The Journal of Mathematical Neuroscience
a SpringerOpen Journal

RESEARCH   Open Access

CrossMark

A Stochastic Version of the Jansen and Rit Neural Mass Model: Analysis and Numerics

**Markus Ableidinger[1] · Evelyn Buckwar[1] ·
Harald Hinterleitner[1]**

Abstract Neural mass models provide a useful framework for modelling mesoscopic
neural dynamics and in this article we consider the Jansen and Rit neural mass model
(JR-NMM). We formulate a stochastic version of it which arises by incorporating random input and has the structure of a damped stochastic Hamiltonian system with nonlinear displacement. We then investigate path properties and moment bounds of the
model. Moreover, we study the asymptotic behaviour of the model and provide longtime stability results by establishing the geometric ergodicity of the system, which
means that the system—independently of the initial values—always converges to an
invariant measure. In the last part, we simulate the stochastic JR-NMM by an efficient numerical scheme based on a splitting approach which preserves the qualitative
behaviour of the solution.

Keywords Jansen and Rit neural mass model · Stochastic Hamiltonian system ·
Asymptotic behaviour · Stochastic splitting schemes

1 Introduction

Neural mass models have been studied as models describing coarse grained activity
of large populations of neurons [1–7] since the 1970s. They have successfully been
used to fit neuroimaging data, understanding EEG rhythms [8] or epileptic brain dynamics [9], and are now also a major building block in the Virtual Brain [10]. For a

✉ H. Hinterleitner
Harald.Hinterleitner@jku.at

M. Ableidinger
Markus.Ableidinger@jku.at

E. Buckwar
Evelyn.Buckwar@jku.at

[1]   Johannes Kepler University Linz, Altenberger Straße 69, Linz, 4040, Austria

Springer

summary on their history, applications and an outlook on their future possible use, we refer to [11]. In general, neural mass models can be derived as a mean-field limit from microscopic models [12] and involve just a few state variables such as average membrane potentials and average population firing rates.

In this article, we focus on the Jansen and Rit neural mass model (JR-NMM) [13], which has been introduced as a model in the context of human electroencephalography (EEG) rhythms and visual evoked potentials [14]. It dates back to the work of Lopes da Silva and Van Rotterdam [3, 5, 15]. The JR-NMM is a biologically motivated convolution-based model of a neuronal population involving two subpopulations, i.e. excitatory and inhibitory interneurons forming feedback loops, which can describe background activity, alpha activity, sporadic and also rhythmic epileptic activity.

The original JR-NMM is formulated as a set of three coupled second-order nonlinear ordinary differential equations (ODEs), i.e. these constitute a system of coupled nonlinear oscillators, often rewritten as the six-dimensional system of first-order equations. After introducing this system in Sect. 2, we rewrite the system in the format of classical mechanics, that is, as a damped Hamiltonian system with a nonlinear displacement. Furthermore, in most of the literature, the JR-NMM includes a term representing extrinsic input or background noise, which essentially is done by declaring that input function to be a stochastic process. Mathematically this implies that the solution process of the ODE system then also is a stochastic process inheriting the analytical properties of the input process and requiring some framework of stochastic analysis for its mathematical treatment. In Sect. 3 we discuss options for such a framework and in this article we choose to formulate a stochastic JR-NMM as a stochastic differential equation (SDE) with additive noise, in particular a stochastic damped Hamiltonian system with a nonlinear term. Systems of SDEs of this or similar form are well studied in the molecular dynamics literature, where they are often called Langevin equations.[1] In this article we provide a range of results employing various techniques available in the framework of stochastic analysis developed for SDEs: In Sect. 4 we establish basic properties of the SDE such as moment bounds and bounds on the path behaviour. Section 5 augments existing analysis of the dynamics of the deterministic JR-NMM, in particular we consider stochastic versions of equilibrium solutions, i.e. invariant measures, as well as the long-time behaviour of solutions of the SDE with respect to this invariant measure. These results may be interpreted as starting points for studies of phenomenological stochastic bifurcations or noise-induced transitions. Finally, in Sect. 6, we present efficient numerical methods designed for stochastic Hamiltonian problems and show that these numerical methods, which represent discrete stochastic systems for any fixed step-size, respect the properties previously established for the SDE system (subject to mild conditions on the step-size). Thus the resulting numerical methods will not only be quite efficient for future computational studies with the stochastic JR-NMM, they also provide *reliable* computational results.

---

[1]The term 'Langevin equation' goes back to Langevin's model of Brownian motion (1908) describing the velocity of a Brownian particle in a fluid by what is now called an Ornstein–Uhlenbeck process. However, for example, in the molecular dynamics literature the term refers to a class of models describing the interaction of heavy particles with light ones [16], whereas in some of the physics literature it may also refer to SDEs in general as in [17].

2 Description of the Original Jansen and Rit Neural Mass Model

A detailed summary of the model derivation from the neuroscientific point of view can be found in [18–20]. The main neural population, the excitatory and inhibitory interneurons, are in each case described by both a second-order ordinary differential operator, which transforms the mean incoming firing rate into the mean membrane potential, and a nonlinear function, which transforms the mean membrane potential into the mean output firing rate. For $t \in [0, T]$ with $T \in \mathbb{R}^+$, the JR-NMM proposed in [13] consists of three coupled nonlinear ODEs of second order

$$
\begin{aligned}
\ddot{x}_0(t) &= Aa\,\mathrm{Sigm}\big(x_1(t) - x_2(t)\big) - 2a\dot{x}_0(t) - a^2 x_0(t), \\
\ddot{x}_1(t) &= Aa\big[p(t) + C_2\,\mathrm{Sigm}\big(C_1 x_0(t)\big)\big] - 2a\dot{x}_1(t) - a^2 x_1(t), \\
\ddot{x}_2(t) &= BbC_4\,\mathrm{Sigm}\big(C_3 x_0(t)\big) - 2b\dot{x}_2(t) - b^2 x_2(t),
\end{aligned}
\tag{1}
$$

which can be written as the six-dimensional first-order ODE system

$$
\begin{aligned}
\dot{x}_0(t) &= x_3(t), \\
\dot{x}_1(t) &= x_4(t), \\
\dot{x}_2(t) &= x_5(t), \\
\dot{x}_3(t) &= Aa\,\mathrm{Sigm}\big(x_1(t) - x_2(t)\big) - 2ax_3(t) - a^2 x_0(t), \\
\dot{x}_4(t) &= Aa\big[p(t) + C_2\,\mathrm{Sigm}\big(C_1 x_0(t)\big)\big] - 2ax_4(t) - a^2 x_1(t), \\
\dot{x}_5(t) &= BbC_4\,\mathrm{Sigm}\big(C_3 x_0(t)\big) - 2bx_5(t) - b^2 x_2(t),
\end{aligned}
\tag{2}
$$

with initial value $(x_0(0), \dots, x_5(0))^T = \mathbf{x_0} \in \mathbb{R}^6$. Here, x_i for $i \in \{0, 1, 2\}$ describe the mean postsynaptic potentials of distinct neuronal populations. The *output signal* $y(t) := x_1(t) - x_2(t)$ describes the average membrane potential of the main family, i.e. the principal neurons of the JR-NMM (see [18, 19, 21]). The function $p : [0, T] \to \mathbb{R}$ describes the external input which may originate both from external sources or the activity of neighbouring neural populations. We will discuss the mathematical modelling of p in more detail at the end of this section. The sigmoid function $\mathrm{Sigm} : \mathbb{R} \to [0, v_{\max}]$, $v_{\max} > 0$ (as suggested in [4]) is given by

$$
\mathrm{Sigm}(x) := \frac{v_{\max}}{1 + e^{r(v_0 - x)}}
$$

and works as a gain function transforming the average membrane potential of a neural population into an average firing rate (see [22, 23]). The constant $v_{\max}$ denotes the maximum firing rate of the neural population, $v_0 \in \mathbb{R}$ is the value for which 50% of the maximum firing rate is attained and $r > 0$ determines the slope of the sigmoid function at v_0.

System (2) includes 11 parameters A, B, a, b, C_1, C_2, C_3, C_4, $v_{\max}$, r, v_0 and typical values for these parameters, taken from [13, 19], are given in Table 1. The parameters A, B, a and b model basic features of postsynaptic potentials. In particular, A and B denote the excitatory and inhibitory synaptic gain, respectively, and

Table 1 Typical values established in the original JR-NMM [13] taken from [19]

Parameter	Description	Typical value
A	Average excitatory synaptic gain	3.25 mV
B	Average inhibitory synaptic gain	22 mV
a^{-1}	Time constant of excitatory postsynaptic potential	10 ms
b^{-1}	Time constant of inhibitory postsynaptic potential	20 ms
C	Average number of synapses between the populations	135
C_1	Avg. no. of syn. established by principal neurons on excitatory interneurons	C
C_2	Avg. no. of syn. established by excitatory interneurons on principal neurons	0.8 C
C_3	Avg. no. of syn. established by principal neurons on inhibitory interneurons	0.25 C
C_4	Avg. no. of syn. established by inhibitory interneurons on principal neurons	0.25 C
v_{max}	Maximum firing rate of the neural populations (max. of sigmoid fct.)	$5\,s^{-1}$
v_0	Value for which 50% of the maximum firing rate is attained	6 mV
r	Slope of the sigmoid function at v_0	$0.56\,\text{mV}^{-1}$

a^{-1} and b^{-1} are corresponding time constants. The connectivity constants C_i for $i \in \{1, 2, 3, 4\}$, modelling the interactions between the main population and interneurons, are assumed to be proportional to a single parameter C which characterises the average number of synapses between populations (see [13]). The solution behaviour of System (2) depends sensitively on the values of the parameters (we refer to the bifurcation analyses in [18, 19, 24]). Especially, changes in the connectivity constants C_i can result in drastic changes of the solution path.

Subsequently, we will employ the Hamiltonian formulation of classical mechanics to study coupled oscillators such as System (1) or (2). Let $Q := (x_0, x_1, x_2)^T$ and $P := (x_3, x_4, x_5)^T$ denote three-dimensional vectors, then System (2) can be written as a *damped Hamiltonian system with nonlinear displacement*,

$$
\begin{aligned}
\frac{dQ}{dt} &= \nabla_P H(Q, P), \\
\frac{dP}{dt} &= -\nabla_Q H(Q, P) - 2\Gamma P + G(t, Q).
\end{aligned}
\tag{3}
$$

In this formulation, the system consists of a *Hamiltonian part* with Hamiltonian function $H : \mathbb{R}^6 \to \mathbb{R}_0^+$,

$$
H(Q, P) := \frac{1}{2}\big(\|P\|_{\mathbb{R}^3}^2 + \|\Gamma Q\|_{\mathbb{R}^3}^2\big),
$$

a *damping part* with damping matrix $\Gamma = \text{diag}[a, a, b] \in \mathbb{R}^{3 \times 3}$, and a *nonlinear part* given by the function $G : [0, T] \times \mathbb{R}^3 \to \mathbb{R}^3$, with

$$
G(t, Q) := \big(Aa\,\text{Sigm}(x_1 - x_2),\, Aa\big[p(t) + C_2\,\text{Sigm}(C_1 x_0)\big],\, BbC_4\,\text{Sigm}(C_3 x_0)\big)^T.
$$

The Hamiltonian $H(Q, P)$ may be interpreted as the total energy of an electrical RCL parallel resonant circuit; see [22]. In particular, $H(Q, P)$ is proportional to the

sum of the inductive and capacitive energy of the neuronal population, respectively. If the input $p(t)$ is a bounded deterministic function, the solution curve and the total energy $H(Q, P)$ are bounded (this is an immediate result of Theorem 4.2 in Sect. 4) and the time change in the total energy is given by

$$\frac{d}{dt}H(Q, P) = -2P^T \Gamma P + P^T G(t, Q).$$

In the original paper by Jansen and Rit [13], the external input $p(t)$ has been used to represent spontaneous background noise as well as peak-like functions for generating evoked potentials. In the latter case the extrinsic input has been modelled as a deterministic periodic function (see also [25]) and with this type of input, the solution of the System (1) (or (2) or (3)) remains a deterministic function and the mathematical background to treat it is deterministic analysis. In the former case, i.e. when $p(t)$ represents spontaneous background noise and is modelled as a stochastic process, the mathematical background immediately changes to be stochastic analysis. In particular, the solution of the Systems (1), (2) or (3) becomes a stochastic process and it inherits the mathematical properties of the input process $p(t)$. Within stochastic analysis, (1), (2) or (3) may be interpreted in different frameworks, with consequences depending on the specific choices of $p(t)$.

(i) Random Ordinary Differential Equation (RODE) framework: RODEs are pathwise ODEs involving a stochastic process in their right-hand side, i.e. for a sufficiently smooth function $f : \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and an m-dimensional stochastic process $\xi(t)$, a d-dimensional system of RODEs is given by

$$\dot{x}(t) = f\big(\xi(t), x(t)\big),$$

with an appropriate initial value. One may then choose the stochastic input process for example as a Wiener process or a coloured noise process, these processes exist in the classical sense and have almost surely continuous paths. In this framework standard deterministic analysis for e.g. guaranteeing existence and uniqueness of solutions can be applied pathwise; see for example [26], Chap. 1. However, the solution of this equation inherits the smoothness properties of the driving stochastic process $\xi(t)$, independent of the smoothness of the function f. Analysis of properties and dynamics of solutions of RODEs may be performed pathwise by standard analysis techniques, bearing in mind that the low smoothness of the solutions limits the applicability of many classical results, such as Taylor's theorem. We further refer to [27] for relevant results concerning random dynamical systems. Another consequence concerns the numerical treatment: as the order of convergence of classical numerical schemes for ODEs is determined by the smoothness of the solution of that ODE, when such schemes are applied pathwise to RODEs, they usually converge with a lower order than their expected one. In particular, they converge with order at most $1/2$ when the input process is chosen as the Wiener process or a coloured noise process, as their paths are only Hölder continuous of order less than $1/2$. We refer to [28] and its references for further information on numerical methods specifically designed for RODEs.

(ii) Stochastic Differential Equation framework: If one were to choose the stochastic input process in an RODE as above as a Gaussian white noise process, one would need to deal with the fact that such a process exists only in the sense of *generalised stochastic processes*; see [29], Sect. 3.2, or [30], Appendix I. In particular, Gaussian white noise is usually interpreted as the (generalised) derivative of the Wiener process, which itself is almost surely nowhere differentiable in the classical sense. It is much more convenient to work in the classical stochastic analysis framework designed to deal with differential equations 'subject to (white) noise' and interpret Systems (1), (2) or (3) as a stochastic differential equation; see also [29], Sect. 4.1. A considerable amount of results concerning analysis, dynamics, numerics, statistics, etc. of SDEs is available and for stochastic numerics we refer for example to [31], which also treats SDEs driven by coloured noise.

3 Jansen and Rit Neural Mass Model as a Damped Stochastic Hamiltonian System with Nonlinear Displacement

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space together with the filtration $\{\mathcal{F}_t\}_{t \in [0,T]}$ which is right-continuous and complete. We extend the model of System (2) by allowing perturbation terms such as $p(t)$ not only in $x_1(t)$ but in both $x_0(t)$ and $x_2(t)$ as well. For this purpose, we define the functions $\mu_i : [0, T] \to \mathbb{R}$ and $\sigma_i : [0, T] \to \mathbb{R}^+$ for $i \in \{3, 4, 5\}$. The functions μ_i will be used for representing deterministic input whereas σ_i will be used for scaling the stochastic components. In an analogous way to the exposition concerning stochastic oscillators in [32], Chap. 8, or in [33], Chap. 14.2, we symbolically introduce Gaussian white noise $\dot{W}_i$ representing the stochastic input into Eq. (1) as follows:

$$
\begin{aligned}
dX_0(t) &= X_3(t)\,dt, \\
dX_1(t) &= X_4(t)\,dt, \\
dX_2(t) &= X_5(t)\,dt, \\
dX_3(t) &= \big[Aa\big[\mu_3(t) + \mathrm{Sigm}\big(X_1(t) - X_2(t)\big)\big] - 2aX_3(t) - a^2 X_0(t)\big]\,dt \\
&\quad + \sigma_3(t)\,dW_3(t), \\
dX_4(t) &= \big[Aa\big[\mu_4(t) + C_2\,\mathrm{Sigm}\big(C_1 X_0(t)\big)\big] - 2aX_4(t) - a^2 X_1(t)\big]\,dt \\
&\quad + \sigma_4(t)\,dW_4(t), \\
dX_5(t) &= \big[Bb\big[\mu_5(t) + C_4\,\mathrm{Sigm}\big(C_3 X_0(t)\big)\big] - 2bX_5(t) - b^2 X_2(t)\big]\,dt \\
&\quad + \sigma_5(t)\,dW_5(t),
\end{aligned}
\tag{4}
$$

with deterministic initial value $(X_0(0), \dots, X_5(0))^T = \mathbf{X_0} \in \mathbb{R}^6$. Here, the processes $W_i(t)$ for $i \in \{3, 4, 5\}$ are independent, $\mathcal{F}_t$-adapted Wiener processes on $(\Omega, \mathcal{F}, \mathbb{P})$. Note that as the system above is an additive noise system the Itô and Stratonovich interpretations of that SDE system coincide. As for the deterministic case in Sect. 2,

we can use the (Q, P)-notation of classical mechanics

$$dQ(t) = \nabla_P H(Q, P) \, dt,$$
$$dP(t) = \big(-\nabla_Q H(Q, P) - 2\Gamma P + G(t, Q)\big) \, dt + \Sigma(t) \, dW(t), \tag{5}$$

with initial values

$$Q(0) = \big(X_0(0), X_1(0), X_2(0)\big)^T = Q_0 \in \mathbb{R}^3 \quad \text{and}$$
$$P(0) = \big(X_3(0), X_4(0), X_5(0)\big)^T = P_0 \in \mathbb{R}^3,$$

diffusion matrix

$$\Sigma(t) = \text{diag}\big[\sigma_3(t), \sigma_4(t), \sigma_5(t)\big] \in \mathbb{R}^{3\times 3},$$

and nonlinear displacement

$$G(t, Q) := \begin{pmatrix} Aa[\mu_3(t) + \text{Sigm}(X_1 - X_2)] \\ Aa[\mu_4(t) + C_2 \text{Sigm}(C_1 X_0)] \\ Bb[\mu_5(t) + C_4 \text{Sigm}(C_3 X_0)] \end{pmatrix}.$$

As before, we define the output signal as $Y(t) = X_1(t) - X_2(t)$.

Systems of the type (5), typically called Langevin equations, have received considerable attention in the literature of molecular dynamics (see [16] for an overview). In particular, the long-time properties of such systems have been intensively studied in [34–36]. We employ these techniques in Sect. 5 to study the long-time behaviour of System (5).

We briefly discuss the existence of a solution of Eq. (5). As the sigmoid function Sigm is globally Lipschitz continuous, the existence and pathwise uniqueness of an $\mathcal{F}_t$-adapted solution is a standard result; see e.g. in [29], Theorem 6.2.2. In particular, Q is continuously differentiable. In the current context, it makes sense to assume that the functions μ_i and σ_i are smooth and bounded which we will do in the following.

We simulate the solution of Eq. (5) with the splitting integrator (24) proposed in Sect. 6 and illustrate the output signal in Fig. 1. The coefficients and the noise components are chosen in such a way that the simulation results of [14] for varying connectivity constants C can be reproduced. The numerical values for the parameters are given in Table 1. For the deterministic part of the external inputs we set $\mu_3 = \mu_5 = 0$ and $\mu_4 = 220$, for the diffusion components we set $\sigma_3 = \sigma_5 = 10$ and $\sigma_4 = 1,000$ such that 'weak noise' is acting on the components X_3 and X_5; X_4 receives a stronger noise input. As in the original paper [14] we see (noisy) α-rhythm-like behaviour as well as spiking behaviour for varying connectivity constants C. In Fig. 2 we provide an illustration of changes in the system behaviour induced by including noise with plots of the phase portrait of the output signal for the case $C = 135$ and $C = 270$. The top two pictures show simulations of y of System (2), i.e. without noise, where the solution curves quickly converge towards a limit cycle. The bottom two pictures show a path of Y of System (5) and in particular for $C = 135$, the behaviour of the path is markedly different from the deterministic case.
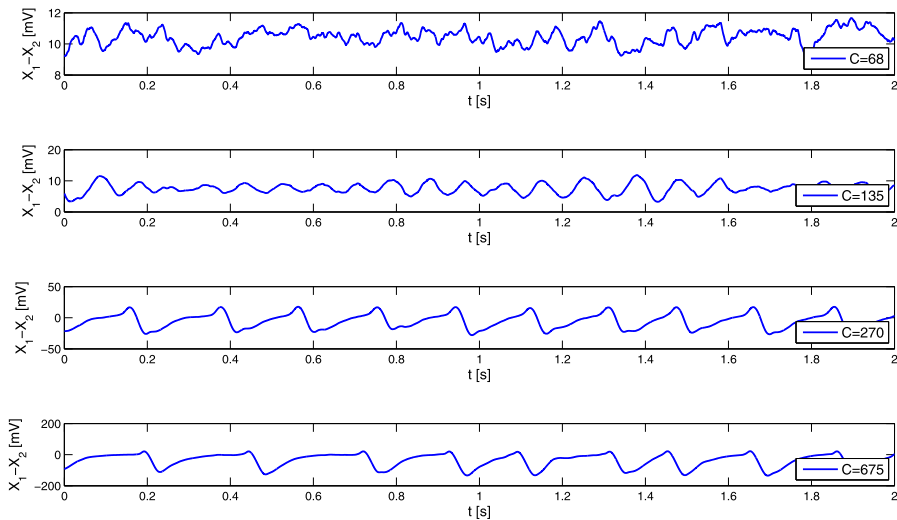
Trajectories of the output signal Y for varying coupling parameter C

Fig. 1 Output signal Y

4 Moment Bounds and Path Behaviour

We have already mentioned in Sect. 2 that the solution paths of Eq. (1) take values in a bounded set. It is natural to ask in which sense this behaviour transfers to the stochastic setting. We answer this question via a twofold strategy. On the one hand we will study the time evolution of the moments of the solution, which describes the average behaviour of all solution paths. On the other hand we will study the behaviour on the level of single paths and estimate the probability that a specific path exceeds a given threshold. Before we study these qualitative properties of Eq. (5) we provide a convolution-based representation for the Q-component of Eq. (5) which simplifies the corresponding calculations considerably.

4.1 Convolution-Based Representation of the JR-NMM

In this section we rewrite Eq. (5) using $X = (Q, P)^T$ as

$$dX(t) = \big(MX(t) + N\big(t, X(t)\big)\big)\, dt + S(t)\, dW(t), \tag{6}$$

where

$$M = \begin{pmatrix} \mathbb{O}_3 & \mathbb{I}_3 \\ -\Gamma^2 & -2\Gamma \end{pmatrix},$$

$$N\big(t, X(t)\big) = \begin{pmatrix} 0_3 \\ G(t, Q(t)) \end{pmatrix} \quad \text{and}$$

$$S(t) = \begin{pmatrix} \mathbb{O}_3 \\ \Sigma(t) \end{pmatrix}.$$

Phase portrait of y for the deterministic System (2)

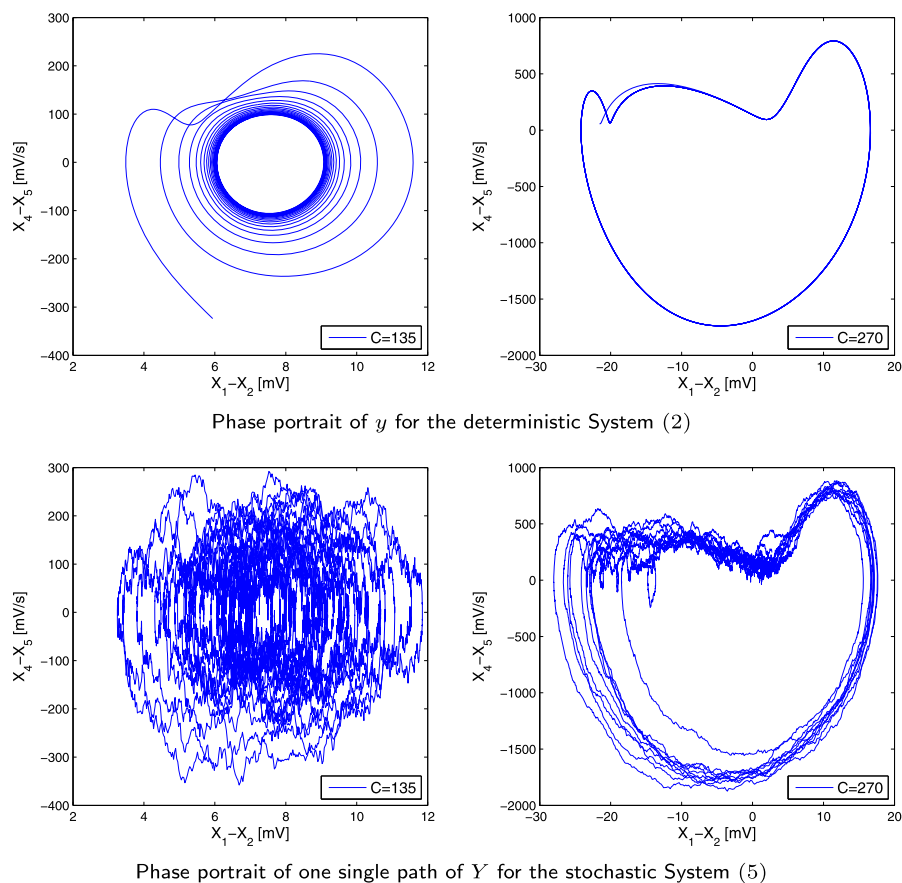Phase portrait of one single path of Y for the stochastic System (5)

Fig. 2 Phase portraits

Here, we denote by $\mathbb{O}_3, \mathbb{I}_3 \in \mathbb{R}^{3\times3}$ the zero and identity matrix, respectively. Moreover, we define $0_3 := (0,0,0)^T$ and $1_3 := (1,1,1)^T$.

Note that M is a block matrix with diagonal submatrices. Hence, we can calculate an explicit expression for the matrix exponential,

$$
e^{Mt} = \begin{pmatrix} e^{-\Gamma t}(\mathbb{I}_3 + \Gamma t) & e^{-\Gamma t}t \\ -\Gamma^2 e^{-\Gamma t}t & e^{-\Gamma t}(\mathbb{I}_3 - \Gamma t) \end{pmatrix}
$$
$$
=: \begin{pmatrix} \vartheta(t) & \kappa(t) \\ \vartheta'(t) & \kappa'(t) \end{pmatrix}. \tag{7}
$$

Obviously, the matrix exponential fulfils $e^{Mt}M = Me^{Mt}$. This allows one to represent solutions of Eq. (5) via the following convolution-based formula.

Theorem 4.1 *The component Q of the unique solution of Eq. (5) solves for $t \in [0, T]$ the integral equation*

$$Q(t) = \vartheta(t)Q_0 + \kappa(t)P_0 + \int_0^t \kappa(t-s)G\big(s, Q(s)\big)\,ds$$

$$+ \int_0^t \kappa(t-s)\Sigma(s)\,dW(s). \tag{8}$$

We call Eq. (8) the convolution-based representation of Q in Eq. (5).

Proof Applying Itô's formula ([29], Theorem 5.3.8) to the JR-NMM in Eq. (6) and using the commutativity of M and e^{Mt} we obtain

$$de^{-Mt}X(t) = e^{-Mt}\,dX(t) - Me^{-Mt}X(t)\,dt$$

$$= e^{-Mt}\big(N\big(t, X(t)\big)\,dt + S(t)\,dW(t)\big),$$

which reads in integral form

$$X(t) = e^{Mt}X(0) + \int_0^t e^{M(t-s)}N\big(s, X(s)\big)\,ds + \int_0^t e^{M(t-s)}S(s)\,dW(s).$$

Since the nonlinear part N only depends on Q the equation for Q is given by Eq. (8). □

Remark 1 From the latter proof we also get a convolution-based representation for P, however, this formula depends on Q. Indeed, for $t \in [0, T]$,

$$P(t) = \vartheta'(t)Q_0 + \kappa'(t)P_0 + \int_0^t \kappa'(t-s)G\big(s, Q(s)\big)\,ds + \int_0^t \kappa'(t-s)\Sigma(s)\,dW(s).$$

Remark 2 System (1) has originally been deduced by using convolutions of impulse response functions with functions of the output from other subpopulations within the neural mass (see [13, 18, 37, 38]). These response functions have the same shape as the kernel κ in Eq. (8), which thus can be interpreted as the stochastic version of this kernel representation.

4.2 Moment Bounds

Using representation (8) we provide bounds on the first and second moment of Q; analogous results can be derived for P. In the remainder of this section we will perform various componentwise calculations and estimations. For ease and consistency of notation we define the following: Let $x, y \in \mathbb{R}^n$, then $x \leq_\odot y$ denotes $x_i \leq y_i$ for all $1 \leq i \leq n$. Furthermore, for $U, V \in \mathbb{R}^{n \times k}$ we denote the *Hadamard product* of U and V as $U \odot V$, which is defined as the elementwise product (see [39, 40]) such that each element of the $n \times k$ matrix $U \odot V$ is given as

$$(U \odot V)_{ij} = U_{ij}V_{ij} \quad \text{for } 1 \leq i \leq n, 1 \leq j \leq k.$$

In addition, we define $U^{2\odot} := U \odot U$ and $U^{(1/2)\odot}$ as the elementwise root with $(U^{(1/2)\odot})_{ij} = \sqrt{U_{ij}}$.

Theorem 4.2 *Let $\mu_i : [0, T] \to \mathbb{R}^+$ for $i \in \{3, 4, 5\}$ be nonnegative functions bounded by $\mu_{i,\max} \in \mathbb{R}^+$, respectively, and $C_G := (Aa(\mu_{3,\max} + v_{\max}), Aa(\mu_{4,\max} + C_2 v_{\max}), Bb(\mu_{5,\max} + C_4 v_{\max}))^T$. Then $\mathbb{E}[Q(t)]$ is bounded in each component by*

$$\vartheta(t)Q_0 + \kappa(t)P_0 \leq_\odot \mathbb{E}[Q(t)] \leq_\odot \vartheta(t)Q_0 + \kappa(t)P_0 + \Gamma^{-2}(\mathbb{I}_3 - \vartheta(t))C_G.$$

Proof We write

$$Q(t) = \underbrace{\vartheta(t)Q_0 + \kappa(t)P_0}_{=:u(t)} + \underbrace{\int_0^t \kappa(t-s)G(s, Q(s))\,ds}_{=:v(t)} + \underbrace{\int_0^t \kappa(t-s)\Sigma(s)\,dW(s)}_{=:w(t)}.$$

Note that $\mathbb{E}[u(t)] = u(t)$ and that the expectation of an Itô integral is zero, i.e. $\mathbb{E}[w(t)] = 0_6$. Recall that $\text{Sigm} : \mathbb{R} \to [0, v_{\max}]$, thus $0_3 \leq_\odot G(t, Q(t)) \leq_\odot C_G$ and also $0_3 \leq_\odot \mathbb{E}[G(t, Q(t))] \leq_\odot C_G$. Applying the latter bounds to $\mathbb{E}[v(t)] = \int_0^t \kappa(t-s)\mathbb{E}[G(s, Q(s))]\,ds$ and integration of κ yield the desired estimates.   □

Obviously, the bounds provided by Theorem 4.2 also hold for the deterministic equation (3) which justifies our claim in the introduction.

Remark 3 The upper bound depends linearly on $\mu_{i,\max}$ and the connectivity constants C_i whereas $u(t)$ decays exponentially fast towards 0_3. In particular,

$$0_3 \leq_\odot \lim_{t \to \infty} \mathbb{E}[Q(t)] \leq_\odot \Gamma^{-2}C_G.$$

Similar calculations can be done for the second moments of the components of $Q(t)$. We obtain the following result.

Theorem 4.3 *Let the assumptions of Theorem 4.2 hold and assume $\Sigma(t)$ to be a constant matrix, $\Sigma \in \mathbb{R}^{3 \times 3}$. We define for $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ the function $\mathbf{1}_\odot^+(x) := (\mathbf{1}^+(x_1), \mathbf{1}^+(x_2), \mathbf{1}^+(x_3))^T$, where $\mathbf{1}^+$ denotes the indicator function of the set $\mathbb{R}^+$. Using the functions u, v and w from Theorem 4.2, a bound for the second moment of each component of $Q(t)$ reads*

$$\mathbb{E}[Q^{2\odot}(t)]$$
$$\leq_\odot u^{2\odot}(t) + 2u(t) \odot \mathbf{1}_\odot^+(u(t)) \odot \Gamma^{-2}(\mathbb{I}_3 - \vartheta(t))C_G$$
$$+ \left[\Gamma^{-2}(\mathbb{I}_3 - \vartheta(t))C_G + \frac{1}{2}\Gamma^{-3/2}\Sigma(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \vartheta^2(t))^{1/2}1_3\right]^{2\odot}.$$

In particular,

$$\lim_{t \to \infty} \mathbb{E}[Q^{2\odot}(t)] \leq_\odot \left(\Gamma^{-2}C_G + \frac{1}{2}\Gamma^{-3/2}\Sigma 1_3\right)^{2\odot}.$$
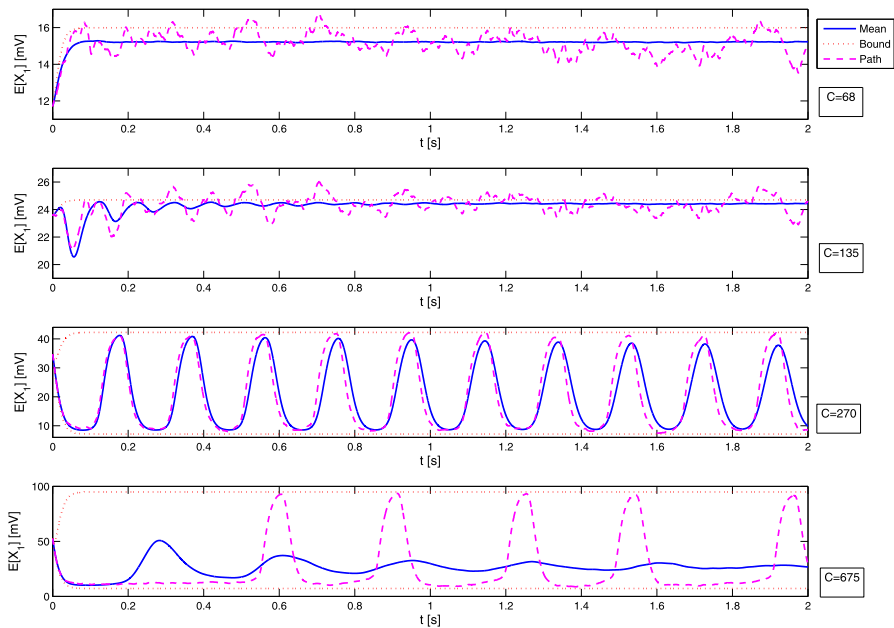
Fig. 3 Time evolution of $\mathbb{E}[X_1]$

Proof From the proof of Theorem 4.2 it immediately follows that

$$\mathbb{E}\big[u^{2\odot}(t)\big] = u^{2\odot}(t), \qquad \mathbb{E}\big[v^{2\odot}(t)\big] \leq \Gamma^{-4}\big(\mathbb{I}_3 - \vartheta(t)\big)^2 C_G^{2\odot} \quad \text{and} \tag{9a}$$

$$\mathbb{E}\big[w^{2\odot}(t)\big] = \frac{1}{4}\Gamma^{-3}\Sigma^2\big(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \vartheta^2(t)\big)1_3. \tag{9b}$$

The last equality can be shown by applying the Itô isometry. For notational simplicity we omit the dependence on t in the following. By using the Cauchy–Schwarz inequality we bound

$$\mathbb{E}[v \odot w] \leq_{\odot} \big(\mathbb{E}\big[v^{2\odot}\big] \odot \mathbb{E}\big[w^{2\odot}\big]\big)^{(1/2)\odot}.$$

Applying the bounds (9a)-(9b), the desired result follows from

$$\begin{aligned}
\mathbb{E}\big[Q^{2\odot}\big] &= u^{2\odot} + \mathbb{E}\big[v^{2\odot}\big] + \mathbb{E}\big[w^{2\odot}\big] + 2u \odot \mathbb{E}[v] + 2\mathbb{E}[v \odot w] \\
&\leq_{\odot} u^{2\odot} + 2u \odot \mathbf{1}_{\odot}^{+}(u) \odot \mathbb{E}[v] \\
&\quad + \big[\big(\mathbb{E}\big[v^{2\odot}\big]\big)^{(1/2)\odot} + \big(\mathbb{E}\big[w^{2\odot}\big]\big)^{(1/2)\odot}\big]^{2\odot}. \qquad \square
\end{aligned}$$

In Fig. 3 we employ Monte Carlo simulation to estimate $\mathbb{E}[X_1(t)]$ for varying coupling parameter C. The results for the second moment $\mathbb{E}[X_1^2(t)]$ are essentially the same; see Fig. 4. Similar results can be obtained for X_2 and X_3. The numerical approximations of the expectation (blue curves) stay well within the theoretical bounds
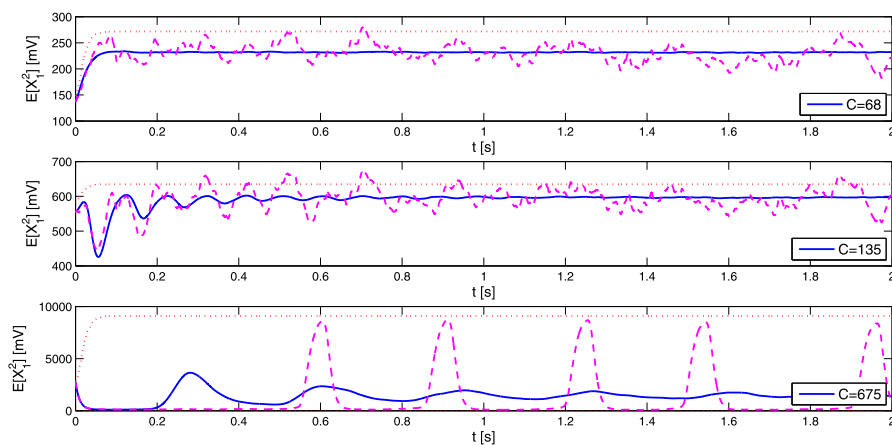
Fig. 4 Time evolution of $\mathbb{E}[X_1^2]$

(red curves), whereas single trajectories (purple curves) of course may exceed the bounds of the average. Note that, for $C = 68, 135$ and 675, the approximations of $\mathbb{E}[X_1(t)]$ rapidly converge towards fixed values for growing t. The same behaviour can be observed for $C = 270$ on larger time scales. We will give a theoretical explanation for this phenomenon in Sect. 5 when we study the long-time behaviour of Eq. (5).

4.3 Pathwise Bounds

Theorem 4.2 states that on average the solution of Eq. (5) stays in some bounded set. However, the theorem gives no information for single solution paths, which can in principle attain arbitrarily large values with positive probability; see Lemma A.2 in the Appendix. In this section we want to quantify the probability of such large values. The following theorem provides an upper bound on the escape probability of the components of Q, i.e. the probability that for $i \in \{0, 1, 2\}$ the solution X_i is larger than a given threshold $x_i^{th} \in \mathbb{R}^+$.

Theorem 4.4 *Let the assumptions of Theorem 4.3 hold. For fixed $t \in [0, T]$ we define a Gaussian random vector $Y(t) = (Y_0(t), Y_1(t), Y_2(t))$ with*

$$\mathbb{E}[Y(t)] = u(t) + \Gamma^{-2}(\mathbb{I}_3 - \vartheta(t))C_G \quad and$$

$$\text{Cov}[Y(t)] = \frac{1}{4}\Gamma^{-3}\Sigma^2(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \vartheta^2(t)),$$

where its components $Y_i(t)$ are independent. Let $F_{Y_i(t)}$ denote the cumulative distribution function of $Y_i(t)$. Then the probability that the components $X_i(t)$ for $i \in \{0, 1, 2\}$ exceed the given thresholds $x_i^{th} \in \mathbb{R}^+$ is bounded by

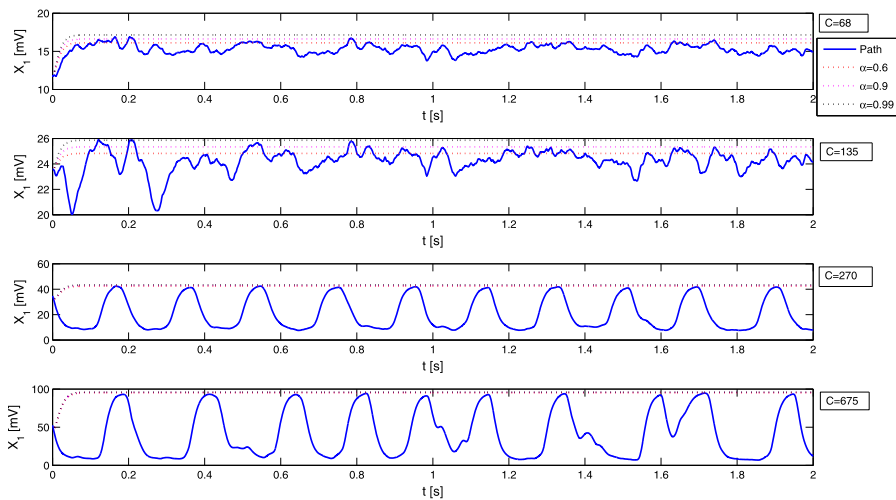$$\mathbb{P}(X_i(t) \geq x_i^{th}) \leq 1 - F_{Y_i(t)}(x_i^{th}).$$

Fig. 5 Pathwise bounds of X_1

Proof From Eq. (8), the bound on G and again integrating κ, we immediately see that each path of Q is bounded by the stochastic process Y defined by

$$Y(t) = u(t) + \Gamma^{-2}\big(\mathbb{I}_3 - \vartheta(t)\big)C_G + \int_0^t \kappa(t-s)\Sigma\, dW(s).$$

The process $Y(t)$ is Gaussian distributed with mean $u(t) + \Gamma^{-2}(\mathbb{I}_3 - \vartheta(t))C_G$ and covariance matrix $\frac{1}{4}\Gamma^{-3}\Sigma^2(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \vartheta^2(t))$, which were calculated in Theorems 4.2 and 4.3. Then, for each $t \in [0, T]$,

$$\mathbb{P}\big(X_i(t) \geq x_i^{th}\big) \leq \mathbb{P}\big(Y_i(t) \geq x_i^{th}\big) = 1 - F_{Y_i(t)}\big(x_i^{th}\big). \qquad \square$$

Remark 4 Theorem 4.4 can, for example, be used for calibration of the noise parameters in Σ. Let $\Sigma = \mathrm{diag}[\sigma_3, \sigma_4, \sigma_5]$. Suppose we want to choose σ_3 such that the corresponding component $X_0(t)$ stays below some given threshold x_0^{th} with high probability α. Then a suitable choice of σ_3 is implicitly given by $F_{Y_0(t)}(x_0^{th}) = \alpha$.

In Fig. 5 we illustrate numerical trajectories of Eq. (5) and the corresponding bounds for varying levels of α, i.e. for a given time point t, the probability of $X_1(t)$ to be below the red, purple and black curve is at least 60, 90 and 99 percent, respectively.

5 Long-Time Behaviour and Stationary Solutions

A further property of interest of an SDE concerns the asymptotic behaviour of solution trajectories. The classical approach in ODE theory for analysing the long-time behaviour of ODE systems is to study the stability of equilibrium solutions and limit cycles. Even in the simplest case of constant input $p(t) = p \in \mathbb{R}$, the deterministic

equation (3) can possess several equilibrium solutions (both stable and unstable) as well as limit cycles with typically nontrivial basins of attraction (again we refer to the bifurcation analyses in [18, 19, 24]). Thus, the choice of the initial value can have large impact on the long-time behaviour of the solution curves of Eq. (3). From a practical point of view, this fact may be problematic, as it is not at all obvious how to estimate the initial value of the P-component.

In this section, we will analyse a stochastic counterpart of equilibrium solutions, more precisely invariant measures, and study the long-time asymptotics of Eq. (5). Our main tool is the theory of ergodic Markov processes, for convenience of the reader we recapitulate the basic definitions.

Let $(X(t))_{t\in[0,T]}$ be the solution process of Eq. (5). Standard stochastic analysis shows that $X(t)$ is a Markov process and the corresponding transition probability $\mathbb{P}_t(\mathcal{A}, x)$, i.e. the probability that $X(t)$ reaches a Borel-set $\mathcal{A} \subset \mathbb{R}^6$ at time t, when it started in the point $x \in \mathbb{R}^6$ at time $t = 0$, is given by

$$\mathbb{P}_t(\mathcal{A}, x) := \mathbb{P}\big(X(t) \in \mathcal{A} | X(0) = x\big).$$

We use the definition provided in [41], Chap. 2, to characterise invariant measures. For simplicity, let η be a probability measure on $(\mathbb{R}^6, \mathcal{B}(\mathbb{R}^6))$ (in general η can degenerate on some lower dimensional space). The measure η is called *invariant* if

$$\eta(\mathcal{A}) = \int_{\mathbb{R}^6} \mathbb{P}_t(\mathcal{A}, x)\eta(dx) \quad \forall \mathcal{A} \in \mathcal{B}(\mathbb{R}^6), \forall t \in [0, T].$$

In particular, if we set the initial value (Q_0, P_0) to be a random vector with distribution η, then there exists a Markov process $X(t)$ which satisfies Eq. (5) and the distribution of $X(t)$ is η for all $t \in [0, T]$. In this sense, the concept of invariant measures can be seen as a natural extension of stationary solutions of deterministic ODEs.

We are interested in the following questions:

 (i) Does Eq. (5) have an invariant measure?
 (ii) Is the invariant measure unique?
(iii) Do quantities of the type $\mathbb{E}[h(X(t))]$ converge towards stationary values for a suitable class of functions $h : \mathbb{R}^6 \to \mathbb{R}$ and any initial value (Q_0, P_0)?

We answer these questions in two steps: In the first step we will show that Eq. (5) fulfills a Lyapunov condition ensuring the existence of (possibly many) invariant measures. The questions (ii) and (iii) will be answered positively via the concept of geometric ergodicity. In classical mechanics and molecular dynamics, equations with a similar structure as Eq. (5), termed Langevin equations in the corresponding literature (see e.g. [16, 31, 36]), are well studied and we make use of relevant results concerning the long-time behaviour. In particular, we follow the presentation in [34].

As in the bifurcation analyses for Eq. (3) mentioned before, we assume that the deterministic parts of the perturbation as well as the diffusion matrix are constant, i.e.

$$\mu_i(t) := \mu_i \in \mathbb{R}, \quad \text{for } i \in \{3, 4, 5\} \quad \text{and} \quad \Sigma(t) = \Sigma \in \mathbb{R}^{3\times 3}. \tag{10}$$

Thus, G does not depend on t and we simply write $G(Q)$.

5.1 Existence of Invariant Measures and Geometric Ergodicity

The existence of invariant measures for Eq. (5) can be established by finding a suitable Lyapunov function. Heuristically speaking, the existence of a Lyapunov function ensures both that the solution trajectories stay in some bounded domain (except for some rare excursions) and in the case of excursions, the trajectories return to the bounded set. The following lemma shows that a perturbed version of the Hamiltonian H in Eq. (5) can act as a Lyapunov function (see [42]).

Lemma 5.1 *Assume $a, b > 0$ and let for $n \in \mathbb{N}$*

$$V_n(Q, P) := \left(1 + \frac{1}{2} \|P\|_{\mathbb{R}^3}^2 + \frac{3}{2} \|\Gamma Q\|_{\mathbb{R}^3}^2 + \langle P, \Gamma Q \rangle \right)^n, \quad n \in \mathbb{N}.$$

Then V_n is a Lyapunov function for Eq. (5) in the following sense:

(i) $V_n \geq 1$ *and* $V_n \to \infty$ *for* $\|(Q, P)^T\|_{\mathbb{R}^6} \to \infty$,
(ii) $\exists \alpha_n < 0, \beta_n > 0$ *such that*

$$\mathcal{L} V_n \leq \alpha_n V_n + \beta_n,$$

where $\mathcal{L}$ denotes the generator of Eq. (5),

$$\mathcal{L} := P^T \nabla_Q + \left[-Q^T \Gamma^2 - 2P^T \Gamma + G(Q)^T \right] \nabla_P + \frac{1}{2} \sum_{i=3}^{5} \sigma_i^2 \frac{\partial^2}{\partial X_i^2}.$$

Here, ∇_Q and ∇_P denote the gradient with respect to the Q and P component, respectively.

Proof For $a, b > 0$, Property (i) is satisfied by construction and we only have to prove (ii). In a first step we set $n = 1$ and analyse the action of $\mathcal{L}$ on V_1. Note that V_1 is quadratic, therefore the second derivatives in $\mathcal{L}$ result in constants. Since $P^T \Gamma^2 Q = Q^T \Gamma^2 P$, we obtain

$$\mathcal{L} V_1 = P^T \left(3\Gamma^2 Q + \Gamma P \right)$$

$$+ \left[-Q^T \Gamma^2 - 2P^T \Gamma + G(Q)^T \right] (P + \Gamma Q) + \frac{1}{2} \sum_{i=3}^{5} \sigma_i^2$$

$$= -P^T \Gamma P - Q^T \Gamma^3 Q + \frac{1}{2} \sum_{i=3}^{5} \sigma_i^2 + G(Q)^T (P + \Gamma Q). \qquad (11)$$

The first two terms in Eq. (11) are quadratic and non-positive. As V_1 is quadratic and non-negative, there always exist constants $\widetilde{\alpha} < 0$ and $\widetilde{\beta} > 0$ such that the first three terms in Eq. (11) can be bounded in the following way:

$$-P^T \Gamma P - Q^T \Gamma^3 Q + \frac{1}{2} \sum_{i=3}^{5} \sigma_i^2 \leq \widetilde{\alpha} V_1(Q, P) + \widetilde{\beta}.$$

Furthermore, Young's inequality implies

$$G(Q)^T (P + \Gamma Q) \leq \frac{\epsilon}{2} \left(\|P\|_{\mathbb{R}^3}^2 + \|\Gamma Q\|_{\mathbb{R}^3}^2 \right) + \frac{1}{2\epsilon} \|G(Q)\|_{\mathbb{R}^3}^2,$$

where $\epsilon > 0$ can be chosen arbitrarily small. Thus, there exist $\widetilde{\alpha} < \alpha < 0$ and $\beta > 0$ sufficiently large such that

$$\mathcal{L} V_1(Q, P) \leq \alpha V_1(Q, P) + \beta. \tag{12}$$

In a second step, we will extend this procedure to V_n for $n \in \mathbb{N}$. For brevity we use the notation

$$\mathcal{L}_2 := \frac{1}{2} \sum_{i=3}^{5} \sigma_i^2 \frac{\partial^2}{\partial X_i^2} \quad \text{and} \quad \mathcal{L}_1 := \mathcal{L} - \mathcal{L}_2.$$

$\mathcal{L}_1$ is a first-order differential operator and the action of $\mathcal{L}_1$ on V_n can be bounded in the following way by using Eq. (12):

$$\mathcal{L}_1 V_n = n V_{n-1} \mathcal{L}_1 V_1 \leq \alpha n V_n + n \beta V_{n-1}.$$

Applying $\mathcal{L}_2$ to V_n leads to

$$\mathcal{L}_2 V_n = \frac{n}{2} \sum_{i=3}^{5} \sigma_i^2 \left(V_{n-1} \frac{\partial^2 V_1}{\partial X_i^2} + (n-1) V_{n-2} \left(\frac{\partial V_1}{\partial X_i} \right)^2 \right)$$

$$= \frac{n}{2} \sum_{i=3}^{5} \sigma_i^2 \left(V_{n-1} + (n-1) V_{n-2} (X_i + \Gamma_{i-2,i-2} X_{i-3})^2 \right),$$

where $\Gamma_{i,j}$ denotes the entry of Γ at the intersection of the ith row and jth column. Note that

$$\sum_{i=3}^{5} \sigma_i^2 (X_i + \Gamma_{i-2,i-2} X_{i-3})^2 \leq \max\{\sigma_3^2, \sigma_4^2, \sigma_5^2\} \|\Gamma Q + P\|_{\mathbb{R}^3}^2.$$

This implies that there exist $c, c_n > 0$ such that

$$\frac{n(n-1)}{2} V_{n-2} \sum_{i=3}^{5} \sigma_i^2 (X_i + \Gamma_{i-2,i-2} X_{i-3})^2 \leq c V_{n-1},$$

and thus

$$\mathcal{L}_2 V_n \leq c_n V_{n-1}.$$

As a consequence, there exist $\alpha_n < 0$ (possibly close to zero) and $\beta_n > 0$ such that

$$\mathcal{L} V_n = \mathcal{L}_1 V_n + \mathcal{L}_2 V_n \leq \alpha n V_n + (n\beta + c_n) V_{n-1} \leq \alpha_n V_n + \beta_n,$$

and Property (ii) follows. □

Lemma 5.1 has two immediate consequences. First, applying Itô's formula on V_n we obtain the following bounds (see [34]).

Corollary 5.2 *Let Assumption* (10) *hold and* $s, t \in [0, T]$ *with* $t \geq s$. *Then*

$$\mathbb{E}\big[V_n\big(Q(t), P(t)\big)|\mathcal{F}(s)\big] \leq e^{-|\alpha_n|(t-s)} V_n\big(Q(s), P(s)\big) + \frac{\beta_n}{|\alpha_n|}\big(1 - e^{-|\alpha_n|(t-s)}\big).$$

In particular,

$$\mathbb{E}\big[V_n\big(Q(t), P(t)\big)\big] \leq e^{-|\alpha_n|t} V_n\big(Q(0), P(0)\big) + \frac{\beta_n}{|\alpha_n|}\big(1 - e^{-|\alpha_n|t}\big).$$

Second, the existence of a Lyapunov function ensures the existence of an invariant measure (see e.g. [43], Corollary 1.11).

Corollary 5.3 *Let Assumption* (10) *hold and let* $X(t)$ *denote the solution of Eq.* (5). *Then there exists an invariant measure* $\boldsymbol{\eta}$ *of* $X(t)$.

Lemma 5.1 does not give any information on the uniqueness of the invariant measure. If we further assume that the three Wiener processes W_i act on all components of P, i.e. $\sigma_i > 0$ for $i \in \{3, 4, 5\}$, we can establish the uniqueness of the invariant measure. Furthermore, the Markov process X fulfills the property of geometric ergodicity in the sense of [34]. We give a modification of the result in [34], Theorem 3.2, including the nonlinear function G.

Theorem 5.4 *Let* $\sigma_i > 0$ *for all* $i \in \{3, 4, 5\}$. *The Markov process* $X(t)$ *defined by Eq.* (5) *has a unique invariant measure* $\boldsymbol{\eta}$ *on* $\mathbb{R}^6$. *Furthermore, let*

$$\mathcal{H}_n := \big\{h : \mathbb{R}^6 \to \mathbb{R} \text{ Borel-measurable with } |h| \leq V_n\big\}.$$

Then for any $n \in \mathbb{N}$ *and any initial value* $X(0) = (Q_0, P_0)$ *there exist positive constants* C_n, λ_n *such that*

$$\left|\mathbb{E}\big[h\big(X(t)\big)\big] - \int_{\mathbb{R}^6} h \, d\boldsymbol{\eta}\right| \leq C_n V_n\big(X(0)\big) e^{-\lambda_n t} \quad \forall h \in \mathcal{H}_n, \forall t \geq 0.$$

Proof The proof is the same as in [34]. The Lyapunov condition has been established in Lemma 5.1, the corresponding results for the necessary smoothness of the transition probabilities and the irreducibility of the Markov process are given in the Appendix in Lemma A.2 and A.1. Both lemmas rely on the assumption that $\sigma_i > 0$ for $i \in \{3, 4, 5\}$. □

Theorem 5.4 has two implications for the numerical simulation of Eq. (5). First, the actual choice of the initial value is insignificant as the impact of the initial value on the distribution of $X(t)$ diminishes exponentially fast for growing t and an appropriate approximation of the system behaviour should be obtained with any choice of (Q_0, P_0) provided that the system is simulated on a large enough time horizon.
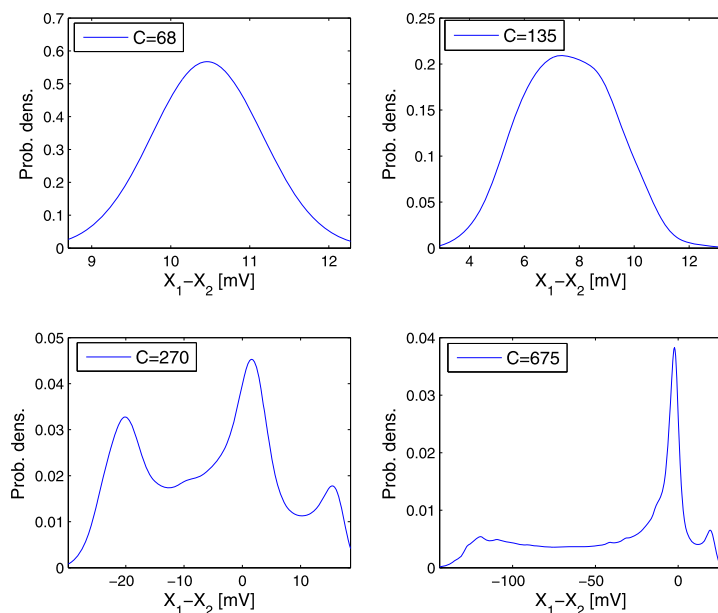
Fig. 6 Densities of the invariant measure of Y

Second, due to the correspondence of the time averages and "space averages" of ergodic systems (see [41], Theorem 3.2.4.), one can estimate quantities of the type $\mathbb{E}[h(X(t))]$ (for t sufficiently large) by computing the time average of a single path of $X(t)$ on a large time horizon instead of using Monte Carlo estimation which requires one to compute a large number of paths of $X(t)$. Of course, both aspects hold only true if the numerical method reproduces the geometric ergodicity of the original system (see Sect. 6).

The computation of the invariant measure for nonlinear systems is highly nontrivial. One possibility would be to solve the corresponding Fokker–Planck equation, which is a six-dimensional nonlinear PDE. A standard alternative is to use stochastic simulation techniques to approximate the marginal densities of η. Several possibilities have been proposed in the literature how to estimate the distribution of the solution of SDEs; see e.g. [44] for an approach based on Malliavin calculus and kernel density estimation (see [45], Chap. 2). We use the latter approach where we choose the kernel functions to be Gaussian. The numerical samples are calculated as a long-time simulation of a single path with the splitting integrator Eq. (24) proposed in Sect. 6. In Fig. 6 we compare approximations of the stationary probability density of the output signal Y for varying coupling parameter C. We observe the change from unimodal densities (for $C = 68, 135$) to multi-modal densities $C = 270$ and to the peak-like structure $C = 675$. This behaviour can be interpreted as a phenomenological stochastic bifurcation as discussed, for example, in [46] or a noise-induced transition (see [47]).

6 Numerical Simulation

In order to obtain an approximation of Eq. (5) which accurately reproduces the qualitative behaviour, it is highly important to construct numerical integrators which on the one hand fulfil the properties of Eq. (5) derived in Sects. 4 and 5, and on the other hand are computationally efficient such that large ensembles of trajectories can be calculated in reasonable time.

We want to emphasise that the difficulty does not lie in the construction of a mean-square convergent integrator for Eq. (5). In fact, as the coefficient functions of Eq. (5) are globally Lipschitz continuous, any standard integrator (e.g. the Euler–Maruyama method) converges in the mean-square sense. However, it has already been shown for linear stochastic oscillators that the Euler–Maruyama method does not preserve second moment properties of that system [48] and it is expected that this negative result extends to nonlinear stochastic oscillators as well. The splitting methods fulfil the following properties:

(i) The methods preserve the moment bounds proposed in Theorems 4.2 and 4.3. Furthermore, for $\Sigma = \mathbb{O}_3$, the numerical method preserves the bounds of the exact solution.
(ii) The Markov process generated by the numerical method is geometrically ergodic and fulfils a Lyapunov condition under very mild step-size restrictions.

6.1 Splitting Integrators for the JR-NMM

For convenience of the reader we provide a brief introduction to splitting integrators. Further details can be found, for example, in the classical monograph [49], Chapter II, for the deterministic case and [35, 50] for stochastic Langevin-type equations.

The main idea of splitting integrators is the following: Assume for simplicity we want to approximate a deterministic ODE system

$$\dot{y} = f(y), \qquad y(0) = y_0 \in \mathbb{R}^n$$

for which the function $f : \mathbb{R}^n \to \mathbb{R}^n$ can be written as

$$f(y) = \sum_{i=1}^{d} f^{[j]}(y), \quad \text{where } f^{[j]} : \mathbb{R}^n \to \mathbb{R}^n \text{ for } j \in \{1, \dots, d\}.$$

Of course, there can be several possibilities to decompose f. The goal is to choose $f^{[j]}$ in such a way that the subsystems

$$\dot{y} = f^{[j]}(y), \quad j \in \{1, \dots, d\}, \tag{13}$$

can be solved exactly. Let $\varphi_t^{[j]}(y_0)$ denote the exact flow of the Subsystem (13) with initial value y_0. Then the following compositions of flows define integrators of deter-

ministic order one (Lie–Trotter splitting) and two (Strang splitting):

$$\Psi_{\Delta t}^{\text{LT}}(y) = \big(\varphi_{\Delta t}^{[1]} \circ \cdots \circ \varphi_{\Delta t}^{[d]}\big)(y),$$

$$\Psi_{\Delta t}^{\text{S}}(y) = \big(\varphi_{\Delta t/2}^{[1]} \circ \cdots \circ \varphi_{\Delta t/2}^{[d-1]} \circ \varphi_{\Delta t}^{[d]} \circ \varphi_{\Delta t/2}^{[d-1]} \circ \cdots \circ \varphi_{\Delta t/2}^{[1]}\big)(y).$$

This strategy can be extended to the stochastic setting (see [35] and the references therein for splitting integrators in the field of molecular dynamics, [50] for quasi-symplectic splitting integrators, [51] for variational integrators based on a splitting approach and [52] for splitting integrators in a Lie group setting). In particular, splitting integrators have been applied efficiently to Langevin equations with a similar structure as Eq. (5), see [35, 50], thus we extend this approach here to Eq. (5).

The main step in the construction of splitting integrators is to choose a suitable decomposition of the coefficient functions of Eq. (5). The right-hand side decomposes into three rather distinct parts: First, a damped, linear oscillatory part, second, a nonlinear and non-autonomous coupling part which does not depend on the P-component, and third, a stochastic part which does only arise in the P-component. Therefore,

$$\begin{pmatrix} dQ \\ dP \end{pmatrix} = \underbrace{\begin{pmatrix} \nabla_P H(Q, P) \\ -\nabla_Q H(Q, P) - 2\Gamma P \end{pmatrix}}_{\text{damp. lin. oscill.}} dt + \underbrace{\begin{pmatrix} 0_3 \\ G(t, Q) \end{pmatrix}}_{\text{nonlin. coupl.}} dt + \underbrace{\begin{pmatrix} 0_3 \\ \Sigma(t)\,dW(t) \end{pmatrix}}_{\text{stochastic}}$$

with the nonlinear term given by

$$G(t, Q) = G^{\text{I}}(t) + G^{\text{II}}(Q) := \begin{pmatrix} Aa\mu_3(t) \\ Aa\mu_4(t) \\ Bb\mu_5(t) \end{pmatrix} + \begin{pmatrix} Aa\,\text{Sigm}(X_1 - X_2) \\ AaC_2\,\text{Sigm}(C_1 X_0) \\ BbC_4\,\text{Sigm}(C_3 X_0) \end{pmatrix}.$$

It makes sense to split the linear and the nonlinear drift contributions, thus providing two options to incorporate the stochastic noise. This yields two different sets of subsystems and therefore two different sets of numerical methods. In the first case, the stochastic subsystem defines a general Ornstein–Uhlenbeck process and we denote the corresponding splitting integrator as *Ornstein–Uhlenbeck integrator*. In the second case, the stochastic subsystem defines a Wiener process with drift and we denote the splitting integrator as *Wiener integrator*.

In the following let $0 = t_0 < \cdots < t_N = T$ with $N \in \mathbb{N}$ be an equidistant partition of $[0, T]$ with step-size Δt.

6.2 Ornstein–Uhlenbeck Integrator

The first variant is to include the stochastic contribution into the linear oscillator part, which gives rise to the two following subsystems:

$$\begin{pmatrix} dQ^{[1]} \\ dP^{[1]} \end{pmatrix} = \begin{pmatrix} \nabla_{P^{[1]}} H(Q^{[1]}, P^{[1]}) \\ -\nabla_{Q^{[1]}} H(Q^{[1]}, P^{[1]}) - 2\Gamma P^{[1]} \end{pmatrix} dt + \begin{pmatrix} 0_3 \\ \Sigma(t)\,dW(t) \end{pmatrix}, \qquad (14a)$$

$$\begin{pmatrix} dQ^{[2]} \\ dP^{[2]} \end{pmatrix} = \begin{pmatrix} 0_3 \\ G(t, Q^{[2]}) \end{pmatrix} dt. \qquad (14b)$$

For both subsystems we can easily derive explicit representations of the exact solutions which can be used directly for the numerical simulation. Subsystem (14a) is a six-dimensional Ornstein–Uhlenbeck process. Let $X^{[1]}(t_i) = (Q^{[1]}(t_i), P^{[1]}(t_i))^T$ denote the solution of Eq. (14a) at time point t_i for $i \in \{0, \ldots, N-1\}$, then the exact solution at time point $t_{i+1} > t_i$ can be represented as

$$X^{[1]}(t_{i+1}) = e^{M\Delta t} X^{[1]}(t_i) + \int_{t_i}^{t_{i+1}} \begin{pmatrix} \kappa(t_{i+1} - s)\Sigma(s)\,dW(s) \\ \kappa'(t_{i+1} - s)\Sigma(s)\,dW(s) \end{pmatrix}, \qquad (15)$$

where $e^{M\Delta t}$ is defined in Eq. (7). $X^{[1]}$ is a Gaussian process with conditional expectation

$$\mathbb{E}\big[X^{[1]}(t_{i+1})|\mathcal{F}_{t_i}\big] = e^{M\Delta t} X^{[1]}(t_i)$$

and the conditional covariance matrix (see [29], Theorem 8.2.6)

$$\begin{aligned}
\mathrm{Cov}(t_{i+1}) &:= \mathrm{Cov}\big[X^{[1]}(t_{i+1}), X^{[1]}(t_{i+1})|\mathcal{F}_{t_i}\big] \\
&= \mathbb{E}\big[\big(X^{[1]}(t_{i+1}) - \mathbb{E}[X^{[1]}(t_{i+1})]\big)\big(X^{[1]}(t_{i+1}) - \mathbb{E}[X^{[1]}(t_{i+1})]\big)^T|\mathcal{F}_{t_i}\big] \\
&= \int_{t_i}^{t_{i+1}} e^{M(t_{i+1}-s)} \Sigma^2(s)\big(e^{M(t_{i+1}-s)}\big)^T ds.
\end{aligned}$$

In particular, the integral term in Eq. (15) can be simulated exactly. Indeed, it is Gaussian distributed with mean zero and covariance matrix $\mathrm{Cov}(t_{i+1})$, which is for $t \geq t_i$ given as the unique solution of the matrix-valued ODE

$$\frac{d\,\mathrm{Cov}(t)}{dt} = M\,\mathrm{Cov}(t) + \mathrm{Cov}(t)M^T + \begin{pmatrix} \mathbb{O}_3 & \mathbb{O}_3 \\ \mathbb{O}_3 & \Sigma^2(t) \end{pmatrix}, \qquad (16)$$

$$\mathrm{Cov}(t_i) = \mathbb{O}_6.$$

In the special case of a constant diffusion matrix $\Sigma(t) = \Sigma \in \mathbb{R}^{3\times3}$, the exact solution of Eq. (16) can be explicitly calculated for $t \geq 0$ as

$$\begin{aligned}
&\mathrm{Cov}(t_i + t) \\
&= \begin{pmatrix} \frac{1}{4}\Gamma^{-3}\Sigma^2(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \vartheta^2(t)) & \frac{1}{2}\Sigma^2\kappa^2(t) \\ \frac{1}{2}\Sigma^2\kappa^2(t) & \frac{1}{4}\Gamma^{-1}\Sigma^2(\mathbb{I}_3 + \kappa(t)\vartheta'(t) - \kappa'^2(t)) \end{pmatrix}.
\end{aligned}$$

In general, Eq. (16) has to be solved by numerical approximation, however, it only needs to be precomputed once for the step-size Δt. In either case, we obtain

$$X^{[1]}(t_{i+1}) = e^{M\Delta t} X^{[1]}(t_i) + \xi_i(\Delta t), \qquad (17)$$

where $\xi_i(\Delta t)$ are iid six-dimensional Gaussian random vectors with expectation $\mathbb{E}[\xi_i(\Delta t)] = 0_6$ and covariance matrix $\mathrm{Cov}(\Delta t)$.

Subsystem (14b) is a deterministic system and the solution can be obtained by integration with respect to time. As before let $X^{[2]}(t_i) = (Q^{[2]}(t_i), P^{[2]}(t_i))^T$ denote

the solution of Eq. (14b) at time point t_i, then the exact solution at time point $t_{i+1} > t_i$ is given by

$$Q^{[2]}(t_{i+1}) = Q^{[2]}(t_i),$$

$$P^{[2]}(t_{i+1}) = P^{[2]}(t_i) + \int_{t_i}^{t_{i+1}} G(s, Q^{[2]}(s)) \, ds \tag{18}$$

$$= P^{[2]}(t_i) + \Delta t \, G^{\mathrm{II}}(Q^{[2]}(t_i)) + \int_{t_i}^{t_{i+1}} G^{\mathrm{I}}(s) \, ds,$$

where we assume that the last integral can be calculated exactly.

Now, let $\varphi_t^{\mathrm{ou},[1]}$ and $\varphi_t^{\mathrm{ou},[2]}$ denote the exact flows of Eq. (14a) and (14b) given via Eq. (17) and (18), respectively. Let $x \in \mathbb{R}^6$, then a one-step integrator is defined by the composition of the flows

$$\psi_{\Delta t}^{\mathrm{ou}}(x) = \left(\varphi_{\Delta t}^{\mathrm{ou},[1]} \circ \varphi_{\Delta t}^{\mathrm{ou},[2]}\right)(x). \tag{19}$$

6.3 Wiener Integrator

The second possibility is to include the stochastic terms into the nonlinear contribution yielding the subsystems

$$\begin{pmatrix} dQ^{[1]} \\ dP^{[1]} \end{pmatrix} = \begin{pmatrix} \nabla_{P^{[1]}} H(Q^{[1]}, P^{[1]}) \\ -\nabla_{Q^{[1]}} H(Q^{[1]}, P^{[1]}) - 2\Gamma P^{[1]} \end{pmatrix} dt, \tag{20a}$$

$$\begin{pmatrix} dQ^{[2]} \\ dP^{[2]} \end{pmatrix} = \begin{pmatrix} 0 \\ G(t, Q^{[2]}) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \Sigma(t) \, dW(t) \end{pmatrix}. \tag{20b}$$

Subsystem (20a) is a deterministic system. Let $X^{[1]}(t_i) = (Q^{[1]}(t_i), P^{[1]}(t_i))^T$ denote the solution of Eq. (20a) at time point t_i, then the exact solution at time point t_{i+1} is given by

$$X^{[1]}(t_{i+1}) = e^{M \Delta t} X^{[1]}(t_i). \tag{21}$$

The solution of subsystem (20b) is—by definition—given by

$$Q^{[2]}(t_{i+1}) = Q^{[2]}(t_i),$$

$$P^{[2]}(t_{i+1}) = P^{[2]}(t_i) + \Delta t \, G^{\mathrm{II}}(Q^{[2]}(t_i)) \tag{22}$$

$$+ \int_{t_i}^{t_{i+1}} G^{\mathrm{I}}(s) \, ds + \int_{t_i}^{t_{i+1}} \Sigma(s) \, dW(s),$$

where the last term can be simulated exactly as a three-dimensional Gaussian random vector with zero mean and covariance matrix $\int_{t_i}^{t_{i+1}} \Sigma^2(s) \, ds$. In the case of a constant diffusion matrix $\Sigma(t) = \Sigma \in \mathbb{R}^{3 \times 3}$, the covariance matrix is given as $\Delta t \, \Sigma^2$.

In analogy to the considerations above let $\varphi_t^{\mathrm{w},[1]}$ and $\varphi_t^{\mathrm{w},[2]}$ denote the exact flows of Eq. (20a) and (20b) given via Eq. (21) and (22), respectively. Then, for $x \in \mathbb{R}^6$, a

one-step integrator for Eq. (5) is given by

$$\psi_{\Delta t}^{\text{w}}(x) = \left(\varphi_{\Delta t}^{\text{w},[1]} \circ \varphi_{\Delta t}^{\text{w},[2]}\right)(x). \tag{23}$$

6.4 Order of Convergence and Strang Splitting

As the noise in Eq. (5) is additive, standard integrators such as the Euler–Maruyama method converge with mean-square order one. The same holds true for the splitting integrators constructed above.

Theorem 6.1 *Let $0 = t_0 < \cdots < t_N = T$ be an equidistant partition of $[0, T]$ with step-size Δt, and let $X^{\text{ou}}(t_i)$ and $X^{\text{w}}(t_i)$ denote the numerical solutions defined by Eq. (19) and (23) at time point t_i starting at initial value $(Q_0, P_0) \in \mathbb{R}^6$. Then the one-step methods defined in Eq. (19) and (23) are of mean-square order one, i.e. there exist constants $C_1, C_2 > 0$ such that for sufficiently small Δt the inequalities*

$$\mathbb{E}\left[\left\|X(t_i) - X^{\text{ou}}(t_i)\right\|_{\mathbb{R}^6}^2\right]^{1/2} \le C_1 \Delta t,$$

$$\mathbb{E}\left[\left\|X(t_i) - X^{\text{w}}(t_i)\right\|_{\mathbb{R}^6}^2\right]^{1/2} \le C_2 \Delta t,$$

hold for all time points t_i.

Proof The result can be proved in the same way as in [50], Lemma 2.1.. □

For deterministic ODE systems the convergence order of splitting methods can be increased by using composition based on fractional steps (see e.g. [49], Chapter II). We will illustrate this approach for the method based on the subsystems (20a) and (20b), the other method can be treated analogously. Using a Strang splitting we can compose the integrator

$$\psi_{\Delta t}^{\text{w}}(x) = \left(\varphi_{\Delta t/2}^{\text{w},[1]} \circ \varphi_{\Delta t}^{\text{w},[2]} \circ \varphi_{\Delta t/2}^{\text{w},[1]}\right)(x), \quad x \in \mathbb{R}^6. \tag{24}$$

For $\Sigma = \mathbb{O}_3$, Eq. (24) is a second-order method for the deterministic Eq. (2), however, the mean-square order of Eq. (24) is still one. To increase the mean-square order one has to include higher-order stochastic integrals to reproduce the interactions of the Subsystems (20a) and (20b) (see [50], Sect. 2, for details). Note that even without including the higher-order stochastic integrals the Strang splitting integrator given by Eq. (24) performs considerably better in our numerical simulations than the Lie–Trotter methods, thus we recommend to use this type of integrator. We have not yet studied the reason for this improved performance, but expect that the symmetry of the Strang splitting or the weak noise acting on the system may contribute.

We illustrate the mean-square convergence of our proposed methods in Fig. 7 and compare the Strang splitting Eq. (24) with the standard Euler–Maruyama method for the coupling parameters $C = 68$ and $C = 135$. As expected, both methods have mean-square order one, however, for $C = 135$ the mean-square error (MSE) of the splitting method is significantly smaller than the MSE of the Euler–Maruyama method. Obviously, one might use smaller step-sizes for the Euler–Maruyama method; however, this quickly becomes highly inefficient, e.g. for the JR-NMM for multiple populations
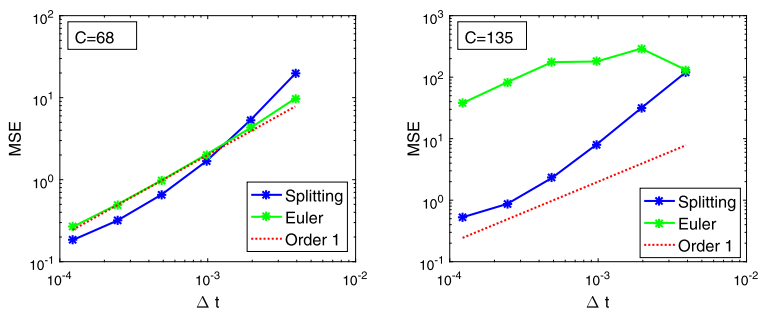
Fig. 7 Mean-square convergence of the splitting method

using Strang splitting with $\Delta t = 10^{-4}$

using Euler–Maruyama with $\Delta t = 10^{-4}$

using Strang splitting with $\Delta t = 10^{-3}$

using Euler–Maruyama with $\Delta t = 10^{-3}$

using Strang splitting with $\Delta t = 2 \cdot 10^{-3}$

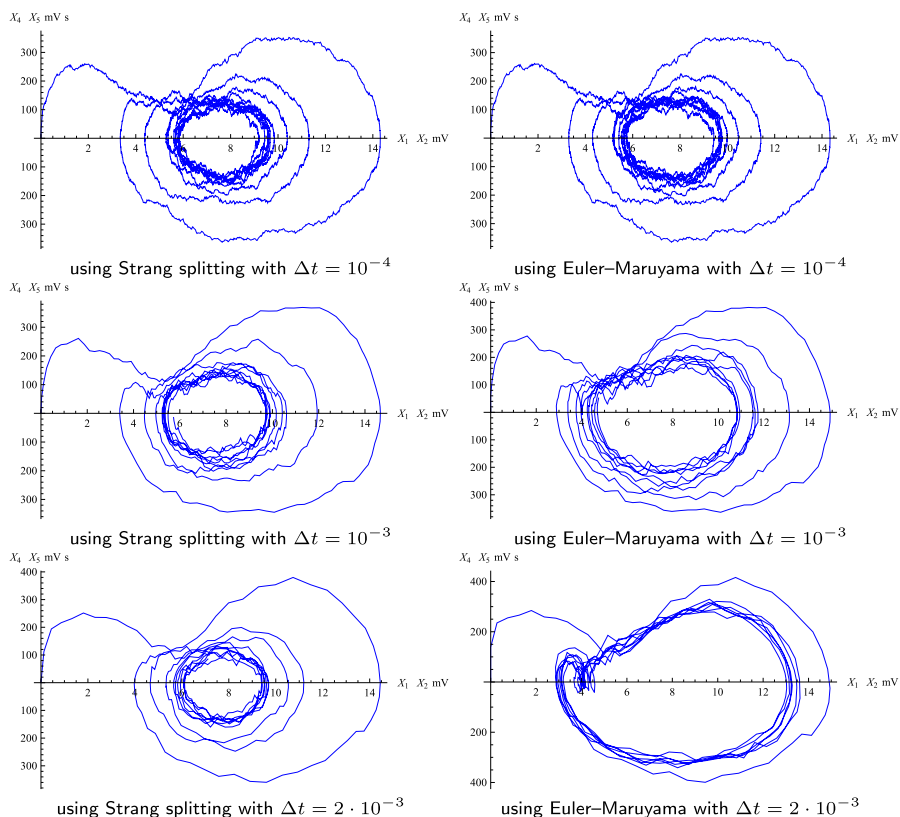using Euler–Maruyama with $\Delta t = 2 \cdot 10^{-3}$

Fig. 8 Phase portrait of one single path of Y

or when the Euler–Maruyama method is embedded in a continuous-time particle filter.

Figs. 8 and 9 also demonstrate the efficiency of the splitting scheme, as the *correct* (as not only observed from both methods with small step-sizes, but also based on our analysis of the method's properties) results can still be produced with much larger
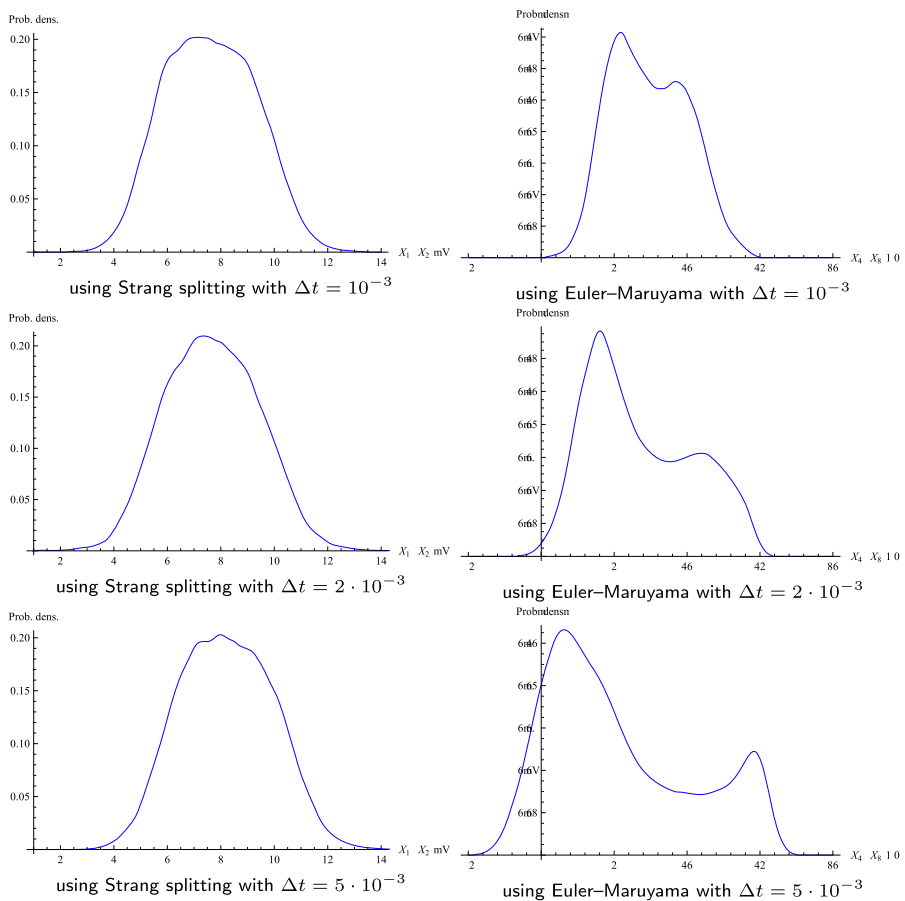
Fig. 9 Densities of the invariant measure of Y

step-sizes than those required for the Euler–Maruyama method. The other important feature of the proposed method is its reliability. Figure 8 shows several plots of the phase portrait of one single path of the output Y, with the splitting method and the Euler–Maruyama method and different step-sizes. It can be observed that the phase portrait obtained with the latter method changes markedly with increasing step-size. These phase portraits have been computed for the coupling parameter $C = 135$, initial value $X(0) = 0_6$, $\sigma_3 = \sigma_5 = 1$, $\sigma_4 = 200$ and $\Delta t \in \{10^{-4}, 10^{-3}, 2 \cdot 10^{-3}\}$. Figure 9 corresponds to the upper right plot in Fig. 6, which itself can be interpreted as a computational study of a phenomenological stochastic bifurcation for varying coupling parameter C. It shows the densities of the invariant measure of Y for $C = 135$, $\sigma_3 = \sigma_5 = 10$ and $\sigma_4 = 10^3$ and compares the Strang splitting scheme with the Euler–Maruyama method over the time-step-sizes $\Delta t \in \{10^{-3}, 2 \cdot 10^{-3}, 5 \cdot 10^{-3}\}$. The Euler–Maruyama method with moderately small step-sizes would report a change from a unimodal to a bimodal density for the parameter $C = 135$, whereas the correct value of C for this change to happen should be much larger.

6.5 Moment Bounds and Geometric Ergodicity

The following two lemmas represent the properties presented in Sect. 4 for the numerical approximation schemes defined by Eq. (19) and Eq. (23). Let $X^{\mathrm{ou}} = (Q^{\mathrm{ou}}, P^{\mathrm{ou}})$ and $X^{\mathrm{w}} = (Q^{\mathrm{w}}, P^{\mathrm{w}})$ denote the numerical solutions defined by Eq. (19) and Eq. (23), respectively. We start with proving analogous bounds to those in Theorem 4.2 for the expected value of Q^{ou} and Q^{w}. It is well known already in the deterministic setting that the Euler scheme does not preserve such properties, see [49], Chap. 1, in the stochastic case negative results for the Euler–Maruyama method for (simple) stochastic oscillators have been observed in [48]. Note that the following two lemmas also hold when commuting the compositions in Eq. (19) and Eq. (23).

Lemma 6.2 *Let $\mu_j : [0, T] \to \mathbb{R}^+$ for $j \in \{3, 4, 5\}$ be non-negative functions bounded by $\mu_{j,\max} \in \mathbb{R}^+$, respectively. Then for $i \in \{0, \ldots, N\}$, $\mathbb{E}[Q^{\mathrm{w}}(t_i)]$ (and also $\mathbb{E}[Q^{\mathrm{ou}}(t_i)]$) is bounded in each component by*

$$\vartheta(t_i)Q_0 + \kappa(t_i)P_0 \leq_\odot \mathbb{E}[Q^{\mathrm{w}}(t_i)] \leq_\odot \vartheta(t_i)Q_0 + \kappa(t_i)P_0 + \Gamma^{-2}C_G.$$

Proof We prove the result for the numerical method ψ^{w}, ψ^{ou} can be treated analogously. Bearing in mind the notation in Sect. 4, we obtain from Eq. (23) that

$$\mathbb{E}[X^{\mathrm{w}}(t_i)] = e^{M\Delta t}\mathbb{E}[X^{\mathrm{w}}(t_{i-1})] + e^{M\Delta t}\mathbb{E}[N(t_{i-1}, X^{\mathrm{w}}(t_{i-1}))]\Delta t$$

$$= e^{Mt_i}X(0) + \sum_{k=1}^{i} e^{Mt_k}\mathbb{E}[N(t_{i-k}, X^{\mathrm{w}}(t_{i-k}))]\Delta t,$$

and in particular its Q-component reads

$$\mathbb{E}[Q^{\mathrm{w}}(t_i)] = \vartheta(t_i)Q_0 + \kappa(t_i)P_0 + \sum_{k=1}^{i} \kappa(t_k)\mathbb{E}[G(t_{i-k}, X^{\mathrm{w}}(t_{i-k}))]\Delta t. \qquad (25)$$

From the proof of Theorem 4.2 we obtain $0_3 \leq_\odot \mathbb{E}[G(t_{i-k}, X^{\mathrm{w}}(t_{i-k}))] \leq_\odot C_G$. Obviously, the lower bound of $\mathbb{E}[Q^{\mathrm{w}}(t_i)]$ is fulfilled for any time step-size Δt. To prove the upper bound it remains to show that

$$\sum_{k=1}^{i} \kappa(t_k)\Delta t \leq_\odot \Gamma^{-2}.$$

From the decomposition

$$\sum_{k=1}^{i} \kappa(t_k)\Delta t = \left(\mathbb{I}_3 - e^{-\Gamma\Delta t}\right)^{-1}\left(\mathbb{I}_3 - e^{-\Gamma t_i}\right)\kappa(\Delta t)\Delta t - \kappa(\Delta t)\kappa(t_i)$$

$$+ e^{-\Gamma\Delta t}\sum_{k=1}^{i} \kappa(t_k)\Delta t$$

we derive the formula

$$\sum_{k=1}^{i} \kappa(t_k)\Delta t = \big(\mathbb{I}_3 - e^{-\Gamma\Delta t}\big)^{-2}\big(\mathbb{I}_3 - e^{-\Gamma t_i}\big)\kappa(\Delta t)\Delta t$$

$$- \big(\mathbb{I}_3 - e^{-\Gamma\Delta t}\big)^{-1}\kappa(\Delta t)\kappa(t_i),$$

which is in each component smaller than Γ^{-2} for any time step-size Δt. $\qquad\square$

Remark 5 Another viewpoint of Eq. (23) is to apply the rectangle method (using the left boundary point of the integral) in order to approximate the convolution-based formula Eq. (8) in the form

$$Q(t + \Delta t) = \vartheta(\Delta t)Q(t) + \kappa(\Delta t)P(t)$$

$$+ \int_t^{t+\Delta t} \kappa(t + \Delta t - s)G\big(s, Q(s)\big)\,ds$$

$$+ \int_t^{t+\Delta t} \kappa(t + \Delta t - s)\Sigma(s)\,dW(s).$$

Moreover, Eq. (25) permits better insight into the distinction of the numerical schemes: The sum in Eq. (25) corresponds to the rectangle method in order to approximate the convolution integral $\mathbb{E}[v(t)]$ defined in the proof of Theorem 4.2, where the right boundary point is used in each approximation interval. Analogously, when commuting the composition in Eq. (23) one obtains the rectangle method evaluating the left boundary points. In the case of the Strang splitting scheme given by Eq. (24), the function κ is evaluated at the midpoints $(t_k + t_{k+1})/2$.

Remark 6 It can be shown analogously that the second moment $\mathbb{E}[(Q^{\mathrm{w}}(t_i))^{2\odot}]$ (and also $\mathbb{E}[(Q^{\mathrm{ou}}(t_i))^{2\odot}]$) is bounded by

$$\mathbb{E}\big[\big(Q^{\mathrm{w}}(t_i)\big)^{2\odot}\big] \leq_\odot u^{2\odot}(t_i) + 2u(t_i) \odot \mathbf{1}^+_\odot\big(u(t_i)\big) \odot \Gamma^{-2}C_G$$

$$+ \left(\Gamma^{-2}C_G + \frac{1}{2}\Gamma^{-3/2}\Sigma 1_3\right)^{2\odot}.$$

The last point we discuss in this article is the geometric ergodicity of the discrete Markov processes X^{ou} and X^{w} defined by Eq. (19) and (23). In analogy to Sect. 5 we assume that Assumption (10) holds. Due to the global Lipschitz continuity of the coefficients of Eq. (5), one would expect that standard numerical methods such as the Euler–Maruyama method are again geometrically ergodic for small enough stepsizes Δt (see [34], Theorem 7.3). The advantage of our proposed splitting integrators is that we can directly prove a discrete analog of Lemma 5.1, i.e. a discrete Lyapunov condition for the same Lyapunov function under very mild restrictions on Δt. We formulate the result for the Wiener integrator, the Ornstein–Uhlenbeck integrator can be treated analogously.

Lemma 6.3 *Let $0 < t_0 \cdots < t_N = T$ be an equidistant partition of $[0, T]$ with step-size $\Delta t < 1/(2\|\Gamma\|_{L^\infty})$ and let X^{w} denote the numerical solutions defined by Eq. (23). Then the functional $V(X) := V_1(Q, P)$ defined in Lemma 5.1 is a Lyapunov function for X^{w}, i.e. there exist constants $\alpha \in (0, 1)$ and $\beta \geq 0$ such that*

$$\mathbb{E}\big[V\big(X^{\mathrm{w}}(t_{i+1})\big)|\mathcal{F}_{t_i}\big] \leq \alpha V\big(X^{\mathrm{w}}(t_i)\big) + \beta.$$

Proof For the sake of simplicity we set $a = b$, which implies $e^{-\Gamma t}x = e^{-at}x$ for any $x \in \mathbb{R}^3$. Furthermore, we denote $Q := Q(t_i)$, $P := P(t_i)$ to shorten notation. The one-step approximations $Q(t_{i+1})$ and $P(t_{i+1})$ can be written as

$$\Gamma Q(t_{i+1}) = \Gamma e^{-a\Delta t}(\mathbb{I}_3 + \Gamma\Delta t)Q + \Gamma e^{-a\Delta t}\Delta t\big(P + \Delta t G(Q) + \Sigma\xi\big),$$

$$P(t_{i+1}) = e^{-a\Delta t}(\Gamma Q + P) + e^{-a\Delta t}\Delta t G(Q) + e^{-a\Delta t}\Sigma\xi - \Gamma Q(t_{i+1}),$$

where $\xi \sim \mathcal{N}(0_3, \mathbb{I}_3)$ is a three-dimensional Gaussian vector independent of $\mathcal{F}_{t_i}$. By elementary calculations and application of Young's inequality we obtain

$$\frac{1}{2}\big\|P(t_{i+1}) + \Gamma Q(t_{i+1})\big\|_{\mathbb{R}^3}^2$$

$$\leq \frac{1}{2}e^{-2a\Delta t}\|\Gamma Q + P\|_{\mathbb{R}^3}^2 + \frac{\epsilon}{2}e^{-4a\Delta t}\Delta t^2\|\Gamma Q + P\|_{\mathbb{R}^3}^2$$

$$+ \frac{1}{2\epsilon}\big\|G(Q)\big\|_{\mathbb{R}^3}^2 + \frac{1}{2}\big\|e^{-a\Delta t}\big(\Delta t G(Q) + \Sigma\xi\big)\big\|_{\mathbb{R}^3}^2$$

$$+ e^{-2a\Delta t}\langle\Gamma Q + P, \Sigma\xi\rangle_{\mathbb{R}^3},$$

where $\epsilon > 0$ is a parameter which can be freely chosen. Thus, one can find $C_1 > 0$ sufficiently large such that

$$\frac{1}{2}\mathbb{E}\big[\big\|P(t_{i+1}) + \Gamma Q(t_{i+1})\big\|_{\mathbb{R}^3}^2|\mathcal{F}_{t_i}\big]$$

$$\leq \frac{1}{2}e^{-2a\Delta t}\|\Gamma Q + P\|_{\mathbb{R}^3}^2 + \frac{1}{2\epsilon}\big\|G(Q)\big\|_{\mathbb{R}^3}^2$$

$$+ \frac{\epsilon}{2}e^{-4a\Delta t}\Delta t^2\|\Gamma Q + P\|_{\mathbb{R}^3}^2 + C_1.$$

In the same spirit we can find $C_2 > 0$ such that

$$\mathbb{E}\big[\big\|\Gamma Q(t_{i+1})\big\|_{\mathbb{R}^3}^2|\mathcal{F}_{t_i}\big]$$

$$\leq e^{-2a\Delta t}\|\Gamma Q\|_{\mathbb{R}^3}^2 + a^2 e^{-2a\Delta t}\Delta t^2\|\Gamma Q + P\|_{\mathbb{R}^3}^2$$

$$+ \tilde{\epsilon}e^{-2a\Delta t}\Delta t\|\Gamma Q\|_{\mathbb{R}^3}^2 + \frac{1}{\tilde{\epsilon}}a^2 e^{-2a\Delta t}\Delta t\|\Gamma Q + p\|_{\mathbb{R}^3}^2 + \frac{1}{\hat{\epsilon}}\big\|G(Q)\big\|_{\mathbb{R}^3}^2$$

$$+ \hat{\epsilon}a^2 e^{-4a\Delta t}\Delta t^4\big(\|\Gamma Q\|_{\mathbb{R}^3}^2 + a^2\Delta t^2\|\Gamma Q + P\|_{\mathbb{R}^3}^2\big) + C_2$$

with free parameters $\tilde{\epsilon}, \hat{\epsilon} > 0$. Combining the bounds above we can find a suitable $\alpha \in (0, 1)$ if for any given Δt there exists a choice $\tilde{\epsilon}^*$ such that

$$e^{-2a\Delta t}\left(1 + \epsilon\Delta t^2 + 2a^2\Delta t^2 + \frac{2}{\tilde{\epsilon}^*}a^2\Delta t + \hat{\epsilon}a^4 e^{-2a\Delta t}\Delta t^6\right) < 1, \qquad (26a)$$

$$e^{-2a\Delta t}\left(1 + \tilde{\epsilon}^*\Delta t + \hat{\epsilon}a^2 e^{-2a\Delta t}\Delta t^2\right) < 1. \qquad (26b)$$

Note that ϵ and $\hat{\epsilon}$ can be chosen arbitrarily small, therefore the corresponding terms can be neglected. Now let $\Delta t < 1/(2a)$, then Eq. (26a) and (26b) are fulfilled for

$$\frac{2a^2 - 4a^3\Delta t}{2a - 4a^2\Delta t + 4a^3\Delta t} < \tilde{\epsilon}^* < \frac{2a}{1 - 2a\Delta t},$$

which implies the result. $\square$

In analogy to Sect. 5, geometric ergodicity of the (discrete) Markov processes X^w and X^{ou} can be established by proving smoothness of the transition probabilities and irreducibility of the processes. Both properties can be proven in exactly the same way as in [34], Corollary 7.4, thus we only sketch the proof for X^w:

(i) Smoothness of the transition probability densities: Due to Assumption (10) the transition probability of two (or more) consecutive steps $\psi_{\Delta t}^w \circ \psi_{\Delta t}^w$ of our integrator has a smooth density.
(ii) Irreducibility: As in the time-continuous case in Sect. 5 we have to establish a reachability condition, i.e. the numerical method starting at $x \in \mathbb{R}^6$ can reach any $y \in \mathbb{R}^6$ after a fixed number of steps. For our splitting method, two consecutive steps are sufficient to reach any point y by suitably choosing the vectors $\xi(\Delta t)$ such that

$$y = \left(\psi_{\Delta t}^w \circ \psi_{\Delta t}^w\right)(x). \qquad (27)$$

In fact, Eq. (27) is a six-dimensional system of equations with six degrees of freedom (three Gaussian random variables for each step $\psi_{\Delta t}^w$) which can always be solved under Assumption 10.

To summarise, the numerical approximations X^{ou} and X^w are geometrically ergodic with respect to a unique invariant measure $\eta_{\Delta t}^{ou}$ and $\eta_{\Delta t}^w$ under mild restrictions on the time-step-size Δt. Furthermore, as X^{ou} and X^w converge towards X in the mean-square sense, $\eta_{\Delta t}^{ou}$ and $\eta_{\Delta t}^w$ are convergent approximations of the original invariant measure η (see [53], Theorem 3.3, for details). Thus, our numerical approximations of the marginal densities in Sect. 5 (see Fig. 6) are supported by the theory.

7 Summary and Conclusions

We proposed a version of the original JR-NMM incorporating random input, as a stochastic Hamiltonian system with nonlinear displacement, and discussed a range

of properties based on results available in the framework of stochastic analysis, in particular properties such as moment bounds and the existence of invariant measures. The latter represent a step towards analysing the dynamical properties of a stochastic formulation of the JR-NMM. Furthermore, we presented an efficient numerical scheme based on a splitting approach which preserves the qualitative behaviour of the solution of the system. We have also discussed the advantages of applying such a scheme designed according to the obtained features of the stochastic JR-NMM for future computational studies in contrast to applying other numerical methods such as the Euler–Maruyama scheme. By a suitable introduction of noise our results can be generalised to both the extension of the JR-NMM to multiple populations [37, 54–57] and the extension to multiple areas, e.g. the 2-column model in [13] or the multi-area neural mass model in [56].

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed equally.

Appendix

For completeness we give auxiliary results establishing the necessary smoothness properties of the transition probability densities as well as the irreducibility of the solution process X of Eq. (5). The proofs are in principle the same as in [34]. Although the Langevin equation treated there did not involve a nonlinear displacement such as the sigmoid function, the smoothness and boundedness of the sigmoid function allows us to use the same arguments. The following lemma establishes the smoothness properties of the transition probability densities by proving the hypoellipticity of the generator $\mathcal{L}$. For further details we refer to [34, 36].

Lemma A.1 *The generator $\mathcal{L}$, its formal adjoint $\mathcal{L}^\dagger$, $\partial/\partial t - \mathcal{L}$ and $\partial/\partial t - \mathcal{L}^\dagger$ are hypoelliptic.*

Proof In order to apply Hörmander's theorem we have to show that the Lie algebras based on the operators in question have maximal rank 6 resp. 7 (see [36] and for a more general discussion [58], Chap. 2). We show this for the generator $\mathcal{L}$, the other operators can be treated analogously. Using the notation from Sect. 4.1, we consider the following six-dimensional vector fields:

$$f_0(X) = MX + N(X), \qquad g_i = \begin{pmatrix} 0_3 \\ \Sigma_{(:,i)} \end{pmatrix} \quad \text{for } i \in \{1, 2, 3\},$$

where $\Sigma_{(:,i)}$ denotes the ith column of Σ. For six-dimensional vector fields $U, V : \mathbb{R}^6 \to \mathbb{R}^6$ with Jacobians $\partial U, \partial V$ we define the Lie bracket $[U, V] := (\partial U)V - (\partial V)U$. According to [34] it suffices to show that

$$\dim\big(\text{span}\{g_1, g_2, g_3, [f_0, g_1], [f_0, g_2], [f_0, g_3]\}\big) = 6.$$

As

$$(\partial f_0)g_i = \begin{pmatrix} \Sigma_{(:,i)} \\ -2\Gamma\Sigma_{(:,i)} \end{pmatrix} \quad \text{and} \quad (\partial g_i)f_0 = 0_6,$$

the statement directly follows from Assumption (10) and $\sigma_i > 0$ for $i \in \{3, 4, 5\}$. Note that the nonlinear term N does not play any role in the computation as its Jacobian is only non-zero at the derivatives corresponding to the Q-component, which are multiplied with the first three components of the vectors g_i. However, these three components are always zero. $\qquad\square$

Irreducibility can be established via a control type argument. For this purpose, let $\mathbb{P}_t(\mathcal{A}, x)$ denote the transition probabilities of the Markov process X.

Lemma A.2 *The solution process $X = (Q, P)$ of Eq. (5) is irreducible, i.e. for arbitrary open sets $\mathcal{A} \in \mathbb{R}^6$, initial values $x \in \mathbb{R}^6$ and $t \in [0, T]$*

$$\mathbb{P}_t(\mathcal{A}, x) > 0. \tag{28}$$

Proof It suffices to consider open neighbourhoods $\mathcal{A} = B_\epsilon(\mathbf{X}_\tau)$ of arbitrary terminal values $\mathbf{X}_\tau \in \mathbb{R}^6$ for some time point $\tau \in [0, T]$. Now, Property (28) can be established in the following way: Let $\widehat{X}(t)$ be the solution of the controlled JR-NMM

$$d\widehat{X}(t) = d\begin{pmatrix} \widehat{Q}(t) \\ \widehat{P}(t) \end{pmatrix} = f_0(\widehat{X})\,dt + \begin{pmatrix} \mathbb{0}_3 \\ \Sigma \end{pmatrix} d\widehat{W}(t), \tag{29}$$

where $\widehat{W} : [0, T] \to \mathbb{R}^3$ is a continuously differentiable function. If

$$\mathbb{P}\Big(\sup_{0 \le \tau \le T} \big\| X(\tau) - \widehat{X}(\tau) \big\| < \epsilon \Big) > 0, \tag{30}$$

then Property (28) holds as long as $\widehat{X}(\tau) = \mathbf{X}_\tau$. Now, let $\widehat{Q}(t)$ be a smooth curve such that

$$\begin{pmatrix} \widehat{Q}(0) \\ \frac{d\widehat{Q}}{dt}(0) \end{pmatrix} = x \quad \text{and} \quad \begin{pmatrix} \widehat{Q}(\tau) \\ \frac{d\widehat{Q}}{dt}(\tau) \end{pmatrix} = \mathbf{X}_\tau.$$

The control $\widehat{W}(t)$ is given via the second-order differential equation

$$\frac{d\widehat{W}}{dt} = \Sigma^{-1}\Big(\frac{d^2\widehat{Q}}{dt^2} + 2\Gamma\frac{d\widehat{Q}}{dt} + \Gamma^2\widehat{Q} + G(\widehat{Q}) \Big),$$

which has a unique solution as G is globally Lipschitz. Consequently, the point $\mathbf{X}_\tau$ is reachable for solutions of the controlled System (29). Condition (30) (and therefore irreducibility) is now an immediate consequence of the Stroock–Varadhan Support Theorem (see [59]).   □

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Wilson HR, Cowan JD. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J. 1972;12(1):1–24.
2. Wilson HR, Cowan JD. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik. 1973;13(2):55–80.
3. Lopes Da Silva FH, Hoeks A, Smits H, Zetterberg LH. Model of brain rhythmic activity. Kybernetik. 1974;15(1):27–37.
4. Freeman WJ. Mass action in the nervous system. New York: Academic Press; 1975.
5. Lopes Da Silva FH, Van Rotterdam A, Barts P, Van Heusden E, Burr W. Models of neuronal populations: the basic mechanisms of rhythmicity. Prog Brain Res. 1976;45:281–308.
6. Freeman W. Models of the dynamics of neural populations. Electroencephalogr Clin Neurophysiol, Suppl. 1977;34:9–18.
7. Zetterberg LH, Kristiansson L, Mossberg K. Performance of a model for a local neuron population. Biol Cybern. 1978;31(1):15–26.
8. Sotero RC, Trujillo-Barreto NJ, Iturria-Medina Y, Carbonell FF, Jimenez JC. Realistically coupled neural mass models can generate eeg rhythms. Neural Comput. 2007;19(2):478–512.
9. Wendling F, Benquet P, Bartolomei F, Jirsa V. Computational models of epileptiform activity. J Neurosci Methods. 2016;260:233–51.
10. Sanz-Leon P, Knock SA, Spiegler A, Jirsa VK. Mathematical framework for large-scale brain network modeling in the Virtual Brain. NeuroImage. 2015;111:385–430.
11. Coombes S, Byrne A. Next generation neural mass models. arXiv:1607.06251 (2016).
12. Faugeras O, Touboul J, Cessac B. A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. Front Comput Neurosci. 2009;3(1):1–28.
13. Jansen BH, Rit VG. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern. 1995;73(4):357–66.
14. Jansen BH, Zouridakis G, Brandt ME. A neurophysiologically-based mathematical model of flash visual evoked potentials. Biol Cybern. 1993;68(3):275–83.
15. van Rotterdam A, Lopes Da Silva FH, Van den Ende J, Viergever MA, Hermans AJ. A model of the spatial-temporal characteristics of the alpha rhythm. Bull Math Biol. 1982;44(2):283–305.
16. Leimkuhler B, Matthews C. Molecular dynamics. Interdisciplinary applied mathematics, vol. 39. Cham: Springer; 2015. p. 443. With deterministic and stochastic numerical methods.
17. Coffey WT, Kalmykov YT, Waldron JT. The Langevin equation. With applications to stochastic problems in physics, chemistry and electrical engineering. 3rd ed. Hackensack: World Scientific; 2012. p. 827.
18. Grimbert F, Faugeras O. Bifurcation analysis of Jansen's neural mass model. Neural Comput. 2006;18(12):3052–68.
19. Touboul J, Wendling F, Chauvel P, Faugeras O. Neural mass activity, bifurcations, and epilepsy. Neural Comput. 2011;23(12):3232–86.
20. Knösche T. Jansen-Rit model (2013).
21. Kandel ER, Schwartz JH, Jessell TM. Principles of neural science. vol. 4. New York: McGraw-Hill; 2000.
22. Dayan P, Abbott LF. Theoretical neuroscience. vol. 806. Cambridge: MIT Press; 2001.

23. Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal dynamics: from single neurons to networks and models of cognition. Cambridge: Cambridge University Press; 2014.
24. Spiegler A, Kiebel SJ, Atay FM, Knösche TR. Bifurcation analysis of neural mass models: impact of extrinsic inputs and dendritic time constants. NeuroImage. 2010;52(3):1041–58.
25. Spiegler A, Knösche TR, Schwab K, Haueisen J, Atay FM. Modeling brain resonance phenomena using a neural mass model. PLoS Comput Biol. 2011;7(12):1002298.
26. Khasminskii R. Stochastic stability of differential equations. Stochastic modelling and applied probability, vol. 66. Heidelberg: Springer; 2012.
27. Arnold L. Random dynamical systems. Berlin Heidelberg: Springer; 1998.
28. Asai Y, Kloeden PE. Multi-step methods for random odes driven by Itô diffusions. J Comput Appl Math. 2016;294:210–24.
29. Arnold L. Stochastic differential equations: theory and applications. New York: Wiley; 1974. p. 228. Translated from the German.
30. Yaglom AM. An introduction to the theory of stationary random functions. Englewood Cliffs: Prentice Hall International; 1962.
31. Milstein GN, Tretyakov MV. Stochastic numerics for mathematical physics. Scientific computation. Berlin: Springer; 2004.
32. Mao X. Stochastic differential equations and applications. Amsterdam: Elsevier; 2007.
33. Klebaner FC. Introduction to stochastic calculus with applications. vol. 57. Singapore: World Scientific; 2005.
34. Mattingly JC, Stuart AM, Higham DJ. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. Stoch Process Appl. 2002;101(2):185–232. doi:10.1016/S0304-4149(02)00150-3.
35. Leimkuhler B, Matthews C, Stoltz G. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. IMA J Numer Anal. 2016;36(1):13–79.
36. Rey-Bellet L. Ergodic properties of Markov processes. In: Open quantum systems. II. Lecture notes in math. vol. 1881. Berlin: Springer; 2006. p. 1–39.
37. David O, Friston KJ. A neural mass model for meg/eeg: coupling and neuronal dynamics. NeuroImage. 2003;20(3):1743–55.
38. Freestone DR, Kuhlmann L, Chong M, Nesic D, Grayden DB, Aram P, Postoyan R, Cook MJ. Patient-specific neural mass modelling: stochastic and deterministic methods. In: Recent advances in predicting and preventing epileptic seizures. 2013. p. 63–82.
39. Horn RA, Johnson CR. Matrix analysis. 2nd ed. Cambridge: Cambridge University Press; 2013. p. 643.
40. Lütkepohl H. Handbook of matrices. Chichester: Wiley; 1996. p. 304.
41. Da Prato G, Zabczyk J. Ergodicity for infinite-dimensional systems. London mathematical society lecture note series. vol. 229. Cambridge: Cambridge University Press; 1996. p. 339.
42. Sanz-Serna JM, Stuart AM. Ergodicity of dissipative differential equations subject to random impulses. J Differ Equ. 1999;155(2):262–84. doi:10.1006/jdeq.1998.3594.
43. Baňas L, Brzeźniak Z, Neklyudov M, Prohl A. Stochastic ferromagnetism: analysis and numerics. vol. 58. Berlin: de Gruyter; 2014.
44. Kohatsu-Higa A, Pettersson R. Variance reduction methods for simulation of densities on Wiener space. SIAM J Numer Anal. 2002;40(2):431–50. doi:10.1137/S0036142901385507.
45. Pons O. Functional estimation for density, regression models and processes. Hackensack: World Scientific; 2011. p. 199. doi:10.1142/9789814343749.
46. Tateno T, Pakdaman K. Random dynamics of the Morris–Lecar neural model. Chaos, Interdiscip J Nonlinear Sci. 2004;14(3):511–30.
47. Horsthemke W, Lefever R. Noise-induced transitions in physics, chemistry, and biology. In: Noise-induced transitions: theory and applications in physics, chemistry, and biology. 1984. p. 164–200.
48. Strømmen Melbø AH, Higham DJ. Numerical simulation of a linear stochastic oscillator with additive noise. Appl Numer Math. 2004;51(1):89–99.
49. Hairer E, Lubich C, Wanner G. Geometric numerical integration. Heidelberg: Springer; 2006.
50. Milstein GN, Tretyakov MV. Quasi-symplectic methods for Langevin-type equations. IMA J Numer Anal. 2003;23(4):593–626.
51. Bou-Rabee N, Owhadi H. Long-run accuracy of variational integrators in the stochastic context. SIAM J Numer Anal. 2010;48(1):278–97.
52. Misawa T. A Lie algebraic approach to numerical integration of stochastic differential equations. SIAM J Sci Comput. 2001;23(3):866–90. doi:10.1137/S106482750037024X.

53. Shardlow T, Stuart AM. A perturbation theory for ergodic Markov chains and application to numerical approximations. SIAM J Numer Anal. 2000;37(4):1120–37. doi:10.1137/S0036142998337235.

54. Wendling F, Bellanger J, Bartolomei F, Chauvel P. Relevance of nonlinear lumped-parameter models in the analysis of depth-eeg epileptic signals. Biol Cybern. 2000;83(4):367–78.

55. Wendling F, Bartolomei F, Bellanger J, Chauvel P. Epileptic fast activity can be explained by a model of impaired gabaergic dendritic inhibition. Eur J Neurosci. 2002;15(9):1499–508.

56. Babajani-Feremi A, Soltanian-Zadeh H. Multi-area neural mass modeling of EEG and MEG signals. NeuroImage. 2010;52(3):793–811.

57. Wang P, Knösche TR. A realistic neural mass model of the cortex with laminar-specific connections and synaptic plasticity - evaluation with auditory habituation. PLoS ONE. 2013;8(10):77876.

58. Nualart D. The Malliavin calculus and related topics. 2nd ed. Probability and its applications (New York). Berlin: Springer; 2006. p. 382.

59. Stroock DW. Lectures on topics in stochastic differential equations. Tata institute of fundamental research lectures on mathematics and physics. vol. 68. Bombay: Tata; 1982. p. 93. By Springer-Verlag, Berlin-New York. With notes by Satyajit Karmakar.