

Approximation of an Integral Markov Process Arising in the Approximation of Stochastic Differential Equation

Mohammad Rahman

Department of Mathematics & Statistics, University of North Florida, Jacksonville, FL-32224, USA
Email: mrahman@unf.edu

How to cite this paper: Rahman, M. (2022) Approximation of an Integral Markov Process Arising in the Approximation of Stochastic Differential Equation. *Advances in Pure Mathematics*, 12, 29-47.
<https://doi.org/10.4236/apm.2022.121003>

Received: November 3, 2021

Accepted: January 23, 2022

Published: January 26, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We provide the derivation of a new formula for the approximation of an integral Markov process arising in the approximation of stochastic differential equations. This formula extends an existing formula derived in [1]. We have shown numerically that the leading order approximation of the differential equation with noise by solving an associated averaged problem and estimating the difference between them and the result is illustrated through some examples.

Keywords

Variance, Markov Process, Parametric Noise, Differential Equation, Approximations

1. Introduction

Nonlinear ordinary and partial differential equations arise in various fields of sciences, particularly fluid mechanics, solid state physics, plasma physics, nonlinear optics, and mathematical biology. Powerful numerical methods and its implementation can be obtained in [2]-[7]. However, nonlinear differential equations (DE) with parametric noise play a significant role in a range of application areas, including engineering, physics, mechanics, epidemiology, and neuroscience. It is important to mention that noisy systems can be modeled in several ways: for example, Langevin's equation describes a linear physical system to which white noise is added, and the linear theory for it has been extended to nonlinear stochastic differential equations with additive white noise [8]. Another approach is to derive models, such as Markov chains, for the systems state variables as being random, and then use the method of probability for analysis, see

[9]. Third approach was derived by averaging nonlinear oscillatory systems. In this approach, parameters in the system are allowed to be random processes, and the method based on averaging and ergodic theory provides useful predictions from the model [10]. Indeed, solutions of differential or dynamical systems are functions, say of time. If they are in addition random, we must describe both randomness and time dependence simultaneously. Thus we refer to the system as either a random process, or a stochastic process. A complete understanding of DE theory with perturbed noise requires familiarity with advanced probability and stochastic processes (see [8] [11]).

In this paper, our approach to modeling randomness in differential or dynamical system is through allowing parameters b in a system to be a random process. For example, in the case of differential equation we write $\dot{x}(t) = f(t, \omega, x(t), b(y(t)))$, indicating that the parameters can change with some random process $y(t)$. The resulting solutions $x(t, b)$ will also be a random process. This approach will be based on assuming noise process faster time scale than the system time scale. This work has been dedicated to the question in particular when y is a discrete-space Markov process. Most recent contributions have aimed in general at relaxing the assumptions on y [1] [12]. This result is known as Functional Central Limit Theorem [13] [14] [15].

In Section 2 we reviewed and discussed limit of variance for discrete-space Markov Chain [1]. In Section 3, a new equivalent expression for a limit of a variance is given for circulant n -state markov chain. In Section 3, stochastic approximation and numerical simulation were discussed. It is observed that computer simulations of this type of stochastic ordinary differential equation with standard methods have some issues needed to be explored so that the reader will be benefitted while solving these type problems numerically.

2. Discrete-Space Markov Chains

We consider the case where $Y = \{y_1, \dots, y_n\}$, and let

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \varphi(\underline{y}) = \begin{bmatrix} \varphi(y_1) \\ \vdots \\ \varphi(y_n) \end{bmatrix}, \quad \text{and} \quad \underline{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

The transition probability $P(\Delta t, y, dy')$ can be represented by an $n \times n$ matrix P such that $P_{i,j} = P(\Delta t, y_i, y_j)$, denotes the probability that $y(\Delta t) = y_j$, if $y(0) = y_i$ (P depends on Δt but this dependence is omitted from the notation for simplicity). The matrix P satisfies the following properties.

- P is nonnegative (its entries are probabilities).
- P is stochastic *i.e.*, $P\underline{e} = \underline{e}$ (one of the y_j must be the outcome of a transition from y_i).

Then $\lambda_1 = 1$, is an eigenvalue of P and $1 \leq \rho(P) \leq \|P\|_\infty = 1$, shows that all other eigenvalues have modulus at most 1. We shall assume that

- P is irreducible, *i.e.*, any state y_j can be eventually be reached in a finite number of steps with a nonzero probability (this is the case if $P > 0$). This

implies that $\lambda_1 = 1 = \rho(P)$, has multiplicity one (e.g., see [16]).

- P is aperiodic (or acyclic, e.g., not a permutation). This implies that $|\lambda_j| < 1$ for $j = 2, \dots, n$ (e.g., see [9]; the chain is then called regular).

The above conditions guarantee the existence of a unique vector $\underline{v} > 0$ such that

$$\underline{v}^T \underline{e} = 1, \tag{2.1}$$

and $\lim_{N \rightarrow \infty} P^N = \underline{e} \underline{v}^T$. The vector \underline{v} is the unique positive left eigenvector associated to $\lambda_1 = 1$, satisfying (2.1). It is natural to consider the (spectral) decomposition

$$P = \underline{e} \underline{v}^T + S, \tag{2.2}$$

where

$$\underline{v}^T S = \underline{0}^T, \quad S \underline{e} = \underline{0}, \tag{2.3}$$

and $\rho(\tilde{P}) < 1$. From Chapman-Kolmogorov equation and the homogeneity property (see, [1]) we have

$$P(2\Delta t, y_i, y_j) = \sum_{k=1}^n P(\Delta t, y_i, y_k) P(\Delta t, y_k, y_j) = \sum_{k=1}^n P_{i,k} P_{k,j} = P_{i,j}^2,$$

and by induction

$$P(N\Delta t, y_i, y_j) = P_{i,j}^N = (\underline{e} \underline{v}^T + S^N)_{i,j} \approx (\underline{e} \underline{v}^T)_{i,j} = v_j,$$

as $N \rightarrow \infty$, independently of i (i.e., y_i). Thus

$$\underline{v} = \rho_{\Delta t}(\underline{y}), \tag{2.4}$$

defines the limit distribution. An explicit expression of the coefficients of \underline{v} in terms of the coefficients of P can be found in [[17], p. 21].

The relation of the expected value yields

$$\begin{aligned} E_{y(0)=y_j} [\varphi(y(N\Delta t))] &= \sum_{j=1}^n \varphi(y_j) P(N\Delta t, y_j, y_i) \\ &= \sum_{j=1}^n (P^N)_{i,j} \varphi(y_j) \\ &= (P^N \varphi(\underline{y}))_i, \end{aligned} \tag{2.5}$$

i.e.,

$$E_{y(0)=\underline{y}} [\varphi(y(N\Delta t))] = \begin{bmatrix} E_{y(0)=y_1} [\varphi(y(N\Delta t))] \\ \vdots \\ E_{y(0)=y_n} [\varphi(y(N\Delta t))] \end{bmatrix} = P^N \varphi(\underline{y}).$$

Then the zero average condition on φ becomes

$$0 = \int_y \varphi(y) \rho(dy) = \sum_{j=1}^n \varphi(y_j) \rho(y_j) = \rho(\underline{y})^T \varphi(\underline{y}) = \underline{v}^T \varphi(\underline{y}). \tag{2.6}$$

The relation (2.6) implies $P\varphi(\mathbf{y}) = \varphi(\mathbf{y})$. Therefore

$$\begin{aligned} \mathcal{R}^{\Delta t,1} \varphi(\underline{y}) &= \Delta t \sum_{N=1}^{\infty} E_{y(0)=\underline{y}} [\varphi(y(N\Delta t))] \\ &= \Delta t \sum_{N=1}^{\infty} P^N \varphi(\underline{y}) \end{aligned} \tag{2.7}$$

$$= \Delta t \sum_{N=1}^{\infty} S^N \varphi(\underline{y}) \tag{2.8}$$

$$= \Delta t (I - S)^{-1} S \varphi(\underline{y}), \tag{2.9}$$

where $\frac{1}{\sqrt{\varepsilon}} \mathcal{R}^{\Delta t,1} \varphi(\underline{y})$ can be interpreted as the expected value of the random variable $\frac{1}{\sqrt{\varepsilon}} \mathcal{R}^{\Delta t,1} \varphi(\underline{y})$ obtained after one transition probability applied to the random variable y (for more detail, see [1]). Note that (2.7) implies

$$\mathcal{R}^{\Delta t,1} \varphi(\underline{y}) - P \mathcal{R}^{\Delta t,1} \varphi(\underline{y}) = \Delta t P \varphi(\underline{y}).$$

Because of (2.3) we also have

$$\underline{y}^T \mathcal{R}^{\Delta t,1} \varphi(\underline{y}) = 0, \tag{2.10}$$

With $V = \text{diag}(\underline{y})$ we obtain

$$\begin{aligned} \sigma_{\Delta t}^2 &= 2 \sum_{j=1}^n \varphi(y_j) \rho(y_j) (\mathcal{R}^{\Delta t,1} \varphi(\underline{y}))_j + \Delta t \sum_{j=1}^n \varphi^2(y_j) \rho(y_j) \\ &= 2 \Delta t \varphi(\underline{y})^T V (I - S)^{-1} S \varphi(\underline{y}) + \Delta t \varphi(\underline{y})^T V \varphi(\underline{y}) \\ &= \Delta t \varphi(\underline{y})^T V (I - S)^{-1} (I + S) \varphi(\underline{y}), \end{aligned} \tag{2.11}$$

as $t \rightarrow \infty$. The expression (2.11) represents the limit of a variance and thus expected to be nonnegative.

Two-State Markov Chain

Let $n = 2$ and

$$P = P(\Delta t) = I + \Delta t Q, \quad Q = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}, \tag{2.12}$$

with $0 < a, b < 1$, $0 < \Delta t < \min\left(\frac{1}{a}, \frac{1}{b}\right) \leq \frac{2}{a+b}$. Then (2.2) holds with

$$\begin{aligned} \underline{y} &= \frac{1}{a+b} \begin{bmatrix} b \\ a \end{bmatrix} \\ \text{and } S &= \frac{1 - \Delta t(a+b)}{a+b} \begin{bmatrix} a & -a \\ -b & b \end{bmatrix}. \end{aligned}$$

As a result $V = \frac{1}{a+b} \begin{bmatrix} b \\ a \end{bmatrix}$ and

$$\Delta t V (I - S)^{-1} (I + S) = \frac{1}{(a+b)^3} \begin{bmatrix} b(2a + \Delta t(b^2 - a^2)) & -2ab(1 - \Delta t(a+b)) \\ -2ab(1 - \Delta t(a+b)) & a(2b + \Delta t(a^2 - b^2)) \end{bmatrix} \tag{2.13}$$

is (symmetric) positive definite for any choice $0 < a, b < 1$ and $0 < \Delta t < \frac{2}{a+b}$.

A simplified expression for (2.11) is obtained using (2.6), *i.e.*,

$$\varphi(\underline{y}) = \begin{bmatrix} 1 \\ -\frac{b}{a} \end{bmatrix} \varphi(y_1).$$

We obtain

$$\begin{aligned} \sigma_\varepsilon^2 &= \frac{b(2-\Delta t(a+b))}{a(a+b)} (\varphi(y_1))^2 \\ &= \frac{ab(2-\Delta t(a+b))}{(a+b)^3} (\varphi(y_1) - \varphi(y_2))^2. \end{aligned} \quad (2.14)$$

3. Circulant n -State Markov Chain

If P is doubly stochastic (*i.e.*, P and P^T are stochastic) and irreducible aperiodic then

$$\underline{v} = \frac{1}{n} \underline{e},$$

i.e., the limit distribution is uniform and

$$D = \frac{1}{n} I. \quad (3.16)$$

Stochastic Toeplitz, hence circulant, matrices constitute an example of doubly stochastic matrices. Let

$$P = \begin{bmatrix} a_1 & a_2 & \ddots & a_n \\ a_n & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & a_2 \\ a_2 & \ddots & a_n & a_1 \end{bmatrix},$$

with $a_i \geq 0$ and $\sum_{j=1}^n a_j = 1$. A sufficient condition for P to be irreducible and aperiodic is for two consecutive a_j to be nonzero (*i.e.*, positive, see e.g. [[18], p. 5]). The *symbol* of P is the polynomial

$$p(z) = \sum_{j=1}^n a_j z^{j-1}.$$

Since

$$P \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-1} \end{bmatrix} = p(z) \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-1} \end{bmatrix},$$

for $z^n = 1$, P admits the spectral decomposition

$$P = V \Lambda V^{-1} = V \Lambda V^H,$$

with

$$V = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & z_1 & \cdots & z_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_1^{n-1} & \cdots & z_{n-1}^{n-1} \end{bmatrix} = \left[\frac{1}{\sqrt{n}} \underline{e} \mid \tilde{V} \right], \tag{3.17}$$

$$\Lambda = \begin{bmatrix} 1 & & & \\ & p(z_1) & & \\ & & \ddots & \\ & & & p(z_{n-1}) \end{bmatrix} = \left[\frac{1}{\sqrt{n}} \mid \tilde{\Lambda} \right], \tag{3.18}$$

where $z_j = e^{\frac{2i\pi j}{n}} = z_j^j$. Multiplication of a vector by the matrix V^H performs a (normalized) discrete Fast Fourier Transform (FFT), while multiplication by V results in the inverse (normalized) discrete FFT.

We write $P = \frac{1}{n} \underline{e} \underline{e}^T + \tilde{V} \tilde{\Lambda} \tilde{V}^H = \frac{1}{n} \underline{e} \underline{e}^T + \tilde{P}$, with $\tilde{P} = \tilde{V} \tilde{\Lambda} \tilde{V}^H$. Since

$|p(z_k)| < 1$ for $k = 1, \dots, n-1$ the matrix $I - \tilde{\Lambda}$ is nonsingular. The matrix $\tilde{V} \tilde{V}^H$ represents the (orthogonal) projection onto $\text{Span}\{\underline{y}\}^\perp = \text{Span}\{\underline{e}\}^\perp$.

Hence $\underline{v}^T \varphi(\underline{y}) = 0$ implies $\varphi(\underline{y}) = \tilde{V} \tilde{V}^H \varphi(\underline{y})$. Then

$$\begin{aligned} (I + \tilde{P})\varphi(\underline{y}) &= (I + \tilde{V} \tilde{\Lambda} \tilde{V}^H) \tilde{V} \tilde{V}^H \varphi(\underline{y}) \\ &= \tilde{V} (I + \tilde{\Lambda}) \tilde{V}^H \varphi(\underline{y}) \\ &= \tilde{V} (I - \tilde{\Lambda}) (I - \tilde{\Lambda})^{-1} (I + \tilde{\Lambda}) \tilde{V}^H \varphi(\underline{y}) \\ &= (I - \tilde{P}) \tilde{V} (I - \tilde{\Lambda})^{-1} (I + \tilde{\Lambda}) \tilde{V}^H \varphi(\underline{y}). \end{aligned} \tag{3.19}$$

Now

$$\begin{aligned} (\tilde{V}^H \varphi(\underline{y}))_k &= \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & \overline{z_k} & \cdots & \overline{z_k}^{n-1} \end{bmatrix} \begin{bmatrix} \varphi(y_1) \\ \varphi(y_2) \\ \vdots \\ \varphi(y_n) \end{bmatrix} \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(y_j) \overline{z_k}^{j-1} \\ &= \overline{\Phi(z_k)}, \end{aligned}$$

for $k = 1, \dots, n-1$ ($\varphi(y)$ is real), with

$$\Phi(z) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(y_j) z^{j-1}. \tag{3.20}$$

From (3.19) and (3.16) we obtain

$$\begin{aligned} \sigma_{\Delta t}^2 &= \frac{\Delta t}{n} (\tilde{V}^H \varphi(\underline{y}))^H (I - \tilde{\Lambda})^{-1} (I + \tilde{\Lambda}) (\tilde{V}^H \varphi(\underline{y})) \\ &= \frac{\Delta t}{n} \sum_{k=1}^{n-1} \frac{1 + p(z_k)}{1 - p(z_k)} |\Phi(z_k)|^2 \end{aligned} \tag{3.21}$$

$$= \frac{\Delta t}{n} \sum_{k=1}^{n-1} \Re \left(\frac{1 + p(z_k)}{1 - p(z_k)} \right) |\Phi(z_k)|^2 \tag{3.22}$$

$$= \frac{\Delta t}{n} \sum_{k=1}^{n-1} \frac{1 - |p(z_k)|^2}{|1 - p(z_k)|^2} |\Phi(z_k)|^2, \quad (3.23)$$

since $\sigma_{\Delta t}^2$ is real and $\overline{p(z_k)} = p(\overline{z_k})$, for $k = 1, \dots, n-1$. The condition $|p(z_k)| < 1$ for $k = 1, \dots, n-1$ clearly shows that (3.23) is nonnegative.

3.1. Example: Uniform Distribution

If $a_j = \frac{1}{n}$, $j = 1, \dots, n$, (i.e., $P = \frac{1}{n} \underline{e} \underline{e}^T$) we obtain

$$p(z_k) = \frac{1}{n} \sum_{j=1}^n z_k^{j-1} = \frac{1}{n} \frac{1 - z_k^n}{1 - z_k} = 0$$

for $k = 1, \dots, n-1$. Parseval's identity yields

$$\begin{aligned} \sum_{k=1}^{n-1} |\Phi(z_k)|^2 &= \sum_{k=0}^{n-1} |\Phi(z_k)|^2 = \sum_{k=0}^{n-1} \left| (V^H \varphi(\underline{y}))_k \right|^2 \\ &= \varphi(\underline{y})^T V V^H \varphi(\underline{y}) = \varphi(\underline{y})^T \varphi(\underline{y}) = \sum_{j=1}^n |\varphi(y_j)|^2. \end{aligned}$$

Then (3.23) reduces to

$$\sigma_{\Delta t}^2 = \frac{\Delta t}{n} \sum_{k=1}^{n-1} |\Phi(z_k)|^2 = \frac{\Delta t}{n} \sum_{j=1}^n |\varphi(y_j)|^2 \approx \Delta t \int_Y \varphi^2(y) \rho_{\Delta t}(dy), \quad \text{as } n \rightarrow \infty.$$

3.2. Example: Big World Transition Probability

Assume now that the probability $y_i \rightarrow y_j$, $j \neq i$, is independent of j but distinct from the probability $y_i \rightarrow y_i$, i.e.,

$$a_1 = 1 - a\Delta t, \quad a_2 = \dots = a_n = \frac{a\Delta t}{n-1},$$

with $0 < a < 1/\Delta t$. Then

$$p(z) = 1 - a\Delta t + \frac{a\Delta t}{n-1} (z + \dots + z^{n-1}) = 1 - a\Delta t + \frac{a\Delta t}{n-1} \frac{z - z^n}{1 - z}$$

for $z \neq 1$. In particular

$$p(z_k) = 1 - a\Delta t + \frac{a\Delta t}{n-1} \frac{z_k - 1}{1 - z_k} = 1 - \frac{na\Delta t}{n-1},$$

for $k = 1, \dots, n-1$. Therefore, by Parseval again,

$$\begin{aligned} \sigma_{\Delta t}^2 &= \frac{2 - \frac{na\Delta t}{n-1}}{\frac{na}{n-1}} \frac{1}{n} \sum_{j=1}^n |\varphi(y_j)|^2 \\ &\approx \frac{2 - \Delta t a}{a} \int_Y \varphi^2(y) \rho_{\Delta t}(dy) \quad \text{as } n \rightarrow \infty \\ &\approx \frac{2}{a} \int_Y \varphi^2(y) \rho_{\Delta t}(dy) \quad \text{as } \Delta t \rightarrow 0. \end{aligned} \quad (3.24)$$

3.3. Example: Small World Transition Probability

A case of particular interest in the study of the transmission of a signal around a

cyclic biological chain corresponds to

$$a_1 = 1 - a\Delta t, \quad a_2 = a\Delta t, \quad a_3 = \dots = a_n = 0,$$

with $0 < a < 1/\Delta t$. The quantity a_2 represents the transition probability of a state y_j to the next state y_{j+1} . Then $p(z) = 1 - a\Delta t + az\Delta t$, and

$$\frac{1 - |p(e^{iy})|^2}{|1 - p(e^{iy})|^2} = \frac{1 - (1 - a\Delta t + a\Delta t)^2 - (a\Delta t)^2}{(a\Delta t - a\Delta t)^2 + (a\Delta t)^2} = \frac{1 - a\Delta t}{a\Delta t},$$

for $0 < y < 2\pi$. Therefore

$$\sigma_{\Delta t}^2 = \frac{1 - a\Delta t}{a} \frac{1}{n} \sum_{j=1}^n |\varphi(y_j)|^2 \tag{3.25}$$

$$\approx \frac{1 - a\Delta t}{a} \int_Y \varphi^2(y) \rho_{\Delta t}(dy) \quad \text{as } n \rightarrow \infty$$

$$\approx \frac{1}{a} \int_Y \varphi^2(y) \rho_{\Delta t}(dy), \quad \text{as } \Delta t \rightarrow 0. \tag{3.26}$$

For $n = 2$, (3.25) reduces to $\frac{1 - \Delta t a}{a} \frac{\varphi(y_1)^2 + \varphi(y_2)^2}{2}$. Note that (3.26) is half

of (3.24), as could be expected from a less dispersive signal.

In the following section our approach to modeling randomness in dynamical systems is through allowing parameters in a system to be a random process. This approach will determine when solution of this type of stochastic problem do or do not persist when the system is perturbed.

4. Mathematical Derivation for Numerical Approximation of Stochastic Differential Equations (SDE)

For the simplicity, consider

$$\begin{cases} \dot{x}(t) = f(t, \omega, x(t), y), \\ y = y(t/\varepsilon), \\ x(0) = x_0, \end{cases} \tag{27}$$

The average system is defined by the differential equation

$$\begin{cases} \dot{\bar{x}}(t) = \bar{f}(t, \omega, \bar{x}(t)), \\ \bar{x}(0) = x_0, \end{cases} \tag{28}$$

We want to compute the deviation of the perturbed system to be average one. We consider $\tilde{x} = x - \bar{x}$, then

$$\begin{aligned} \dot{\tilde{x}} &= \dot{x} - \dot{\bar{x}} \\ &= f(t, \omega, \bar{x} + \tilde{x}, y) - \bar{f}(t, \omega, \bar{x}) \\ &= f(t, \omega, \bar{x}, y) + f(t, \omega, \bar{x}, y)\tilde{x} + \dots - \bar{f}(t, \omega, \bar{x}) \\ &\approx \bar{f}_x(t, \omega, \bar{x})\tilde{x} + (f(t, \omega, \bar{x}, y) - \bar{f}(t, \omega, \bar{x})). \end{aligned}$$

Note that $\tilde{x}(0) = 0, \Rightarrow$

$$\tilde{x} = \int_0^t \bar{f}_x(s, \omega, \bar{x}(s)) \tilde{x} ds + \int_0^t [f(s, \omega, \bar{x}(s), y(s/\varepsilon)) - \bar{f}(s, \omega, \bar{x}(s))] ds.$$

It follows from limit theorem of stochastic processes (see, [2]),

$$\frac{1}{\sqrt{\varepsilon}} \int_0^t [f(s, \omega, \bar{x}(s), y(s/\varepsilon)) - \bar{f}(s, \omega, \bar{x}(s))] ds \approx N(0, \sigma^2(t)).$$

Detailed derivation of $\sigma^2(t)$ is shown in Section 2. The stochastic processes

$$\frac{x(t) - \bar{x}(t)}{\sqrt{\varepsilon}} \approx \tilde{x}(t),$$

converge in expected sense to the solution $\tilde{x}(t)$ that is the solution to the integral equation

$$\tilde{x}(t) = \int_0^t \bar{f}_x(s, \omega, \bar{x}(s)) \tilde{x}(s) ds + \sigma(t).$$

The distribution of $x(t)$ is close to the distribution of the stochastic processes $\bar{x}(t) + \sqrt{\varepsilon} \tilde{x}(t)$ in the sense that

$$E(x(t)) = E(\bar{x}(t) + \sqrt{\varepsilon} \tilde{x}(t)).$$

Thus $x(t) \approx \bar{x}(t) + \sqrt{\varepsilon} \tilde{x}(t) + o(\sqrt{\varepsilon})$, as $\varepsilon \rightarrow 0$. Thus the nature of this convergence and the sense in which the error in the formula of the expansion x is small are in the sense of distributions over some time interval.

4.1. Forward Euler Scheme for SDE

The Euler method to approximate the analytic solution of the IVP

$$\begin{cases} \dot{x}(t) = f(t, x, y(t/\varepsilon)), \\ x(0) = x_0. \end{cases} \tag{4.29}$$

We can derive an entire family of discrete numerical methods (including the Euler method) by truncating the Taylor series and utilizing Taylor’s Theorem. First, we rewrite the Taylor series expansion in differential form,

$$\begin{aligned} x(t+h) &= x(t) + \frac{h}{1!} x'(t) + \frac{h^2}{2!} y''(t) + \dots + \frac{h^n}{n!} y^n(t) + \dots \\ &= x(t) + hf(t, x, y(t/\varepsilon)) + \frac{h^2}{2} f'(t, x, y(t/\varepsilon)) f(t, x, y(t/\varepsilon)) + \dots \\ &\quad + \frac{h^3}{6} (f''(f, f) + f'(f'(f))) + \dots \end{aligned}$$

The Euler method is found by truncating Taylor series at the first derivative, giving

$$x(t+h) = x(t) + hf(t, x, y(t/\varepsilon)) + o(h^2), \tag{4.30}$$

where $o(h^2)$ is the error term, or Taylors remainder term, which is of order h^2 . So if we define $x_n = x(t+h)$, and $x_{n+1} = x(t+h)$, we obtain

$$x_{n+1} = x_n + hf(n, x_n, y(n/\varepsilon)),$$

which has local error term (error at each step) of $o(h^2)$, giving a global error $o(h)$.

4.2. Discrete Approximation for SDE

Numerical schemes for solving SDE can be classified into either explicit or implicit methods. Explicit methods compute approximations that are dependent on previous approximations only, whereas the implicit methods compute approximations that are dependent on previous and current approximations. The Euler method presented earlier is also known as the explicit Euler method. The explicit Euler method is defined as

$$x_{i+1} = x_i + hf(t_i, x_i, y_{ij/\varepsilon}), \tag{4.31}$$

where $x_i = x(t_i)$ and $x_{i+1} = x(t_{i+h})$. Here h is the step size. This method has a global error $\approx o(h)$. The implicit method is defined as

$$x_{i+1} = x_i + hf(t_i, x_{i+1}, y_{(i+1)/\varepsilon}), \tag{4.32}$$

where $x_i = x(t_i)$, and $x_{i+1} = x(t_i + h)$. This method has a global error $\approx o(h)$.

4.3. Example 1

We write the algorithm for the following equation, which we will be used as a test equation since it has an analytic solution.

$$\begin{cases} \dot{x}(t) = y(t/\varepsilon), \\ x(0) = 0, \end{cases} \tag{4.33}$$

over $[0,1]$ and then

$$x(t+h) = x(t) + hy(t/\varepsilon) + o(h). \text{ Now,}$$

- 1) choose a step $h = (1-0)/N$. Set $x_n = 0 + nh$, $n = 0 : N$.
- 2) Generate approximation x_n to from the following recursion:

$$\begin{aligned} x_n &= 0 + h \sum_{k=0}^{n-1} y(kh/\varepsilon) \\ &= h \sum_{k=0}^{n-1} y(kh/\varepsilon) \\ &= hn\bar{y} + h \sum_{l=0}^{h(n-1)/\varepsilon} (y(l) - \bar{y}) \\ &= hn\bar{y} + H\sqrt{\varepsilon} \frac{1}{1/\sqrt{\varepsilon}} \sum_{l=0}^{h(n-1)/\varepsilon} (y(l) - \bar{y}), \end{aligned}$$

for $n = 0 : N - 1$.

As $\varepsilon \rightarrow 0$,

$$x_n \rightarrow hn\bar{y} + H\sqrt{\varepsilon}W(hn) + Error,$$

where $h = H\varepsilon$ in our numerical simulation and W is realized by the normally distributed random number whose expectation and variance are 0 and 1, respectively. A computer simulation of the distribution of solution of $x(1, \varepsilon)$ using Euler without max step size, ode15s without max step, and ode15s using max step size are given in **Figure 1**.

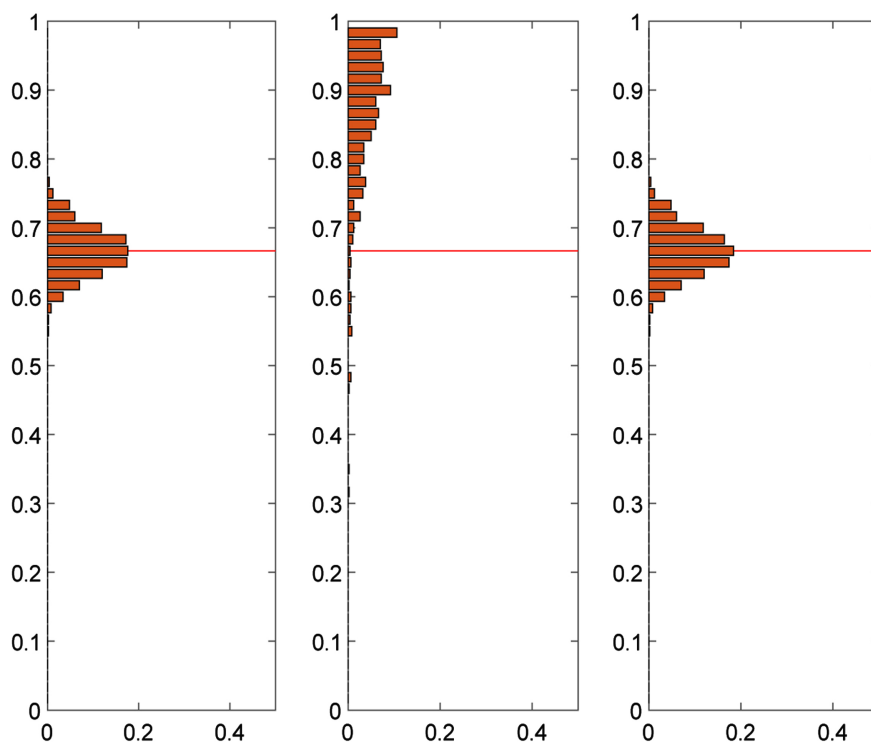


Figure 1. Horizontal bar plot: x vs. bins. This figure shows the distribution of $x(1, \varepsilon)$ represented by the vertical axis of 500 sample paths versus bins in the horizontal axis at the final time $t = 1$ of the system (4.33). In this case, $x(0) = 1$, ≈ 0.67 , $0 \leq t \leq 1$, and $H = 0.1$, $\varepsilon = 0.01$. The left figure shows the distribution of $x(1, \varepsilon)$ using Euler without max step size. The middle figure shows the distribution of $x(1, \varepsilon)$ using ode15s solver without max step size. The right figure shows the distribution of $x(1, \varepsilon)$ using ode15s using max step size.

4.4. Example 2

We consider a stochastic process $x_\varepsilon(t)$ is defined by the differential equation

$$\begin{cases} \dot{x}(t) = -tx \left(y\left(\frac{t}{\varepsilon}\right) - 0.5 \right) \equiv f(t, \varepsilon, x, y), \\ x(0) = 1, \end{cases} \quad (4.34)$$

with

$$\begin{aligned} Y &= \{0, 1\}, \\ P &= \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}, \\ Q &= \begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{pmatrix}, \end{aligned}$$

and limiting distribution

$$\rho = \{1/3, 2/3\}$$

We want to solve (4.34) numerically over time interval $[0, 1]$ using

1) Forward Euler with smaller time scale.

2) evaluate $\bar{x} + \sqrt{\varepsilon} \tilde{x}$, where

$$\dot{\tilde{x}} = -t\tilde{x} \left(\frac{1}{3}y_1 + \frac{2}{3}y_2 - 0.5 \right) \equiv \bar{f}(t, \varepsilon, x), \tag{4.35}$$

and

$$d\tilde{x} = \bar{f}_x(t, \varepsilon, x)\tilde{x}dt + \sigma(t)dW. \tag{4.36}$$

3) compare 2. with 1., where $\bar{x}(t)$ and $\tilde{x}(t)$ are the solutions of (4.35) (using large time scale) and (4.36) respectively. And $\sigma(t)$ is computed using (2.15) along with ρ, f, \bar{f} , and $dW = \sqrt{dt}N(0,1)$. Comparison of the averaged system to the perturbed system as well as error convergence are illustrated in **Figure 2** and **Figure 3**.

4.5. Example 3

We consider a stochastic process $x_\varepsilon(t)$ is defined by the differential equation

$$\begin{cases} \dot{x}(t) = 1/2 + \cos\left(y\left(\frac{t}{\varepsilon}\right)\right)\cos(x) \equiv f(t, \varepsilon, x, y), \\ x(0) = 1, \end{cases} \tag{4.37}$$

with

$$Y = \{0,1\},$$

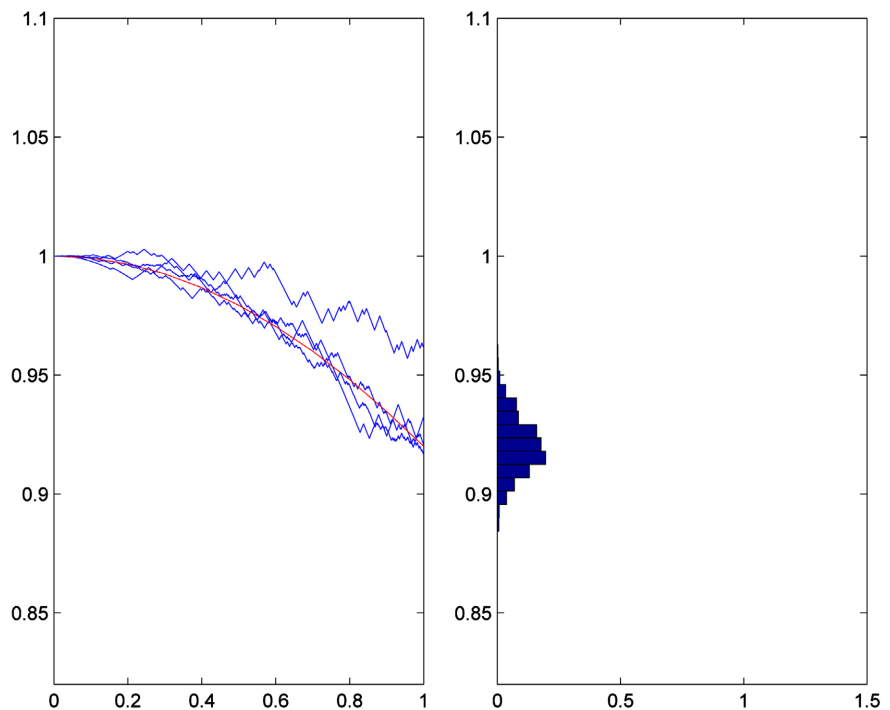


Figure 2. Left figure: x and \bar{x} vs. t . Comparison of the averaged system to the perturbed system. Red curve shows the solution of the averaged system; other curves are solutions of the perturbed system with different ε . Right figure: horizontal bar plot, where the vertical axis represents the distribution of solutions for 500 sample paths of the system (4.34) at the final time $t = 1$. In this case, $\varepsilon = 0.01$, $0 \leq t \leq 1$, y is generated with two state Markov process with $\bar{y} = 0.67$.

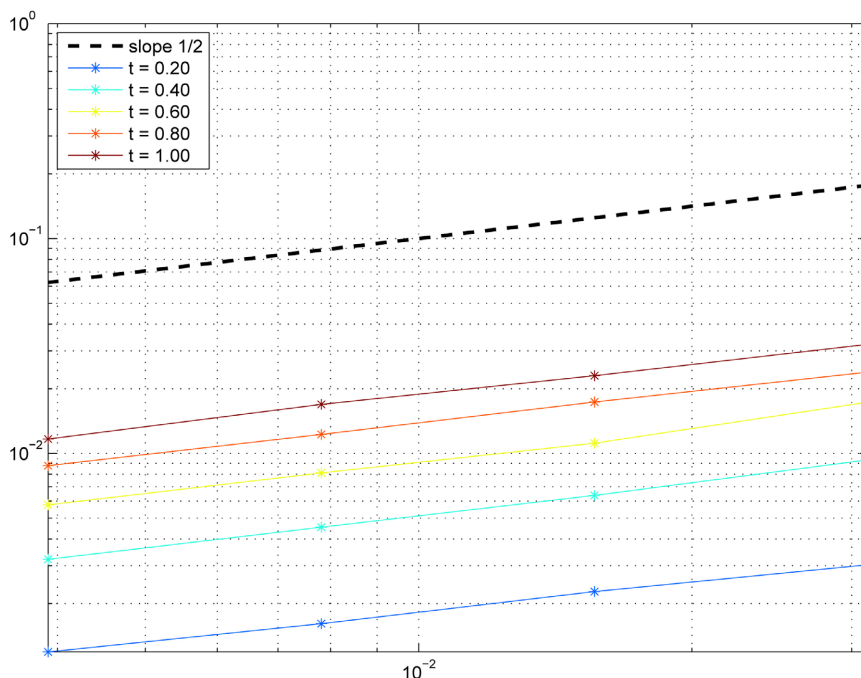


Figure 3. Error plots: $E|x(t) - (\bar{x}(t) + \sqrt{\epsilon}\tilde{x}(t))|$ represents vertical axis vs ϵ , which represents horizontal axis. Dashed line is the line with the reference slope 1/2. Graphs are drawn on log-log scale.

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix},$$

$$Q = \begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{pmatrix},$$

and limiting distribution

$$\rho = \{1/3, 2/3\}$$

We want to solve (4.37) numerically over time interval $[0,1]$ using

- 1) Forward Euler with smaller time scale.
- 2) evaluate $\bar{x} + \sqrt{\epsilon}\tilde{x}$ (using smaller time scale, h), where

$$\dot{\bar{x}} = 1/2 + \left(\frac{1}{3}\cos 0 + \frac{2}{3}\cos 1\right)\cos(\bar{x}) \equiv \bar{f}(t, \epsilon, x), \tag{4.38}$$

and

$$d\tilde{x} = \bar{f}_x(t, \epsilon, x)\tilde{x}dt + \sigma(t)dW. \tag{4.39}$$

3) compare 2. with 1., where $\bar{x}(x)$ and $\tilde{x}(t)$ is the solution of (4.38) (using large time scale), and (4.39) respectively. And $\sigma(t)$ is computed using (2.15) along with ρ, f, \bar{f} and $dW = \sqrt{dt}N(0,1)$. Comparison of the averaged system to the perturbed system as well as error convergence are illustrated in **Figure 4** and **Figure 5**.

4.6. Example 4

We consider a stochastic process $x_\epsilon(t)$ is defined by the differential equation

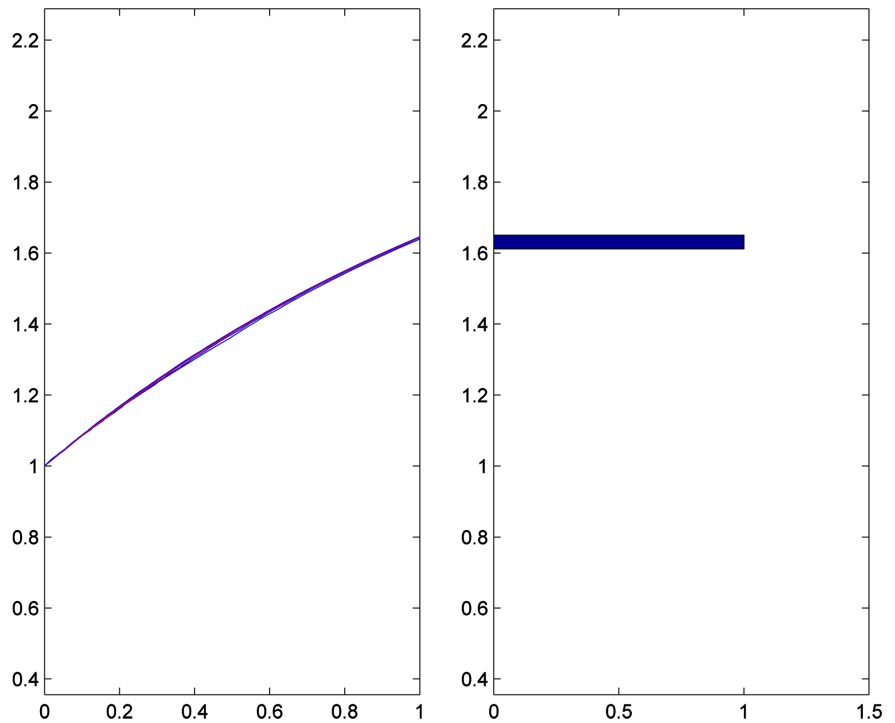


Figure 4. Left figure: x and \bar{x} vs. t . Comparison of the averaged system to the perturbed system. Red curve shows the solution of the averaged system; other curves are solutions of the perturbed system with different ε . Right figure: horizontal bar plot, where the vertical axis represents the distribution of solutions for 500 sample paths of the system (4.37) at the final time $t = 1$. Here $\varepsilon = 0.01$, $0 \leq t \leq 1$, y is generated with two state Markov process with $\bar{y} = 0.67$.

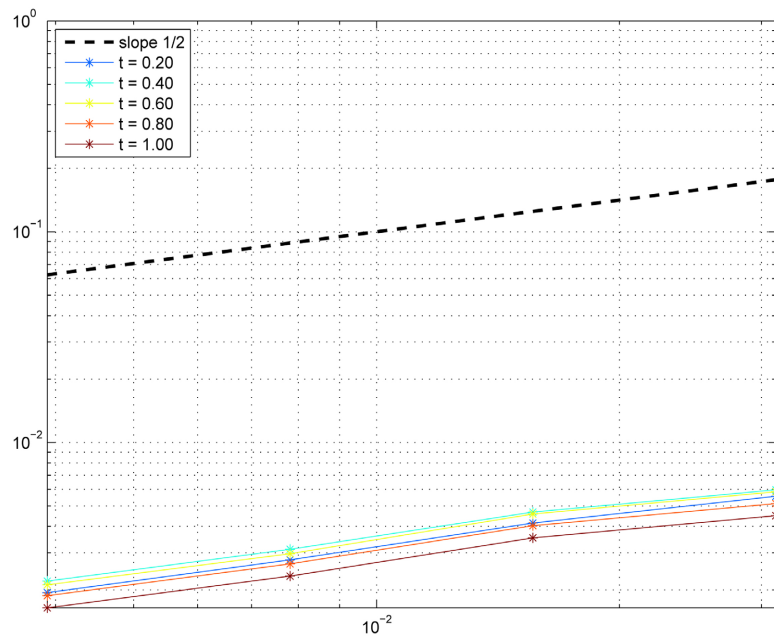


Figure 5. Error plots: $E \left| x(t) - \left(\bar{x}(t) + \sqrt{\varepsilon} \tilde{x}(t) \right) \right|$ represents vertical axis vs ε , which represents horizontal axis. Dashed line is the line with slope 1/2. Graphs are drawn on log-log scale.

$$\begin{cases} \dot{x}(t) = 1 + y\left(\frac{t}{\varepsilon}\right) - \left(3 + y\left(\frac{t}{\varepsilon}\right)\right)\cos(x) \equiv f(t, \varepsilon, x, y) \\ x(0) = 1, \end{cases} \quad (4.40)$$

with

$$\begin{aligned} Y &= \{0, 1\}, \\ P &= \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}, \\ Q &= \begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{pmatrix}, \end{aligned}$$

and limiting distribution

$$\rho = \{1/3, 2/3\}.$$

We want to solve (4.40) numerically over time interval $[0, 1]$ using

- 1) Forward Euler with smaller time scale.
- 2) evaluate $\bar{x} + \sqrt{\varepsilon}\tilde{x}$ (using smaller time scale, h), where

$$\dot{\tilde{x}} = 1 + y - (3 + y)\cos(x) \equiv \bar{f}(t, \varepsilon, x). \quad (4.41)$$

and

$$d\tilde{x} = \bar{f}_x(t, \varepsilon, x)\cos(\bar{x})\tilde{x}dt + \sigma(t)dW. \quad (4.42)$$

3) compare 2. with 1. where $\bar{x}(t)$ and $\tilde{x}(t)$ is the solution of (4.41) (using large time scale) and (4.42) respectively. And $\sigma(t)$ is computed using (2.15) along with ρ, f, \bar{f} , and $dW = \sqrt{dt}N(0, 1)$. Comparison of the averaged system to the perturbed system as well as error convergence are illustrated in **Figure 6** and **Figure 7**.

4.7. Strong Convergence

In the examples above, directly simulated solution $x_\varepsilon(t)$ with smaller time scale matches more closely to the solution $\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t)$, (where $\bar{x}(t)$ solved using larger time scale and $\tilde{x}(t)$ is the solution of $d\tilde{x} = \bar{f}_x(t, y, \bar{x})\tilde{x}dt + \sigma(t)dW$, using Euler with large time scale H) as ε is decreased, the convergence seems to take place. Using $E\left|x_\varepsilon(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))\right|$, where E denotes the expected value, leads the concept of strong convergence. A method is said to have strong order of convergence equal to m if there exists a constants K such that

$$E\left|x_\varepsilon(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))\right| \leq K(\varepsilon)^m, \quad (4.43)$$

and ε is sufficiently small. It can be shown that perturbation method has strong order of convergence $m = 1$. In our numerical tests, we will focus on the error at the end point $t = t_{final}$, so let

$$Error^{strong} = E\left|x_\varepsilon(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))\right|. \quad (4.44)$$

If the bound in (4.43) holds with $m = 1$ at any fixed point in $[0, t_{final}]$, then it certainly holds at the end point, so we have

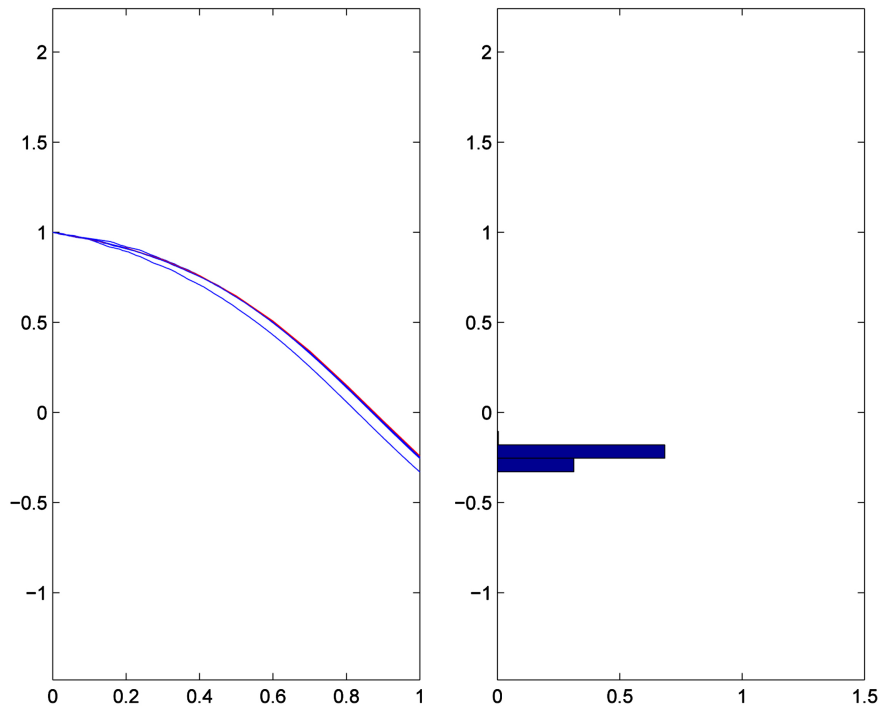


Figure 6. Left figure: x and \bar{x} vs. t . Comparison of the averaged system to the perturbed system. Red curve shows the solution of the averaged system; other curves are solutions of the perturbed system with different ε . Right figure: horizontal bar plot, where the vertical axis represents the distribution of solutions for 500 sample paths of the system (4.40) at the final time $t = 1$. In this case, $\varepsilon = 0.01$, $0 \leq t \leq 1$, y is generated with two state Markov process with $\bar{y} = 0.67$.

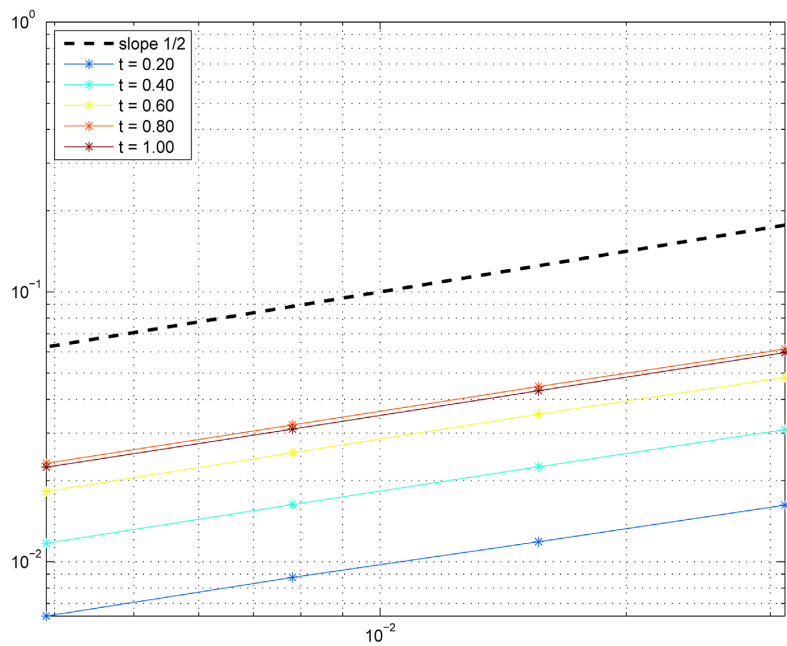


Figure 7. Error plots: The vertical axis represents $E|x(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))|$ vs ε , which represents the horizontal axis. Dashed line is the line with the reference slope 1/2. Graphs are drawn on log-log scale.

$$Error^{strong} \leq K\varepsilon^1, \quad (4.45)$$

for sufficiently small ε . It is shown that perturbation method has strong order of convergence $m = 1$. While experimenting the error $Error^{strong}$, we implicitly assumed that number of other sources of error are negligible, including error arising from approximating an expected value by sample mean, inherent error in the random generator, and floating point roundoff errors. For a typical computation the sampling error is likely to be the most significant of these three. In preparing the programs for these simulations we found that some experimentation is required to make the number of samples sufficiently large and the time step is sufficiently small for the predicted order of convergence to be observable. The sampling error decays like $1/\sqrt{n}$, where n is the number of sample paths used. A study in ([19]) indicates that as step size decreases, the lack of independence in the samples from a random generator typically degrades the computation before rounding errors becomes significant.

Although the definition of strong convergence involves an expected value, it has implications for individual simulations. The Markov inequality says that if a random variable ξ has a finite expected value, then for any $a > 0$ the probability that $|\xi| \geq a$ is bounded above by $E|\xi|/a$, that is,

$$P(|\xi| > a) \leq \frac{E|\xi|}{a}.$$

Hence taking $a = \varepsilon^{1/2}$, we see that perturbation method's strong convergence of order $m = 1$ is

$$P\left(\left|x_\varepsilon(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))\right| \geq \varepsilon^{1/2}\right) \leq K\varepsilon^{1/2},$$

or, equivalently,

$$P\left(\left|x_\varepsilon(t) - (\bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t))\right| < \varepsilon^{1/2}\right) \geq 1 - K\varepsilon^{1/2}.$$

This shows that the error at a fixed point in $[0, t_{final}]$ is small with probability close to 1.

4.8. Conclusion

Stochastic approximations for the process $\dot{x}(t)$ with parametric noise can be used to analyze aspects of noise. The result of the analysis is an approximation of the form $x(t) \approx \bar{x}(t) + \sqrt{\varepsilon}\tilde{x}(t) + o(\sqrt{\varepsilon})$ for $x(t)$ of the system. This can be used to evaluate the impact of parametric noise in the neural network. First of all, the system can be averaged. The system for \bar{x} may or may not be analytically solvable but we develop numerical method for its solution. Second, the next order term solves a linear system forced by a Gaussian process, whose statistics depends on the nature of noise in the model.

Our systems are characterized by some system components which combine very fast and very slow behavior. These systems require adaptable step-size, as only in certain phases they require very small step size. It is important to use in-

tegration method that allows an efficient step size control. A system is called stiff when integrated with an explicit algorithm and a local error tolerance 10^{-n} , the step size of the algorithm is forced down to below a value indicated by the local error estimate due to constraints imposed on it by the limited size of the numerical stable region. Ode15s is a variable-order solver based on numerical differentiations formulas, optionally uses the backward differentiations formulas (also known as Gear's method) like ode113, ode15s is multi-step solver. If one suspects that the problem is stiff or if ode45 fails or is very inefficient, try ode15s. But this slow vs. fast time scale problem, if we do not use max step size for the ode15s, the method does not give the correct result which is reflected in **Figure 1**.

Future work will address systems involving noisy neural network and the impact of noise on a node's information processing capability which is determined by its signal-to-noise ratio which can be estimated by spectral methods.

Acknowledgements

The author would like to thank Prof. Samir K. Bhowmik for his opinion and discussion. The author also thanks the unanimous referees for their careful reading the manuscript and useful suggestions that improved the paper significantly.

Conflicts of Interest

The author declares no conflict of interest regarding the publication of the paper.

References

- [1] Rahman, M. and Welfert, B. (2013) Functional Central Limit Theorem for Markov Processes and Chains. *Journal of Probability and Statistical Science (JPSS)*, **11**, 111-127.
- [2] Battal Gazi Karakoc, S. and Zeybek, H. (2016) Solitary Wave Solutions of the GRLW Equation Using Septic B Spline Collocation Method. *Applied Mathematics and Computation*, **289**, 159-172. <https://doi.org/10.1016/j.amc.2016.05.021>
- [3] Zeybek, H. and Battal Gazi Karakoc, S. (2017) Application of the Collocation Method with B-Splines to the GEW Equation. *Electronic Transactions on Numerical Analysis*, **46**, 77-88.
- [4] Battal Gazi Karakoc, S., Geyikli, T. and Bashana, A. (2013) A Numerical Solution of the Modified Regularized Long Wave MRLW Equation Using Quartic B Splines. *TWMS Journal of Applied and Engineering Mathematics*, **3**, 231-244. <https://doi.org/10.1186/1687-2770-2013-27>
- [5] Turgut, A.K. and Battal Gazi Karakoc, S. (2018) A Numerical Technique Based on Collocation Method for Solving Modified Kawahara Equation. *Journal of Ocean Engineering and Science*, **3**, 67-75. <https://doi.org/10.1016/j.joes.2017.12.004>
- [6] Turgut, A.K., Battal Gazi Karakoc, S. and Triki, H. (2016) Numerical Simulation for Treatment of Dispersive Shallow Water Waves with Rosenau KdV Equation. *The European Physical Journal Plus*, **131**, 1-15. <https://doi.org/10.1140/epjp/i2016-16356-3>

-
- [7] Bhowmik, S. and Rahman, M. (2019) Stability and Accuracy Analysis of Theta Scheme for a Convolutional Integro-Differential Equation. *Differential Equations and Dynamical Systems*, **28**, 633-646. <https://doi.org/10.1007/s12591-019-00476-w>
- [8] Kloeden, P.E. and Platen, E. (1999) Numerical Solution of Stochastic Differential Equations, Applications of Mathematics. Vol. 23, Corrected Third Printing, Springer-Verlag, Berlin.
- [9] Karlin, S. and Taylor, H.W. (1975) A First Course in Stochastic Processes. Academic Press, New York. <https://doi.org/10.1016/B978-0-08-057041-9.50005-2>
- [10] Gikhmann (1973) Differential Equations with Random Functions. AMS Transl. 12.
- [11] Borodin, A.N. and Salminen, P. (2002) Handbook Brownian Motion-Facts and Formulae. 2nd Edition, Probability and Its Applications, Birkhäuser. <https://doi.org/10.1007/978-3-0348-8163-0>
- [12] Rahman, M. (2018) Asymptotic Estimate of Variance with Applications to Stochastic Differential Equations Arises in Mathematical Neuroscience. *Communications in Statistics—Theory and Methods*, **27**, 289-306. <https://doi.org/10.1080/03610926.2017.1303729>
- [13] Skorokhod, A.V. (2000) On Randomly Perturbed Linear Oscillating Mechanical Systems. *Ukrainian Mathematical Journal*, **52**, 1483-1495. <https://doi.org/10.1023/A:1010392421925>
- [14] Skorokhod, A.V., Hoppensteadt, F.C. and Salehi, H. (2002) Random Perturbations Methods with Applications in Science and Engineering, Applied Mathematical Sciences. Vol. 150, Springer-Verlag, New York. <https://doi.org/10.1007/b98905>
- [15] Bhattacharya, R.N. (1982) Functional Central Limit Theorem for Markov Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **60**, 185-201. <https://doi.org/10.1007/BF00531822>
- [16] Berman, A. and Plemmons, R.J. (1994) Nonnegative Matrices in the Mathematical Sciences. Classics in Applied Mathematics, SIAM, Philadelphia. <https://doi.org/10.1137/1.9781611971262>
- [17] Romanovsky, V. (1970) Discrete Markov Chain. Wolters-Noordhoff Publishing, Groningen.
- [18] Gordin, M.I. and Lifsic, B.A. (1978) The Central Limit Theorem for Stationary Processes. *Soviet Mathematics—Doklady*, **19**, 392-394.
- [19] Komory, Y., Sato, Y. and Mitsui, T. (1994) Some Issues in Discrete Approximate Solution for Stochastic Differential Equations. *Computers & Mathematics with Applications*, **28**, 269-278. [https://doi.org/10.1016/0898-1221\(94\)00197-9](https://doi.org/10.1016/0898-1221(94)00197-9)