Applied Mathematical Modelling 163 (2027) 117236

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Parameter estimation in stochastic differential equations via Wiener chaos expansion and stochastic gradient descent

Francisco Delgado-Vences [a,*], Jose Julian Pavon-Español [b], Arelly Ornelas [c]

[a] Universitat Autónoma de Barcelona, Departament de Matemátiques, Edifici C, Facultat de Ciéncies, 080193, Bellaterra, Spain
[b] Facultad de Ciencias - Universidad Nacional Autónoma de México. Ciudad Universitaria, Ciudad de México, México
[c] Secihti - Centro Interdisciplinario de Ciencias Marinas. IPN., La Paz, BCS., México

ARTICLE INFO

Keywords:
Stochastic gradient descent
Parameter estimation
Wiener chaos expansion
Geometric Brownian motion
Microbial growth dynamics

ABSTRACT

This study addresses the inverse problem of parameter estimation for Stochastic Differential Equations by minimizing a regularized discrepancy functional via Stochastic Gradient Descent. To achieve computational efficiency, we leverage the Wiener Chaos Expansion, a spectral decomposition technique that projects the stochastic solution onto an orthogonal basis of Hermite polynomials. This transformation effectively maps the stochastic dynamics into a hierarchical system of deterministic functions, termed the *propagator*. By reducing the stochastic inference task to a deterministic optimization problem, our framework circumvents the heavy computational burden and sampling requirements of traditional simulation-based methods like Markov chain Monte Carlo or Maximum Likelihood Estimation. The robustness and scalability of the proposed approach are demonstrated through numerical experiments on various non-linear Stochastic Differential Equations, including models for individual biological growth. Results show that the Wiener Chaos Expansion-Stochastic Gradient Descent framework provides accurate parameter recovery even from discrete, noisy observations, offering a significant paradigm shift in the efficient modeling of complex stochastic systems.

1. Introduction

Mathematical optimization is fundamental to applied mathematics, influencing areas like optimal control theory, inverse problems, and the training of modern deep neural networks. The main aim is usually described as the minimization of a continuously differentiable objective function $f : \mathbb{R}^d \to \mathbb{R}$. While some problems can be solved analytically, the overwhelming majority of practical scenarios demand iterative numerical approaches. Among these, first-order methods—specifically Gradient Descent (GD) and its stochastic counterpart, Stochastic Gradient Descent (SGD)— are very important due to their implementation simplicity and remarkable scalability in high-dimensional spaces.

This research focuses on the intersection of these optimization techniques with the theory of Stochastic Differential Equations (SDEs). Specifically, we address the inverse problem of parameter estimation in SDEs by reformulating it as a deterministic optimization task via the Wiener Chaos Expansion.

* Corresponding author.
E-mail addresses: FranciscoJavier.Delgado@uab.cat (F. Delgado-Vences), julian.pavon2@ciencias.unam.mx (J.J. Pavon-Español), aornelasv@ipn.mx (A. Ornelas).

https://doi.org/10.1016/j.apm.2026.117236
Received 27 March 2026; Received in revised form 27 July 2026; Accepted 29 July 2026
Available online 31 July 2026
0307-904X/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1.1. Deterministic gradient descent

The concept of gradient-based optimization originates from Cauchy's steepest descent method for simultaneous equations. In the context of unconstrained minimization, the Gradient Descent algorithm creates a sequence of iterates represented as $\{x_k\}_{k\geq 0}$ through the specified update rule:

$$x_{k+1} = x_k - \eta_k \nabla f(x_k), \tag{1}$$

where $\eta_k > 0$ denotes the step size (or learning rate). The theoretical convergence of this method is well-established; for convex functions with Lipschitz continuous gradients (L-smooth), GD converges at a rate of $O(1/k)$ (see [1,2]).

However, the efficiency of standard GD is limited by the condition number of the Hessian matrix. To mitigate this issue, Polyak introduced the heavy-ball method, or momentum, using the history of update vectors to accelerate convergence. Further theoretical breakthroughs were made by Nesterov, who proved that by employing a specific momentum schedule, one can achieve a convergence rate of $O(1/k^2)$, which is optimal for the class of first-order methods on smooth convex functions. Despite these advances, the computational cost of GD becomes prohibitive in large-scale settings where the objective function decomposes into a finite sum, $f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$, requiring n gradient evaluations per iteration.

1.2. Stochastic gradient descent

To tackle the computational challenges inherent in full-batch gradient calculation, [3] introduced the Stochastic Approximation method. When applied to the optimization process, this concept results in Stochastic Gradient Descent (SGD). Rather than computing the exact gradient, SGD provides an estimate by utilizing a single sample (or a small mini-batch) i_k chosen uniformly at random:

$$x_{k+1} = x_k - \eta_k \nabla f_{i_k}(x_k). \tag{2}$$

It is essential to note that the stochastic gradient serves as an unbiased estimator of the true gradient, satisfying $\mathbb{E}[\nabla f_{i_k}(x_k)] = \nabla f(x_k)$.

The convergence analysis of SGD differs fundamentally from GD due to the introduction of variance. While GD converges to the exact minimizer with a constant step size, SGD will eventually oscillate around the minimizer within a "noise ball" proportional to the variance of the gradients and the step size [4,5]. To guarantee almost sure convergence to the optimum, the step size sequence must satisfy the Robbins-Monro conditions:

$$\sum_{k=1}^{\infty} \eta_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \eta_k^2 < \infty. \tag{3}$$

Under these conditions, SGD achieves a sublinear convergence rate of $O(1/k)$ for strongly convex functions. While asymptotically slower than GD, the per-iteration cost of SGD is independent of the dataset size n, rendering it superior for large-scale problems.

1.3. Comparison with variational data assimilation methods

In the context of parameter estimation for Stochastic Differential Equations (SDEs), it is pertinent to contrast our proposed framework with widely adopted variational data assimilation methods, such as weak-constraint 4D-Var and stochastic optimal control [6,7]. Variational approaches typically retain the stochastic nature of the system, formulating the inverse problem by treating the intrinsic process noise as an additional control variable. Under this paradigm, finding the optimal parameters and state trajectories requires minimizing a cost functional that balances the discrepancy with the observations and the magnitude of the model error. However, performing this optimization generally relies on the continuous adjoint method, which entails computationally demanding forward and backward time integrations of the stochastic dynamics [8].

In contrast, the methodology presented in this study fundamentally circumvents the need for backward stochastic integration. By leveraging the Wiener Chaos Expansion (WCE), we project the stochastic solution onto a basis of Hermite polynomials, effectively determinizing the system. As previously described, the stochastic dynamics are decomposed into a hierarchical system of deterministic ordinary differential equations, referred to as the propagator. Consequently, the stochastic parameter learning task is rigorously reduced to a deterministic optimization problem. This spectral transformation allows us to compute gradients analytically and apply Stochastic Gradient Descent (SGD) directly on the propagator. By doing so, our framework avoids the severe computational burden and the complex non-convexity landscapes often associated with traditional simulation-based or adjoint-based variational methods, while ensuring robust convergence properties for the parameter estimation.

1.4. Overview of the proposed methodology

While the preceding sections outline the foundational mathematical concepts, the primary contribution of this manuscript is the synthesis of these distinct tools into a unified, highly efficient framework for SDE parameter estimation. Specifically, we address the challenge of inferring the unknown drift and diffusion parameters, denoted by θ, from discrete noisy measurements. This is a critical task when modeling highly non-linear dynamics, such as those governing individual biological growth.

Traditional inference frameworks for SDEs, such as Maximum Likelihood Estimation (MLE) and Markov Chain Monte Carlo (MCMC) methods, often rely on intensive Monte Carlo simulations to approximate the transition densities or the likelihood surface. These approaches become computationally prohibitive as the complexity of the system or the dimension of the parameter space

increases. In contrast, our framework leverages the Wiener Chaos Expansion to transform the stochastic inverse problem into a hierarchical system of deterministic equations. By doing so, we eliminate the need for repetitive path simulations, providing a more efficient optimization landscape for Stochastic Gradient Descent (SGD) to operate on.

Instead of relying on standard maximum likelihood estimators or particle-based filters, our approach reformulates the problem through three fundamental steps:

1. **Spectral Determinization:** Rather than dealing with random trajectories directly, we apply the Wiener Chaos Expansion to the underlying SDE. By projecting the stochastic process onto an orthogonal basis of Hermite polynomials, we truncate and transform the original stochastic system into a coupled, hierarchical system of deterministic Ordinary Differential Equations (ODEs). The resulting deterministic system is what we refer to as the *propagator*.
2. **Formulation of the Discrepancy Functional:** We cast the parameter inference as a deterministically driven inverse problem. We construct a Tikhonov-Phillips regularized cost functional that quantifies the error between the statistical properties of the experimental data (specifically, the empirical mean and variance) and the corresponding zeroth and higher-order states of our deterministic propagator.
3. **Optimization via SGD:** To minimize this functional and recover the optimal parameters θ, we analytically compute the gradients directly from the propagator equations. We then deploy a Stochastic Gradient Descent algorithm to iteratively navigate the optimization landscape.

By executing this sequence, we effectively translate a complex stochastic inference problem into a classical machine learning optimization task. This paradigm shift not only avoids the computational bottlenecks of Monte Carlo simulations but also leverages the robust convergence properties of gradient-based methods on deterministic, regularized loss landscapes.

The remainder of this manuscript is organized as follows. In Section 2 we briefly review the inverse problem and in particular the ill-posed inverse problems. Section 3 formalizes the theoretical framework for differential operators subjected to noise, establishing the stochastic inverse problem as a well-posed minimization task grounded in Tikhonov-Phillips regularization. Section 4 bridges this abstract theory with classical models, demonstrating the construction of the deterministic propagator and the corresponding objective functions for both SDEs and Stochastic Partial Differential Equations (SPDEs). Section 5 details the core algorithmic contribution: it introduces a modified empirical loss function incorporating the quadratic variation of the data, explicitly computes the analytical gradients from the propagator, and discusses the convergence guarantees of the SGD algorithm utilizing the Barzilai-Borwein step-size method. Section 6 validates the proposed framework through numerical experiments on a foundational toy model, the Ornstein-Uhlenbeck process. In Section 6.2 we apply and validate the proposed method to three examples of SDEs. Finally, Section 7 extends the methodology to more complex, non-linear SDEs with practical relevance, including Geometric Brownian Motion, and the stochastic logistic equation for biological growth.

Supplementary Material: For the sake of brevity and to maintain a sharp focus on the practical implementation and numerical validation of the method, the comprehensive background on Malliavin calculus, and the derivations of the Wiener Chaos Expansion have been omitted from the main text. Additionally, while the core methodology is focused on SDEs, the theoretical extension of this framework to Stochastic Partial Differential Equations (SPDEs) is also thoroughly explored. All these materials are available as supplementary material in our extended preprint [9].

2. On inverse problems

2.1. Inverse problems and regularization

A prominent domain where these iterative methods have found fundamental application is the regularization of ill-posed inverse problems. In this setting, the objective is to recover an unknown parameter $x \in \mathcal{G}$ from noisy measurements $y^\delta \in \mathcal{H}$, modeled by the operator equation:

$$\mathcal{A}x = y^\delta, \tag{4}$$

where $\mathcal{A} : \mathcal{G} \to \mathcal{H}$ is a forward operator (potentially non-linear). Following the classical framework of inverse problems, we assume that $\mathcal{G}$ and $\mathcal{H}$ are Hilbert spaces. Consequently, these spaces are equipped with their respective inner products, and the norms $\|\cdot\|_{\mathcal{G}}$ and $\|\cdot\|_{\mathcal{H}}$ denote the naturally induced Hilbert space norms.

According to Hadamard's definition, such problems are often *ill-posed*, meaning a solution may not exist, may not be unique, or may not depend continuously on the data. The classical approach to solving ill-posed problems involves minimizing a discrepancy functional, typically the least-squares term $\frac{1}{2}\|\mathcal{A}x - y^\delta\|^2_{\mathcal{H}}$.

To address the instability caused by data noise, this minimization is usually augmented with a regularization term. Specifically, the *Tikhonov-Phillips regularization* searches for a solution x_δ that minimizes the following functional:

$$J_\delta(x) = \frac{1}{2}\|\mathcal{A}x - y^\delta\|^2_{\mathcal{H}} + \frac{\delta}{2}\|x - x_0\|^2_{\mathcal{G}}, \tag{5}$$

where $\delta > 0$ is the regularization parameter that balances the data fit and the penalty term, and $x_0 \in \mathcal{G}$ is a prior estimate (often set to zero). The choice of the norm $\|\cdot\|_{\mathcal{G}}$ ensures that the regularized solution is penalized according to the specific geometric or smoothness properties of the space $\mathcal{G}$.

It is well-established that applying Gradient Descent to this functional yields the celebrated *Landweber iteration* [10], a method known to exhibit a self-regularizing property where the iteration count serves as the regularization parameter (semiconvergence) .

However, for large-scale tomographic or seismic imaging problems where the operator A consists of distinct row blocks representing partial data subsets, computing the full gradient is computationally prohibitive. Here, Stochastic Gradient Descent emerges naturally as the *Randomized Kaczmarz method* (or Algebraic Reconstruction Technique) [11]. By selecting random subsets of equations to update the estimate, SGD acts as an efficient iterative solver that drastically reduces the per-iteration computational cost while maintaining convergence guarantees suitable for large-scale inverse problems [12].

2.2. Some statistical ILL-posed inverse problems

In this subsection and for the sake of completeness, we will review some classical inverse problems. There is a high number of good references for the study of the inverse problems, here we refer to: [13,14] and the references therein.
Set two separable Hilbert spaces $\mathcal{H}$ and $\mathcal{G}$, and define an operator A from $\mathcal{G}$ into $\mathcal{H}$, $A : \mathcal{G} \to \mathcal{H}$.

We consider the problem

$$Ax = y \tag{6}$$

with $y \in \mathcal{H}$ and $x \in \mathcal{G}$. Here, we are assuming that A is a (not necessarily bounded) linear operator acting from a subset X of $\mathcal{G}$ into a subset Y of $\mathcal{H}$. The inverse problem focuses on finding $x \in X$ given y; sometimes y is called the *data*.

Remark 1. In this work, we will be interested in the case where the operator A is a differential operator.

The problem (6) is called *well-posed (in the sense of Hadamard)* on $\mathcal{H}$ and $\mathcal{G}$ if the following three conditions hold:

1. for any $y \in \mathcal{H}$ there exists a solution $x_e \in \mathcal{G}$ to the Eq. (6), i.e., $Range(A) = \mathcal{H}$ (the existence condition);
2. the solution x_e to the Eq. (6) is unique in $\mathcal{G}$ (the uniqueness condition);
3. for any neighbourhood $\mathcal{O}(x_e) \subset \mathcal{G}$ of the solution x_e to the Eq. (6), there is a neighbourhood $\mathcal{O}(y) \subset \mathcal{H}$ of y such that for all $y_\delta \in \mathcal{O}(y)$ the element $A^{-1}y_\delta = x_\delta$ belongs to the neighbourhood $\mathcal{O}(x_e)$, i.e., the operator A^{-1} is continuous (the stability condition).

There are three causes leading to ill-posedness of the Eq. (6):

(i) $Range(A) = \overline{Range(A)} \neq \mathcal{H}$.
(ii) $kernel(A) \neq \{0\}$
(iii) The range of A is not closed in $\mathcal{H}$.

For some discussion on the issues and the possible solutions to each cause see some of the references mentioned before.
The minimization problem (5) is an example of what is called a Tikhonov-Phillips regularizer (cf. [14]).

Definition 1. Let $\delta > 0$ be a given constant. The Tikhonov–Phillips regularized solution $x_\delta \in \mathcal{H}$ is the minimizer of the functional

$$F_\delta(x) = \|Ax - y\|_Y^2 + \delta\|x\|_X^2,$$

provided that a minimizer exists. The parameter $\delta > 0$ is called the regularization parameter.

Remark 2. Note that in (5) by setting $x_0 = 0$, the problem simplifies to finding the solution with the minimum norm (or energy) that fits the data. This is the canonical form of Tikhonov regularization used when no specific prior estimate x_0 is available.

3. Differential operators with a noise

3.1. Framework

Denote by $\mathcal{H}$ a separable Hilbert space, finite or infinite dimensional.

Let $(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{H})$ be a Gaussian probability space. That means that $\mathcal{H}$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ whose elements are zero-mean Gaussian random variables; in such manner that we could define a isonormal Gaussian process $W = \{W(h) : h \in \mathcal{H}\}$. That means W is a centered Gaussian family of random variables such that $\mathbb{E}(W(h)W(g)) = \langle h, g\rangle_{\mathcal{H}}$ for all $h, g \in \mathcal{H}$, (cf. [15]).

We will be interested in the white noise case. That is, when we could interpret the elements of $\mathcal{H}$ as stochastic integrals of functions in $L^2(T, \mathcal{B}(T), \mu)$ with respect to a Gaussian measure on the parameter space T (where $\mathcal{B}(T)$ is the Borel σ-algebra of the set T) with respect to a Gaussian measure μ on the parameter space T, μ is a σ-finite measure without atoms (cf. [15]). Thus, by taking $\mathcal{H}$ finite or infinite dimensional, we could study SDEs and SPDEs using this framework.

We define the Hilbert space $\mathbb{H} = L^2(\Omega; \mathcal{H})$, with norm given by

$$\|x\|_{\mathbb{H}}^2 := \mathbb{E}\left(\|x\|_{\mathcal{H}}^2\right)$$

In this section, we consider the stochastic extension of the inverse problem (6), formulated as the operator equation:

$$\mathcal{A}x = y + \mathcal{B}\eta, \tag{7}$$

where $y \in \mathcal{H}$ represents the observation (or the mean data), $A : \mathcal{G} \to \mathcal{H}$, and $\mathcal{B}\eta$ represents the stochastic perturbation, with η being an isonormal process representing observational noise.

On the Solution: Since the source term involves the random variable η, the solution x is inherently stochastic. Therefore, we seek a solution in the space of square-integrable $\mathcal{G}$-valued random variables, denoted by $\mathbb{G} = L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{G})$. The equality in (7) is understood in the mean-square sense; however, due to the ill-posed nature of the operator $\mathcal{A}$, an exact solution satisfying the equation almost surely may not exist. Consequently, we define the *generalized solution* to (7) as the unique minimizer of the regularized expected discrepancy functional:

$$x_{sol} := \operatorname*{argmin}_{u\in\mathbb{G}} \frac{1}{2}\|\mathcal{A}u - (y + \mathcal{B}\eta)\|^2_{\mathbb{H}}. \tag{8}$$

This formulation naturally leads to the stochastic optimization problem (8), where the solution is an estimator characterized by its Wiener Chaos expansion.

The kind of problems like Eq. (7) cover examples such as stochastic differential equations and stochastic partial differential equations.

We will assume the following set of hypotheses.

Hypothesis 1.

a) Assume $\mathcal{A} : \mathcal{G} \to \mathcal{H}$ is a linear bounded operator and takes values in $\mathcal{H}$.
b) The mean value of $\mathcal{A}x$ is equal to Ax from the deterministic problem (6).

Remark 3.

- Observe that given the stochasticity of η in Eq. (7), the value x that solves the inverse problem is stochastic too. Therefore, we need to consider a solution of the inverse problem that lives in a suitable space of random variables.

Taking expectation in Eq. (7) we have $\mathbb{E}(\mathcal{A}x) = y$. Thus the stochastic problem has as a particular case the deterministic problem (6).

Relying on standard concepts from Malliavin calculus, we assume the following hypotheses.

Hypothesis 2.

1. We have the Wiener-chaos expansion for $\mathcal{A}x$:

$$\mathcal{A}x = \sum_{\alpha\in\mathcal{J}} (\mathcal{A}x)_\alpha\, \xi_\alpha \qquad \text{with } (\mathcal{A}x)_\alpha := \mathbb{E}\big[(\mathcal{A}x)\,\xi_\alpha\big].$$

where $\mathcal{J}$ is the set of multi-indices of finite length and ξ_α are the Fourier-Hermite polynomials, as is standard in Malliavin calculus (see [15]).

Using the fundamental property of the Wiener chaos expansion that the basis polynomials are centered for $|\alpha| > 0$ (i.e., $\mathbb{E}[\xi_\alpha] = 0$), jointly with Hypothesis 1-b), we have that

$$\mathbb{E}(\mathcal{A}x) = \sum_{\alpha\in\mathcal{J}} (\mathcal{A}x)_\alpha\, \mathbb{1}_{|\alpha|=0} =: (\mathcal{A}x)_0 \tag{9}$$

3.2. Stochastic inverse problem

Consider the optimization problem for the stochastic Eq. (7):

$$v(x) = \operatorname{argmin}_{x\in X} \frac{1}{2}\Big\|y - \mathcal{A}x\Big\|^2_{\mathbb{H}}. \tag{10}$$

It is worth noting that the noise term $\mathcal{B}\eta$ from (7) does not explicitly appear inside the residual norm of the optimization functional (10). Since η represents an unobservable stochastic realization, it cannot be directly evaluated within a practical data-mismatch criterion. However, this formulation is rigorously justified under expectation; by invoking the zero-mean property of the isonormal Gaussian process ($\mathbb{E}[\eta] = 0$), the stochastic cross-terms vanish identically, i.e., $\mathbb{E}[\langle \mathcal{A}z, \mathcal{B}\eta\rangle_{\mathcal{H}}] = 0$ for all z. Indeed,

$$\mathbb{E}\big[\|y - \mathcal{A}x\|^2_{\mathcal{H}}\big] = \|\mathcal{A}(x_{\text{true}} - x)\|^2_{\mathcal{H}} + \mathbb{E}\big[\|\mathcal{B}\eta\|^2_{\mathcal{H}}\big] + 2\,\mathbb{E}\big[\langle \mathcal{A}(x_{\text{true}} - x), \mathcal{B}\eta\rangle_{\mathcal{H}}\big]$$

and then the last term is equal to zero. Furthermore, the quadratic noise contribution $\mathbb{E}[\|\mathcal{B}\eta\|^2_{\mathcal{H}}]$ acts merely as an additive constant that does not affect the argmin of the functional, establishing a strict equivalence between the deterministic residual minimization and the underlying stochastic energy landscape.

We have the main result of this section.

Theorem 1. *Assume hypotheses 1 and 2 are satisfied. Thus, the inverse problem (10) is well-posed.*

Proof. We will demonstrate that the stochastic minimization problem (10) is mathematically equivalent to a Tikhonov-Phillips regularization for a deterministic problem, which guarantees its well-posedness.

Let $\mathbb{H}$ denote the appropriate space of square-integrable random variables with values in $\mathcal{H}$. Using the Wiener Chaos Expansion, the discrepancy functional can be expanded as follows:

$$\begin{aligned}\left\|\mathcal{A}x - y\right\|_{\mathbb{H}}^2 &= \mathbb{E}\left[\left\|\mathcal{A}x - y\right\|_{\mathcal{H}}^2\right] = \mathbb{E}\left[\left\|\sum_{\alpha\in\mathcal{J}}(\mathcal{A}x)_\alpha\xi_\alpha - y\right\|_{\mathcal{H}}^2\right] \\ &= \mathbb{E}\left[\left\|\Big((\mathcal{A}x)_{\mathbf{0}} - y\Big) + \sum_{\alpha\in\mathcal{J}\,:\,|\alpha|\geq 1}(\mathcal{A}x)_\alpha\xi_\alpha\right\|_{\mathcal{H}}^2\right],\end{aligned} \tag{11}$$

where we have used the fact that for the multi-index $|\alpha| = 0$, the basis polynomial is $\xi_{\mathbf{0}} \equiv 1$.

Expanding the square of the norm yields three terms:

$$\left\|\mathcal{A}x - y\right\|_{\mathbb{H}}^2 = \left\|(\mathcal{A}x)_{\mathbf{0}} - y\right\|_{\mathcal{H}}^2 + \mathbb{E}\left[\left\|\sum_{\alpha\in\mathcal{J}\,:\,|\alpha|\geq 1}(\mathcal{A}x)_\alpha\xi_\alpha\right\|_{\mathcal{H}}^2\right] + 2\sum_{\alpha\in\mathcal{J}\,:\,|\alpha|\geq 1}\left\langle(\mathcal{A}x)_{\mathbf{0}} - y, (\mathcal{A}x)_\alpha\right\rangle_{\mathcal{H}}\mathbb{E}[\xi_\alpha]. \tag{12}$$

By the properties of the WCE basis, the random variables ξ_α are centered for all $|\alpha| \geq 1$, meaning $\mathbb{E}[\xi_\alpha] = 0$. Consequently, the cross-term vanishes completely.

Furthermore, due to the orthogonality of the basis $\{\xi_\alpha\}_{\alpha\in\mathcal{J}}$, we have $\mathbb{E}[\xi_\alpha\xi_\beta] = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and 0 otherwise. Applying this isometry to the variance term, the expectation of the squared norm simplifies to the sum of the squared norms. Assuming the basis is normalized such that $\mathbb{E}[\xi_\alpha^2] = 1$, we obtain:

$$\left\|\mathcal{A}x - y\right\|_{\mathbb{H}}^2 = \underbrace{\left\|(\mathcal{A}x)_{\mathbf{0}} - y\right\|_{\mathcal{H}}^2}_{\text{Data Fidelity}} + \underbrace{\sum_{\alpha\in\mathcal{J}\,:\,|\alpha|\geq 1}\left\|(\mathcal{A}x)_\alpha\right\|_{\mathcal{H}}^2}_{\text{Regularization Penalty}}. \tag{13}$$

The right-hand side of Eq. (13) takes the exact form of a Tikhonov-Phillips regularized functional for the deterministic forward operator $x \mapsto (\mathcal{A}x)_{\mathbf{0}}$. The first term enforces data fidelity by matching the mean dynamics to the deterministic observations y, while the infinite sum of higher-order chaos coefficients acts as a strongly convex regularization penalty that minimizes the stochastic variance (or energy) of the system. By invoking classical regularization theory for deterministic inverse problems [see, e.g.,14], this structure ensures the existence, uniqueness, and stability of the minimizer, concluding the proof. □

Remark 4. Remark: It is important to note that for the models discussed in Section 4, the parameter-to-solution map $\mathcal{A}(\theta)$ is generally non-linear. Consequently, the global strict convexity and uniqueness stated in Theorem 1 do not hold globally. In these non-linear inverse problems, the objective function is typically only locally strictly convex in a neighborhood of the true parameters, which is sufficient for gradient-based methods like SGD to converge given an appropriate initialization

While the isometric properties of the Wiener Chaos Expansion are standard in stochastic analysis, Theorem 1 explicitly bridges this spectral property with the deterministic regularization theory of inverse problems, demonstrating that the higher-order stochastic fluctuations naturally induce a Tikhonov-Phillips penalty.

Consequently, this theorem serves as the theoretical cornerstone of our methodology. By formally establishing that the spectral projection of the stochastic system inherently regularizes the inverse problem, we justify the transition from a computationally prohibitive stochastic optimization to a tractable deterministic one. Optimizing the resulting deterministic propagator via Stochastic Gradient Descent (SGD) is therefore not merely a heuristic for computational efficiency, but a mathematically rigorous strategy. It guarantees that the parameter estimation remains well-posed and remarkably stable against observational noise, effectively circumventing the severe instability and exhaustive sampling requirements typically associated with direct simulation-based inference.

3.3. Well-posedness of the minimization problem

The minimization of the functional defined in (10) constitutes a variational approach to the inverse problem. To establish the well-posedness of this formulation—specifically the existence, uniqueness, and stability of the minimizer—we draw upon recent advances in Bayesian inverse problems in Banach spaces.

As demonstrated by [16], the regularized inverse problem can be interpreted as determining the mode Maximum A Posteriori (MAP estimator) of a posterior measure. They established that under general Banach space priors, the posterior distribution satisfies a Lipschitz continuity condition with respect to the observed data in the Hellinger metric. This result is crucial for our framework, as the variance term in (13) , $\sum_{|\alpha|\geq 1}\|(\mathcal{A}u)_\alpha\|^2$, acts as a strongly convex regularizer in the probability space.

Furthermore, it was proved in [16] that such formulations exhibit *posterior contraction*, meaning that as the observational noise vanishes (or the amount of data increases), the probability mass of the solution concentrates around the true parameter value. This theoretical guarantee justifies the use of Eq. (13) not merely as a heuristic, but as a well-posed mathematical problem where the stochastic "variance penalty" ensures the stability of the reconstruction even when the forward operator is non-linear.

4. Classical examples of differential operators

4.1. Stochastic differential equations

Suppose that a process X_t is given by the stochastic differential equation (SDE)

$$\begin{aligned} dX(t) &= f(X_t)dt + \sigma(X(t))dB_t, \\ X(0) &= x_0. \end{aligned} \tag{14}$$

We assume that the drift f and the diffusion coefficient σ satisfy suitable conditions to ensure the existence and uniqueness of a *strong solution* to the SDE (14). Specifically, we assume that f and σ satisfy the standard global Lipschitz continuity and linear growth bounds.

In addition, since our goal is to estimate the unknown parameters of the SDE (14), we explicitly denote the dependence of the functions f and σ on these parameters. More precisely, we suppose that $f(x) = f(x;\theta_1)$ and $\sigma(x) = \sigma(x;\theta_2)$, where $\Theta = (\theta_1, \theta_2)$ are the parameters targeted by the inverse problem.

We assume the following structural form:

Hypothesis 3. The functions f and σ can be factorized as:

$$f(x;\Theta) = F(x)g_1(\theta_1), \quad \sigma(x;\Theta) = \Gamma(x)g_2(\theta_2), \tag{15}$$

where g_1 and g_2 are differentiable functions with respect to the parameters θ_1 and θ_2, respectively. Furthermore, to guarantee the well-posedness of the strong solution, the state-dependent functions F and Γ are assumed to be globally Lipschitz continuous and satisfy the linear growth condition.

4.1.1. Wiener Chaos expansion for the SDE

In this section, we present a procedure to obtain the propagator for the SDE (14). We know that the solution X belongs to the space $L^2(\Omega;[0,T])$ then, we can write the solution using the standard Wiener Chaos Expansion:

$$X(t) = \sum_{m\in\mathcal{J}} X_m(t)\xi_m \tag{16}$$

where $X_m(t) = \mathbb{E}[X(t)\xi_m]$, with ξ_m being the Fourier-Hermite polynomials and $\mathcal{J}$ the set of multi-indices of finite length, as defined previously. The series in (16) converges in $L^2(\Omega)$. Thus, we have

$$\begin{aligned} \mathbb{E}\left[\Big(X(t)\Big)\xi_m\right] &= \mathbb{E}\left[\left(x_0 + \int_0^t f(X(s))ds + \int_0^t \sigma(X(s))dB(s)\right)\xi_m\right] \\ &= \mathbb{E}\left[x_0\xi_m\right] + \mathbb{E}\left[\left(\int_0^t f(X(s))ds\right)\xi_m\right] + \mathbb{E}\left[\left(\int_0^t \sigma(X(s))dB(s)\right)\xi_m\right] \\ &= x_0\mathbb{1}_{|m|=0} + \int_0^t g_1(\theta_1)\mathbb{E}\Big[F(X(s))\xi_m\Big]ds + \mathbb{E}\left[\left(\int_0^t \sigma(X(s))dB(s)\right)\xi_m\right] \\ &= x_0\mathbb{1}_{|m|=0} + g_1(\theta_1)\int_0^t F_m(X(s))ds + g_2(\theta_2)\mathbb{E}\left[\left(\int_0^t \Gamma(X(s))dB(s)\right)\xi_m\right]. \end{aligned} \tag{17}$$

For the last integral in (17), we rely on the standard properties of the Skorohod integral and Hypothesis 3. Indeed, we know that the integrand is adapted to the generated filtration $\mathcal{F}_t$ generated by the noise $B(t), 0 \le t \le T$. Thus, the Skorohod integral is reduced to the usual Itô stochastic integral. In other words, we can write the following

$$\int_0^t \Gamma(X(s))dB_s = \delta\Big(\Gamma(X(\cdot))\mathbb{1}_{[0,t]}(\cdot)\Big)$$

Using this expression alongside the duality relationship of the Skorohod integral and the action of the Malliavin derivative as a lowering operator, we obtain:

$$\begin{aligned} \mathbb{E}\left[\Big(\int_0^t \Gamma(X(s))dB_s\Big)\xi_m\right] &= \mathbb{E}\left(\delta\Big(\Gamma(X(\cdot))\mathbb{1}_{[0,t]}(\cdot)\Big)\xi_m\right) = \mathbb{E}\Big(\big\langle\Gamma(X(\cdot))\mathbb{1}_{[0,t]}(\cdot), D\xi_m\big\rangle_{\mathbb{H}}\Big) \\ &= \mathbb{E}\left[\int_0^t \Gamma(X(s))D\xi_m ds\right] = \sum_{i=1}^{\infty}\sqrt{m_i}\int_0^t e_i(s)\mathbb{E}\Big[\xi_{m^-(i)}\sigma(X(s))\Big]ds \\ &= \sum_{i=1}^{\infty}\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^-(i)}(X(s))ds \end{aligned} \tag{18}$$

Plugging together (17)-(18) we have

$$X_m(t) = x_0\mathbb{1}_{\{|m|=0\}} + g_1(\theta_1)\int_0^t F_m(X(s))ds + g_2(\theta_2)\sum_{i=1}^{\infty}\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^-(i)}(X(s))ds \tag{19}$$

Remark 5 (Non-linearity and the Zero-th Chaos Mode)**.** It is important to highlight a structural nuance regarding the zero-th chaos mode, $X_{\mathbf{0}}(t) = \mathbb{E}[X(t)]$, derived from the expectation of the SDE. The assumption that the expected trajectory exactly matches the solution of the underlying deterministic ODE (i.e., the system without noise) relies on the property of the Itô integral, which vanishes in expectation.

However, strictly speaking, this exact correspondence holds only when the drift function $f(x;\Theta)$ is linear affine. For instance, if $f(x) = ax$, then the expectation trivially commutes, yielding $\mathbb{E}[f(X(s))] = aX_{\mathbf{0}}(s)$. Conversely, for non-linear systems such as the stochastic logistic equation, Jensen's inequality dictates that $\mathbb{E}[f(X(s))] \neq f(\mathbb{E}[X(s)])$. In such cases, the zero-th chaos propagator equation does not trivially collapse to the classical deterministic model. Consequently, minimizing the L^2 misfit $\|X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\|^2$ strictly fits the observational data to the projected mean dynamics of the stochastic system, inherently capturing the drift's non-linear interaction with the underlying variance.

Remark 6. Note that depending of every SDE one is interested in approximating using the WCE, the Eq. (19) will be in fact a system of coupled ordinary differential equations. Such a system is called the propagator and can be solved iteratively (see [17] for instance).

Some examples are:

- $F(x) = \alpha_0 x$ and $\Gamma(x) = \alpha_1 + \alpha_2 x$, then, if $\alpha_1 = 0$, $\alpha_2 > 0$ then we get the geometric Brownian motion, and when $\alpha_1 > 0$,$\alpha_2 = 0$ we recover the OU process.
- $F(x) = x(1-x)$ and $\Gamma(x) = \alpha_1 x$, we obtain a multiplicative (affine) logistic SDE.

4.1.2. Inverse problem for the SDE

Let us to write $X(t;\Theta)$ instead of $X(t)$, just to denote the dependence on the parameters, recall that $\Theta = (\theta_1, \theta_2)$.

We define the objective function, $u : \mathbb{R}^2 \to \mathbb{R}$, as

$$u(\Theta) := \mathbb{E}\left(\int_0^T \left[X(t;\Theta) - \frac{1}{N}\sum_{k=1}^{N} \hat{x}^k(t) \right]^2 dt \right). \tag{20}$$

Here, $\Theta = (\theta_1, \theta_2)$ represents the parameters to be estimated, and $X^k(t;\Theta)$ denotes the solution to the SDE with the initial condition $X^k(0;\Theta) = \hat{x}^k(0)$. Furthermore, $\hat{x}^k(t)$ corresponds to the kth observed trajectory over the time interval $[0,T]$. We assume a total of N such available trajectories, which will be used to estimate Θ. We define $\bar{x}(t) = \frac{1}{N}\sum_{k=1}^{N} \hat{x}^k(t)$ as the *empirical mean trajectory of the observations*.

We now use the spectral representation (16) in (20). To make the derivation transparent, we expand the expectation of the squared error. Let $\bar{x}(t) = \frac{1}{N}\sum_{k=1}^{N} \hat{x}_k(t)$ be the empirical mean trajectory. By substituting the expansion, we obtain:

$$\begin{aligned}
u(\Theta) &= \mathbb{E}\left(\int_0^T \left[\sum_{m\in\mathcal{J}} X_m(t;\Theta)\xi_m - \bar{x}(t) \right]^2 dt \right) \\
&= \int_0^T \mathbb{E}\left(\left[\sum_{m\in\mathcal{J}} X_m(t;\Theta)\xi_m - \bar{x}(t) \right]^2 \right) dt \\
&= \int_0^T \mathbb{E}\left(\left[\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right] + \left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\xi_m \right] \right]^2 \right) dt \\
&= \int_0^T \mathbb{E}\Big(\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right]^2 \\
&\quad + 2\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right] \left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\xi_m \right] + \left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\xi_m \right]^2 \Big) dt \\
&= \int_0^T \left(\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right]^2 + 2\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right] \mathbb{E}\left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\xi_m \right] + \mathbb{E}\left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\xi_m \right]^2 \right) dt \\
&= \int_0^T \left(\left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right]^2 + \mathbb{E}\left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} \sum_{\{n\in\mathcal{J}:|n|\geq 1\}} X_m(t;\Theta)X_n(t;\Theta)\xi_m\xi_n \right] \right) dt \\
&= \int_0^T \left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\right]^2 dt + \int_0^T \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} \left[X_m(t;\Theta)\right]^2 dt,
\end{aligned} \tag{21}$$

where in the step separating the zero-th chaos mode from the higher-order modes, we have utilized the fact that $\mathbb{E}[\xi_m] = 0$ for all $|m| \geq 1$. Furthermore, the cross-term vanishes, and the final line is obtained by applying the orthogonality property $\mathbb{E}[\xi_m\xi_n] = \delta_{mn}$.

Remark 7. Observe that from (21) we can reinterpret the function $u(\Theta)$ as

$$u(\Theta) = \text{Square errors between the observations and the mean of } X + \text{Variance of the process } X. \tag{22}$$

Remark 8. We rewrite Eq. (19) as follows, to underline the dependence on the parameters $\Theta = (\theta_1, \theta_2)$, that is

$$X_m(t;\Theta) = x_0 \mathbb{1}_{\{|m|=0\}} + g_1(\theta_1)\int_0^t F_m(X(s;\Theta))ds + g_2(\theta_2)\sum_{i=1}^{\infty}\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^{-(i)}}(X(s;\Theta))ds \tag{23}$$

Remark 9. It is worth noting that by employing the WCE, the original stochastic minimization problem is effectively reformulated as a deterministic optimization task. Consequently, the objective function is now evaluated through the evolution of the deterministic propagator equations.

5. Learning parameters for SDEs

5.1. Stochastic differential equations

Suppose that a process $X(t)$ is given by the stochastic differential equation (SDE)

$$\begin{aligned} dX(t) &= f(X(t))dt + \sigma(X(t))dB_t, \\ X(0) &= x_0. \end{aligned} \tag{24}$$

We assume that the functions f and σ satisfy suitable assumptions to ensure the existence and uniqueness of a solution. In particular, we will assume that f and σ are globally Lipschitz and with linear growth; thus, we will have that $X \in L^2(\Omega;[0,T])$. In addition, since we are looking to estimate some parameters from the SDE (14) then we will make explicit the dependence of the functions f and σ on some parameters. More precisely, we suppose that $f(x) = f(x;\theta_1)$ and $\sigma(x) = \sigma(x,\theta_2)$. Here, (θ_1,θ_2), are the parameters we are interested in.

Using the Hypotheses 3: $f(x;\Theta) = F(x)g_1(\theta_1)$, $\sigma(x;\Theta) = \Gamma(x)g_2(\theta_2)$ we have that

$$\frac{\partial f(x;\Theta)}{\partial \theta_2} = 0, \qquad \text{and} \qquad \frac{\partial \sigma(x,\Theta)}{\partial \theta_1} = 0. \tag{25}$$

Note that since we are interested in estimating the parameters θ_1 and θ_2, thus we assume that the functions $g_1(\theta_1)$ and $g_2(\theta_2)$ are independent of time and deterministic.

5.2. Wiener Chaos expansion for the SDE

It is recalled that the SDE (14) admits the following representation:

$$X(t) = \sum_{m\in\mathcal{J}} X_m(t)\xi_m \tag{26}$$

where $X_m(t) = \mathbb{E}\big(X(t)\xi_m\big)$, solves

$$X_m(t) = x_0\mathbb{1}_{\{|m|=0\}} + g_1(\theta_1)\int_0^t F_m(X(s))ds + g_2(\theta_2)\sum_{i=1}^{\infty}\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^{-(i)}}(X(s))ds \tag{27}$$

5.3. Learning parameters using the WCE

From this point we write $X(t;\Theta)$ instead of $X(t)$, to denote the dependence on the parameters, recall that $\Theta = (\theta_1,\theta_2)$.

We are interested in estimating the parameters $\Theta = (\theta_1,\theta_2)$ using the stochastic gradient descent (SGD) method (see [2,5] for instance)

The SGD method usually estimates the parameters by minimizing an objective function. Suppose that we have observed N trajectories of the process X in the interval $[0,T]$ and denote the kth trajectory by $\hat{x}^k(t)$.

We will assume that, for $T > 0$, we have observed $\hat{x}^k(t)$ in the interval $[0,T]$ in times $0 = t_0 < t_1 \ldots, t_n = T$ with n large enough such that, for $i = 1,\ldots,n$, $\Delta_n := t_i - t_{i-1} = \frac{T}{n}$ is close to zero.

Define the quadratic variation of $\hat{x}^k(t)$, up to time t, as

$$\langle \hat{x}^k, \hat{x}^k\rangle_t := \sum_{i=1}^{n^*}[x^k(t_i) - x^k(t_{i-1})]^2.$$

where n^* is such that $t_{n^*} \le t < t_{n^*+1}$.

In a similar vein, we also define the process

$$\langle \hat{x}^k\rangle_t^2 := \frac{\Delta_n}{2}\sum_{i=1}^{n^*}\left([x^k(t_i)]^2 + [x^k(t_{i-1})]^2\right) \approx \int_0^t [x^k(t)]^2 dt.$$

We recall the *objective function*, also called the *loss function*, $\hat{u} : \mathbb{R}^p \to \mathbb{R}$, as

$$\hat{u}(\Theta) := \mathbb{E}\left(\int_0^T \left[X(t;\Theta) - \bar{x}(t) \right]^2 dt \right), \tag{28}$$

where $\Theta = (\theta_1, \dots, \theta_p)$ are the parameters to be estimated and $X(t;\Theta)$ is the SDE with initial condition $X(0;\Theta) = \hat{x}(0)$ which is the initial condition for all the observed trajectories.

By recalling (21), we have the foundational formulation. However, to accurately estimate the diffusion parameter, this L_2 mean-trajectory formulation is theoretically insufficient. From a theoretical standpoint in stochastic analysis, it is a well-known result that the drift (mean behavior) and the diffusion (variance/quadratic variation) of an SDE operate on fundamentally different scales)(see, e.g., [18], where the drift and diffusion are explicitly decoupled using the martingale property and Itô isometry, respectively).

An L_2 discrepancy function based solely on the mean trajectory is often sufficient to estimate the drift parameters, but it is theoretically inadequate to identify the diffusion parameter σ, as the diffusion signature is inherently hidden in the path fluctuations. Therefore, a variance-matching criterion is strictly required to capture these structural fluctuations. To address this, we introduce the following variance-matching objective function:

$$u(\Theta) = \left(\int_0^T \left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t) \right]^2 dt \right) + \left(\int_0^T \left[\sum_{\{m \in \mathcal{J} : |m| \geq 1\}} \left[X_m(t;\Theta) \right]^2 - \frac{1}{N} \sum_{k=1}^{N} \frac{\langle \hat{x}_k, \hat{x}_k \rangle_t}{\langle \hat{x}_k \rangle_t^2} \right]^2 dt \right) \tag{29}$$

It is crucial to note that the structural shift to the criterion defined in Eq. (29) is fundamentally motivated by stochastic theory rather than mere numerical convenience. Separating the drift fitting from the diffusion fitting is necessary to prevent parameter confounding during the numerical optimization process, explicitly forcing the model's energy fluctuations to align with the empirical quadratic variation observed in the data.

This paper will address the case of $p = 2$, which means we will estimate two parameters, however, transitioning to the general case is a straightforward process.

Remark 10. It's important to highlight that the recent modification in the last term, ensures the model's energy fluctuations align with the empirical quadratic variation observed in the data.

Remark 11. To underline the dependence on the parameters $\Theta = (\theta_1, \theta_2)$, we rewrite the Eq. (19) as follows,

$$X_m(t;\Theta) = x_0 \mathbb{1}_{\{|m|=0\}} + g_1(\theta_1) \int_0^t F_m(X(s;\Theta))ds + g_2(\theta_2) \sum_{i=1}^{\infty} \sqrt{m_i} \int_0^t e_i(s) \Gamma_{m^-(i)}(X(s;\Theta))ds. \tag{30}$$

Where $F_m(X(s;\Theta)) = \mathbb{E}\left[F(X(s;\Theta))\, \xi_m \right]$ and $\Gamma_m(X(s;\Theta)) = \mathbb{E}\left[\Gamma(X(s;\Theta))\, \xi_{m^-(i)} \right]$, where $m^-(i)$ denotes the multi-index with the i-th component reduced by one.

Remark 12. Observe that by using the WCE we have now the problem of minimizing the function u is now equivalent to minimizing the *deterministic* sum of equations given by (29), which is derived from the propagator.

At this point we calculate the gradient of the function $u : \mathbb{R}^2 \to \mathbb{R}$ which is defined at the point $\Theta = (\theta_1, \theta_2)$ as

$$V(\Theta) = \nabla u(\Theta) = \begin{bmatrix} \frac{\partial u}{\partial \theta_1}(\Theta) \\ \frac{\partial u}{\partial \theta_2}(\Theta) \end{bmatrix}.$$

Thus, we write the partial derivative from the last expression and we have that for $j = 1, 2$ they are

$$\begin{aligned} \frac{\partial u}{\partial \theta_j}(\Theta) = & 2\left(\int_0^T \left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t) \right] \cdot \frac{\partial X_{\mathbf{0}}(t;\Theta)}{\partial \theta_j} dt \right) \\ & + 4\left(\int_0^T \left[\sum_{\{m \in \mathcal{J} : |m| \geq 1\}} \left[X_m(t;\Theta) \right]^2 - \frac{1}{N} \sum_{k=1}^{N} \frac{\langle \hat{x}^k, \hat{x}^k \rangle_t}{\langle \hat{x}^k \rangle_t^2} \right] \cdot \sum_{\{m \in \mathcal{J} : |m| \geq 1\}} X_{\mathbf{m}}(t;\Theta) \frac{\partial X_{\mathbf{m}}(t;\Theta)}{\partial \theta_j} dt \right) \end{aligned} \tag{31}$$

For any $m \in \mathcal{J}$ we calculate $\frac{\partial X_m^k(t;\Theta)}{\partial \theta_j}$ from (30). At this point we use the Hypotheses 3; more precisely, we use (25), this let us

$$\frac{\partial X_m^k(t;\Theta)}{\partial \theta_1} = g_1'(\theta_1) \int_0^t X(s;\Theta))ds + g_1(\theta_1) \int_0^t \frac{\partial F_m(X(s;\Theta))}{\partial \theta_1} ds, \tag{32}$$

$$\frac{\partial X_m(t;\Theta)}{\partial \theta_2} = g_2'(\theta_2) \sum_{i=1}^{\infty} \sqrt{m_i} \int_0^t e_i(s) \Gamma_{m^-(i)}(X(s;\Theta))ds + g_2(\theta_2) \sum_{i=1}^{\infty} \sqrt{m_i} \int_0^t e_i(s) \frac{\partial \Gamma_{m^-(i)}(X(s;\Theta))}{\partial \theta_2} ds. \tag{33}$$

Then, using (25) we obtain the partial derivatives of u:

$$\frac{\partial u}{\partial \theta_1}(\Theta) = 2g_1'(\theta_1) \int_0^T \left[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t) \right] \cdot \left(\int_0^t F_{\mathbf{0}}(X(s;\Theta))ds \right) dt$$

$$
\begin{aligned}
&+ 2g_1(\theta_1)\int_0^T \Big[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\Big]\cdot\left(\int_0^t \frac{\partial F_{\mathbf{0}}(X(s;\Theta))}{\partial\theta_1}ds\right)dt\\
&+ 2g_1'(\theta_1)\int_0^T \left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}}\Big[X_m(t;\Theta)\Big]^2 - \frac{1}{N}\sum_{k=1}^N \frac{\langle \hat{x}^k,\hat{x}^k\rangle_t}{\langle \hat{x}^k\rangle_t^2}\right]\cdot\left(\int_0^t F_m(X(s;\Theta))ds\right)dt\\
&+ 2g_1(\theta_1)\int_0^T \left[\sum_{\{m\in\mathcal{J}:|m|\geq 1\}}\Big[X_m(t;\Theta)\Big]^2 - \frac{1}{N}\sum_{k=1}^N \frac{\langle \hat{x}^k,\hat{x}^k\rangle_t}{\langle \hat{x}^k\rangle_t^2}\right]\cdot\left(\int_0^t \frac{\partial F_m(X(s;\Theta))}{\partial\theta_1}ds\right)dt,
\end{aligned}
\tag{34}
$$

and

$$
\begin{aligned}
\frac{\partial u}{\partial\theta_2}(\Theta) &= 2g_2'(\theta_2)\int_0^T \Big[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\Big]\cdot\left(\sum_{i=1}^\infty\sqrt{0}\int_0^t e_i(s)\Gamma_{\mathbf{0}^{-(i)}}(X(s;\Theta))ds\right)dt\\
&+ 2g_2(\theta_2)\int_0^T \Big[X_{\mathbf{0}}(t;\Theta) - \bar{x}(t)\Big]\cdot\left(\sum_{i=1}^\infty\sqrt{0}\int_0^t e_i(s)\frac{\partial\Gamma_{\mathbf{0}^{-(i)}}(X(s;\Theta))}{\partial\theta_2}ds\right)dt\\
&+ 2g_2'(\theta_2)\int_0^T \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\cdot\left(\sum_{i=1}^\infty\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^{-(i)}}(X(s;\Theta))ds\right)dt\\
&+ 2g_2(\theta_2)\int_0^T \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m^k(t;\Theta)\cdot\left(\sum_{i=1}^\infty\sqrt{m_i}\int_0^t e_i(s)\frac{\partial\Gamma_{m^{-(i)}}(X^k(s;\Theta))}{\partial\theta_2}ds\right)dt\\
&= 2g_2'(\theta_2)\int_0^T \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\cdot\left(\sum_{i=1}^\infty\sqrt{m_i}\int_0^t e_i(s)\Gamma_{m^{-(i)}}(X(s;\Theta))ds\right)dt\\
&+ 2g_2(\theta_2)\int_0^T \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m(t;\Theta)\cdot\left(\sum_{i=1}^\infty\sqrt{m_i}\int_0^t e_i(s)\frac{\partial\Gamma_{m^{-(i)}}(X(s;\Theta))}{\partial\theta_2}ds\right)dt.
\end{aligned}
\tag{35}
$$

It is well-known that $u(\Theta)$ decreases fastest if one goes from, an initial, $\Theta\in\mathbb{R}^2$, in the direction of the negative Gradient of u at $\Theta - \nabla u(\Theta)$ (see for instance [2]). Thus, it follows that it is possible to start with a guess $\Theta_0\in\mathbb{R}^2$ and consider the sequence

$$\Theta_{k+1} = \Theta_k - \gamma\nabla u(\Theta_k),$$

for a small enough step size or *learning rate* $\gamma\in\mathbb{R}^+$. Then, $u(\Theta_k)\geq u(\Theta_{k+1})$

We will consider γ as in the Barzilai–Borwein method (see [19]):

$$\gamma_k = \frac{\Big|\big(\Theta_k - \Theta_{k-1}\big)^T\big[\nabla u(\Theta_k) - \nabla u(\Theta_{k-1})\big]\Big|}{\big\|\nabla u(\Theta_k) - \nabla u(\Theta_{k-1})\big\|^2}$$

With some additional hypotheses and this choice of γ, we have that the convergence to a local minimum can be guaranteed. (see [2,20,21]).

Remark 13. In our numerical implementation, the gradients of the objective function are approximated using finite differences to facilitate the optimization process.

5.3.1. Convergence of stochastic gradient

To ensure convergence using the Barzilai–Borwein method, certain hypotheses must be satisfied. Let the objective function be defined as the empirical risk $u(\Theta) = \frac{1}{N}\sum_{i=1}^N u_i(\Theta)$, where u_i are suitable convex functions that depend on the parameter Θ. Under classical assumptions, for the SGD algorithm to converge with explicit non-asymptotic bounds [2], each component u_i must satisfy the following properties:

1. The function u_i is μ-strongly convex. This means there exists $\mu > 0$ such that for all Θ and Λ:

$$u_i(\Theta)\geq u_i(\Lambda) + \nabla u_i(\Lambda)^T(\Theta - \Lambda) + \frac{\mu}{2}\|\Theta - \Lambda\|_2^2.$$

2. The gradient of u_i is L-Lipschitz continuous. This means there exists $L > 0$ such that:

$$\|\nabla u_i(\Theta) - \nabla u_i(\Lambda)\|_2 \leq L\|\Theta - \Lambda\|_2.$$

First, observe that the core component of the objective function is the L^2 error of the mean trajectory:

$$\int_0^T\left[X_{\mathbf{0}}(t;\Theta) - \frac{1}{N}\sum_{k=1}^N\hat{x}^k(t)\right]^2 dt.$$

This component is convex only if the difference $X_{\mathbf{0}}(t;\Theta) - \frac{1}{N}\sum_{k=1}^N\hat{x}^k(t)$ is positive, which occurs only in specific cases.

To verify the Lipschitz condition for the gradient, we introduce the notation $\bar{\hat{x}}(t) = \frac{1}{N}\sum_{k=1}^{N}\hat{x}^k(t)$. Evaluating the difference of the gradients for two distinct parameter vectors Θ_1 and Θ_2 yields:

$$\int_0^T \left[X_{\mathbf{0}}(t;\Theta_1) - \bar{\hat{x}}(t)\right]\frac{\partial X_{\mathbf{0}}(t;\Theta_1)}{\partial\theta_j}dt - \int_0^T \left[X_{\mathbf{0}}(t;\Theta_2) - \bar{\hat{x}}(t)\right]\frac{\partial X_{\mathbf{0}}(t;\Theta_2)}{\partial\theta_j}dt$$
$$= \int_0^T \left(\left[X_{\mathbf{0}}(t;\Theta_1) - \bar{\hat{x}}(t)\right]\left[\frac{\partial X_{\mathbf{0}}(t;\Theta_1)}{\partial\theta_j} - \frac{\partial X_{\mathbf{0}}(t;\Theta_2)}{\partial\theta_j}\right] + \left[X_{\mathbf{0}}(t;\Theta_1) - X_{\mathbf{0}}(t;\Theta_2)\right]\frac{\partial X_{\mathbf{0}}(t;\Theta_2)}{\partial\theta_j}\right)dt. \tag{36}$$

From Eq. (36), it is evident that we must assume $X_{\mathbf{0}}$ has continuous second-order derivatives with respect to Θ, and that the parameter space Θ is limited to a compact, bounded interval. Under these conditions, $X_{\mathbf{0}}$ and its derivatives are bounded. Consequently, both the difference of the states $\left|X_{\mathbf{0}}(\Theta_1) - X_{\mathbf{0}}(\Theta_2)\right|$ and the difference of their derivatives can be strictly bounded by a constant multiple of $\|\Theta_1 - \Theta_2\|_2$. By applying the triangle inequality and the Cauchy-Schwarz inequality to the integral, the gradient of the objective function satisfies the L-Lipschitz condition.

However, due to the highly non-linear nature of the parameter-to-solution map in general SDEs (such as the stochastic logistic equation presented in our applications), the associated loss function is not globally strictly convex and may exhibit multiple spurious local minima. Therefore, claiming global strong convexity is too restrictive and theoretically inaccurate for our general non-linear framework.

While global convergence cannot be theoretically guaranteed for general non-linear SDEs, empirical observations of the objective function surface reveal the existence of a clear basin of attraction. This implies that the objective function is locally strictly convex within a neighborhood of the true underlying parameters, Θ^*. Consequently, the convergence toward the physically meaningful global optimum observed in our numerical experiments relies on two critical factors:

- **Local Convexity and Initialization:** The algorithm requires a reasonably bounded initial guess close enough to the ground-truth parameters to operate within the locally convex basin of attraction.
- **SGD Escape Dynamics:** The intrinsic stochastic noise generated by the Stochastic Gradient Descent (SGD) algorithm actively helps the optimizer escape from local traps and saddle points—a phenomenon well-documented in the non-convex optimization literature [22,23].

Because we cannot guarantee a globally convex optimization landscape in the non-linear case, we rely instead on the injectivity of the structural map to ensure the parametric identifiability of the inverse problem, thereby ensuring that the global minimizer is strictly unique when it is reached.

Theorem 2. *We define the map:*

$$\Phi : \Theta \to \left(X_0(\cdot;\Theta), \sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m^2(\cdot;\Theta)\right). \tag{37}$$

The inverse problem is parametrically identifiable (i.e., the global minimizer of the loss function is strictly unique) when the map Φ is injective. Here, the loss function is defined as $u(\theta) = \|X_0(\cdot;\Theta) - \bar{x}\|^2_{L^2[0,T]} + \|\sum_{\{m\in\mathcal{J}:|m|\geq 1\}} X_m^2(\cdot;\Theta) - h\|^2_{L^2[0,T]}$ with $h(t) := \frac{1}{N}\sum_{k=1}^{N}\frac{\langle \hat{x}^k, \hat{x}^k\rangle_t}{\langle \hat{x}^k\rangle_t^2}$.

Therefore, sufficient conditions for the map Φ to be injective are:

1. *g_1 and g_2 are injective.*
2. *$X(t)$ is smooth with respect to θ_1. Additionally, $F(x)$ is non-negative (or non-positive), and $F(X(s))$ is non-zero for a set of positive measure.*
3. *$\Gamma(x) \geq c > 0$ almost everywhere for some constant c, in the support of the process $X(t)$.*
4. *The Jacobian of the map Φ is non-degenerate.*

Proof. First, assuming the functions g_1 and g_2 are injective, it is sufficient to prove the result for the underlying SDE:

$$dX(t) = \eta_1 F(X(t))dt + \eta_2\Gamma(X(t))dB(t). \tag{38}$$

We define the variation process $Y^{\eta_1}(t) = \partial_{\eta_1}X(t)$. By using the dominated convergence theorem and differentiating the equation for $X(t)$, we obtain:

$$dY^{\eta_1}(t) = \left(F(X(t)) + \eta_1 F'(X(t))Y^{\eta_1}(t)\right)dt + \eta_2\Gamma'(X(t))Y^{\eta_1}(t)dB(t), \quad Y^{\eta_1}(0) = 0. \tag{39}$$

We then introduce the integrating factor:

$$Z(t) := \exp\left(\int_0^t \left[\eta_1 F'(X(u)) - \frac{1}{2}\eta_2^2\Gamma'(X(u))^2\right]du + \int_0^t \eta_2\Gamma'(X(u))dB(u)\right). \tag{40}$$

This process satisfies the dynamics:

$$dZ(t) = Z(t)\left(\eta_1 F'(X(t))dt + \eta_2\Gamma'(X(t))dB(t)\right), \tag{41}$$
$$Z(0) = 1. \tag{42}$$

By the variation of parameters formula, the solution is given by $Y(t) = Z(t)\int_0^t Z^{-1}(s)F(X(s))ds$. If F is non-negative (or non-positive), then Y inherits the same sign as F. Consequently, $\mathbb{E}[Y(t)] > 0$ (or $\mathbb{E}[Y(t)] < 0$ depending on the sign of X) as long as F is not identically zero. This implies that the first observable $X_0(t;\eta) = \mathbb{E}[X(t)]$ is strictly monotonic with respect to η_1.

We follow a different, asymptotic approach for the variance. By applying Itô's lemma, we have:

$$X^2(t) = x_0^2 + 2\eta_1\int_0^t X(s)F(X(s))ds + \eta_2^2\int_0^t \Gamma^2(X(s))ds + 2\eta_2\int_0^t X(s)\Gamma(X(s))dB(s). \tag{43}$$

Taking the expectation and differentiating with respect to t yields $\frac{d}{dt}\mathbb{E}[X^2(t)] = 2\eta_1\mathbb{E}[X(t)F(X(t))] + \eta_2^2\mathbb{E}[\Gamma^2(X(t))]$. Applying the same procedure to the first moment gives $\frac{d}{dt}\mathbb{E}[X(t)] = \eta_1\mathbb{E}[F(X(t))]$. Thus, the derivative of the variance is:

$$\frac{d}{dt}\mathrm{Var}(X(t)) = 2\eta_1\mathrm{Cov}(X(t), F(X(t))) + \eta_2^2\mathbb{E}[\Gamma^2(X(t))]. \tag{44}$$

Because the paths are continuous and thanks to the dominated convergence theorem, taking the limit as $t \to 0^+$ makes the covariance term vanish (since the initial condition is deterministic). We obtain:

$$\lim_{t\to 0^+}\frac{d}{dt}\mathrm{Var}(X(t)) = \eta_2^2\Gamma^2(X(0)). \tag{45}$$

Given that $\mathrm{Var}(X(0)) = 0$, we can approximate the variance for small t as $\mathrm{Var}(X(t)) = \eta_2^2\Gamma^2(X(0))t + o(t)$. Suppose two parameters, η_2 and ι_2, produce the exact same variance in a given interval $[0,T]$. Then we must have:

$$\eta_2^2\Gamma^2(X(0)) = \lim_{t\to 0^+}\frac{\mathrm{Var}(X(t))}{t} = \iota_2^2\Gamma^2(X(0)). \tag{46}$$

Since by hypothesis $\Gamma(X(0)) > 0$, this strict equality implies $\eta_2 = \iota_2$. This proves injectivity with respect to the diffusion coefficient. Finally, the third hypothesis guarantees the strict positivity required for this identifiability condition, which completes the proof. □

These hypotheses can be relaxed when the drift F is an affine function, such as $F(x) = ax + b$. Under this assumption, the first moment X_0 becomes independent of the diffusion parameter θ_2 (i.e., $\partial_{\theta_2}X_0 = 0$), which yields a triangular Jacobian matrix for the map Φ. Consequently, the second and fourth conditions are inherently satisfied, and we only need to verify the third condition to guarantee parametric identifiability and the uniqueness of the global minimizer.

6. Numerical experiments and applications

In this section, we provide a numerical simulation of SDEs using its Wiener-chaos expansion as was described in previous sections. Indeed, solving the system given by Eqs. (19) , and (30), we fixed the coefficients for the Wiener-chaos expansion of the stochastic processes X.

Therefore, we present Algorithm 1, based on the definition in (29) and the stochastic gradient descent method detailed in the previous section. We apply this exact algorithm to the Ornstein–Uhlenbeck (OU) process; but in this case, we estimate only the drift parameter, while the diffusion parameter is held fixed.

Algorithm 1 Gradient descent.

Require: $\{x^k\}_{k=1}^N$ trajectories of the Process observed in the interval $[0,T]$ in times $0 = t_0 < t_1 \ldots, t_n = T$ and $\tilde{\theta}_{1,0}$ as first estimation of θ_1

1: Calculate $\tilde{\sigma} = \frac{1}{N}\sum_{k=1}^N \frac{\langle \hat{x}^k, \hat{x}^k\rangle_t}{\langle \hat{x}^k\rangle_t^2}$

2: Initialize $\Theta_2 = \Theta_1 - \nabla u(\Theta_1)$, where $\Theta_1 = \begin{bmatrix}\tilde{\theta}_{10}\\ \tilde{\sigma}\end{bmatrix}$.

3: **while** $m \le m_{\max}$ **do**

4: Calculate $\Theta_{k+1} = \Theta_k - \gamma_k\nabla u(\Theta_k)$, where $\gamma_k = \frac{\left|(\Theta_k - \Theta_{k-1})^T[\nabla u(\Theta_k) - \nabla u(\Theta_{k-1})]\right|}{\|\nabla u(\Theta_k) - \nabla u(\Theta_{k-1})\|^2}$

5: **end while**

6: **return** $\Theta_{m_{\max}}$

6.1. The Ornstein-Uhlenbeck process

Set the Hilbert space $\mathcal{H} = L^2(0,T)$ and consider the Ornstein-Uhlenbeck (OU) process given by the SDE

$$dx(t) = -ax(t)dt + \sigma dB(t), \quad \text{with initial condition } x(0) = x^0, \tag{47}$$

We have its integral representation

$$x(t) = x^0 - a\int_0^t x(s)ds + \sigma\int_0^t dB(s). \tag{48}$$

Observe that in this case $\theta_1 = a$ and $\theta_2 = \sigma$. Furthermore,

$$f(x : \theta_1) = -\theta_1 x \qquad \text{and} \qquad \sigma(x;\theta_2) = \theta_2$$

Note that for this process, the drift $F(x) = (\theta_2 - x)$ and diffusion $\Gamma(x) = 1$ satisfy the Lipschitz and linear growth conditions stated in Hypothesis 1, guaranteeing a unique strong solution.

Thus, we have that $x \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, this implies that we can write

$$x(t) = \sum_{\alpha\in\mathcal{J}} X_\alpha(t)\xi_\alpha, \qquad \text{with } X_\alpha(t) = \mathbb{E}(x(t)\,\xi_\alpha),$$

thus, using the duality between the Malliavin derivative operator D and δ we get

$$\begin{aligned}
X_\alpha(t) &= \mathbb{E}\Big[\Big(x^0 - a\int_0^t x(s)ds + \sigma\int_0^t dB(s)\Big)\xi_\alpha\Big] \\
&= x^0\mathbb{1}_{|\alpha|=0} - a\int_0^t \mathbb{E}\Big(x(s)ds\xi_\alpha\Big) + \sigma\mathbb{E}\Big[\Big(\int_0^t dB(s)\Big)\xi_\alpha\Big] \\
&= x^0\mathbb{1}_{|\alpha|=0} - a\int_0^t X_\alpha(s)ds + \sigma\mathbb{E}\Big[\big\langle \mathbb{1}_{[0,t]}(\cdot), D\xi_\alpha\big\rangle_{\mathcal{H}}\Big] \\
&= x^0\mathbb{1}_{|\alpha|=0} - a\int_0^t X_\alpha(s)ds + \sigma\int_0^t \sum_{i,j=1}^{\infty}\sqrt{\alpha_{i,j}}\,\mathbb{E}\big(\xi_{\alpha^-(i,j)}\big)ds \\
&= x^0\mathbb{1}_{|\alpha|=0} - a\int_0^t X_\alpha(s)ds + \sigma t\sum_{i,j=1}^{\infty}\sqrt{\alpha_{i,j}}\;\mathbb{1}_{|\alpha^-(i,j)|=0}.
\end{aligned} \tag{49}$$

From this expression, we now calculate the propagator for the following three cases $|\alpha| = 0$, $|\alpha| = 1$ and $|\alpha| > 1$.

For $|\alpha| = 0$ denote the propagator as $X_0 = X_{|\alpha|=0}$, observe that in this case the indicator function $\mathbb{1}_{|\alpha^-(i,j)|=0}$ is equal to zero, then from (49) we have the ordinary differential equation

$$\frac{dX_0(t)}{dt} = -aX_0(t),$$

with initial condition x^0. This means that the solution $X_0(t)$ is

$$X_0(t) = x^0\exp(-at). \tag{50}$$

For the case $|\alpha| = 1$ denote the propagator as $X_1 = X_{|\alpha|=1}$.

Then, from (49) and given that $|\alpha| = 1$ then we have $\sqrt{\alpha_i} = 1$ for some i, thus we can write for such i

$$\begin{aligned}
\frac{dX_1(t)}{dt} &= -aX_1(t) + \sigma\sum_{i=1}^{\infty}\sqrt{\alpha_i}\;e_i(t)\mathbb{1}_{\{i=1\}}\;\mathbb{1}_{\{|\alpha^-(i)|=0\}} \\
&= -aX_1(t) + \sigma e_i(t)
\end{aligned} \tag{51}$$

where the ODE has zero initial condition. Thus, we have that

$$X_{1_i}(t) = \sigma e^{-at}\int_0^t e^{as}e_i(s)ds. \tag{52}$$

For the case $|\alpha| > 1$ we have the ODE with zero initial condition

$$\frac{dX_\alpha(t)}{dt} = -aX_\alpha(t), \tag{53}$$

with the trivial solution $X_\alpha(t) \equiv 0$.

This implies that the WCE for the OU process could be written as

$$x(t) = \sum_{\alpha\in\mathcal{J}} X_\alpha(t)\xi_\alpha = X_0(t) + \sum_{k=1}^{\infty} X_{1_k}(t)\xi_{\alpha_k} \quad \text{with} \quad \xi_{\alpha_k} := \int_0^T e_k(s)dB(s) \tag{54}$$

6.1.1. SGD and numerical validation for the OU process

For the OU process we will use the function

$$u(\Theta) = \left(\int_0^T\left[X_{\mathbf{0}}(t;\Theta) - \frac{1}{N}\sum_{k=1}^{N}\hat{x}^k(t)\right]^2 dt\right) \tag{55}$$

to estimate the drift parameter.

We present the validation results in Table 1, where it is included the actual parameters, then the initial conditions for three different estimations for each couple of real parameters and then the final estimation.

Table 1
Validation of the proposed method: estimation results for the Ornstein-Uhlenbeck (OU) process under various initial conditions.

True Parameters (α, σ)	Initial Condition $(\tilde{\alpha}_0, \tilde{\sigma})$	Final Estimation $\tilde{\alpha}_n$
(1.70, 0.150)	(2.00, 0.13310)	1.715620
	(0.50, 0.13310)	1.663870
	(1.01, 0.13310)	1.712660
(0.29, 0.100)	(2.00, 0.08801)	0.289461
	(0.50, 0.08801)	0.289451
	(0.37, 0.08801)	0.289462
(1.26, 0.076)	(1.55, 0.06833)	1.238580
	(0.80, 0.06833)	1.233540
	(0.45, 0.06833)	1.231500

6.2. Geometric Brownian motion

Consider the SDE

$$dX_t = \alpha X_t dt + \sigma X_t dB(t). \tag{56}$$

with $X_0 = x_0$. The process solution of this equation is the so-called the Geometric Brownian Motion (GBM). This process serves as the canonical model for continuous-time stochastic processes where the variable is constrained to be strictly positive, most notably in the modeling of asset prices within mathematical finance. It was famously adopted by [24] and later underpinned the Black-Scholes option pricing framework [25].

We also note that for this process, the drift $F(x) = ax$ and diffusion $\Gamma(x) = \sigma x$ satisfy the Lipschitz and linear growth conditions stated in Hypothesis 1, guaranteeing a unique strong solution.

We have the unknown parameters α and σ, and we want to estimate them. We employ the Wiener chaos method utilizing a truncation of the WCE with $L = 1,000$ and $p = 8$.

Following [26] the propagator for this SDE is calculated as follows. For the case when $|m| = 0$, we denote by X_0 the solution of the propagator which solves the differential equation

$$X_0(t) = x_0 + a\int_0^t X_0(s)ds, \quad \text{which is given by } X_0(t) = x_0 \exp(at). \tag{57}$$

For the case $|m| \geq 1$ and using the vectors from Table 5 in [26], X_m^i satisfies

$$X_m^i(t) = a\int_0^t X_m^i(s)ds + \sigma \sum_j \sqrt{m_j^i} \int_0^t e_j(s)\, X_{m^{-}(j)}^i(s)\, ds, \tag{58}$$

where $m^{i-}(j)$ is the ith vector with the jth entry minus 1. We should numerically solve this system of coupled ODEs, in increasing order of $|m|$.

6.3. Stochastic logistic equation

The classical logistic growth model, originally proposed by Verhulst, describes the dynamics of a population $N(t)$ constrained by a carrying capacity K. For any positive initial condition $P(0) > 0$, the population monotonically approaches the equilibrium state, K, ensuring persistence. However, biological systems are inevitably subject to environmental fluctuations—such as temperature variations or resource availability—that render the growth rate time-dependent and stochastic. To model this, the parameter r is typically perturbed by a white noise term, $r \to r + \sigma \dot{B}(t)$, where σ denotes the noise intensity and $\dot{B}(t)$ is Gaussian white noise. This transformation leads to the *Stochastic Logistic Differential Equation* (59) which is interpreted in the Itô sense:

$$\left.\begin{aligned} dP(t) &= rP(t)\Big[1 - P(t)\Big]dt + \sigma P(t) dB(t), \quad t > 0 \\ P(0) &= p_0. \end{aligned}\right\} \tag{59}$$

where r represents the intrinsic growth rate. In addition, the diffusion term $\sigma P(t) dB(t)$ is multiplicative (linear in $P(t)$), reflecting that environmental noise has a greater absolute impact on larger populations while vanishing when the population is zero (an absorbing state). We seek to estimate the parameters r and σ using the proposed method.

As before, we calculate the propagator following [26]. We denote by P_0 the solution for the case $|m| = 0$. Thus, P_0 solves the differential equation

$$P_0(t) = p_0 + r\int_0^t P_0(s)\big[1 - P_0(s)\big]ds, \tag{60}$$

Table 2
Validation of the proposed method: estimation results for the geometric Brownian motion under various initial conditions.

True Parameters (α, σ)	Initial Condition $(\tilde{\alpha}_0, \tilde{\sigma})$	Final Estimation $(\tilde{\alpha}_n, \tilde{\sigma}_n)$
(0.35, 0.17)	(0.40, 0.151048)	(0.364751, 0.168444)
	(0.20, 0.151048)	(0.364814, 0.156101)
	(0.11, 0.151048)	(0.364836, 0.156445)
(0.63, 0.06)	(0.40, 0.056324)	(0.638756, 0.056059)
	(0.20, 0.056324)	(0.641096, 0.055972)
	(0.80, 0.056324)	(0.638752, 0.055874)
(0.90, 0.11)	(0.40, 0.104350)	(0.917715, 0.095657)
	(0.25, 0.104350)	(0.987138, 0.082628)
	(0.55, 0.104350)	(0.914632, 0.097236)

which has as a solution

$$P_0(t) = \frac{p_0}{1 + \frac{1-p_0}{p_0} e^{-rt}}. \tag{61}$$

We note that while the diffusion coefficient $\Gamma(x) = \sigma x$ satisfies the global Lipschitz and linear growth conditions, the drift $F(x) = rx(1-x)$ does not satisfy them globally on $\mathbb{R}$. However, the existence of a unique global strong solution can still be guaranteed. By employing localization arguments based on stopping times or by constructing appropriate Lyapunov functions, one can prove that the solution exists and remains strictly positive for all $t \geq 0$ (see, e.g., [27]). Then, we still have that $P(t) \in L^2(\Omega; [0, T])$.

We now consider the case where $|m| = 1$, and we will use the vectors from [26, Table 5]. Here we will apply a truncation as follows. For the non-linear logistic drift, the exact Galerkin projection of the quadratic term P_t^2 onto the jth chaos mode introduces the third-order structure tensor $c_{ikj} = \mathbb{E}[\xi_i \xi_k \xi_j]$. While c_{ikj} vanishes exactly when the sum of the polynomial degrees is odd, it yields non-zero contributions for even-order combinations. However, the products of higher-order purely stochastic coefficients ($P_i P_k$) are typically negligible compared to the mean-field cross-terms ($2P_1 P_j$). To maintain computational tractability and avoid the curse of dimensionality associated with the 3D tensor, we employ a standard first-order truncation for the non-linear interactions, approximating the drift of dP_j as $\approx \alpha P_j(1 - 2P_0)$. This linearized approximation effectively captures the macro-scale biological saturation while preserving the efficiency of the SGD algorithm. Thus P_m^i satisfies

$$P_m^i(t) = r \int_0^t P_m^i(s)\big[1 - 2P_0(s)\big] ds + \sigma \sum_j \int_0^t e_j(s)\, P^i_{m^{i-}(j)}(s)\, ds, \tag{62}$$

where $m^{i-}(j)$ is the ith vector with the jth entry minus 1.

Finally, for the case with $|m| > 1$ and using the vectors from Table 5 in [26], P_m^i satisfies

$$P_m^i(t) = r \int_0^t P_m^i(s)\big[1 - 2P_0(s)\big] ds + \sigma \sum_j \sqrt{m_j^i} \int_0^t e_j(s)\, P^i_{m^-(j)}(s)\, ds, \tag{63}$$

where m_j^i denotes the jth entry of the ith vector.

We observe, as in the precedent example, that the functions $P_m^i(t)$ with $|m| \geq 1$ will be solved numerically using expressions (62) and (63) in increasing order.

6.4. Numerical validation: GBM and logistic SDE

We provide a brief overview of the numerical findings for the examples discussed in the previous section. The results for the Geometric Brownian Motion and the Stochastic Logistic equation are summarized in Tables 2 and 3, respectively.

As illustrated in the convergence plots (Fig. 1) and detailed in Tables 2 and 3, the proposed parameter estimation method performs well across various initial conditions. The graphical results show that the Stochastic Gradient Descent algorithm effectively guides the estimated drift parameters toward their actual values within a few iterations. Furthermore, the numerical findings indicate that the final estimations for both the Geometric Brownian Motion and the Stochastic Logistic process are generally accurate. While there are some minor discrepancies between the estimated and the true parameters, which is expected in stochastic numerical approximations, the overall results are satisfactory and provide a reliable validation of the proposed method.

7. Application to empirical data

In this section, we demonstrate the practical applicability and versatility of our parameter estimation framework by utilizing two distinct real-world datasets. First, we validate the method on financial market data using a canonical linear model, the Geometric

Table 3
Validation of the proposed method: estimation results for the Stochastic logistic process under various initial conditions.

True Parameters (α, σ)	Initial Condition $(\tilde{\alpha}_0, \tilde{\sigma})$	Final Estimation $(\tilde{\alpha}_n, \tilde{\sigma}_n)$
(0.68, 0.09)	(0.55, 0.081055)	(0.673395, 0.083288)
	(0.95, 0.081055)	(0.675159, 0.084003)
	(0.12, 0.081055)	(0.674867, 0.089809)
(0.30, 0.05)	(1.01, 0.044360)	(0.319553, 0.044723)
	(0.10, 0.044360)	(0.302488, 0.096654)
	(0.50, 0.044360)	(0.302418, 0.064682)
(1.20, 0.20)	(0.57, 0.179743)	(1.210950, 0.256058)
	(0.67, 0.179743)	(1.214110, 0.243503)
	(0.95, 0.179743)	(1.210810, 0.235354)

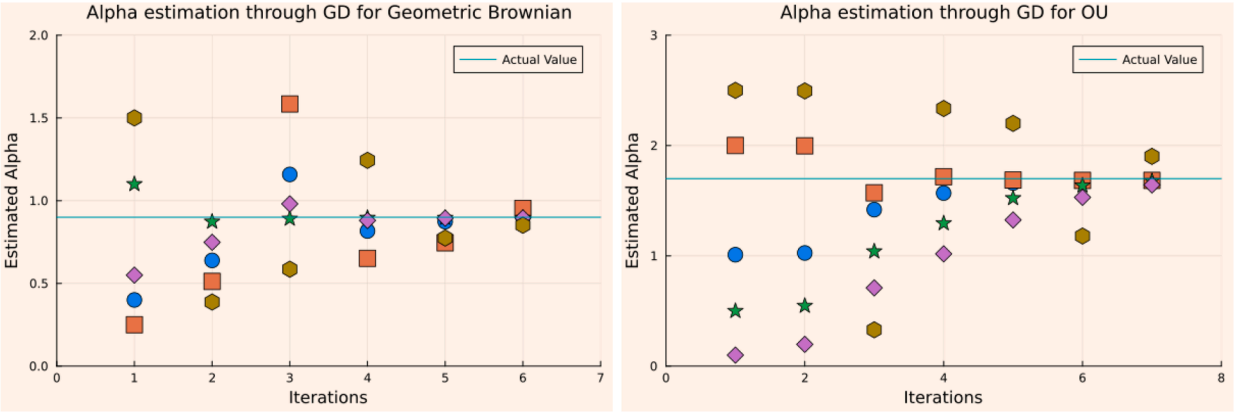

Fig. 1. Convergence analysis of the gradient Descent algorithm for estimating the drift parameter α in the Geometric Brownian Motion and the OU process. The horizontal line indicates the ground truth value ($\alpha = 0.9$ for the GMB and $\alpha = 1.7$ for the OU). Distinct markers represent the evolution of the parameter estimate starting from different 5 initial guesses. As shown, all trajectories converge rapidly toward the true value within approximately 6 iterations, demonstrating the robustness and stability of the Wiener Chaos-based optimization framework.

Brownian Motion (GBM). Subsequently, we apply the algorithm to characterize individual biological growth, employing the nonlinear stochastic logistic differential equation. This dual application highlights the algorithm's robustness across different domains, scaling, and underlying stochastic dynamics.

7.1. Empirical validation: S&P 500 index

To assess the practical applicability and robust performance of the proposed stochastic gradient descent (SGD) approach coupled with the Wiener Chaos Expansion (WCE), we initially applied the methodology to empirical financial market data. The dataset selected consists of daily adjusted closing prices for the S&P 500 index (^GSPC), representing a proxy for the broad US equity market.

Based on quantitative finance standards, the underlying dynamics were modeled using a Geometric Brownian Motion (GBM) specification. The observed period spans a full calendar year, from January 1, 2017, to December 31, 2017. This interval was selected to provide a sufficiently noisy real-world scenario to test the algorithm's learning capabilities, yet remain within a standard stable GBM regime.

Consistent with the procedure used for synthetic data, the volatility parameter (σ) was pre-estimated analytically via standard Maximum Likelihood Estimation (MLE) of the log-returns. Subsequently, the WCE-SGD algorithm was deployed to learn the drift parameter (μ) by minimizing the functional error between the expansion propagator and the empirical trajectory.

The visual results of this validation are presented in Fig. 2. As illustrated, the methodology yields an accurate reconstruction of the market dynamics. The empirical trajectory of the S&P 500 index is remarkably contained within the 95% confidence bands generated by the simulations based on the WCE-learned parameters. Furthermore, the numerical convergence of μ showed excellent agreement with the benchmark MLE value ($\hat{\mu} = 0.1589636$ with the WCE, and $\hat{\mu}_{MLE} = 0.1725922$). Consequently, the method is considered to function with high precision on contrasted real-world datasets, successfully capturing the stochastic and deterministic components of the underlying linear process.

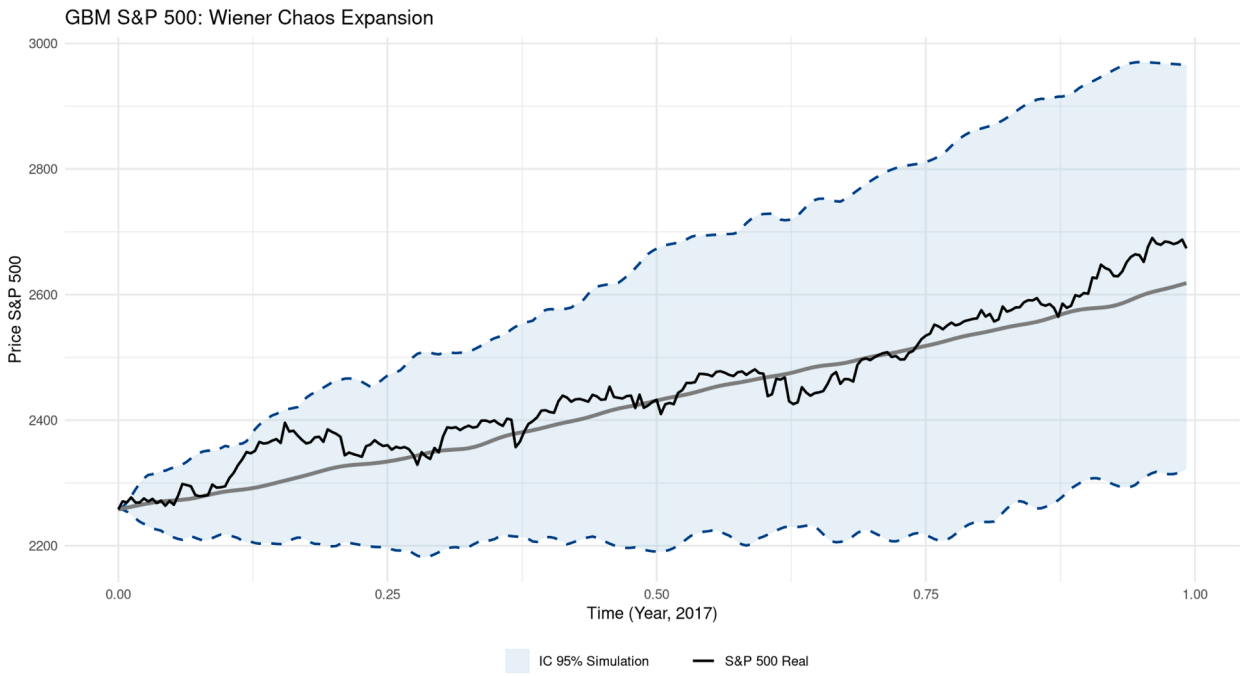

Fig. 2. GBM parameter learning validation on S&P 500 index data (2017). Comparison between the daily adjusted closing prices (solid black line) and stochastic simulations based on parameters learned via the proposed WCE-SGD method ($\hat{\mu} \approx 0.1590$, $\hat{\sigma} \approx 0.0674$). The solid gray line represents the mean of 100 simulations, while the shaded blue area indicates the 95% empirical confidence interval. The close fit demonstrates the successful application of the methodology to noisy empirical financial data.

7.2. Empirical validation: microbial growth dynamics

Having validated the approach on a linear model, we now extend the application to a highly non-linear system where multi-trajectory data is available.

7.2.1. Data acquisition and scientific relevance

The biological dataset, integrated into the `growthcurver` R package, [28,29], comprises high-resolution microbial growth curves obtained via automated spectrophotometry. The raw data were recorded using a 96-well microtiter plate reader, which measures the Optical Density at 600 nm (OD_{600}) at regular intervals over a 24 h period. This experimental setup allows for the simultaneous monitoring of multiple replicates, capturing the complete life cycle of the microbial population, including the lag, exponential, and stationary phases.

The primary utility of this dataset is to provide a robust framework for the computational modeling of population dynamics. By applying non-linear regression models (such as the logistic growth equation), researchers can derive critical kinetic parameters, namely the carrying capacity (K), the intrinsic growth rate (α), and the initial population size (P_0).

7.2.2. Parameter estimation via WCE-SGD

Modeling this biological phenomenon requires addressing the intrinsic multiplicative noise and the non-linear saturation limit (K). To isolate the stochastic dynamics from the spatial scale, we implemented a hybrid estimation strategy.

First, a deterministic standardization was performed. The cross-well mean trajectory of the raw optical density data was computed, and a standard Levenberg-Marquardt Non-Linear Least Squares (NLS) algorithm was utilized to estimate the empirical carrying capacity ($\hat{K}$) and the initial population state. The entire dataset was then normalized by dividing it by $\hat{K}$, mapping the dynamic state variable X_t strictly to the domain $[0, 1]$, which mathematically stabilizes the stochastic logistic equation $dX_t = \alpha X_t(1 - X_t)dt + \sigma X_t dB(t)$.

Second, the volatility parameter (σ) was evaluated analytically by computing the mean quadratic variation across all individual normalized well trajectories. This captures the true experimental stochasticity. Finally, the WCE-SGD algorithm was deployed on the normalized mean trajectory to estimate the intrinsic growth rate (α). To prevent numerical instability during optimization, the Barzilai-Borwein step size was constrained using dynamic dampening (gradient clipping), ensuring smooth convergence to the global minimum despite the non-linear topology of the objective functional.

The outcomes, presented in Fig. 3, confirm that the estimated parameters successfully reconstruct the experimental behavior. By re-scaling the simulations back to the original biological limits (K), the model accurately reflects both the deterministic macroscopic trend (the sigmoidal drift) and the microscopic biological variance observed across the wells.

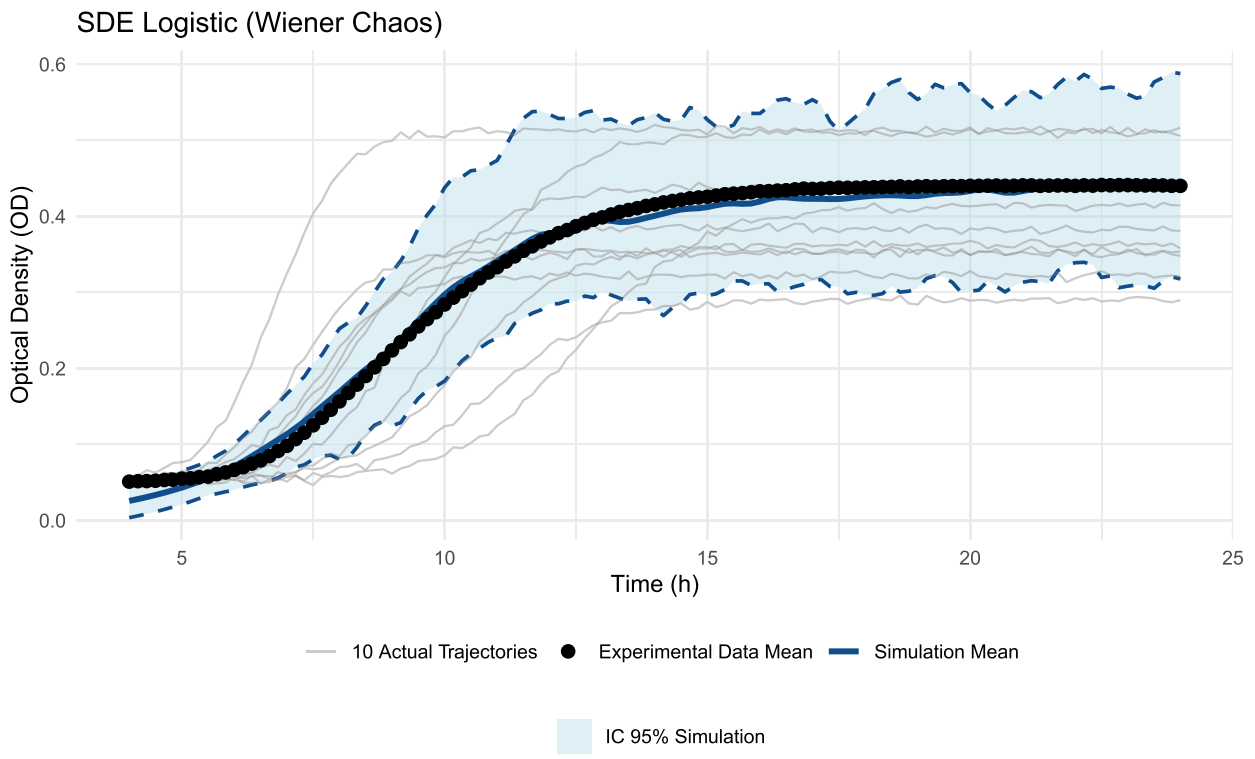

Fig. 3. Stochastic logistic parameter learning on microbial growth data. Comparison between empirical optical density trajectories (gray lines) from a 96-well microtiter plate and stochastic simulations driven by parameters learned via the WCE-SGD framework. The data was pre-standardized to estimate the carrying capacity K. The methodology accurately captures both the asymmetrical, non-linear drift of the exponential phase and the inherent biological noise, encapsulating the individual trajectories within the 95% confidence bands (shaded area).

8. Concluding remarks

In this work, we have developed and validated a novel computational framework for the inverse problem of parameter estimation in Stochastic Differential Equations (SDEs). The core innovation of our approach lies in the integration of the Wiener Chaos Expansion (WCE) with Stochastic Gradient Descent (SGD), which effectively transforms the challenge of stochastic parameter learning into a deterministic optimization problem.

By projecting the solution of the SDE onto the Wiener chaos space, we decomposed the stochastic trajectory into a system of deterministic ordinary differential equations, referred to as the *propagator*. This spectral decomposition allowed us to construct a robust objective function that simultaneously targets the deterministic trend (via the zeroth-order chaos mode) and the stochastic fluctuations (via higher-order modes and quadratic variation). Consequently, we were able to decouple the estimation of drift and diffusion parameters, addressing a common numerical difficulty in stochastic inverse problems.

The numerical experiments conducted on a diverse set of synthetic benchmarks (including the Ornstein-Uhlenbeck process, Geometric Brownian Motion, and the Stochastic Logistic Equation) demonstrated the algorithm's versatility and convergence properties. In all cases, the proposed method successfully recovered the underlying parameters with high accuracy, even in the presence of multiplicative noise and non-linearities.

Crucially, the extension of this framework to empirical data underscores its practical utility for modeling complex real-world systems. By accurately identifying the drift and volatility components of the S&P 500 index within a real-market context, the WCE-SGD approach proved its efficacy in capturing financial market stochasticity. Furthermore, its application to high-resolution microbial growth data demonstrated the method's capacity to handle non-linear biological saturation limits and intrinsic experimental variance using the stochastic logistic model. These dual validations confirm that the methodology represents a significant theoretical advancement in spectral decomposition and offers a computationally efficient alternative for parameter learning. While its performance is demonstrated here through specific examples in ecology, biology, and finance, the underlying framework is fundamentally versatile and applicable across a broad spectrum of scientific disciplines where stochastic modeling is required.

Future research will prioritize the extension of this methodology to a broader range of high-frequency financial instruments, further refining the coupling between the WCE and SGD under volatile market conditions. Additionally, we aim to develop a robust numerical implementation for Stochastic Partial Differential Equations.

Another key avenue of exploration involves the application of this approach to higher-dimensional systems and coupled SDEs. In such contexts, the inherent scalability of the SGD-WCE framework is expected to offer substantial computational advantages

and improved tractability over traditional Maximum Likelihood Estimation (MLE) or computationally intensive Bayesian inference methods.

CRediT authorship contribution statement

Francisco Delgado-Vences: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization; **Jose Julian Pavon-Español:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis; **Arelly Ornelas:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their sincere gratitude to the Associate Editor and the anonymous reviewers for their rigorous, constructive, and insightful assessment of our manuscript. Their comments and suggestions have been invaluable in improving the mathematical rigor, clarity, and overall quality of this work.

This work was supported by the Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI) through the Basic and Frontier Science Project No. CBF-2026-I-3214.

FDV was supported by Ministerio de Ciencia e Innovación , Gobierno de España, project ref. PID2021-123733NB-I00.

Software and codes

The numerical experiments have been implemented in on an average laptop (ASUS Vivobook X512DA with Processor AMD Ryzen 5 3500U and Radeon Graphics, 1200 Mhz, 4 Cores, 8 Logical Processors). The code can be found under the following link: https://github.com/JJulianPavon/GradientDescentSDE

Data availability

Data will be made available on request.

References

[1] A. Beck, First-Order Methods in Optimization, MOS-SIAM Series on Optimization, SIAM, Philadelphia, 2017. https://doi.org/10.1137/1.9781611974997
[2] Y. Nesterov, Lectures on Convex Optimization, 2nd edition, Vol. 137 of Springer Optimization and Its Applications, Springer, Cham, 2018. https://doi.org/10.1007/978-3-031-05793-9_15
[3] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Stat. 22 (3) (1951) 400–407.
[4] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010, Springer, 2010, pp. 177–186.
[5] L. Bottou, F.E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, SIAM Rev. 60 (2) (2018) 223–311. https://doi.org/10.1137/16M1080173
[6] Y. Trémolet, Accounting for an imperfect model in 4D-Var, Q. J. R. Meteorol. Soc. J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr. 132 (621) (2006) 2483–2504. https://doi.org/10.1256/qj.05.224
[7] S. Reich, C. Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, Cambridge, 2015. https://doi.org/10.1017/CBO9781107706804
[8] A.M. Stuart, Inverse problems: a Bayesian perspective, Acta Numer. 19 (2010) 451–559.
[9] F. Delgado-Vences, J.J. Pavon-Español, A. Ornelas, Parameter estimation in stochastic differential equations via wiener chaos expansion and stochastic gradient descent, 2026. arXiv:2603.27019 [math.OC]
[10] L. Landweber, An iteration formula for Fredholm integral equations of the first kind, Am. J. Math. 73 (3) (1951) 615–624.
[11] T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, J. Complex. 25 (1) (2009) 14–25.
[12] D. Needell, R. Ward, N. Srebro, Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm, Math. Program. 155 (1-2) (2016) 549–573.
[13] W. Dörfler, M. Hochbruck, J. Köhler, A. Rieder, R. Schnaubelt, C. Wieners, What is an inverse and ILL-posed problem?, in: Wave Phenomena: Mathematical Analysis and Numerical Approximation, Springer International Publishing, Cham, 2023, pp. 281–284. https://doi.org/10.1007/978-3-031-05793-9
[14] A.H. Hasanoğlu, V.G. Romanov, Introduction to Inverse Problems for Differential Equations, 2 edition, Springer Cham, 2021. https://doi.org/10.1007/978-3-030-79427-9
[15] D. Nualart, The Malliavin Calculus and Related Topics, Probability and Its Applications, 2 edition, Springer, Heidelberg, 2006.
[16] D. Chen, J. Li, Y. Zhang, A posterior contraction for Bayesian inverse problems in Banach spaces, Inverse Probl. 40 (4) (2024) 045011. https://doi.org/10.1088/1361-6420/ad278e
[17] S.V. Lototsky, B.L. Rozovsky, Stochastic Partial Differential Equations, Universitext, Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-58647-2
[18] S. Krumscheid, G.A. Pavliotis, S. Kalliadasis, Semiparametric drift and diffusion estimation for multiscale diffusions, Multiscale Model. Simul. 11 (2) (2013) 442–473.
[19] R. Fletcher, On the Barzilai-Borwein method, in: Optimization and Control with Applications, Springer, 2005, pp. 235–256. https://doi.org/10.1007/0-387-24255-4_10
[20] C. Tan, S. Ma, Y.-H. Dai, Y. Qian, Barzilai-Borwein step size for stochastic gradient descent, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, Curran Associates Inc., Red Hook, NY, USA, 2016, pp. 685–693.

[21] G. Garrigos, R.M. Gower, Handbook of Convergence Theorems for (Stochastic) Gradient Methods, (2023). arXiv preprint arXiv:2301.11235, https://arxiv.org/abs/2301.11235.
[22] R. Ge, F. Huang, C. Jin, Y. Yuan, Escaping from saddle points—online stochastic gradient for tensor decomposition, in: Conference on Learning Theory, PMLR, 2015, pp. 797–842.
[23] C. Jin, R. Ge, P. Netrapalli, S.M. Kakade, M.I. Jordan, How to escape saddle points efficiently, in: International Conference on Machine Learning, PMLR, 2017, pp. 1724–1732.
[24] P.A. Samuelson, Rational theory of warrant pricing, Ind. Manag. Rev. 6 (2) (1965) 13–31.
[25] F. Black, M. Scholes, The pricing of options and corporate liabilities, J. Polit. Econ. 81 (3) (1973) 637–654.
[26] F. Delgado-Vences, G.A. Salcedo-Varela, F. Baltazar-Larios, Simulating diffusion bridges using the Wiener chaos expansion, Methodol. Comput. Appl. Probab. 28 (2026) 60. https://doi.org/10.1007/s11009-026-10292-9
[27] D. Jiang, C. Ji, X. Li, D. O'Regan, Analysis of autonomous Lotka–Volterra competition systems with random perturbation, J. Math. Anal. Appl. 390 (2) (2012) 582–595.
[28] K. Sprouffske, A. Wagner, Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves, BMC Bioinform. 17 (1) (2016) 172. https://doi.org/10.1186/s12859-016-1016-7
[29] K. Sprouffske, Growthcurver: Simple Metrics to Summarize Growth Curves, 2020. R package version 0.3.1, https://CRAN.R-project.org/package=growthcurver.