, t) = \psi(x, t) & \partial\Omega \times [0, T]. \end{cases} \quad (9)$$

where  $\alpha = \frac{1}{2}\sigma^2$ ,  $\gamma = -1$  and  $\psi(x, t)$  is the boundary conditions. The zero-coupon bond price in the Vašíček pricing model is given by [17]:

$$u(x, t) = e^{A(T-t) + xC(T-t)}, \quad (10)$$

where  $C(t) = -\frac{1}{\lambda}(1 - e^{-\lambda t})$  and  $A(t) = \frac{4\lambda^2\beta - 3\sigma^2}{4\lambda^3} + \frac{\sigma^2 - 2\lambda^2\beta}{2\lambda^2}t + \frac{\sigma^2 - \lambda^2\beta}{\lambda^3}e^{-\lambda t} - \frac{\sigma^2}{4\lambda^3}e^{-2\lambda t}$ .

### 3.3. Lie Group Analysis

Groups, which mathematically characterize symmetries, describe transformations that preserve certain invariances. Formally, a group  $(G, \cdot)$  is defined as a set  $G$  equipped with a binary operation  $\cdot$  that satisfies the properties of associativity, contains an identity element  $e \in G$ , and ensures the existence of an inverse element  $g^{-1}$  for each  $g \in G$  [31]. When groups are also differentiable manifolds, they are referred to as Lie groups, which are crucial in analyzing continuous symmetries [32]. Lie group analysis provides a powerful tool for studying symmetry, conservation laws, and dynamic systems of equations [12, 13]. The goal of Lie group analysis is to identify the symmetries of an equation, especially those transformations under Lie group actions that leave the equation invariant. Revealing these symmetries can lead to deriving conservation laws [11], simplifying the solution process, and reducing computational complexity [16, 17].

## 4. Lie Symmetry Net

In this section, we introduce Lie symmetry net (LSN). In Section 4.1, we briefly derive the corresponding conservation law from the Lie symmetry operators of the target equations, which in turn lead to the Lie symmetry risk of LSN. In Section 4.2, we discuss the structure risk minimization of LSN based on the Lie symmetry risk.

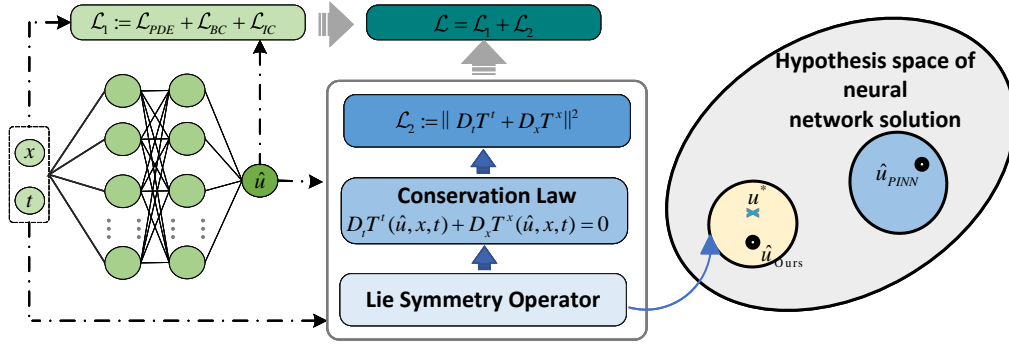
### 4.1. Lie symmetry in equations

This subsection presents the Lie symmetry operators for the Black-Scholes equation and Vašíček equation, and derive the corresponding conservation laws, which allows us to define the associated Lie symmetry risk.

**Lie symmetry operator.** Lie symmetry operator is a major mathematical tool for characterizing the symmetry in PDEs (see Definition A.3) [12]. Below are the Lie symmetry operators for the Black-Scholes and Vašíček equations.

*Black-Scholes equation.* The Lie symmetry operators [13, 11] of BS Equation (8) are given by the vector field

$$\begin{aligned} G_\phi &= \phi(t, x) \frac{\partial}{\partial u}, \quad G_1 = \frac{\partial}{\partial t}, \quad G_2 = x \frac{\partial}{\partial x}, \quad G_3 = u \frac{\partial}{\partial u}, \\ G_4 &= 2t \frac{\partial}{\partial t} + (\ln x + Zt)x \frac{\partial}{\partial x} + 2rtu \frac{\partial}{\partial u}, \quad G_5 = \sigma^2 tx \frac{\partial}{\partial x} + (\ln x - Zt)u \frac{\partial}{\partial u}, \\ G_6 &= 2\sigma^2 t^2 \frac{\partial}{\partial t} + 2\sigma^2 tx \ln x \frac{\partial}{\partial x} + ((\ln x - Zt)^2 + 2\sigma^2 rt^2 - \sigma^2 t)u \frac{\partial}{\partial u}, \end{aligned} \quad (11)$$



**Figure 1:** Schematic Diagram of Lie Symmetry Network Architecture. **Left:** The incorporation of a Lie symmetry block into a PINN architecture. **Right:** The hypothesis space of PINN, denoted as  $\hat{u}_{PINN}$  (blue circle), is a subset of the broader neural network solution space (gray circle).  $u^*$  represents the exact solution. The yellow circle signifies the hypothesis space that satisfies symmetry conditions. The incorporation of Lie symmetry blocks produces our solution  $\hat{u}_{ours}$ . In simpler terms, being closer to the space where the exact solution exists leads to fewer errors. The non-intersection between " $\hat{u}_{PINN}$ " and " $\hat{u}_{ours}$ " in the figure is a specific case provided to enhance clarity in our exposition.

where  $Z = r - \sigma^2/2$ ,  $\phi(t, x)$  is an arbitrary solution to Equation (8) without any boundary condition or initial condition. The first symmetry  $G_\phi$  is an infinite-dimensional symmetry, arising as a consequence of linearity. These Lie symmetry operators span an infinite-dimensional Lie group vector space [11].

*Vařiček equation.* The Lie symmetry operators [17] of Vařiček Equation (9) are given by the vector field

$$\begin{aligned} G_\phi &= \phi(t, x) \frac{\partial}{\partial u}, G_1 = \frac{\partial}{\partial t}, \\ G_2 &= e^{2\lambda t} \frac{\partial}{\partial t} + \frac{e^{2\lambda t}}{\lambda} (\lambda^2 x - 2\alpha\gamma - \beta\lambda^2) \frac{\partial}{\partial x} + \frac{ue^{2\lambda t}}{\alpha\lambda^2} (\alpha^2\gamma^2 + 2\alpha\beta\gamma\lambda^2 - \alpha\lambda^3 - 3\alpha\gamma\lambda^2 x + \lambda^4(\beta - x)^2) \frac{\partial}{\partial u}, \\ G_3 &= e^{-2\lambda t} \left[ -\frac{\partial}{\partial t} + \frac{1}{\lambda} (\lambda^2(x - \beta) - 2\alpha\gamma) \frac{\partial}{\partial x} + \frac{\gamma u}{\lambda^2} (\lambda^2 x - \alpha\gamma) \frac{\partial}{\partial u} \right], \\ G_4 &= e^{\lambda t} \left[ \frac{\partial}{\partial x} + \frac{u}{\alpha\lambda} (-\alpha\gamma - \beta\lambda^2 + \lambda^2 x) \frac{\partial}{\partial u} \right], G_5 = e^{-\lambda t} \left[ \frac{\partial}{\partial x} + \frac{\gamma u}{\lambda} \frac{\partial}{\partial u} \right], G_6 = u \frac{\partial}{\partial u}. \end{aligned} \quad (12)$$

These Lie symmetry operators not only provide a deeper insight into the structure of the PDEs but also form the foundation for deriving conservation laws associated with these equations.

**Conservation law.** Similar to many symmetries in physics, the Lie symmetry can be transformed to conservation laws [11, 17, 20, 21]. In this paper, we interpret the Lie symmetry point operators as the following conservation laws: regardless of how the space  $x$ , time  $t$ , and exact solution  $u$  vary, the conservation vector  $(T^t, T^x)$  corresponding to the Lie point symmetry remains zero, *i.e.*,

$$D_t T^t(u, x, t) + D_x T^x(u, x, t) = 0, \quad (13)$$

where the  $D$  represents the partial derivative with respect to time  $t$  or space  $x$ , and  $(T^t, T^x)$  represents the conservation vector subject to the symmetry condition (*i.e.*, Lie point symmetry operator)  $G$ , such that the action of  $G$  on the conservation vector satisfies  $G(T^t, T^x) = 0$  [11]. It is worth noting that this Lie symmetry risk depends on the specific conservation law. Next, we present the conserved quantities for the BS and the Vařiček equation as examples.

*Black-Scholes equation.* We can derive the conservation law of the operator  $G_2$  (see Equation (11)) of BS equation as follows [11]:

$$\begin{cases} T_2^t(u, x, t) = -\frac{\partial u}{\partial x} l(t) + \frac{\mathcal{A}}{x} + \frac{2\mathcal{B}u}{\sigma^2 x} e^{-rt}, \\ T_2^x(u, x, t) = \frac{\partial u}{\partial t} l(t) + u \frac{\partial l(t)}{\partial t} + g(t) - \mathcal{B}u e^{-rt} + \mathcal{B} \left( \frac{\partial u}{\partial x} + \frac{2ru}{\sigma^2 x} \right) x e^{-rt}, \end{cases} \quad (14)$$

where  $\mathcal{A}$  and  $\mathcal{B}$  are arbitrary constants, and  $l(t)$  and  $g(t)$  are arbitrary functions with respect to  $t$ . Unless stated otherwise, consider  $\mathcal{A} = \mathcal{B} = 1$ ,  $l(t) = t$ , and  $g(t) = t^2$ .

*Vašiček equation.* To ascertain the Lie conservation law operators for the Vašiček Equation (9), it is necessary to analyze its adjoint equation, as follows

$$\frac{\partial v}{\partial t} - \alpha \frac{\partial v}{\partial x^2} - \lambda(x - \beta) \frac{\partial v}{\partial x} - (\lambda + \gamma x)v = 0. \quad (15)$$

where  $v \neq 0$  is a new dependent variable  $v = e^{pt+qx}$  with  $p = \alpha q^2 - \lambda\beta q + \lambda$  and  $q = -\frac{\gamma}{\lambda}$  (Here, only one example is presented for illustration purposes, although there exist numerous solutions to this set of adjoint equation). For illustrative purposes, we choose the relatively simple operator  $G_5$  and  $G_6$  as examples of the Vašiček Equation (12) and provide the corresponding conserved quantities [17]:

$$\left\{ \begin{array}{l} T_5^t(u, x, t) = \frac{1}{\gamma\lambda} e^{-\lambda t} v \left( \lambda \frac{\partial u}{\partial x} - \gamma u \right), \\ T_5^x(u, x, t) = \frac{1}{\gamma\lambda} e^{-\lambda t} \left\{ \gamma u \left( \alpha \frac{\partial v}{\partial x} - \beta \lambda v \right) - \alpha \frac{\partial u}{\partial x} \left( \gamma v + \lambda \frac{\partial v}{\partial x} \right) - \lambda \frac{\partial u}{\partial t} v \right\}; \end{array} \right. \left\{ \begin{array}{l} T_6^t(u, x, t) = uv, \\ T_6^x(u, x, t) = \alpha \frac{\partial u}{\partial x} v - u \left\{ \lambda(x - \beta)v + \alpha \frac{\partial v}{\partial x} \right\}. \end{array} \right.$$

We can then define the Lie conservation residual  $\mathcal{R}_{Lie}$  according to Equation (13) to evaluate the extent to which the Lie symmetry is realised at a specific point in the data space.

**Definition 4.1** (Lie conservation residual). Combining the Lie symmetry operator and the conservation law, we define the Lie conservation residual of  $\hat{u}$  as follows:

$$\mathcal{R}_{Lie}[\hat{u}] = D_t T^t(\hat{u}) + D_x T^x(\hat{u}), \quad (16)$$

where the  $D$  represents the partial derivative with respect to  $t$  or  $x$ , and  $(T^t, T^x)$  represents the conservation vector subject to the symmetry condition  $G$ .

We can aggregate the Lie symmetry residuals over the entire data space to obtain the Lie symmetry risk, which characterises the degree to which the Lie symmetry is realised from a global perspective.

**Definition 4.2** (Lie symmetry risk). According to the expression in Equation (16), the definition of Lie symmetry risk is provided as follows:

$$\mathcal{L}_{Lie}[\hat{u}_\theta](x, t) = \int_{\Omega \times [0, T]} |\mathcal{R}_{Lie}[\hat{u}_\theta](x, t)|^2 dx dt,$$

where  $\hat{u}_\theta$  denotes the network output, and  $\theta$  denotes the network parameters.

*Remark 4.3.* The Lie symmetry risk  $\mathcal{L}_{Lie}$  focuses solely on learning the symmetry of the problem without taking into account the underlying physical laws of the problem.

This Lie symmetry risk is defined over the data distribution, which is unknown in practice. We thus resort to defining the *Empirical Lie symmetry risk*  $\hat{\mathcal{L}}_{Lie}$  as an approximation of the Lie symmetry risk  $\mathcal{L}_{Lie}$ ,

**Definition 4.4** (Empirical Lie symmetry risk). Summing up the Lie symmetry operators at  $N_i$  discrete points provides an approximation to the Lie symmetry risk.

$$\hat{\mathcal{L}}_{Lie}(\theta, S) := \frac{1}{N_i} \sum_{n=1}^{N_i} |\mathcal{R}_{Lie}[\hat{u}_\theta](x_i^n, t_i^n)|^2, \quad (17)$$

where  $S = \{(x_i^n, t_i^n)\}_{n=1}^{N_i}$  represents the set of these  $N_i$  discrete points.

## 4.2. Structure Risk Minimisation

In this subsection, we present the structure risk minimization of LSN based on the Lie symmetry risk. For simplicity, we rewrite the above Black-Scholes Equation (8) and the Vařiček Equation (9) as the Equation (1) (detail reference Appendix A.2), where

$$\mathcal{L}[u] := \frac{1}{2}\sigma(x)^2 \frac{\partial^2 u(x, t)}{\partial x^2} + \mu(x) \frac{\partial u(x, t)}{\partial x} + v(x)u(x, t),$$

is a differential operator with respect to three bounded affine functions  $\sigma(x)$ ,  $\mu(x)$  and  $v(x)$  (for BS equation:  $\sigma(x) = \sigma x$ ,  $\mu(x) = rx$ ,  $v(x) = -r$  and  $\varphi(x) = \max(x - K, 0)$ ; for Vařiček equation  $\sigma(x) = \sqrt{2\alpha} = \sigma$ ,  $\mu(x) = \lambda(\beta - x) = -x$ ,  $v(x) = \gamma x$  and  $\varphi(x) = 1$ ).

**Data fitting residuals.** The following functions  $\mathcal{R}_j$  ( $j = \{i, s, t\}$ ) characterise how well the LSN is fitting the sampled data according to Equation (1), for  $\forall \hat{u} \in C^2(\mathbb{R}^d)$

$$\begin{aligned} \mathcal{R}_i[\hat{u}](x, t) &= \frac{\partial \hat{u}(x, t)}{\partial t} - \mathcal{L}[\hat{u}](x, t) & (x, t) \in \Omega \times [0, T], \\ \mathcal{R}_s[\hat{u}](y, t) &= \hat{u}(y, t) - \psi(y, t) & (y, t) \in \partial\Omega \times [0, T], \\ \mathcal{R}_t[\hat{u}](x) &= \hat{u}(0, x) - \varphi(x) & x \in \Omega. \end{aligned} \quad (18)$$

We then define the population risk  $\mathcal{L}_1$  for fitting the sampled data, based on the aforementioned “residuals” as below,

$$\mathcal{L}_1[\hat{u}_\theta](x, t) = \mathcal{L}_{PDE}[\hat{u}_\theta](x, t) + \mathcal{L}_{BC}[\hat{u}_\theta](x, t) + \mathcal{L}_{IC}[\hat{u}_\theta](x, t), \quad (19)$$

where

$$\begin{aligned} \mathcal{L}_{PDE}[\hat{u}_\theta](x, t) &= \int_{\Omega \times [0, T]} |\mathcal{R}_i[\hat{u}_\theta](x, t)|^2 dx dt, \\ \mathcal{L}_{BC}[\hat{u}_\theta](x, t) &= \int_{\partial\Omega \times [0, T]} |\mathcal{R}_s[\hat{u}_\theta](x, t)|^2 dx dt, \\ \mathcal{L}_{IC}[\hat{u}_\theta](x, 0) &= \int_{\Omega} |\mathcal{R}_t[\hat{u}_\theta](x)|^2 dx. \end{aligned}$$

Similarly, this population risk is defined over the data distribution, which is unknown in practice. We thus resort to defining the empirical loss function  $\hat{\mathcal{L}}_1$  to approximate the population risk as follows,

$$\hat{\mathcal{L}}_1[\hat{u}_\theta](x, t) = \hat{\mathcal{L}}_{PDE}[\hat{u}_\theta](x, t) + \hat{\mathcal{L}}_{BC}[\hat{u}_\theta](x, t) + \hat{\mathcal{L}}_{IC}[\hat{u}_\theta](x, t),$$

where

$$\begin{aligned} \hat{\mathcal{L}}_{PDE}(\theta, S) &:= \frac{1}{N_i} \sum_{n=1}^{N_i} |\mathcal{R}_i[\hat{u}_\theta](x_i^n, t_i^n)|^2, \\ \hat{\mathcal{L}}_{BC}(\theta, S) &:= \frac{1}{N_s} \sum_{n=1}^{N_s} |\mathcal{R}_s[\hat{u}_\theta](x_s^n, t_s^n)|^2, \\ \hat{\mathcal{L}}_{IC}(\theta, S) &:= \frac{1}{N_t} \sum_{n=1}^{N_t} |\mathcal{R}_t[\hat{u}_\theta](x_t^n)|^2. \end{aligned} \quad (20)$$

Here  $S = \left\{ \{(x_i^n, t_i^n)\}_n^{N_i}, \{(x_s^n, t_s^n)\}_n^{N_s}, \{x_t^n\}_n^{N_t} \right\}$  is the dataset by Gaussian sampling within the domain  $\Omega \times [0, T]$ .

*Remark 4.5.*  $\mathcal{L}_1$  essentially corresponds to the Physics-Informed Neural Networks (PINNs) (please refer to Section 3), which accurately estimates the majority of the training data based on the inherent physical laws of the problem. However, it does not explicitly consider the symmetry within.

**Structural risk of LSN.** Eventually, we may define the structure risk  $\mathcal{E}(\theta)$  of LSN as follows.



**Definition 4.6** (Structure risk of LSN). The structure risk of LSN is defined as follows

$$\begin{aligned}\mathcal{E}(\theta) &:= \lambda_1 \mathcal{L}_{PDE}[\hat{u}_\theta](x, t) + \lambda_2 \mathcal{L}_{BC}[\hat{u}_\theta](x, t) + \lambda_3 \mathcal{L}_{IC}[\hat{u}_\theta](x, 0) + \lambda_4 \mathcal{L}_{Lie}[\hat{u}_\theta](x, t) \\ &= \lambda \mathcal{L}_1[\hat{u}_\theta](x, t) + \lambda_4 \mathcal{L}_2[\hat{u}_\theta](x, t).\end{aligned}\quad (21)$$

Here,  $\lambda \mathcal{L}_1 := \lambda_1 \mathcal{L}_{PDE} + \lambda_2 \mathcal{L}_{BC} + \lambda_3 \mathcal{L}_{IC}$  is defined as in Equation (19),  $\mathcal{L}_2 := \mathcal{L}_{Lie}$  in Definition 4.2, and  $\lambda_i$  ( $i = 1, \dots, 4$ ) are the hyperparameters.

*Remark 4.7.* The structural loss of LSN synergistically integrates  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , aiming to learn both the inherent physical laws of the problem and its symmetry, ensuring a comprehensive understanding within the framework of the problem.

Correspondingly, we provide an empirical approximation of the structural risk of LSN as follows.

**Definition 4.8** (Empirical risk of LSN). The empirical loss of LSN is defined as follows

$$\hat{\mathcal{E}}(\theta, S) := \lambda_1 \hat{\mathcal{L}}_{PDE}(\theta, S_t) + \lambda_2 \hat{\mathcal{L}}_{BC}(\theta, S_s) + \lambda_3 \hat{\mathcal{L}}_{IC}(\theta, S_t) + \lambda_4 \hat{\mathcal{L}}_{Lie}(\theta, S_t), \quad (22)$$

where  $S = \left\{ S_i \{(x_i^n, t_i^n)\}_n^{N_i}, S_s \{(x_s^n, t_s^n)\}_n^{N_s}, S_t \{(x_t^n)\}_n^{N_t} \right\}$  are the training data sets.

We train LSN by solving the following minimisation problem,

$$\theta^* = \arg \min_{\theta} \lambda_1 \hat{\mathcal{L}}_{PDE}(\theta, S_t) + \lambda_2 \hat{\mathcal{L}}_{BC}(\theta, S_s) + \lambda_3 \hat{\mathcal{L}}_{IC}(\theta, S_t) + \lambda_4 \hat{\mathcal{L}}_{Lie}(\theta, S_t),$$

and the minimum  $\hat{u}_{\theta^*}$  corresponds to the well-trained LSN.

## 5. Experiments

In this section, we conduct three main experiments. In Section 5.1, we perform an ablation study on LSN, comparing it to PINNs [10] under different equation parameters and using large-scale data to show LSN's performance improvement over the baseline. In Section 5.2, we evaluate LSN against baseline algorithms like IPINNs [25], sfPINNs [26], fIPINNs [26], and LPS [27], validating LSN's superior performance. In Section 5.3, we validate the general applicability of LSN by extending it to the Vařiček model and showing the adaptability of different Lie symmetry operators. We start by providing a brief introduction to the parameter settings for all experiments, with specific parameters for each experiment detailed in their respective sections.

**Data.** The small-scale experiments employ a training set 50 internally scattered points and 2000 points randomly placed at the boundaries, while the test set consists of 2,500 (or 200) uniformly sampled points. The large-scale experiments employ a training set of 2000 internally scattered points and 8000 points randomly placed at the boundaries, while the test set consists of 2,500 (or 200) uniformly sampled points.

**Neural architecture and optimiser.** The LSN network employs a fully connected architecture, consisting of 9 layers with each layer having a width of 50 neurons. The tanh function is used as the activation function. For optimization, we choose Adam with an initial learning rate of 0.001 and a learning rate decay factor  $\Gamma$ .

**Equation parameters.** For the parameters of the Black-Scholes equation, we set them to  $K = 10$ ,  $x \in [0, 20]$ , and  $t \in [0, 1]$ , following the conventions established in existing literature [33]. For the parameters of the Vařiček Equation, we set them to  $x \in [0, 1]$ ,  $t \in [0, 1]$ ,  $\lambda = 0.7$ ,  $\beta = 0.08$ ,  $\gamma = -1$ ,  $\sigma = 0.03$  and  $\alpha = \frac{1}{2}\sigma^2$ .

**Evaluation.** The evaluation metrics include relative test error (see Definition A.4) and conservation error  $\hat{\mathcal{L}}_{Lie}$ .

### 5.1. Comparison with PINNs

**Experimental design.** We conduct comparative experiments between LSN and baseline PINNs under the following four sets of hyperparameter setups.

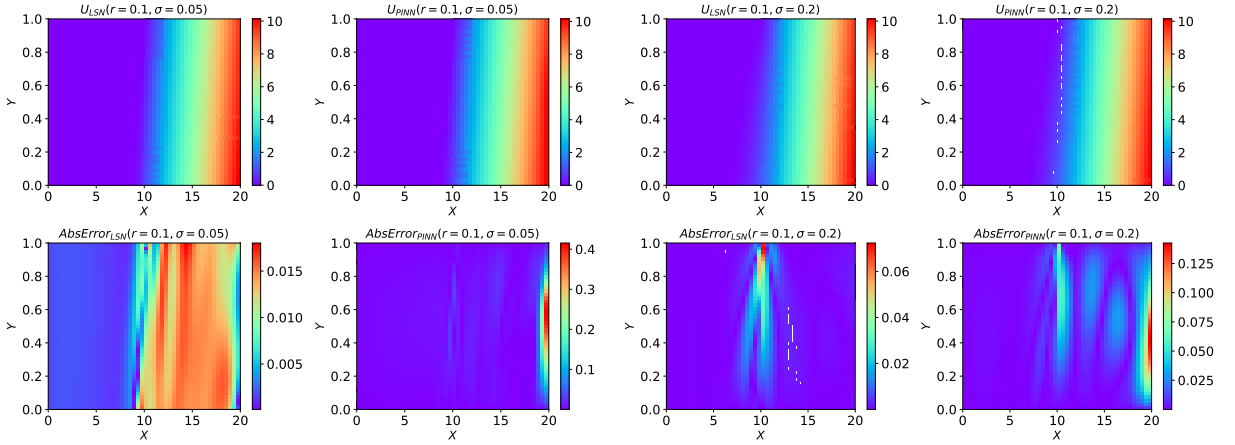
For the first configuration, we chose  $r = 0.1$ ,  $\sigma = 0.05$ , a learning rate decay rate  $\Gamma = 0.99$ , and *Iterations* = 50,000, the weight for LSN's loss function was set as  $\lambda_1 = 0.001$ ,  $\lambda_2 = 1 - \lambda_1 = 0.999$ ,  $\lambda_3 = 0.001$  and  $\lambda_4 = 0.001$ , while for PINNs, the weight was set as  $\lambda_1 = 0.001$ ,  $\lambda_2 = 0.999$  and  $\lambda_3 = 0.001$ . For the second configuration, we select  $r = 0.1$ ,  $\sigma = 0.2$ , a learning rate decay rate of  $\Gamma = 0.95$ , and *Iterations* = 90,000. The weights for the loss functions remains the same as the first configuration. For the third setting, the experimental parameter settings for the small dataset under



**Table 1**

Experimental Parameter Configuration. Values in parentheses represent weights for the enlarged dataset, while values not in parentheses represent weights for the small dataset.

RFR ( $r$ )	volatility ( $\sigma$ )	weight				learning rate ( $lr$ )	iteration
		$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$		
0.1	0.4	0.001(0.001)	1(1)	0.1(0.1)	1(0.1)	0.001	200,000
0.1	0.5	0.0001(0.0001)	1(1)	0.1(0.1)	1(1)	0.001	200,000
0.11	0.4	0.001(0.001)	1(1)	0.1(0.1)	10(0.1)	0.001	200,000



**Figure 2:** Visual representations of numerical solutions obtained using LSN and PINNs, along with absolute errors compared to the exact solution. The configuration of parameters is as follows: (1)  $r = 0.1$ ,  $\sigma = 0.05$ ,  $\Gamma = 0.99$ , and  $Iterations = 50,000$  (the left two column); (2)  $r = 0.1$ ,  $\sigma = 0.2$ ,  $\Gamma = 0.95$ , and  $Iterations = 90,000$  (the right column).

other parameters are given in the Table 1. Regarding the fourth setting, the experimental parameter settings for the enlarged dataset are provided in the Table 1, with values in parentheses.

We visualize the numerical solutions obtained on the test set using these two sets of hyperparameters in Figure 2. In identical experimental configurations, LSN outperforms PINNs, achieving a point-wise error magnitude of  $10^{-2}$  in contrast to  $10^{-1}$  observed with PINNs.

The error curves for LSN and PINNs with respect to the number of training steps for the third set of parameters are presented in Figure 3. It can be observed from Figure 3 that the conservation error of PINNs is of the order  $10^{-1}$ , whereas LSN can achieve an error on the magnitude  $10^{-4}$ . Furthermore, the relative error of LSN also consistently remains lower than that of PINNs.

In the fourth set of experiments, we increase the number of data points to 10k. It can be observed from Figure 4 that increasing the number of data points improves the accuracy of both PINNs and LSN. Notably, the parameters  $r = 0.11$ ,  $\sigma = 0.4$ , the test error magnitude of PINNs and reaches  $10^{-3}$  and  $10^{-3}$ , respectively.

To provide a more intuitive demonstration of the superiority of LSN, we consider the test accuracy of the method under different equation parameters, as shown in Table 2. Compared with vanilla PINNs, LSN can reduce the relative test error by up to 7 times, with an average improvement of 2-4 times.

## 5.2. Comparison with state-of-the-art methods

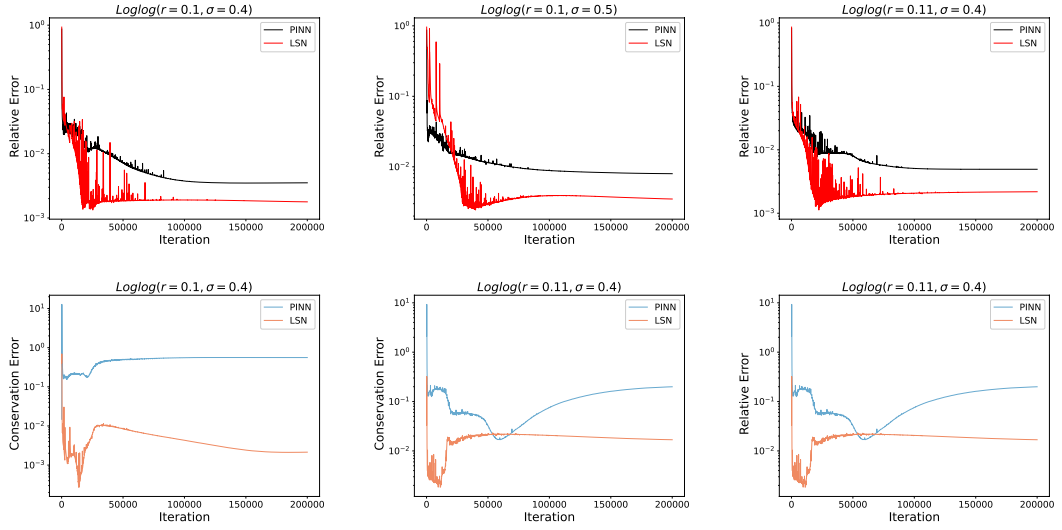
We conduct comparative experiments between LSN and several state-of-the-art methods including IPINNs [25], sfPINNs [26], ffPINNs [26] and LPS [27], under different equation parameter setups (i.e., different risk-free rate and volatility), following Shinde and Takale [34], Ankudinova and Ehrhardt [35], Bai et al. [25].

**Experimental design.** All methods share the following hyperparameter setup, i.e., learning rate  $lr = 0.001$  and learning rate decay rate  $\Gamma = 0.95$ . The training steps is set as 80,000 and 200,000, depending on the speed of convergence.

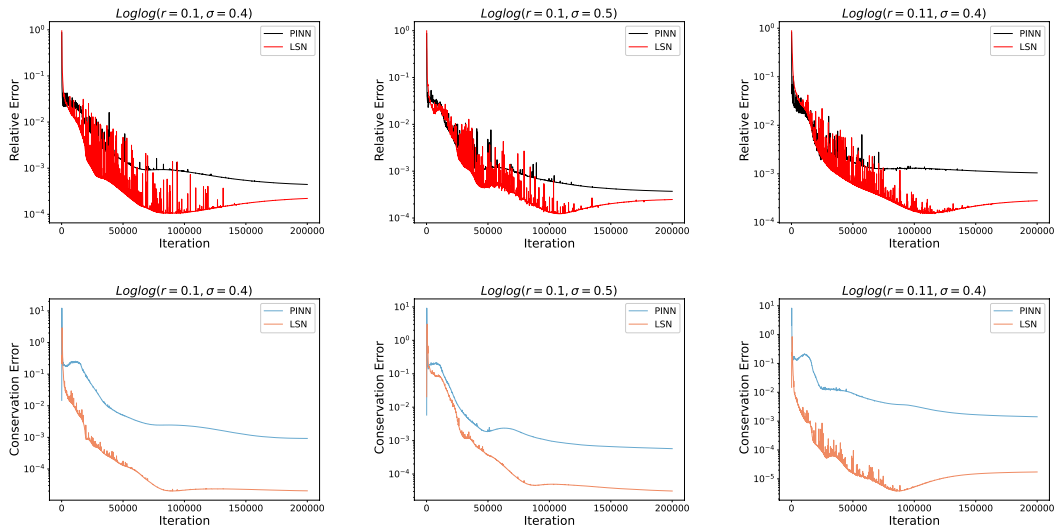
**Table 2**

Comparisons of relative test error of LSN and PINNs after 80,000 training iterations. Here  $\Gamma$  represents the rate of learning rate decay, while "Factor" represents the ratio of the test error of PINNs to that of LSN.

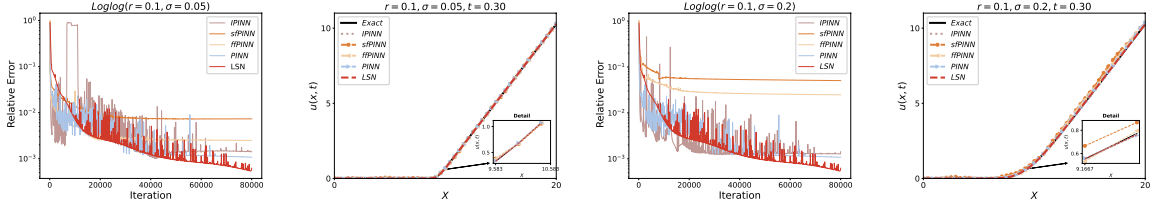
RFR ( $r$ )	Volatility ( $\sigma$ )	Relative test error		Factor
		PINNs	LSN	
0.1( $\Gamma = 0.99$ )	0.05	$3.1 \times 10^{-3}$	$4.5 \times 10^{-4}$	<b>6.9</b>
0.1( $\Gamma = 0.95$ )	0.05	$1.1 \times 10^{-3}$	$5.4 \times 10^{-4}$	2.0
0.1( $\Gamma = 0.95$ )	0.2	$5.5 \times 10^{-3}$	$1.6 \times 10^{-3}$	3.4
0.1( $\Gamma = 0.95$ )	0.4	$3.5 \times 10^{-3}$	$1.3 \times 10^{-3}$	2.6
0.1( $\Gamma = 0.95$ )	0.5	$8.0 \times 10^{-3}$	$2.4 \times 10^{-3}$	3.3
0.11( $\Gamma = 0.95$ )	0.4	$5.0 \times 10^{-3}$	$1.1 \times 10^{-3}$	4.4



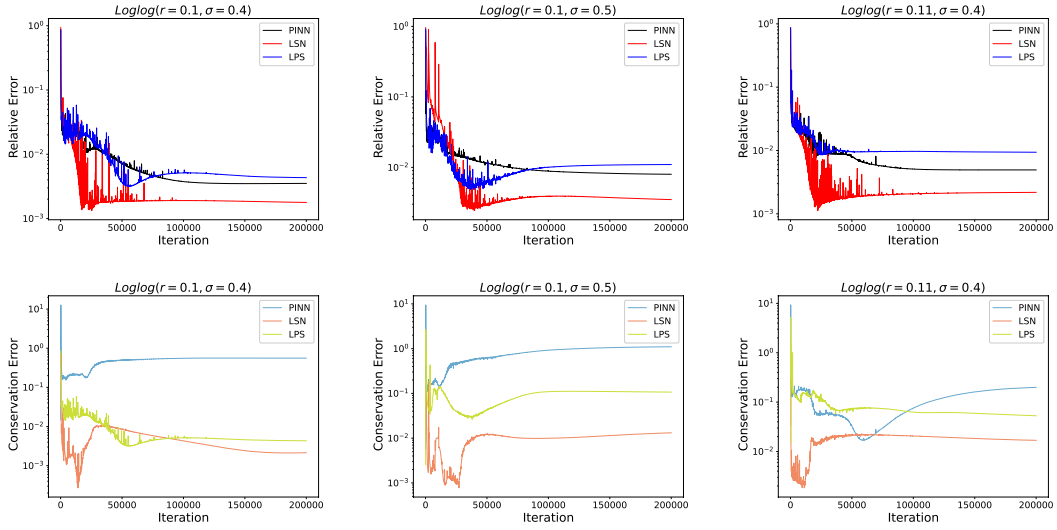
**Figure 3:** Log-log relative test error curves of PINNs and LSN under the third parameters configuration.



**Figure 4:** Log-log relative test error curves of PINNs and LSN under the third parameters configuration.



**Figure 5:** Log-log relative test error curves and function approximation results of LSN, PINNs, and PINNs variants. The error curves are shown in the first and third rows, while the function approximation results at  $t = 0.30$  are presented in the second and fourth rows.



**Figure 6:** Log-log relative test error curves of PINNs, LSN and LPS under the fourth parameters configuration.

Figure 5 provides Log-log relative test error curves and function approximation results of LSN, PINNs, and PINNs variants. The results demonstrate that the relative error of LSN consistently remains below those of vanilla PINNs and their variants across different experimental settings. Both sPINNs and fPINNs exhibit unsatisfactory performance under certain parameters, occasionally performing even worse than vanilla PINNs. This underperformance may be attributed to the fact that sPINNs and fPINNs are more suited to scenarios with sinusoidal-form solutions, thereby failing to effectively approximate the complex solution of the Black-Scholes equation [26].

We conduct comparative experiments among LSN, LPS, and PINNs using the same weights, as shown in Figure 6. Specifically, LSN and LPS share the same weights  $\lambda_i$  for  $i = 1, \dots, 4$ , while PINNs share the same weights  $\lambda_i$  for  $i = 1, \dots, 3$  as LSN and LPS but with  $\lambda_4 = 0$ . The experiments demonstrate that LSN outperforms both PINNs and LPS. Additionally, LPS exhibits overall superior performance compared to PINNs when early stopping is employed.

For a more fine-grained comparison between LPS and LSN, we further finetune the weights  $l_i$  ( $i = 1, \dots, 4$ ) of the loss function of LPS under different configurations. Notably,  $l_i$  ( $i = 1, \dots, 3$ ) in LPS serve the same purpose as  $\lambda_i$  ( $i = 1, \dots, 3$ ) in LSN, while  $l_4$  in LPS determines the weight of the symmetry residuals, and  $\lambda_4$  in LSN determines the weight of the residuals of the conservation laws corresponding to the Lie symmetry operators. To illustrate the specific process of weight tuning for LPS, consider the example with  $r = 0.1$  and  $\sigma = 0.4$ , as shown in Figure 7. We start by fixing all weights to 1 and then traverse  $l_1$  values from  $[10, 1, 0.1, 0.01, 0.001, 0.0001]$  in descending order, selecting the best value of  $l_1 = 1$ . Similarly, we traverse  $l_2$  values and find that  $l_2$  performs well in the range of 0.1 to 10. We then further subdivide this range into  $[10, 4, 2, 1, 0.5, 0.25, 0.1]$  for experimentation and select the best value for  $l_2$ , which is fixed thereafter. This process is repeated for finetuning other parameters of LPS.

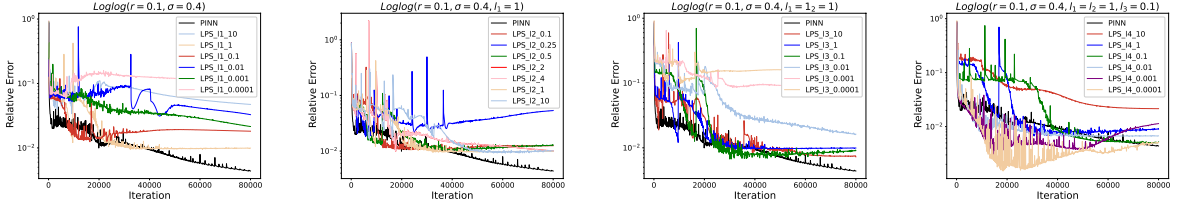


Figure 7: Log-log relative test error curves of PINNs and LPS.

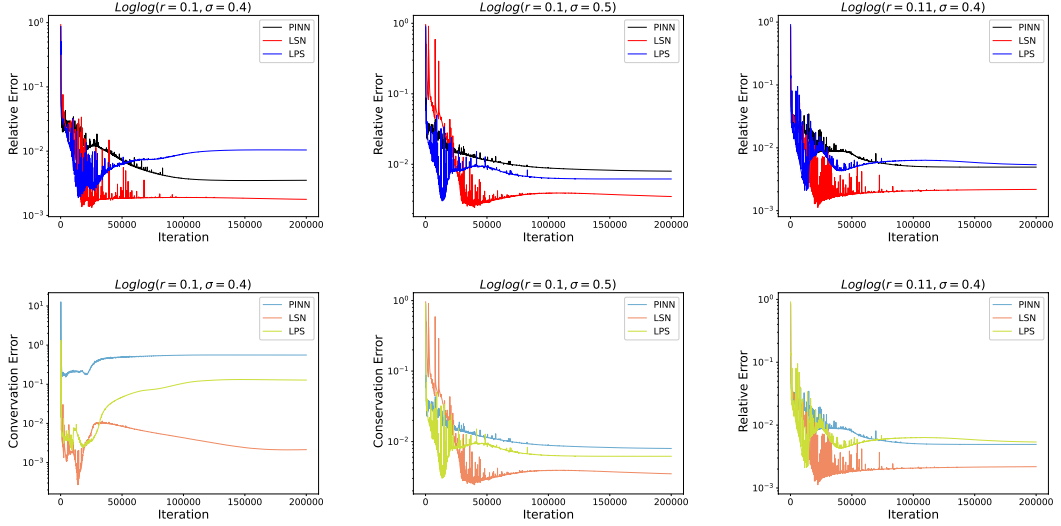


Figure 8: Log-log Error Curves Over Training Steps: PINNs vs. LSN vs. LPS. After individually tuning the weights for LPS, the performance of LSN is evaluated on the testing dataset using two metrics.

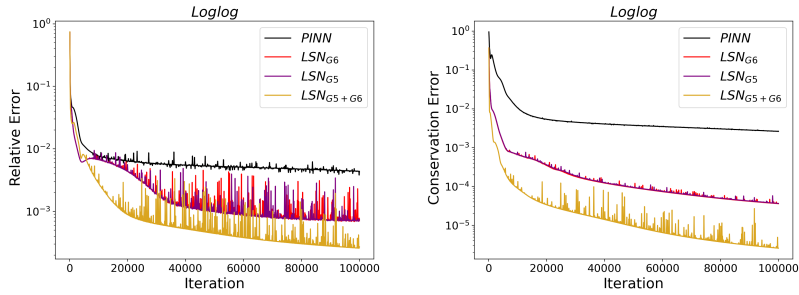


Figure 9: Log-log relative test error curves of PINNs solving the Vařicek equation. Here  $LSN_{G_i}$ ,  $i = 5, 6$  represents LSN with a single Lie symmetry operator, respectively, while " $LSN_{G_5+G_6}$ " represents LSN with the combination of two operators.

After finetuning the weights of the loss function of LPS, and using the previously set weights for LSN and PINNs, we validate the performance of LSN, PINNs, and the finetuned LPS model. The results, as shown in Figure 8, show that after extensive weight tuning, LPS can achieve a lower relative test error than PINNs with early stopping but is still outperformed by LSN in terms of both relative test error and conservation law error.

### 5.3. Experiments with different operator combinations

To demonstrate the general applicability of LSN, we apply LSN to solve the Vařiček equation and under a single and multiple operator combinations.

**Experimental design.** In this experiment, the parameters of the Vařiček Model are set as follows:  $\alpha = 0.03$ ,  $\beta = 0.08$ ,  $\gamma = -1$ ,  $\sigma = 0.03$ ,  $\Omega = 1$  and  $T = 1$ . The dataset comprises 500 internal points and 200 boundary points. The neural network architecture is designed with a depth of 2 layers and a width of 10 neurons. The training iteration is 100,000, with a learning rate of  $lr = 0.001$  and a learning rate decay factor of  $\Gamma = 0.95$ .

As shown in Figure 9, we extend LSN to the Vařiček equation, demonstrating its general applicability to different problems. We observe that although the use of a single operator alone yields significant improvements over PINNs, the performance enhancement achieved through combined operators is even more substantial. This indicates the flexibility of our method: we can effectively employ both single operators and multiple operator combinations.

## 6. Conclusion

This paper proposes a Lie symmetry net (LSN) to solve differential equations for modeling financial market dynamics by exploiting the intrinsic symmetry in the data. The Lie symmetry of these equations is interpreted as several conservation laws. A Lie symmetry residual is defined to measure how well these conservation laws are realised at specific points in the data space, which is then integrated over the entire data space to form a Lie symmetry risk. This risk helps create a structural risk that incorporates a "data fitting" risk. Our LSN is optimized under the structural risk minimization framework. Extensive experiments demonstrate the effectiveness and scalability of our algorithm, showing that the test error is reduced by over an order of magnitude.

## Broader Impacts and Future Work

This paper aims to develop AI-driven, symmetry-aware DE simulators to model financial market dynamics, which may also contribute to scientific discovery and engineering. This paper also pioneers the realization of Lie symmetries by maintaining the corresponding conservation laws, presenting a universal, off-the-shelf solution that is not limited to PINNs or the Black-Scholes equation, but can be extended to a wide range of backbones and differential equations. For future work, we will consider the incorporation of symmetries into network architecture.

## Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under the grant 12071069, the National Key R&D Program of China under the grant 2021YFA1003400, the Science and Technology Development Planning of Jilin Province under the grant YDZJ202201ZYT573 and the Fundamental Research Funds for the Central Universities under the grant 2412022ZD032.

## References

- [1] P. Del Moral, P. Del Moral, Feynman-kac formulae, Springer, 2004.
- [2] M. R. Rodrigo, R. S. Mamon, An alternative approach to solving the black-scholes equation with time-varying parameters, *Applied Mathematics Letters* 19 (2006) 398–402.
- [3] V. Goodman, J. Stampfli, The mathematics of finance, American Mathematical Society (2001).
- [4] R. C. Merton, Theory of rational option pricing, *The Bell Journal of economics and management science* (1973) 141–183.
- [5] F. Black, M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy* (1973) 637–654.
- [6] N. Privault, Short Rates and Bond Pricing, 2022, p. 479–512. URL: <http://dx.doi.org/10.1201/9781003298670-17>. doi:10.1201/9781003298670-17.
- [7] R. Valkov, Fitted finite volume method for a generalized black-scholes equation transformed on finite interval, *Numerical Algorithms* 65 (2014) 195–220.
- [8] M. K. Kadalbajoo, L. P. Tripathi, A. Kumar, A cubic b-spline collocation method for a numerical solution of the generalized black-scholes equation, *Mathematical and Computer Modelling* 55 (2012) 1483–1505.
- [9] J. Huang, Z. Cen, Cubic spline method for a generalized black-scholes equation, *Mathematical Problems in Engineering* 2014 (2014) 1–7.
- [10] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational physics* 378 (2019) 686–707.
- [11] R. Edelstein, K. Govinder, Conservation laws for the black-scholes equation, *Nonlinear Analysis: Real World Applications* 10 (2009) 3372–3380.
- [12] A. Paliathanasis, K. Krishnakumar, K. Tamizhmani, P. G. Leach, Lie symmetry analysis of the black-scholes-merton model for european options with stochastic volatility, *Mathematics* 4 (2016) 28.
- [13] R. K. Gazizov, N. H. Ibragimov, Lie symmetry analysis of differential equations in finance, *Nonlinear Dynamics* 17 (1998) 387–407.
- [14] G. Gaeta, Symmetry of stochastic non-variational differential equations, *Physics Reports* 686 (2017) 1–62.
- [15] R. Kozlov, Symmetries of systems of stochastic differential equations with diffusion matrices of full rank, *Journal of Physics A: Mathematical and Theoretical* 43 (2010) 245201.
- [16] S. C. S. Rao, et al., High-order numerical method for generalized black-scholes model, *Procedia Computer Science* 80 (2016) 1765–1776.
- [17] C. M. Khalique, T. Motsepa, Lie symmetries, group-invariant solutions and conservation laws of the vasicek pricing equation of mathematical finance, *Physica A: Statistical Mechanics and its Applications* 505 (2018) 871–879.
- [18] L. Yang, T. Gao, Y. Lu, J. Duan, T. Liu, Neural network stochastic differential equation models with applications to financial data forecasting, *Applied Mathematical Modelling* 115 (2023) 279–299.
- [19] T. Cohen, et al., Equivariant convolutional networks, Ph.D. thesis, Taco Cohen, 2021.
- [20] Y. S. Özkan, E. Yaşar, A. R. Seadawy, A third-order nonlinear schrödinger equation: the exact solutions, group-invariant solutions and conservation laws, *Journal of Taibah University for Science* 14 (2020) 585–597.
- [21] A. H. Kara, F. M. Mahomed, A basis of conservation laws for partial differential equations, *Journal of Nonlinear Mathematical Physics* 9 (2002) 60–72.
- [22] A. H. Kara, F. M. Mahomed, Relationship between symmetries and conservation laws, *International Journal of Theoretical Physics* 39 (2000) 23–40.
- [23] Y. Xie, Y. Ma, Y. Wang, Automatic boundary fitting framework of boundary dependent physics-informed neural network solving partial differential equation with complex boundary conditions, *Computer Methods in Applied Mechanics and Engineering* 414 (2023) 116139.
- [24] J. Shawe-Taylor, P. L. Bartlett, R. C. Williamson, M. Anthony, Structural risk minimization over data-dependent hierarchies, *IEEE transactions on Information Theory* 44 (1998) 1926–1940.
- [25] Y. Bai, T. Chaolu, S. Bilige, The application of improved physics-informed neural network (ipinn) method in finance, *Nonlinear Dynamics* 107 (2022) 3655–3667.
- [26] J. C. Wong, C. Ooi, A. Gupta, Y.-S. Ong, Learning in sinusoidal spaces with physics-informed neural networks, *IEEE Transactions on Artificial Intelligence* (2022).
- [27] T. Akhound-Sadeqh, L. Perreault-Levasseur, J. Brandstetter, M. Welling, S. Ravanbakhsh, Lie point symmetry and physics-informed networks, *Advances in Neural Information Processing Systems* 36 (2024).
- [28] M. Koc, I. Boztosun, D. Boztosun, On the numerical solution of black-scholes equation, in: *International Workshop on MeshFree Methods*, 2003, p. 1.
- [29] L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via deepnet based on the universal approximation theorem of operators, *Nature machine intelligence* 3 (2021) 218–229.
- [30] S. Janson, J. Tysk, Feynman-kac formulas for black-scholes-type operators, *Bulletin of the London Mathematical Society* 38 (2006) 269–282.
- [31] J. J. Rotman, An introduction to the theory of groups, volume 148, Springer Science & Business Media, 2012.
- [32] A. W. Knap, A. W. Knap, Lie groups beyond an introduction, volume 140, Springer, 1996.
- [33] Y. Bai, T. Chaolu, S. Bilige, The application of improved physics-informed neural network (ipinn) method in finance, *Nonlinear Dynamics* 107 (2022) 3655–3667.
- [34] A. Shinde, K. Takale, Study of black-scholes model and its applications, *Procedia Engineering* 38 (2012) 270–279.
- [35] J. Ankudinova, M. Ehrhardt, On the numerical solution of nonlinear black-scholes equations, *Computers & Mathematics with Applications* 56 (2008) 799–812.
- [36] U. Hassler, Ito's lemma, *Stochastic Processes and Calculus: An Elementary Introduction with Applications* (2016) 239–258.
- [37] F. C. Klebaner, Introduction to Stochastic Calculus with Applications, 2012. URL: <http://dx.doi.org/10.1142/p821>. doi:10.1142/p821.
- [38] T. De Ryck, S. Mishra, Error analysis for physics-informed neural networks (pinns) approximating kolmogorov pdes, *Advances in Computational Mathematics* 48 (2022) 79.

- [39] J. G. Cervera, Solution of the black-scholes equation using artificial neural networks, in: Journal of Physics: Conference Series, volume 1221, IOP Publishing, 2019, p. 012044.



## A. Appendix

The Appendix is divided into three parts: 1) Appendix A.1 provides the necessary definitions and lemmas, 2) Appendix A.2 includes the general form of the PDE, and 3) Appendix A.3 presents the theoretical analysis of LSN, including its approximation and generalization properties.

### A.1. Definitions and Technical Lemmas

In this section, we will present the definitions and lemmas required for our subsequent discussions.

**Definition A.1** (Wiener process). The Wiener process (also known as Brownian motion) is a continuous-time stochastic process commonly used to model random walks. The standard definition of a Wiener process includes several key features:

1. Starting Point: The process starts at  $W_0 = 0$ , indicating that its initial position is zero.
2. Independent Increments: For all  $0 \leq s < t$ , the increments  $W_t - W_s$  are mutually independent. This implies that the process is memory-less, and its future behavior is not influenced by its past.
3. Stationary Increments: For all  $0 \leq s < t$ , the distribution of the increment  $W_t - W_s$  depends only on the time difference  $t - s$ , and is independent of the specific values of  $s$  and  $t$ . Mathematically, this is expressed as  $W_t - W_s \sim \mathcal{N}(0, t - s)$ , where  $\mathcal{N}(0, t - s)$  denotes a normal distribution with mean 0 and variance  $t - s$ .
4. Continuous Paths: The paths of the Wiener process are almost surely continuous. This means that the function  $t \mapsto W_t$  is continuous with probability 1.

**Definition A.2** (European call options). European call options are financial derivatives granting the holder the right, without obligation, to purchase the underlying asset at a predetermined price upon expiration.

**Definition A.3** (Lie symmetry [13]). Consider second-order evolutionary PDEs:

$$u_t - F(t, x, u, u_{(1)}, u_{(2)}) = 0, \quad (23)$$

where  $u$  is a function of independent variables  $t$  and  $x = (x^1, \dots, x^n)$ , and  $u_{(1)}, u_{(2)}$  represent the sets of its first and second-order partial derivatives:  $u_{(1)} = (u_{x^1}, \dots, u_{x^n})$ ,  $u_{(2)} = (u_{x^1 x^1}, u_{x^1 x^2}, \dots, u_{x^n x^n})$ . Transformations of the variables  $t, x, u$  are given by:

$$\bar{t} = f(t, x, u, a), \quad \bar{x}^i = g^i(t, x, u, a), \quad \bar{u} = h(t, x, u, a), \quad i = 1, \dots, n, \quad (24)$$

where these transformations depend on a continuous parameter  $a$ . These are defined as symmetry transformations of Equation (23) if the equation retains its form in the new variables  $\bar{t}, \bar{x}, \bar{u}$ . The collection  $G$  of all such transformations forms a continuous group, meaning  $G$  includes the identity transformation:

$$\bar{t} = t, \quad \bar{x}^i = x^i, \quad \bar{u} = u,$$

the inverse of any transformation in  $G$ , and the composition of any two transformations in  $G$ . This symmetry group  $G$  is also known as the group admitted by Equation (23). According to the Lie group theory, constructing the symmetry group  $G$  is equivalent to determining its infinitesimal transformations:

$$\bar{t} \approx t + a\xi^0(t, x, u), \quad \bar{x}^i \approx x^i + a\xi^i(t, x, u), \quad \bar{u} \approx u + a\eta(t, x, u). \quad (25)$$

For convenience, the infinitesimal transformation Equation (25) can be represented by the operator:

$$X = \xi^0(t, x, u) \frac{\partial}{\partial t} + \xi^i(t, x, u) \frac{\partial}{\partial x^i} + \eta(t, x, u) \frac{\partial}{\partial u}.$$

**Definition A.4** (Relative test error). The relative test error between an approximate solution  $\hat{u}(S)$  and an exact solution  $u^*(S)$  on test data  $S$  is defined as follows:

$$\text{Relative test error} = \left\| \frac{\hat{u}(S) - u^*(S)}{u^*(S)} \right\|.$$

**Lemma A.5** (Itô's lemma [36]). *The general form of Itô's Lemma for a function  $f(t, X_t)$  of time  $t$  and a stochastic process  $X_t$  satisfying a stochastic differential equation is given by*

$$df(t, X_t) = \left( \frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma \frac{\partial f}{\partial x} dW_t.$$

Here  $t$  represents time,  $X_t$  is a stochastic process satisfying a stochastic differential equation  $dX_t = \mu dt + \sigma dW_t$ ,  $f(t, X_t)$  is the function of interest.  $\frac{\partial f}{\partial t}$ ,  $\frac{\partial f}{\partial x}$ , and  $\frac{\partial^2 f}{\partial x^2}$  denote the partial derivatives of  $f$  with respect to time and the state variable  $x$ ,  $\mu$  is the drift coefficient in the SDE,  $\sigma$  is the diffusion coefficient in the SDE and  $dW_t$  is the differential of a Wiener process.

**Lemma A.6** (Dynkin's formula [37]). *For every  $x \in \Omega$ , let  $X^x$  be the solution to a linear PDE Equation (1) with affine  $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ . If  $\varphi \in C^2(\mathbb{R}^d)$  with bounded first partial derivatives, then it holds that  $(\partial_t u)(x, t) = \mathcal{L}[u](x, t)$  where  $u$  is defined as*

$$u(x, t) = \varphi(x) + \mathbb{E} \left[ \int_0^t (F\varphi)(X_\tau^x) d\tau \right], \quad \text{for } x \in \Omega, \quad t \in [0, T], \quad (26)$$

where

$$\begin{aligned} dX_t^x &= \mu(X_t^x)dt + \sigma(X_t^x)dW_t, \quad X_0^x = x, \\ (F\varphi)(X_t^x) &= \sum_{i=1}^d \mu_i(X_t^x)(\partial_i \varphi)(X_t^x) + \frac{1}{2} \sum_{i,j,k=1}^d \sigma_{i,k}(X_t^x) \sigma_{kj}(X_t^x) (\partial_{ij}^2 \varphi)(X_t^x), \end{aligned} \quad (27)$$

where  $W_t$  is a standard  $d$ -dimensional Brownian motion on probability space  $(\Omega, \mathcal{F}, P, (\mathbb{F}_t)_{t \in [0, T]})$ , and  $F$  is the generator of  $X_t^x$ .

**Lemma A.7** ([38]). *Let  $d, L, W \in \mathbb{N}, R \geq 1, L, W \geq 2$ , let  $\mu$  be a probability measure on  $\Omega = [0, 1]^d$ , let  $f : \Omega \rightarrow [-R(W+1), R(W+1)]$  be a function and let  $f_\theta : \Omega \rightarrow \mathbb{R}, \theta \in \Theta$ , be tanh neural networks with at most  $L-1$  hidden layers, width at most  $W$  and weights and biases bounded by  $R$ . For every  $0 < \epsilon < 1$ , it holds for the generalisation and training error Equation (21) that,*

$$\mathbb{P}(\mathcal{E}_G(\theta^*(S)) \leq \epsilon + \mathcal{E}_T(\theta^*(S), S)) \geq 1 - \eta \quad \text{if} \quad N \geq \frac{64d(L+3)^2 W^6 R^4}{\epsilon^4} \ln \left( \frac{4\sqrt[5]{d+4RW}}{\epsilon} \right).$$

## A.2. General PDEs

In this section, we will demonstrate the transformation of the BS equation and the Vařiček equation into a general form, i.e., Equation (1).

**Black-Scholes equation.** As detailed in the main text, the specific expression of the Black-Scholes Equation (8) is

$$\begin{cases} \frac{\partial u_t}{\partial t} + \frac{1}{2} \sigma^2 x_t^2 \frac{\partial^2 u_t}{\partial x_t^2} + r x_t \frac{\partial u_t}{\partial x_t} - r u_t = 0, & (x_t, t_t) \in \Omega \times [0, T], \\ u_t(T, x_t) = \max(x_t - K, 0), & x_t \in \Omega, \\ u_t(t_t, 0) = 0, & t_t \in [0, T]. \end{cases} \quad (28)$$

Let's  $t = T - t_t \in [T, 0]$ ,  $x = x_t \in \Omega$ . Then the BS Equation (28) can be transformed into a more generalised initial-boundary value problem [39],

$$\begin{cases} -\frac{\partial u}{\partial t} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 u}{\partial x^2} + r x \frac{\partial u}{\partial x} - r u = 0, & (x, t) \in \Omega \times [0, T], \\ u(0, x) = \max(x - K, 0), & x \in \Omega \\ V(t, 0) = 0, & t \in [T, 0]. \end{cases} \quad (29)$$

Here  $\mathcal{L}[u]$  in Equation (1) for Equation (29) is  $\mathcal{L}[u] = \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 u(x, t)}{\partial x^2} + r x \frac{\partial u(x, t)}{\partial x} - r u(x, t)$  with  $\sigma(x) = \sigma x$ ,  $\mu(x) = r x$ ,  $v(x) = -r$  and  $\varphi(x) = \max(x - K, 0)$ .

**Vašiček equation [6].** The Vašiček pricing Equation (9) for pricing risk-free bonds  $u(t, x)$  is in the following form:

$$\begin{cases} \frac{\partial u}{\partial t} + \alpha \frac{\partial u}{\partial x^2} + \lambda(\beta - x) \frac{\partial u}{\partial x} + \gamma x u = 0, & \Omega \times [0, T], \\ u(x, T) = 1, & \Omega \times T, \\ u(x, t) = \psi(x, t) & \partial\Omega \times [0, T]. \end{cases} \quad (30)$$

Similarly, we can express the Vašiček equation in a general form as follows:

$$\begin{cases} u_t(x, t) = \mathcal{L}[u], & \text{for all } (x, t) \in \Omega \times [0, T], \\ u(0, x) = \varphi(x), & \text{for all } x \in \Omega, \\ u(y, t) = \psi(y, t), & \text{for all } (y, t) \in \partial\Omega \times [0, T], \end{cases} \quad (31)$$

where  $\mathcal{L}[u]$  in Equation (1) for Equation (30) is  $\mathcal{L}[u] = \alpha u_{xx} + \lambda(\beta - x)u_x + \gamma x u$  with  $\sigma(x) = \sqrt{2\alpha}$ ,  $\mu(x) = \lambda(\beta - x)$ ,  $v(x) = \gamma x$  and  $\varphi(x) = 1$ .

### A.3. Theoretical analysis

Given the wide range of choices for Lie symmetry operators, we use the BS equation with the selected lie operator  $G_2 = x \frac{\partial}{\partial x}$  of Equation (11) as an example to theoretically demonstrate the effectiveness of our method. The corresponding conservation law Equation (13) is as follows,

$$\mathcal{R}_{Lie}[\hat{u}] := D_t T_2^t(\hat{u}) + D_x T_2^x(\hat{u}), \quad (32)$$

where

$$\begin{cases} T_2^t(\hat{u}) = -\hat{u}_x l(t) + \frac{a}{x} + \frac{2b\hat{u}}{\sigma^2 x} e^{-rt}, \\ T_2^x(\hat{u}) = \hat{u}_t l(t) + \hat{u} l'(t) + g(t) - b\hat{u} e^{-rt} + b \left( \hat{u}_x + \frac{2r\hat{u}}{\sigma^2 x} \right) x e^{-rt}. \end{cases} \quad (33)$$

Performing operator calculations with the conserved quantities  $(T_2^t, T_2^x)$  substituted into the Equation (32) yields

$$\begin{cases} D_t T_2^t(\hat{u}) = -\hat{u}_{xt} l(t) - \hat{u}_x l_t(t) + \frac{2b\hat{u}_t}{\sigma^2 x} e^{-rt} - \frac{2rb\hat{u}}{\sigma^2 x} e^{-rt}, \\ D_x T_2^x(\hat{u}) = \hat{u}_{tx} l(t) + \hat{u}_x l_t(t) - b\hat{u}_x e^{-rt} + b \left( \hat{u}_{xx} + \frac{2r\hat{u}_x}{\sigma^2 x} - \frac{2r\hat{u}}{\sigma^2 x^2} \right) x e^{-rt} + b \left( \hat{u}_x + \frac{2r\hat{u}}{\sigma^2 x} \right) e^{-rt}. \end{cases}$$

Therefore, we have

$$\begin{aligned} D_t T_2^t(\hat{u}) + D_x T_2^x(\hat{u}) &= \frac{2b\hat{u}_t}{\sigma^2 x} e^{-rt} - b\hat{u}_x e^{-rt} + b x \hat{u}_{xx} e^{-rt} + \frac{2rb\hat{u}_x}{\sigma^2} e^{-rt} - \frac{2rb\hat{u}}{\sigma^2 x} e^{-rt} + b\hat{u}_x e^{-rt} \\ &= b x e^{-rt} \hat{u}_{xx} + \frac{2b}{\sigma^2 x} e^{-rt} \hat{u}_t + \frac{2rb}{\sigma^2} e^{-rt} \hat{u}_x - \frac{2rb}{\sigma^2 x} e^{-rt} \hat{u} \\ &= \frac{2be^{-rt}}{\sigma^2 x} \left( \hat{u}_t + \frac{1}{2} \sigma^2 x^2 \hat{u}_{xx} + r x \hat{u}_x - r \hat{u} \right). \end{aligned}$$

Since  $b$  is arbitrarily chosen, let's set  $b = x_{min}$ . Where  $x_{min}$  represents the smallest  $x$ -coordinate among the points in the configuration set. And  $(x, t) \in \Omega \times [0, T]$  represents a bounded interior region, where  $x$  and  $t$  are within the specified domain  $\Omega$  and time interval  $[0, T]$  respectively. Therefore, there exists a positive number  $M > 0$  such that

$$0 < \left| \frac{2be^{-rt}}{\sigma^2 x} \right|^2 < \left\| \frac{2be^{-rt}}{\sigma^2 x} \right\|_\infty^2 = \left\| \frac{2x_{min}e^{-rt}}{\sigma^2 x} \right\|_\infty^2 \leq \left( \frac{2x_{min}e^{-rT}}{\sigma^2 x_{min}} \right)^2 \leq \left( \frac{2e^{-rT}}{\sigma^2} \right)^2 := M. \quad (34)$$

Therefore,

$$\begin{aligned}
 \mathcal{L}_{Lie}[\hat{u}] &= \|\mathcal{R}_{Lie}[\hat{u}]\|^2 = \|D_t T_2^t(\hat{u}) + D_x T_2^x(\hat{u})\|^2 \\
 &= \left\| \frac{2be^{-rt}}{\sigma^2 x} \left( \hat{u}_t + \frac{1}{2} \sigma^2 x^2 \hat{u}_{xx} + rx \hat{u}_x - r \hat{u} \right) \right\|^2 \\
 &\leq M \|\hat{u}_t + \frac{1}{2} \sigma^2 x^2 \hat{u}_{xx} + rx \hat{u}_x - r \hat{u}\|^2 \\
 &= M \|\mathcal{R}_{PDE}[\hat{u}]\|^2.
 \end{aligned} \tag{35}$$

### A.3.1. Approximation error bounds of LSN

The PDE in Equation (1) is a linear parabolic equation with smooth coefficients, and conclusions about the existence of a unique classical solution  $u$  to the equation, which is sufficiently regular, can be derived using standard parabolic theory. If  $u$  is considered a classical solution, then the residual concerning  $u$  should be zero.

$$\mathcal{R}_i[u](x, t) = 0, \quad \mathcal{R}_s[u](y, t) = 0, \quad \mathcal{R}_t[u](x) = 0, \quad \mathcal{R}_{Lie}[u](x, t) = 0, \quad \forall x \in \Omega, \quad y \in \partial\Omega. \tag{36}$$

Here  $\mathcal{R}_{Lie}[u](x, t) = \frac{2be^{-rt}}{\sigma^2 x} \left( u_t + \frac{1}{2} \sigma^2 x^2 u_{xx} + rx u_x - ru \right) = \frac{2be^{-rt}}{\sigma^2 x} \mathcal{R}_i[u](x, t) = 0$  (with  $\frac{2be^{-rt}}{\sigma^2 x} \neq 0$ .) We first list several crucial lemmas used to prove the approximation error of LSN.

**Lemma A.8.** *Let  $T > 0$  and  $\gamma, d, s \in \mathbb{N}$  with  $s \geq 2 + \gamma$ . Suppose  $u \in W^{s, \infty}((0, 1)^d \times [0, T])$  is the solution to a linear PDE (1). Then, for every  $\varepsilon > 0$  there exists a tanh neural network  $\hat{u}^\varepsilon = u_{\hat{\theta}^\varepsilon}$  with two hidden layers of width at most  $\mathcal{O}(\varepsilon^{-d/(s-2-\gamma)})$  such that  $\mathcal{E}(\hat{\theta}^\varepsilon) \leq \varepsilon$ .*

*Proof.* We extend the proof of the Theorem 1 in [38] to the LSN algorithm with regularization terms incorporating Lie symmetries. There exists a tanh neural network  $\hat{u}^\varepsilon$  with two hidden layers of width at most  $\mathcal{O}(\varepsilon^{-d/(s-2-\gamma)})$  such that

$$\|u - \hat{u}^\varepsilon\|_{W^{2, \infty}((0, 1)^d \times [0, T])} \leq \varepsilon.$$

Due to the linearity of PDEs (where Equation (1) is a linear equation with respect to  $u$ ), it immediately follows that  $|\mathcal{R}_i[u]|_{L^2((0, 1)^d \times [0, T])} \leq \varepsilon$  and  $|\mathcal{R}_{Lie}[u]|_{L^2((0, 1)^d \times [0, T])} \leq M |\mathcal{R}_i[u]|_{L^2((0, 1)^d \times [0, T])} \leq \varepsilon$ . By employing a standard trace inequality, one can establish similar bounds for  $\mathcal{R}_s[u]$  and  $\mathcal{R}_t[u]$ . Consequently, it directly follows that  $\mathcal{E}(\hat{\theta}^\varepsilon) \leq \varepsilon$ .  $\square$

This lemma shows that the structure risk of LSN in Equation (21) can converge to zero. To address the challenge of the curse of dimensionality in structure risk of LSN Equation (21) bounds, we will leverage Dynkin's Lemma A.6, which establishes a connection between the linear partial differential Equation (1) and the Itô diffusion stochastic equation. Next, we will extend the proof for PINNs from [38] to LSNs to demonstrate that the loss for LSNs can be made infinitesimally small.

**Lemma A.9.** *Let  $\alpha, \beta, \varpi, \zeta, T > 0$ , and  $p > 2$ . For any  $d \in \mathbb{N}$ , define  $\Omega_d = [0, 1]^d$  and consider  $\varphi_d \in C^5(\mathbb{R}^d)$  with bounded first partial derivatives. Given the probability space  $(\Omega_d \times [0, T], \mathcal{F}, \mu)$ , and let  $u_d \in C^{2, 1}(\Omega_d \times [0, T])$  be a function satisfying*

$$(\partial_t u_d)(x, t) = \mathcal{L}[u_d](x, t), \quad u_d(x, 0) = \varphi_d(x), \quad \mathcal{L}_{Lie}[u_d](x, t) = 0 \quad \text{for all } (x, t) \in \Omega_d \times [0, T].$$

Assume for every  $\xi, \delta, c > 0$ , there exist hyperbolic tangent (tanh) neural networks such that

$$\|\varphi_d - \widehat{\varphi}_{\xi, d}\|_{C^2(D_d)} \leq \xi \quad \text{and} \quad \|\mathcal{F}\varphi - (\widehat{\mathcal{F}\varphi})_{\delta, d}\|_{C^2([-c, c]^d)} \leq \delta. \tag{37}$$

Under these conditions, there exist constants  $C, \lambda > 0$  such that for every  $\varepsilon > 0$  and  $d \in \mathbb{N}$ , a constant  $\rho_d > 0$  (independent of  $\varepsilon$ ) and a tanh neural network  $\Psi_{\varepsilon, d}$  with at most  $C(d\rho_d)^\lambda \varepsilon^{-\max\{5p+3, 2+p+\beta\}}$  neurons and weights that grow at most as  $C(d\rho_d)^\lambda \varepsilon^{-\max\{\zeta, 8p+6\}}$  for  $\varepsilon \rightarrow 0$  can be found such that

$$\begin{aligned}
 &\|\partial_t \Psi_{\varepsilon, d} - \mathcal{L}[\Psi_{\varepsilon, d}]\|_{L^2(\Omega_d \times [0, T])} + \|\Psi_{\varepsilon, d} - u_d\|_{H^1(\Omega_d \times [0, T])} \\
 &+ \|\Psi_{\varepsilon, d} - u_d\|_{L^2(\partial(\Omega_d \times [0, T]))} + \|\mathcal{L}_{Lie}[\Psi_{\varepsilon, d}]\|_{L^2(\Omega_d \times [0, T])} \leq \varepsilon,
 \end{aligned} \tag{38}$$

where  $\rho_d$  is defined as

$$\rho_d := \max_{x \in \Omega_d} \sup_{\substack{s, t \in [0, T] \\ s < t}} \frac{\|X_s^x - X_t^x\|_{\mathcal{L}^q(P, \|\cdot\|_{\mathbb{R}^d})}}{|s - t|^{\frac{1}{p}}} < \infty. \quad (39)$$

In this context,  $X^x$  denotes the solution, following the Itô interpretation, of the stochastic differential equation (SDE) specified by Equation (23). Here  $q > 2$  remains independent of  $d$  and the norm  $\|\cdot\|_{\mathcal{L}^q(P, \|\cdot\|_{\mathbb{R}^d})}$  is defined as follows: Given a measure space  $(\Omega, \mathcal{F}, \mu)$  where  $q > 0$ , for any  $\mathcal{F}/B(\mathbb{R}^d)$ -measurable function  $f : \Omega \rightarrow \mathbb{R}^d$ ,

$$\|f\|_{\mathcal{L}^q(\mu, \|\cdot\|_{\mathbb{R}^d})} := \left[ \int_{\Omega} \|f(\omega)\|_{\mathbb{R}^d}^q \mu(d\omega) \right]^{\frac{1}{q}}. \quad (40)$$

*Proof.* The main proof follows directly from Theorem 2 in De Ryck and Mishra [38], where

$$\|\mathcal{L}_{Lie} [\Psi_{\varepsilon, d}]\|_{L^2(\Omega_d \times [0, T])} \leq M \|\partial_t \Psi_{\varepsilon, d} - \mathcal{L} [\Psi_{\varepsilon, d}]\|_{L^2(\Omega_d \times [0, T])} \leq \varepsilon. \quad (41)$$

□

According to Remark 2 by [38], it is indicated that the assumption conditions in the Lemma A.9 are easily satisfied after modifications for the BS equation.

**Theorem A.10.** Let  $u$  be a classical solution to linear PDE as described in Equation (1) with  $\mu \in C^1(\Omega; \mathbb{R}^d)$  and  $\sigma \in C^2(\Omega; \mathbb{R}^{d \times d})$ , let  $M = \left(\frac{2e^{-rT}}{\sigma^2}\right)^2$ ,  $v \in C^2(\Omega \times [0, T]; \mathbb{R})$ , and define the residuals according Equation (18). Then,

$$\begin{aligned} \|u - v\|_{L^2(\Omega \times [0, T])}^2 &\leq C_1 \left[ \|\mathcal{R}_i[v]\|_{L^2(\Omega \times [0, T])}^2 + \|\mathcal{R}_{lie}[v]\|_{L^2(\Omega \times [0, T])}^2 + \|\mathcal{R}_t[v]\|_{L^2(\Omega)}^2 \right. \\ &\quad \left. + C_2 \|\mathcal{R}_s[v]\|_{L^2(\partial\Omega \times [0, T])} + C_3 \|\mathcal{R}_s[v]\|_{L^2(\partial\Omega \times [0, T])}^2 \right], \end{aligned} \quad (42)$$

where

$$\begin{aligned} C_0 &= 2 \sum_{i,j=1}^d \left\| \partial_{ij} (\sigma \sigma^T)_{ij} \right\|_{L^\infty(\Omega \times [0, T])}, \\ C_1 &= T e^{\left(2C_0 + 2\|\operatorname{div} \mu\|_\infty + 1 + \frac{1}{M} + 2\|v\|_\infty\right)T}, \\ C_2 &= 2 \sum_{i=1}^d \left\| (\sigma \sigma^T \nabla_x [u - v])_i \right\|_{L^2(\partial\Omega \times [0, T])}, \\ C_3 &= 2\|\mu\|_\infty + (1 + M) \sum_{i,j,k=1}^d \left\| \partial_i (\sigma_{ik} \sigma_{jk}) \right\|_{L^\infty(\partial\Omega \times [0, T])}. \end{aligned} \quad (43)$$

*Proof.* Let  $\hat{u} = v - u$ . Integrating  $\mathcal{R}_i[\hat{u}](t, x)$  over  $\Omega$  and rearranging terms gives

$$\frac{d}{dt} \int_{\Omega} |\hat{u}|^2 dx = \frac{1}{2} \int_{\Omega} \operatorname{Trace} (\sigma^2 H_x[\hat{u}]) \hat{u} dx + \int_{\Omega} \mu J_x[\hat{u}] \hat{u} dx + \int_{\Omega} v |\hat{u}|^2 dx + \int_{\Omega} \mathcal{R}_i[\hat{u}] \hat{u} dx, \quad (44)$$

where all integrals are understood as integrals with respect to the Lebesgue measure on  $\Omega$  and  $\partial\Omega$ , and where  $J_x$  represents the Jacobian matrix, which is the transpose of the gradient with respect to the spatial coordinates. Following the derivation by Theorem 4 of [38], we can similarly show that :

for the first term

$$\begin{aligned} &\int_{\Omega} \operatorname{Trace} (\sigma \sigma^T H_x[\hat{u}]) \hat{u} dx \\ &\leq \sum_{i=1}^d \int_{\partial\Omega} \left| (\sigma \sigma^T J_x(\hat{u})^T)_i \hat{u} (\hat{e}_i \cdot \hat{n}) \right| dx - \underbrace{\int_{\Omega} J_x[\hat{u}] \sigma (J_x[\hat{u}] \sigma)^T dx}_{\geq 0} + \frac{c_2}{2} \int_{\partial\Omega} |\mathcal{R}_s[v]|^2 dx + \frac{c_3}{2} \int_{\Omega} \hat{u}^2 dx, \end{aligned} \quad (45)$$

for the second term

$$\int_{\Omega} \mu J_x[\hat{u}] \hat{u} dx \leq \frac{1}{2} \|\operatorname{div} \mu\|_{\infty} \int_{\Omega} \hat{u}^2 dx + \frac{1}{2} \|\mu\|_{\infty} \int_{\partial\Omega} |\mathcal{R}_s[v]|^2 dx, \quad (46)$$

for the fourth term

$$\int_{\Omega} \mathcal{R}_i[\hat{u}] \hat{u} dx \leq \frac{1}{2} \int_{\Omega} \mathcal{R}_i[\hat{u}]^2 dx + \frac{1}{2} \int_{\Omega} \hat{u}^2 dx, \quad (47)$$

where  $\hat{n}$  denotes the unit normal on  $\partial\Omega$ .  $1 \leq i, j, k \leq d$  and

$$\begin{aligned} c_1 &= 2 \sum_{i=1}^d \left\| (\sigma \sigma^T J_x[\hat{u}]^T)_i \right\|_{L^2(\partial\Omega \times [0, T])}, \\ c_2 &= \sum_{i,j,k=1}^d \left\| \partial_i (\sigma_{ik} \sigma_{jk}) \right\|_{L^{\infty}(\partial\Omega \times [0, T])}, \\ c_3 &= \sum_{i,j=1}^d \left\| \partial_{ij} (\sigma \sigma^T)_{ij} \right\|_{L^{\infty}(\Omega \times [0, T])}. \end{aligned} \quad (48)$$

As for the third term of Equation (44), we obtain

$$\int_{\Omega} v |\hat{u}|^2 dx \leq \|v\|_{\infty} \int_{\Omega} |\hat{u}|^2 dx. \quad (49)$$

Integrating Equation (44) over the interval  $[0, \tau] \subset [0, T]$ , using all the previous inequalities together with Hölder's inequality, we find that

$$\begin{aligned} \int_{\Omega} |\hat{u}(x, \tau)|^2 dx &\leq \int_{\Omega} |\mathcal{R}_t[v]|^2 dx + c_1 \left( \int_{\partial\Omega \times [0, T]} |\mathcal{R}_s[v]|^2 dx dt \right)^{1/2} + \int_{\Omega \times [0, T]} |\mathcal{R}_i[\hat{u}]|^2 dx dt \\ &\quad + (c_2 + \|\mu\|_{\infty}) \int_{\partial\Omega \times [0, T]} |\mathcal{R}_s[v]|^2 dx dt + (c_3 + \|\operatorname{div} \mu\|_{\infty} + 1 + \|v\|_{\infty}) \int_{[0, \tau]} \int_{\Omega} |\hat{u}(x, s)|^2 dx dt. \end{aligned} \quad (50)$$

Referring to Equation (44), we can transform operator  $\mathcal{R}_{Lie}[\hat{u}] = \frac{2be^{-rt}}{\sigma^2 x} \left( u_t + \frac{1}{2} \sigma^2 x^2 u_{xx} + rxu_x - ru \right)$  with  $\frac{2be^{-rt}}{\sigma^2 x} \neq 0$ , i.e.,  $\mathcal{R}_i[\hat{u}] = u_t + \frac{1}{2} \sigma^2 x^2 u_{xx} + rxu_x - ru = \frac{\sigma^2 x}{2be^{-rt}} \mathcal{R}_{Lie}[\hat{u}]$  into the following form,

$$\begin{aligned} \frac{d}{dt} \int_{\Omega} |\hat{u}|^2 dx &= \frac{1}{2} \int_{\Omega} \operatorname{Trace} (\sigma^2 H_x[\hat{u}]) \hat{u} dx + \int_{\Omega} \mu J_x[\hat{u}] \hat{u} dx + \int_{\Omega} v |\hat{u}|^2 dx + \int_{\Omega} \frac{\sigma^2 x}{2be^{-rt}} \mathcal{R}_{Lie}[\hat{u}] \hat{u} dx \\ &\leq \frac{1}{2} \int_{\Omega} \operatorname{Trace} (\sigma^2 H_x[\hat{u}]) \hat{u} dx + \int_{\Omega} \mu J_x[\hat{u}] \hat{u} dx + \int_{\Omega} v |\hat{u}|^2 dx + \frac{1}{M} \int_{\Omega} \mathcal{R}_{Lie}[\hat{u}] \hat{u} dx. \end{aligned} \quad (51)$$

The proof is ultimately established by using Using Grönwall's inequality and integrating over  $[0, T]$ .  $\square$

*Remark A.11.* Theorem A.10 states that by optimizing structure risk Equation (21), the network's output can approximate the exact solution, while Lemma A.9 confirms that structure risk can be minimized. This verifies the numerical approximation of the LSN to the exact solution.

### A.3.2. Generalisation error bounds of LSN

We set a general configuration let  $\Omega \subset \mathbb{R}^d$  be compact and let  $u : \Omega \rightarrow \mathbb{R}, u_{\theta} : \Omega \rightarrow \mathbb{R}$  be functions for all  $\theta \in \Theta$ . We consider  $u$  as the exact value of the PDE (1), and  $u_{\theta}$  as the approximation generated by LSN with weights  $\theta$ .

Let  $N \in \mathbb{N}$  be the training set size and let  $S = \{z_1, \dots, z_M\} \in \Omega^N$  be the training set, where each  $z_i$  is independently drawn according to some probability measure  $\mu$  on  $\Omega$ . We define the structure risk and empirical loss as

$$\mathcal{L}(\theta) = \int_{\Omega} |u_{\theta}(z) - u(z)|^2 d\mu(z), \quad \hat{\mathcal{L}}(\theta, S) = \frac{1}{N} \sum_{i=1}^N |u_{\theta}(z_i) - u(z_i)|^2, \quad \theta^*(S) \in \arg \min_{\theta \in \Theta} \hat{\mathcal{L}}(\theta, S). \quad (52)$$

**Lemma A.12.** Let  $d, L, W \in \mathbb{N}$  with  $R \geq 1$ , and define  $M = \left(\frac{2e^{-rT}}{\sigma^2}\right)^2$ . Consider  $u_\theta : [0, 1]^d \rightarrow \mathbb{R}$ , where  $\theta \in \Theta$  representing tanh neural networks with at most  $L - 1$  hidden layers, each with a width of at most  $W$ , and weights and biases bounded by  $R$ . Let  $\mathcal{L}^q$  and  $\hat{\mathcal{L}}^q$  denote the structure risk and empirical error, respectively, for linear general PDEs as in Equation (1). Assume  $\max \{\|\varphi\|_\infty, \|\psi\|_\infty\} \leq \max_{\theta \in \Theta} \|u_\theta\|_\infty$ . Denote by  $\mathfrak{L}^q$  the Lipschitz constant of  $\mathcal{L}^q$ , for  $q = i, t, s$ . Then, it follows that

$$\mathfrak{L}^q \leq 2^{5+2L} C(d+7)^2 L^4 R^{6L-1} W^{6L-6},$$

$$\text{where } C = (1 + M) \max_{x \in D} \left( 1 + \sum_{i=1}^d |\mu(x)_i| + \sum_{i,j=1}^d |(\sigma(x)\sigma(x)^*)_{ij}| \right)^2.$$

*Proof.* Similar to Lemma 16 in [38], we have the following:

$$\begin{aligned} \left| \mathcal{R}_i[u_\theta](t, x) - \mathcal{R}_i[\Phi^\theta](t, x) \right| &\leq |v(x)|_1 |u_\theta - \Psi^\theta|_\infty + (1 + |\mu(x)|_1) \left| J^\theta - J^\theta \right|_\infty + |\sigma(x)\sigma(x)^*|_1 \left| H_x^\theta - H_x^\theta \right|_\infty \\ &\leq 4\alpha (1 + |v(x)|_1 + |\mu(x)|_1 + |\sigma(x)\sigma(x)^*|_1) (d+7) L^2 R^{3L-1} W^{3L-3} 2^L |\theta - \vartheta|_\infty. \end{aligned}$$

And we have

$$\left| \mathcal{R}_{lie}[u_\theta](t, x) - \mathcal{R}_{lie}[\Phi^\theta](t, x) \right| \leq M \left| \mathcal{R}_i[u_\theta](t, x) - \mathcal{R}_i[\Phi^\theta](t, x) \right|, \quad (53)$$

where we let  $\|\cdot\|_p$  denote the vector  $p$ -norm of the vectored version of a general tensor. Next, we set  $\vartheta = 0$  and  $\max \{\|\varphi\|_\infty, \|\psi\|_\infty\} \leq \max_{\theta \in \Theta} \|u_\theta\|_\infty$  for  $q = t, s$  that

$$\begin{aligned} \max_\theta \left\| \mathcal{R}_i[u_\theta] \right\|_\infty &\leq 4\alpha C_1 (d+7) 2^L L^2 R^{3L} W^{3L-3}, \\ \max_\theta \left\| \mathcal{R}_{lie}[u_\theta] \right\|_\infty &\leq 4\alpha C_1 M (d+7) 2^L L^2 R^{3L} W^{3L-3}, \\ \max_\theta \left\| \mathcal{R}_q[u_\theta] \right\|_\infty &\leq 2WR, \end{aligned} \quad (54)$$

where  $C_1 = \max_{x \in \Omega} (1 + |v(x)|_1 + |\mu(x)|_1 + |\sigma(x)\sigma(x)^*|_1)$ . Combining all the previous results yields the bound.  $\square$

We can then obtain the generalization bound of LSN as follows.

**Theorem A.13.** Let  $L, W, N \in \mathbb{N}$ ,  $R \geq 1$ ,  $L, W \geq 2$ ,  $a, b \in \mathbb{R}$  with  $a < b$  and let  $u_\theta : [0, 1]^d \rightarrow \mathbb{R}$ ,  $\theta \in \Theta$ , be tanh neural networks with at most  $L - 1$  hidden layers, width at most  $W$ , and weights and biases bounded by  $R$ . For  $q = i, t, s$ , let  $\mathcal{L}^q$  and  $\hat{\mathcal{L}}^q$  denote the LSN structure risk and training error, respectively, for linear general PDEs as in Equation (1). Let  $c_q > 0$  be such that  $\hat{\mathcal{L}}^q(\theta, S)$ ,  $\mathcal{L}^q(\theta) \in [0, c_q]$ , for all  $\theta \in \Theta$  and  $S \subset \Omega^N$ . Assume  $\max \{\|\varphi\|_\infty, \|\psi\|_\infty\} \leq \max_{\theta \in \Theta} \|u_\theta\|_\infty$  and define the constants

$$C = (1 + M) \max_{x \in \Omega} \left( 1 + \sum_{i=1}^d |v(x)_i| + \sum_{i=1}^d |\mu(x)_i| + \sum_{i,j=1}^d |(\sigma(x)\sigma(x)^*)_{ij}| \right)^2.$$

Then, for any  $\epsilon > 0$ , it holds that

$$\mathcal{L}^q \leq \epsilon + \hat{\mathcal{L}}^q \quad \text{if } M_q \geq \frac{24dL^2W^2c_q^2}{\epsilon^4} \ln \left( 4c_1RW \sqrt[6]{\frac{C(d+7)}{\epsilon^2}} \right). \quad (55)$$

*Proof.* The proof follows the generalization analysis of PINNs [38]. Setting

$$C = (1 + M) \max_{x \in D} \left( 1 + \sum_{i=1}^d |v(x)_i| + \sum_{i=1}^d |\mu(x)_i| + \sum_{i,j=1}^d |(\sigma(x)\sigma(x)^*)_{ij}| \right)^2, \quad (56)$$

we can use Lemma A.12 with  $a \leftarrow R$ ,  $c \leftarrow c_q$ ,  $\mathfrak{L} \leftarrow 2^{5+2L} C^2 (d+7)^2 L^4 R^{6L-1} W^{6L-6}$  and  $k \leftarrow 2dLW^2$  (Lemma A.7). We then arrive at

$$k \ln \left( \frac{4a\mathfrak{L}}{\epsilon^2} \right) + \ln \left( \frac{2c_q}{\epsilon^2} \right) \leq 6kL \ln \left( 4c_qRW \sqrt[6]{\frac{C(d+7)}{\epsilon^2}} \right) = 12dL^2W^2 \ln \left( 4c_qRW \sqrt[6]{\frac{C(d+7)}{\epsilon^2}} \right).$$

$\square$