Simple Nonlinear Models with Rigorous Extreme Events and Heavy Tails

Andrew J Majda and Xin T Tong

May 11, 2018

Abstract

Extreme events and the heavy tail distributions driven by them are ubiquitous in various scientific, engineering and financial research. They are typically associated with stochastic instability caused by hidden unresolved processes. Previous studies have shown that such instability can be modeled by a stochastic damping in conditional Gaussian models. However, these results are mostly obtained through numerical experiments, while a rigorous understanding of the underlying mechanism is sorely lacking. This paper contributes to this issue by establishing a theoretical framework, in which the tail density of conditional Gaussian models can be rigorously determined. In rough words, we show that if the stochastic damping takes negative values, the tail is polynomial; if the stochastic damping is nonnegative but takes value zero, the tail is between exponential and Gaussian. The proof is established by constructing a novel, product-type Lyapunov function, where a Feynman-Kac formula is applied. The same framework also leads to a non-asymptotic large deviation bound for long-time averaging processes.

1 Introduction

With dramatic global climate change in recent years, extreme climate events, along with their destructive power, are observed more often than ever. Severe heatwaves reduce crop harvest, increase forest fire risk and sometimes lead to human casualties. Heavy downpours, as another extreme, can flood large areas and cause significant economic losses [16]. Extreme events are also of great interest in engineering and financial research, because of the underlying risk. Rogue waves, seen as walls of waters of 10 meters high, can easily sink unprepared ships [20]. Credit default of one bank can lead to world-wide financial recession [19]. The capability to model, measure and predict these extreme events has never been so important [26, 9, 42].

Mathematically, extreme events can be viewed as strong anomalies seen in the time series of certain observables. Collectively, they produce an exponential or even polynomial heavy tail in the observable's histogram. They often appear in complex nonlinear models that have stochastic instability. This instability typically comes as a combined effect of many hidden or unresolved processes [9]. Examples of these hidden processes include cloud formation, precipitation and refined scale turbulence [31, 35, 27, 28]. For these processes, only limited direct observations are available. Accurate physical models of them are lacking, or require expensive computation. A better modeling strategy is viewing them as stochastic processes, of which the parameters can be tuned to fit data statistically [39].

The stochastic instability discussed above can be described by the following simple nonlinear model:

$$\begin{aligned} dX_t &= -b(u_t)X_t dt + \sigma_x dW_t, \\ du_t &= h(u_t)dt + dB_t. \end{aligned} \tag{1.1}$$

Here $X_t \in \mathbb{R}^{d_X}$ represents certain observables of a physical model, while its dynamics is affected by a hidden process $u_t \in \mathbb{R}^{d_u}$. For simplicity, we assume throughout the paper that u_t is ergodic, and π is its equilibrium distribution.

In (1.1), the Stochastic instability is represented by the damping rate $b(u_t) \in \mathbb{R}$. The dynamics of X_t is unstable if $b(u_t)$ is zero or negative in an interval of time, strong large spikes will appear in the trajectory of X_t as a consequence, which we can interpret as extreme events. Note that this is a random event that takes place intermittently, since it is triggered by the random realization of the process u_t.

One important feature of model (1.1) is that the dynamics of X_t is linear if u_t is fixed. We can generalize the formulation in (1.1) and maintain this feature. Consider

$$\begin{aligned} dX_t &= -B(u_t)X_t + \Sigma_X dW_t, \\ du_t &= h(u_t)dt + dB_t, \end{aligned} \tag{1.2}$$

where $B(u)$ is a matrix valued function. This is known as a conditional Gaussian system [30]. This formulation can be found in many nonlinear models, such as stochastic parameterization Kalman filter (SPEKF), Lagrangian floater, low order Madden Julian Oscillation model, and turbulent tracers [13, 14, 11, 10, 12, 37]. The conditional Gaussian structure can be exploited for efficient computations. If we apply the vanilla Monte Carlo method to estimate the density of X_t, the necessary sample size is e^{d_X}, which is prohibitive when d_X is large. But knowing that X_t is conditionally Gaussian, it suffices to compute the conditional mean and covariance, of which the computational cost only scale cubically with d_X. This feature can be exploited for high dimensional prediction and data assimilation [11, 13, 14, 10, 12].

Conditional Gaussian model is known to be a good tool for studying extreme events and heavy tail phenomena. In the SPEKF model, an observable x_t is driven by

$$dx_t = (-\gamma(t) + i\omega)x_t dt + f(t)dt + \sigma_x dW_t,$$

where $\gamma(t)$ and $f(t)$ are independent Ornstein-Uhlenbeck (OU) processes modeling the unobservable instability and forcing. This simple three-dimensional nonlinear model was first introduced in [21] for filtering multiscale turbulent signals with hidden instabilities, and later used for filtering, prediction, parameter estimation in the presence of model error [22, 5, 34, 9]. Another example is the turbulent passive tracer. Passive tracers are substance transported by a turbulence. They can reveal many important properties of the underlying turbulence and have important environmental impacts [40, 32, 29]. Mathematically, given a turbulence velocity field V, the passive tracer density $T(x,t)$ follows an advection-diffusion equation:

$$\partial_t T + V \cdot \nabla T = -\gamma_T T + \kappa \Delta T.$$

This dynamic is linear conditioned on V, so T can be interpreted as X_t in (1.2). Numerical evidence indicates that even with a simple zonal sweep V, the passive tracers have extreme events and an exponential-like histogram. This is in accordance with the laboratory observations such as the classical Rayleigh-Bernard convection [6, 23] and readings from the atmosphere [40]. An earlier result of the authors [37] has rigorously explained this phenomenon using a delicate phase resonance.

Despite the extensive success in using conditional Gaussian models for extreme event research, most findings are justified by numerical experiments. The only rigorous result [37] focuses only on a specific passive tracer model. There lacks a rigorous extreme event framework that applies to the general model (1.1) or (1.2). This can be problematic, since extreme events are typically rare and can be very difficult to simulate or observe. Experimental data, therefore, can be inaccurate, especially if the model contains many variables or has complicated nonlinearity.

This paper intends to close this gap by giving concrete criteria that lead to provable heavy tails of X_t. As a result, when a model of type (1.1) or (1.2) is available, we know apriori the tail density of $\|X_t\|$. This will be extremely helpful to the stochastic modeling of extreme events, as it turns a nonparametric problem parametric. Furthermore, we can obtain lower and upper bounds of the shape parameters of these distributions, and in some simple cases, these bounds are sharp. From the reverse perspective, when we only have data of X_t and intend to fit it with a model, the criteria in this paper can be used as guidelines for the choice of the model.

1.1 Main results in a simplified setting

In order to give a quick idea of our main result, consider an unforced SPEKF model with general damping [9]:

$$\begin{aligned} dX_t &= -b(u_t)X_t + dW_t, \\ du_t &= -\gamma u_t dt + dB_t. \end{aligned} \tag{1.3}$$

We assume u_0 follows the equilibrium distribution $\pi = \mathcal{N}(0, \frac{1}{2\gamma})$ and $X_0 = 0$. Our result indicates that the tail of X_t is controlled by some simple properties of the damping function b:

Theorem 1.1. *In model* (1.3)*, suppose the damping on average is positive,* $\mathbb{E}b(u_t) > 0$*, and* b *is Lipschitz, then* $(X_t, u_t) \in \mathbb{R}^{1+1}$ *has an unique equilibrium distribution, under which the tail of* $|X_t|$ *is*

i) polynomial, if b *can take negative values.*

ii) exponential, if $b(u)$ *is nonnegative and takes value* 0 *in an interval.*

iii) between exponential and Gaussian, if b *is nonnegative, and takes value* 0 *at a point.*

iv) Gaussian, if b *is bounded from below.*

The difference between case ii) and iii) is quite subtle but important: in case ii), $b(u_t)$ takes value zero with positive amount of time, while in case iii), $b(u_t)$ can be close to zero,

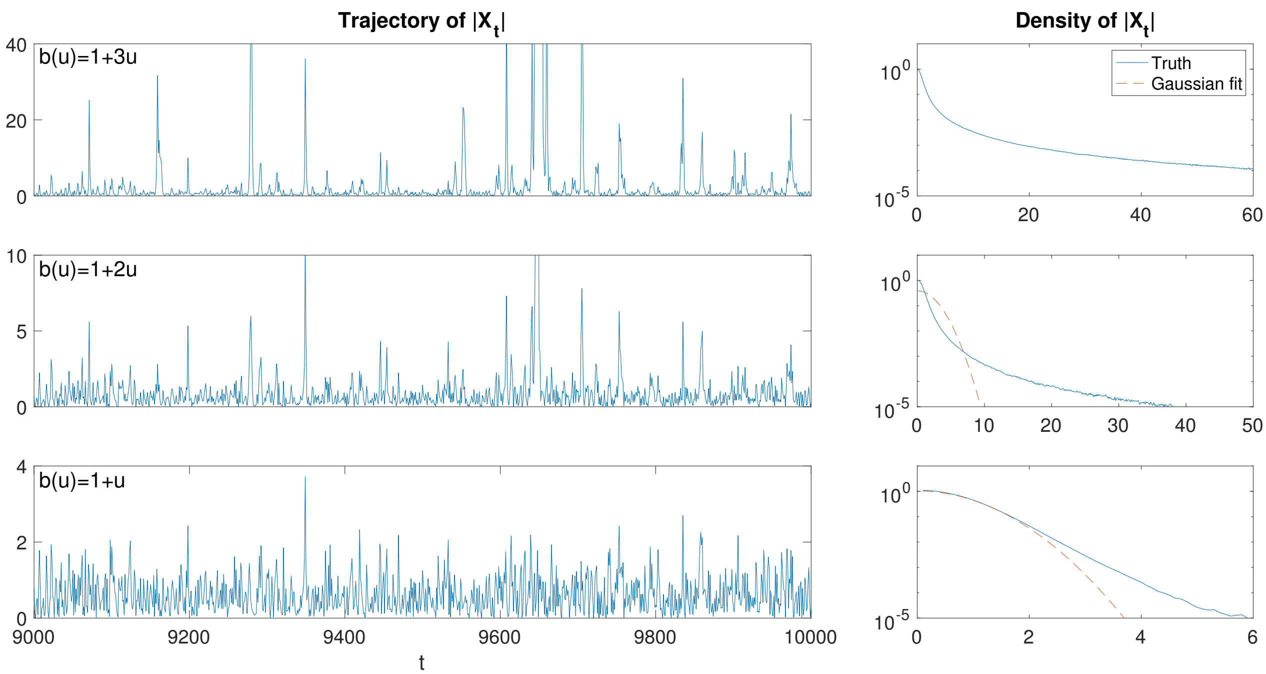

Figure 1.1: Unstable dampings lead to polynomial tails. The damping function being used is labeled at the top left corner of each panel. For the $b(u) = 1 + 3u$ case, the Gaussian fit is invalid since the theoretical variance is infinite.

but takes value zero only at a singular set of times. We also emphasize that this simplified setting can be generalized and include systems (1.1) and (1.2). The general statements can be found in Theorems 2.3, 3.1 and 4.1.

As a quick verification of Theorem 1.1, we conduct several simple numerical experiments. For each experiment, model (1.3) with $\gamma = 2$ is simulated for an extensive length $T = 10^6$. An implicit Euler scheme [25] is implemented for the X_t part with a small time step $\Delta t = 10^{-2}$, so the large anomalies come not as a result of numerical instability. The realization of the unobservable process u_t is kept the same for comparison. We present the trajectory of $|X_t|$ for $t \in [9,000, 10,000]$ to demonstrate the extreme events. We also present the log-density plot based on the histogram of $T/\delta t = 10^8$ data points. A Gaussian density with the same mean and variance is plotted as a reference.

In the first group of experiments, we consider damping functions that can take negative values. Following the example of unforced SPEKF model in [9], we test affine functions $b(u) = 1 + cu$, where $c = 3, 2, 1$. The intercept 1 is necessary for $\mathbb{E}b(u_t) > 0$. The results are presented in Figure 1.1. We can see clearly that the trajectories of $|X_t|$ are filled with strong intermittent extremal anomalies. And with the increment of c, the amplitudes of the extreme events grow exponentially. This can also be seen in the log-density plots, which all have a logarithmic tail profile, while the range increases with c. This indicates the tails are indeed polynomial like. In particular, when $c = 3$, the theoretical variance of X_t is infinite, which we will find out in Section 2.4. The sample variance exceeds 10^5 because of the extremal anomalies. We do not plot the Gaussian density reference as it is invalid.

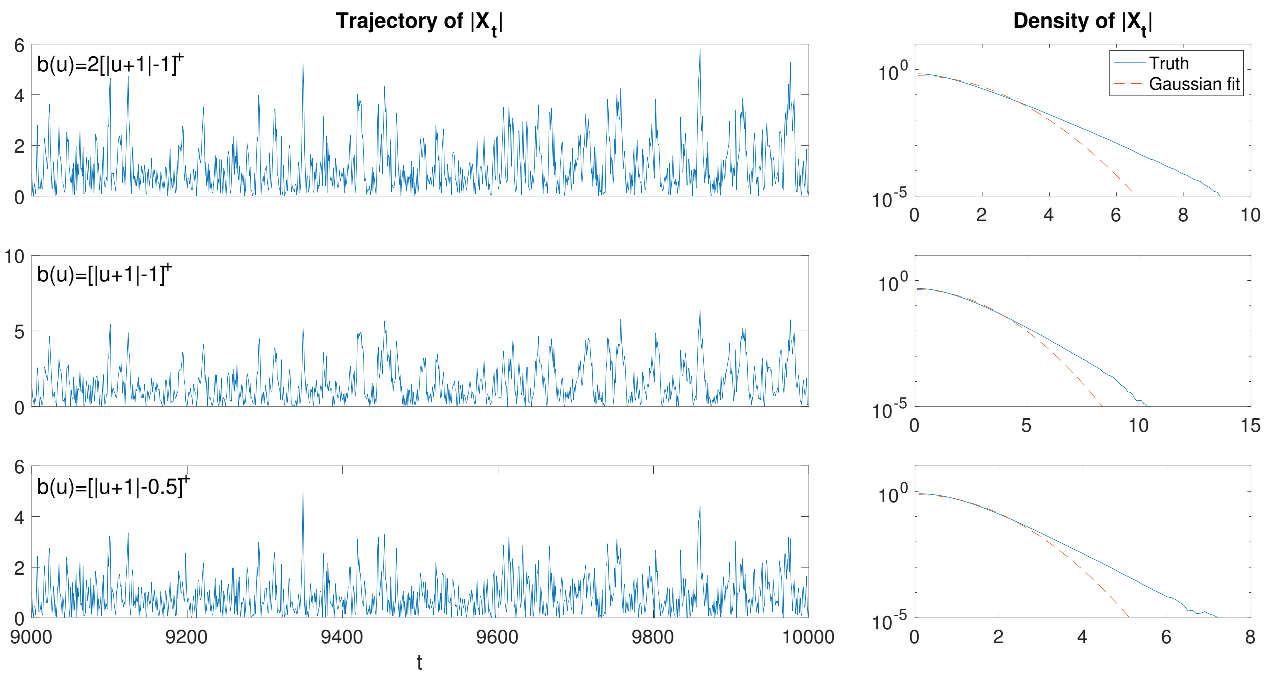

Figure 1.2: Nonnegative dampings that take value zero near $u = -1$ lead to exponential tails. The damping function being used is labeled at the top left corner of each panel.

In the second group of experiments, we consider damping functions that are nonnegative, but take value 0 in intervals. We test with piecewise linear functions

$$b(u) = 2[|u_t + 1| - 1]^+, \quad b(u) = [|u_t + 1| - 1]^+, \quad \text{and } b(u) = [|u_t + 1| - 0.5]^+.$$

Here $[x]^+ := \max\{x, 0\}$ takes the positive part of the input. The results are presented in Figure 1.2. We can see that the trajectories of $|X_t|$ are filled with extreme events of various types. The log-density plots all have a linear profiles, which indicates that the tails are exponential.

In the third group of experiments, we consider damping functions that are nonnegative, but take value 0 only at the origin. We test with functions $b(u) = |u|^c$, where $c = 4, 2, 1$. The results are presented in Figure 1.3. From both the trajectory plots and the density plots, we find that with a larger c, the anomalies last longer, and the tails are more like exponential. And for $c = 1$, the plots are quite similar to the OU case studied next.

In the final experiments, we consider damping functions that are strictly positive

$$b(u) = |u_t|^4 + 1, \quad b(u) = |u_t|^2 + 1, \quad \text{and } b(u) \equiv 1.$$

So in the last experiment, X_t is simply OU. The results are presented in Figure 1.4. We see that the densities are fitted very well with Gaussian approximations. Moreover, the trajectories are all very similar.

As a quick summary, the simulation results are in accordance with the predictions made in Theorem 1.1. We can also see that some simple changes in the damping function can lead to vastly different types of intermittency and heavy tail distributions. In practice, Figure 1.1-1.4 can be used as references for modeling extreme events.

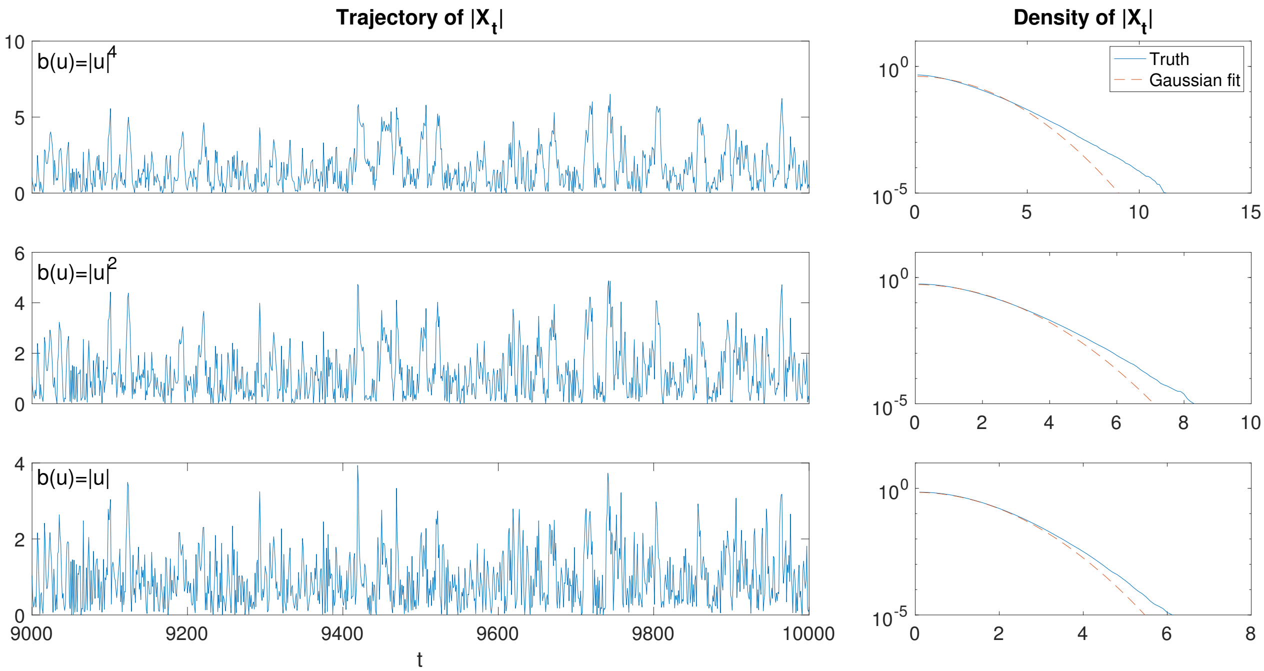

Figure 1.3: Nonnegative dampings that take value zero at the origin lead to tails between exponential and Gaussian. The damping function being used is labeled at the top left corner of each panel.

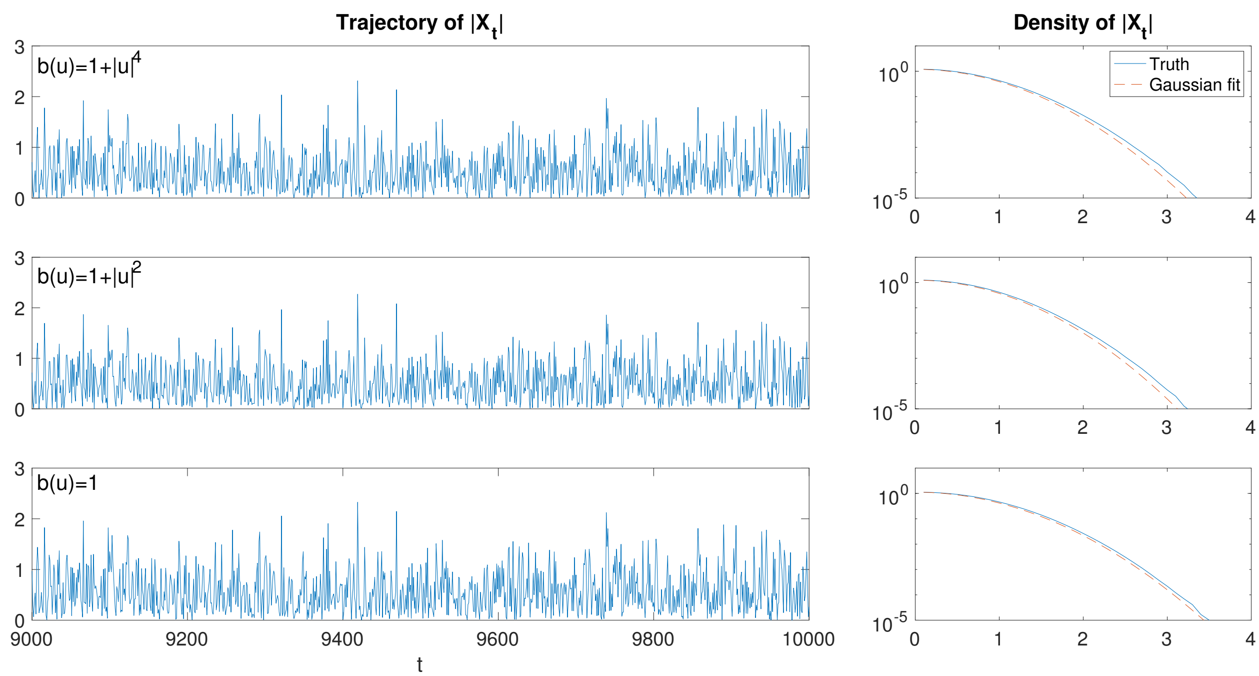

Figure 1.4: Strictly positive dampings lead to Gaussian tails. The damping function being used is labeled at the top left corner of each panel. For the $b(u) = 1$ case, X_t is simply an OU process.

1.2 Moment behaviors

To determine the tail type of X_t, we will consider the moments $\mathbb{E}\|X_t\|^p$ of different power p. By investigating the moments of density functions like $cx^{-p}, \exp(-cx)$ and $\exp(-cx^2)$, we see the moments of a random variable X have very different behavior, depending on the distribution of X:

- Polynomial like: $\mathbb{E}X^p < \infty$ if and only if p is below a threshold.
- Exponential like: $\mathbb{E}X^p < \infty$ for all $p > 0$ and $\log \mathbb{E}X^{2p} \propto 2p \log p$ for large p.
- Gaussian like: $\mathbb{E}X^p < \infty$ and $\log \mathbb{E}X^{2p} \propto p \log p$ for large p.

Such difference can be used to obtain the classification in Theorem 1.1. As we will see, the high moments are very sensitive to the behavior of the damping function b.

Similar moment behaviours can be found in other stochastic models as well. Another way to model stochastic intermittency, is to model X_t as in (1.1), and let u_t be a continuous time Markov jump process [33]. Such a model is known as a Markov switching or regime switching diffusion [4, 15, 41]. It is used in atmospheric science to model complex cloud precipitation, in filtering theory to represent model error, and in financial time series to model hidden market behavior [35, 36, 33, 42].

Markov switching models can also produce heavy tail distributions. In fact, the quadchotomy in Theorem 1.1 has a similar version for finite state Markov switching models in [4]. There are further efforts to generalize this result to infinite state spaces [15, 41], and to investigate the regularity of underlying measures [1, 2]. Yet these results often require a life-death process in the background, which limits their range of application.

While the theoretical result here can be interpreted as an extension of [4], such extension is nontrivial. Stochastic differential equations (SDE) are natural tools when modeling physical processes of continuous values. Approximating them as Markov jump processes is often inappropriate and intractable when the underlying dimension is large. Many physical concepts such as energy dissipation and flow contraction are usually understood only through SDE formulation. The new results in below will reveal the important connection between these physical concepts and the heavy tail phenomena.

Moreover, this paper employs a different analysis framework with the ones used in [4, 15]. In the previous framework, the moment analysis is established by investigating the amount of time the Markov jump process spend in each state. This is difficult to be generalized for an SDE. In this paper, the estimates are constructed by finding novel product type Lyapunov functions. A similar strategy can also be implemented on Markov switching processes to understand complicated geometric ergodicity and multi-scale behaviors [38].

Apart from SDE, moment analysis can also be conducted for stochastic partial differential equations (SPDE) [7, 8, 24]. This has been applied to understand the regularity, growth speed, and localization of the SPDE solutions. So far, these results apply to specific SPDEs, for example the heat equation and the Anderson model. In comparison, our requirements for the SDE are rather general. It will be interesting if the analysis framework developed here can be applied to SPDEs as well.

1.3 Connection to large deviations of trajectory average

Interestingly, our result also leads to a non-asymptotic large deviation bound for trajectory average. Given any function b, by the Birkhoff ergodic theorem, we know

$$t^{-1}\int_0^t b(u_s)ds \overset{t\to\infty}{\longrightarrow} \langle \pi, b\rangle.$$

Such convergence has been used routinely to compute $\langle \pi, b\rangle$, known as the Markov Chain Monte Carlo method. It is natural to ask how does $D_t = t^{-1}\int_0^t b(u_s)ds - \langle \pi, b\rangle$ converge to zero as t becomes large.

To see how does our study of system (1.1) connect with this problem, we let

$$X_t := \exp\left(\int_0^t (b(u_s) - \langle \pi, b\rangle)ds\right). \tag{1.4}$$

Then clearly $X_t > 0$ follows the ordinary differential equation (ODE) $\dot{X}_t = b(u_t)X_t$, and it fits in the formulation (1.1) with $\sigma_x = 0$. A large deviation bound of D_t can be obtained by finding the moments of X_t and then apply the Markov inequality,

$$\mathbb{P}(D_t \geq c) \leq \frac{\mathbb{E}\exp(tpD_t)}{\exp(pct)} = \frac{\mathbb{E}X_t^p}{\exp(pct)}.$$

Corollary 2.8 below implements this idea to asymptotically contractive u_t. Recent results [17, 18] have shown that a large class of diffusion processes, for example over-damped Langevin processes with a convex-at-infinity potential, are asymptotically contractive.

1.4 Paper arrangement and preliminaries

The remainder of this paper is arranged as follow. In Section 2, Theorem 2.3 demonstrates that an unstable damping leads to polynomial tails. As an example, Section 2.4 considers the affine damping in Figure 1.1, where the exact polynomial order of the tail can be found. As another example, Section 2.5 employs our framework to setup an large deviation bound for long time average. Section 3 discusses the scenario where the damping is nonnegative and can take value zero. Theorem 3.1 illustrates the necessary conditions that lead to exponential tails, while Proposition 3.5 considers more general scenarios. Strictly positive damping leads to Gaussian tails is quite well known, but for self-containedness, we give a short proof in Section 4. Lastly, Section 5 discusses how to apply our framework to more general conditional Gaussian systems of type (1.2).

In order to focus on the delivery of the main ideas, we only provide the most important arguments in our discussion. **Most technical verifications are allocated in the appendix.**

In this paper, we use $\|a\|$ to denote the l_2 norm of a vector a, $\langle a, b\rangle$ to denote the inner product of a and b. $\langle \pi, f\rangle = \int f(u)\pi(du)$ is the average of f under the equilibrium measure π. We denote the generator of process (X_t, u_t) as $\mathcal{L}$, which can be written explicitly as

$$\mathcal{L}f(x,u) = -\langle b, \nabla_x f\rangle + \langle h, \nabla_u f\rangle + \frac{1}{2}\mathrm{tr}(\sigma_x^2\nabla_x^2 f + \nabla_u^2 f).$$

In above, ∇_x and ∇_u are the gradients with respect to variables x and u, and ∇_x^2 and ∇_u^2 are the corresponding Hessian matrices. We can also define the Carre du champ operator using $\mathcal{L}$ [3]:

$$\Gamma(f,g) = \frac{1}{2}(\mathcal{L}(fg) - h\mathcal{L}g - g\mathcal{L}f) = \sigma_x^2 \langle \nabla_x g, \nabla_x f\rangle + \langle \nabla_u g, \nabla_u f\rangle.$$

Obviously Γ is bilinear, symmetric and positive. We will also write $\Gamma(g) := \Gamma(g,g)$ for simplicity. One important arithmetic property of Γ is the following chain rule of the generator

$$\mathcal{L}\varphi(g) = \dot{\varphi}(g)\mathcal{L}g + \ddot{\varphi}(g)\Gamma(g). \tag{1.5}$$

The derivation of the formula above and more properties of Γ can be found in [3]. Also, it is worth noting that in our discussion below, we often concern functions of only one variable, that is $f(x,u) = f(x)$ or $f(x,u) = f(u)$. Then some parts of the formulas above will vanish.

The moment function $\|x\|^p$ will naturally be of interest in our discussion. Unfortunately it is not $\mathcal{C}^2$ at the origin when $p < 2$, so $\mathcal{L}$ cannot be applied. To remedy this, we will often use $\mathcal{E}_p(x)$ in below as an surrogate, which is also used in [4, 38]:

Lemma 1.2. *For any $p > 0$, $\mathcal{E}_p(x) = \frac{\|x\|^{p+2}}{1+\|x\|^2} + 1$ is equivalent to $\|x\|^p$ in the following sense:*

$$\frac{1}{2}(\|x\|^p + 1) \leq \mathcal{E}_p(x) \leq \|x\|^p + 1.$$

Moreover, for any $\delta > 0$, there is a $C_\delta > 0$ such that

$$-(pb(u)+\delta)\mathcal{E}_p(x) - C_\delta(|b(u)|+1) \leq \mathcal{L}\mathcal{E}_p(x) \leq -(pb(u)-\delta)\mathcal{E}_p(x) + C_\delta(|b(u)|+1).$$

2 Polynomial tails from unstable dampings

Our first result shows that if the damping is unstable, that is $b(u_*) < 0$ for some u_*, then X_t in (1.1) will have polynomial tails. This involves two parts, showing $\mathbb{E}\|X_t\|^p < \infty$ when $p > 0$ is small enough, and $\mathbb{E}\|X_t\|^p = \infty$ when p is large enough.

To establish the lower bound, that is $\lim_t \mathbb{E}\|X_t\|^p = \infty$ for a large p, it suffices to assume some general regularity and growth conditions on b and h.

Assumption 2.1. *Suppose the following holds for all y, where $C > 0, m \geq 2$ are constants, M_y is a constant that may depend on y:*

$$\|h(x)\| \leq C\|x\|^{m-1} + C, \quad b(x) \leq b(y) + M_y\|x-y\| + M_y\|x-y\|^m.$$

To establish the upper bound, we need in addition that u_t is asymptotically contractive:

Definition 2.2. *Given two distributions μ and ν, we use $d(\mu,\nu)$ to denote the Wasserstein-1 distance between μ and ν, generated by the l_2 norm. Let P_t^u denote the distribution of u_t with $X_0 = x$. We say u_t is asymptotically contractive if there are constants $C_\gamma, \gamma > 0$ such that*

$$d(P_t^u, P_t^v) \leq C_\gamma \exp(-\gamma t)\|u - v\|$$

holds for all u, v and t.

Recent results [17, 18] have shown that a wide range of SDE are asymptotically contractive. For example, if u_t follows the overdamped Langevin dynamics, that is $h(u) = -\nabla H(u)$ in (1.1), and the potential H is strictly convex outside a bounded region, then u_t is asymptotically contractive. If u_t is a stable OU process, this assumption holds naturally.

The general statement of unstable damping leads to polynomial tails is given below.

Theorem 2.3. *Under Assumption 2.1, suppose that $b(u^*) < 0$ for a certain u^*.*

1) If $\sigma_x > 0$, then

$$\lim_{t\to\infty} \mathbb{E}\|X_t\|^p = \infty, \quad \text{for sufficiently large } p.$$

2) If u_t is asymptotically contractive, b has Lipschitz constant $\|b\|_{Lip}$, and the average damping $\langle \pi, b\rangle > 0$, then for any p such that

$$p\pi(b) - \frac{1}{2}p^2 C_\gamma^2 \gamma^{-2}\|b\|_{Lip}^2 > 0,$$

we have

$$\limsup_{t\to\infty} \mathbb{E}\|X_t\|^p < \infty.$$

3) Assuming the conditions of 2), if in addition $\sigma_x > 0$, and h preserves energy, that is for some constant $\lambda > 0, M_\lambda > 0$

$$\langle h(u), u\rangle \le -\lambda\|u\|^2 + M_\lambda,$$

then (X_t, u_t) is geometrically ergodic.

The proof comes as a combination of the results from the next three subsections. The complete proof can be found in the appendix.

Before we move on, we give a quick remark on the average damping condition $\langle \pi, b\rangle > 0$. This is a necessary condition. In the simplified case $\sigma_x = 0$ and $X_0 = x_0$,

$$\mathbb{E}\|X_t\|^p = \|x_0\|^p \mathbb{E}\exp\left(-p\int_0^t b(u_s)ds\right).$$

By Jensen's inequality, the long time damping effect on $\|X_t\|^p$ can be bounded by

$$\mathbb{E}\exp\left(-p\int_0^t b(u_s)ds\right) \ge \exp\left(-p\int_0^t \mathbb{E}b(u_s)ds\right) \overset{t\to\infty}{\approx} \exp\left(-pt\langle \pi, b\rangle\right).$$

So in order for $\mathbb{E}\|X_t\|^p$ to be stable, $\langle \pi, b\rangle$ needs to be positive. On the other hand, Jensen's inequality provides only one side of the estimate. In fact, when b is not strictly positive, the long time damping effect, $\mathbb{E}\exp\left(-p\int_0^t b(u_s)ds\right)$, does not scale as $\exp(-cpt)$ for large p. This is the main mechanism behind the extreme events and heavy tails.

2.1 Building Lyapunov functions

In order to show $\mathbb{E}\|X_t\|^p$ is bounded uniformly in time, we will try to find a Lyapunov function $V(x,u) \approx \|x\|^p$, such that for some $\rho, k_v > 0$, when applying the generator $\mathcal{L}$ of (1.1)

$$\mathcal{L}V(x,u) \leq -\rho V(x,u) + k_v. \tag{2.1}$$

Then applying the Gronwall's inequality and Dynkin's formula, we have

$$\mathbb{E}V(X_t,u_t) \leq e^{-\rho t}\mathbb{E}V(X_0,r_0) + k_v/\rho.$$

Conversely, in order to show $\mathbb{E}\|X_t\|^p \to \infty$ for $t \to \infty$, it suffices to find a function $U(x,u) \approx \|x\|^p$, such that for some $\rho, k_v > 0$

$$\mathcal{L}U(x,u) \geq \rho U(x,u) - k_v.$$

The key to this method is finding the proper V and U. One naive attempt is letting U or V to be $\|x\|^p$. However, this will not be sufficient, since for $p \geq 2$,

$$\mathcal{L}\|x_t\|^p = -pb(u_t)\|x_t\|^p + \frac{1}{2}p(p-1)\|x_t\|^{p-2}.$$

Inequality like (2.1) does not hold because the appearance of $b(u_t)$.

The main idea here is to look for a function that is the product of two parts, one part is a potential that depends on u, the other part is roughly the moment of x:

Lemma 2.4. *Fix $q > 0$ and $\delta > 0$. Assume there are functions $\mathcal{E}_q > 0, f > 0$ such that for some $C_\delta > 0$ and ρ*

$$\begin{gathered}\mathcal{L}\mathcal{E}_q(x) \leq -(qb(u)-\delta)\mathcal{E}_q(x) + C_\delta(1+|b(u)|),\\ \mathcal{L}f(u) - (qb(u)-\delta)f(u) \leq -\rho f(u),\end{gathered} \tag{2.2}$$

then $V(x,u) = f(u)\mathcal{E}_q(x)$ satisfies: $\mathcal{L}V(x,u) \leq -\rho V(x,u) + C_\delta(1+|b(u)|)f(u)$.

The converse is also true. If there are functions $\mathcal{E}_q > 0, g > 0$

$$\begin{gathered}\mathcal{L}\mathcal{E}_q(x) \geq -(qb(u)+\delta)\mathcal{E}_q(x) - C_\delta(1+|b(u)|),\\ \mathcal{L}g(u) - (qb(u)+\delta)g(u) \geq \rho g(u),\end{gathered} \tag{2.3}$$

then $U(x,u) = g(u)\mathcal{E}_q(x)$ satisfies: $\mathcal{L}U(x,u) \geq \rho U(x,u) - C_\delta(1+|b(u)|)g(u)$.

Proof. By the product rule, the generator of $V(x,u)$ is

$$\mathcal{L}V(x,u) = \mathcal{E}_q(x)[\mathcal{L}f(u)] + [\mathcal{L}\mathcal{E}_q(x)]f(u) \quad \text{(or the similar version with } g).$$

Plug in the conditions give us the claim. □

For our purpose, we will let $\mathcal{E}_q = \frac{\|x\|^{q+2}}{1+\|x\|^2} + 1$ which satisfies the requirement of Lemma 2.4, and $\mathcal{E}_q$ is equivalent to $\|x\|^q$ by Lemma 1.2. We don't use $\|x\|^q$ directly because it is not C^2 when $q < 2$.

Then suppose we can find a regular f that satisfies (2.2), we can show $\mathbb{E}\|X_t\|^p$ is finite for $p < q$. Conversely, with a regular g that satisfies (2.3), we can show $\lim_{t\to\infty}\mathbb{E}\|X_t\|^p = \infty$ for $p > q$. This is proved by the following lemma:

Lemma 2.5. *Suppose there is a function f that satisfies* (2.2) *with a $\rho > 0$, and*

$$\limsup_{t\to\infty} \mathbb{E}(1+|b(u_t)|)f(u_t) < M_0, \quad \limsup_{t\to\infty} [\mathbb{E}f(u_t)^{-\frac{1}{\alpha}}]^\alpha < M_\alpha,$$

for any $\alpha > 0$ with an appropriate M_α, then

$$\limsup_{t\to\infty} \mathbb{E}\|X_t\|^p < \frac{2}{\rho} C_\delta M_0 M_{\frac{q-p}{q}}, \quad \forall p < q.$$

Conversely, suppose $\sigma_x \neq 0$ and there is a function $g > 0$ that satisfies (2.3) *with a $\rho > 0$, and*

$$\limsup_{t\to\infty} \mathbb{E}(1+|b(u_t)|)g(u_t) < M_0, \quad \limsup_{t\to\infty} [\mathbb{E}g(u_t)^{\frac{1}{\alpha}}]^\alpha < M_\alpha,$$

for any $\alpha > 0$ with an appropriate M_α, then $\lim_{t\to\infty} \mathbb{E}\|X_t\|^p = \infty$ for any $p > q$.

2.2 Lower bound: constructive verification

Based on Lemma 2.4 and 2.5, in order to show that $\mathbb{E}\|X_t\|^q = \infty$ for a large q, it suffices to find a positive function g such that (2.3) holds.

Let $\eta = q^{-1}\log g$, it is well defined. Then by the chain rule formula (1.5), (2.3) is equivalent to

$$(qb(u) + \delta + \rho)\exp(q\eta) \leq \mathcal{L}\exp(q\eta) = q\exp(q\eta)\langle h, \nabla_u \eta\rangle + \frac{1}{2}\exp(q\eta)(q\mathrm{tr}(\nabla_u^2\eta) + q^2\|\nabla_u\eta\|^2).$$

In other words, we need to find an η such that

$$\big(\mathrm{tr}(\nabla_u^2\eta) + q\|\nabla_u\eta\|^2\big) + 2\langle h, \nabla_u\eta\rangle \geq 2b + 2q^{-1}(\delta+\rho). \tag{2.4}$$

This can be done by an explicit construction, as long as b and h are regular as in Assumption 2.1.

Lemma 2.6. *Suppose $b(u^*) < 0$. Under Assumption 2.1, there is a $c > 0$, so that for sufficiently large q,* (2.4) *holds with*

$$\eta(u) = c\|u - u^*\|^m.$$

Proof. By our assumption, there is an M such that

$$b(u) \leq b(u^*) + M\|u-u^*\| + M\|u-u^*\|^m, \quad \forall u.$$

Then notice that

$$\mathrm{tr}(\nabla_u^2\eta) = m(m-1)cd_u\|u-u^*\|^{m-2} \geq 0,$$
$$\nabla_u\eta = mc(u-u^*)\|u-u^*\|^{m-1}, \quad \|\nabla_u\eta\|^2 = m^2c^2\|u-u^*\|^{2m}.$$

Under Assumption 2.1, by Young's inequality, we can increase M such that

$$\langle h, \nabla_u\eta\rangle \geq -cM\|u-u^*\|^{2m} - cM.$$

So in combine, to show (2.4) it suffices to show, (we ignore the $\mathrm{tr}(\nabla_u^2\eta) \geq 0$ term)

$$(qm^2c^2 - 2cM)\|u-u^*\|^{2m} + (-2b(u^*) - 2cM - 2q^{-1}(\delta+p)) \geq 2cM\|u-u^*\| + 2cM\|u-u^*\|^m.$$

By Young's inequality, this can be achieved by a sufficiently large q and small c. □

2.3 Upper bound: solution from the Feynman Kac formula

To find a f that satisfies (2.2), we let $\theta = q^{-1}\log f$. Then similar to the derivation of (2.4), we find that (2.2) is equivalent to

$$\mathcal{L}\theta(u) \le b(u) - q^{-1}(\rho+\delta) - \frac{1}{2}q\|\nabla_u\theta(u)\|^2. \tag{2.5}$$

Directly solving (2.5) is challenging, since it involves a nonlinear term $\|\nabla_u\theta(u)\|^2$. Here the idea is that we look for θ that is Lipschitz, so with a certain constant M, $\frac{1}{2}\|\nabla_u\theta(u)\|^2 \le M$. Then for (2.5) to hold, it suffices to solve a linear problem:

$$\mathcal{L}\theta(u) \le b(u) - q^{-1}(\rho+\delta) - qM.$$

To solve this, we recall the formula for Cauchy problems. Given a specific $b(u)$, the solution of

$$\mathcal{L}\theta(u) = b(u)$$

exists if only and if $\langle\pi, b\rangle = 0$, and θ is given by the following Feymann Kac's formula

$$\theta(u) = \int_0^\infty \mathbb{E}^u b(u_t)dt.$$

For self-completeness, we verify this fact in Lemma A.2.

For our purpose, it is natural to try

$$\theta(u) = \int_0^\infty \mathbb{E}^u(b(u_t) - \langle\pi, b\rangle)dt. \tag{2.6}$$

Then to verify (2.5), we simply need

$$q^{-1}(\rho+\delta) + \frac{1}{2}q\|\nabla_u\theta(u)\|^2 \le \langle\pi, b\rangle.$$

Note that $\rho+\delta$ can be arbitrarily small positive numbers, and $\langle\pi, b\rangle > 0$ by our assumption, so it suffices to verify that $\|\nabla_u\theta(u)\|$ is bounded globally. We would assume the following assumption for u_t

Lemma 2.7. *Assume that u_t satisfies the asymptotic Lipschitz contraction, and define θ as in* (2.6)*, then*

$$\|\nabla_u\theta(u)\| \le C_\gamma\gamma^{-1}\|b\|_{Lip}.$$

Proof. Note that

$$\begin{aligned}\|\nabla_u\mathbb{E}^u b(u_t)\| &= \sup_{\|v\|=1}\lim_{\epsilon\to 0}\epsilon^{-1}[\mathbb{E}^{u+\epsilon v}b(u_t) - \mathbb{E}^u b(u_t)]\\ &\le \sup_{\|v\|=1}\lim_{\epsilon\to 0}\epsilon^{-1}\|b\|_{Lip}d(P_t^{u+\epsilon v}, P_t^u)\\ &\le \sup_{\|v\|=1}\lim_{\epsilon\to 0}\|b\|_{Lip}C_\gamma\|v\|\exp(-\gamma t) = C_\gamma\|b\|_{Lip}\exp(-\gamma t).\end{aligned}$$

Therefore

$$\|\nabla_u\theta(u)\| = \left\|\int_0^\infty \nabla_u\mathbb{E}^u b(u_t)dt\right\| \le C_\gamma\|b\|_{Lip}\int_0^\infty \exp(-\gamma t)dt = C_\gamma\gamma^{-1}\|b\|_{Lip}. \tag{2.7}$$

□

2.4 Example: unforced SPEKF

Here we consider the unforced SPEKF model (1.3) with an affine damping function

$$\begin{aligned} dX_t &= -b(u_t + m_u)X_t dt + dW_t, \\ du_t &= -\gamma u_t dt + dB_t. \end{aligned} \tag{2.8}$$

We will demonstrate our procedures discussed above by applying them on this simple process. The $\theta(t)$ in Lemma A.2 is given by

$$\theta(u) = b\int_0^\infty dt\mathbb{E}^u u_t = bu\int_0^\infty \exp(-\gamma t)dt = \frac{bu}{\gamma}.$$

From this, we see that Lemma 2.7 is sharp, since u_t is asymptotically contractive with $C_\gamma = 1$, and $\gamma = \gamma$. This suggests us to use $f(u) := \exp(qbu/\gamma)$ in Lemma 2.4. In fact, g can be chosen as the same. Simply note that

$$\mathcal{L}f(u) = (-qbu + \tfrac{1}{2}q^2b^2/\gamma^2)f(u).$$

So by Lemma 2.4, if we let $V(x,u) = \mathcal{E}_q(x)f(u)$, then for any $\delta > 0$, there is a C_δ, so that

$$\begin{aligned} (\tfrac{1}{2}q^2b^2/\gamma^2 - \delta - qm_u)V(x,u) - C_\delta(1+|b(u)|)f(u) &\leq \mathcal{L}V(x,u) \\ &\leq (\tfrac{1}{2}q^2b^2/\gamma^2 + \delta - qm_u)V(x,u) + C_\delta(1+|b(u)|)f(u). \end{aligned}$$

By Lemma 2.5, this means that $\limsup_{t\to\infty}\|X_t\|^q$ is infinite if $q > q_0$, and is finite if $q < q_0$. The threshold here is given by

$$q_0 = \frac{2m_u\gamma^2}{b^2}.$$

2.5 Example: large deviation bound

As another example, we demonstrate how to apply our framework to show the deviation of long time average of u_t is sub-Gaussian.

Corollary 2.8. *Assume that u_t is asymptotically Lipschitz contractive, and it is exponentially integrable*

$$\limsup_{t\to\infty}\mathbb{E}\exp(\alpha\|u_t\|) < \infty, \quad \forall\alpha.$$

Then for any $\delta > 0$, the following large deviation bound holds for certain M_δ

$$\mathbb{P}\left(\frac{1}{t}\int_0^t b(u_s)ds - \langle\pi, b\rangle > c\right) \leq M_\delta \exp\left(-(\tfrac{1}{2}c^2 - \delta)D_M t\right), \quad \forall t > 0,$$

where $D_M = C_\gamma^2\gamma^{-2}\|b\|_{Lip}^2$.

Proof. Consider X_t as defined in (1.4). Since there is no diffusion term in X_t, we consider function

$$V(x,u) := x^q\exp(q\theta(u)),$$

where $\theta(u)$ comes from Lemma A.2, and satisfies

$$\mathcal{L}\theta = b - \langle \pi, b\rangle, \quad \|\nabla_u \theta\| \leq C_\gamma \gamma^{-1}\|b\|_{Lip}.$$

Apply the generator to V, we find

$$\mathcal{L}V(x,u) = \left(-q(b - \langle \pi, b\rangle) + q\mathcal{L}\theta + \frac{1}{2}q^2\|\nabla_u\theta\|^2\right)V(x,u) \leq \frac{1}{2}q^2 D_M V(x,u).$$

So by Dynkin's formula,

$$\mathbb{E}X_t^q \exp(q\theta(u_t)) \leq \exp\left(\frac{1}{2}q^2 t D_M\right)\mathbb{E}\exp(q\theta(u_0)).$$

For any $p < q$, by Hölder's inequality,

$$\begin{aligned}
\mathbb{E}X_t^p &\leq \left(\mathbb{E}X_t^q \exp(q\theta(u_t))\right)^{\frac{p}{q}} \left(\mathbb{E}\exp(q\theta(u_t))^{\frac{-q}{q-p}}\right)^{\frac{q-p}{q}} \\
&\leq \exp(\tfrac{1}{2}qptD_M)\left(\mathbb{E}\exp(q\theta(u_0))^{\frac{p}{q}}\right. \left(\mathbb{E}\exp(q\theta(u_t))^{\frac{-q}{q-p}}\right)^{\frac{q-p}{q}} \leq \exp(\tfrac{1}{2}qptD_M)M_{p,q}.
\end{aligned}$$

The constant $M_{p,q}$ exists because we assume $\limsup_{t\to\infty}\mathbb{E}\exp(\alpha\|u_t\|) < \infty$, and θ is Lipschitz. Therefore if we let

$$D_t := \frac{1}{t}\int_0^t b(u_s)ds - \langle \pi, b\rangle,$$

then

$$\mathbb{P}(D_t \geq c) \leq \frac{\mathbb{E}\exp(tpD_t)}{\exp(ptc)} = \frac{\mathbb{E}X_t^p}{\exp(ptc)} = \exp((\tfrac{1}{2}q - c)ptD_M)M_{p,q}.$$

We pick $p = c, q = c + \frac{2\delta}{c}$ and find our claim. □

3 Exponential tails from nonnegative dampings

This section shows that nonnegative dampings lead to exponential or weaker tails.

Theorem 3.1. *Suppose the following hold:*

- u_t *is asymptotically contractive.*
- *The damping function satisfies* $b(u) \geq 0$ *and* $b(u) = 0$ *when* $\|u - u_*\| \leq \epsilon$ *for some* u_*.
- *The dynamics of* u_t *dissipates the energy centered at* u_*, *that is there are* $\lambda, M_\lambda > 0$ *so that*

$$\langle u - u_*, h(u)\rangle \leq -\lambda\|u - u_*\|^2 + M_\lambda.$$

Then X_t *has exponential-like tails. In particular*

$$\lim_{p\to\infty}\lim_{t\to\infty}\frac{\log \mathbb{E}\|X_t\|^{2p}}{p\log p} = 2.$$

Again, this result consists of two parts. The upper bound comes from Proposition 3.2. The lower bound comes from as the combination of Proposition 3.5 and Corollary 3.7. One can find the detailed verification in the appendix.

Theorem 3.1 doesn't consider the delicate case where $b(u) = 0$ only at a single point. This was mentioned in Theorem 1.1 as case (iii). We only provide a lower bound in Proposition 3.5, indicating the tail is strictly heavier than Gaussian. This is already useful in practice. Also note that the statement of Theorem 1.1 is rigorous, as we only claim that the distribution is between exponential and Gaussian.

3.1 Upper bound

As a matter of fact, it is relatively easy to see that process with nonnegative damping has sub-exponential tails.

Proposition 3.2. *Suppose $b \geq 0$, and the θ in (2.6) is well defined, with $\Gamma(\theta)$ bounded, and the following integrability condition holds for any $\alpha > 0$*

$$\limsup_{t\to\infty} \mathbb{E}\exp(\alpha\theta(u_t)) < \infty, \quad \limsup_{t\to\infty} \mathbb{E}b(u_t)\exp(\alpha\theta(u_t)) < \infty.$$

Then X_t has sub-exponential tails. In particular, for any $\beta \in \mathbb{R}^d$ that

$$\tfrac{1}{2}\|\beta\|^2\|\nabla_u\theta\|^2 + \tfrac{1}{2}\sigma_x^2\|\beta\|^2 - \|\beta\|\langle\pi, b\rangle < 0,$$

then

$$\limsup_{t\to\infty} \mathbb{E}\exp\langle\beta, X_t\rangle < \infty.$$

In particular, this indicates that

$$\limsup_{p\to\infty}\limsup_{t\to\infty} \frac{\log \mathbb{E}\|X_t\|^{2p}}{p\log p} \leq 2.$$

3.2 Lower bound

To show the lower bound requires additional works. First let us define the set of damping function that can yield approximately exponential tails.

Definition 3.3. *We say function $(b, h) \in \mathcal{A}_m$ if there are $g_1, \ldots, g_m$ on $\mathbb{R}^d$ such that*

(1) $b \geq 0$, and $b(u_*) = 0$ for certain u_*.

(2) g_1 satisfies the level-1 constraint $\Gamma(g_1) \geq b$.

(3) g_k satisfies the level-k constraint $\Gamma(g_k) + \mathcal{L}g_{k-1} \geq 0$, for $k = 2, \ldots m$.

(4) $\mathcal{L}g_m \geq -M$ for a constant M.

(5) There are constants M_0 and M_1

$$G_p(u) = \sum_{k=1}^{m} p^{\frac{1}{2^k}} g_k(u) \leq \sqrt{p}M_0, \quad \mathbb{E}G_p(u_0) \geq -\sqrt{p}M_1.$$

(6) Alignment condition: for all $j, k \leq m$, $\Gamma(g_j, g_k) \geq 0$.

The long time damping effect from $b(u_t) \in \mathcal{A}_m$ is revealed by the following lemma. The main message is that $b(u_t)$ creates a weaker long time damping when applied to higher moments of X_t.

Lemma 3.4. *Suppose $(b, h) \in \mathcal{A}_m$, then the following holds under the invariant measure for $p \geq 1$:*

$$\mathbb{E} \exp\left(-2p \int_0^t b(u_s) ds\right) \geq \exp(-2p^{\frac{1}{2^m}} Mt - 2\sqrt{p} M_0 - 2\sqrt{p} M_1).$$

Sub-exponential tails come as a result of this week long time damping.

Proposition 3.5. *Suppose the damping and the drift of u_t, (b, h), belongs to $\mathcal{A}_m$ as in Definition 3.3, then under the equilibrium measure,*

$$\liminf_{p \to \infty} \liminf_{t \to \infty} \frac{\log \mathbb{E}\|X_t\|^{2p}}{p \log p} \geq 2 - \frac{1}{2^m}.$$

In other words, X_t has a tail between exponential and Gaussian.

3.3 Energy dissipation

If b is in $\mathcal{A}_m$ for all m, then Theorem 3.5 indicates the moment of $\|X_t\|$ behaves very much the same as the exponential distribution. In this section we show that this will be the case under the conditions of Theorem 3.1.

On the other hand, note that b is in $\mathcal{A}_m$ for all m does not mean $b \in \mathcal{A}_\infty$. For a $b \in \mathcal{A}_\infty$ to be well defined, one also needs to check the G_p in Definition 3.3 converges as an infinite sum. We will not do this in our analyses. But this will not be very restrictive in application, since in Theorem 3.5, $\frac{1}{2^m} \approx 0$ even with a small m.

Lemma 3.6. *Suppose $b(u_*) = 0$, and for some $m \in \mathbb{Z}^+, C > 0$,*

$$b(u) \leq C\|u - u_*\|^{2^{m+1}-2}. \tag{3.1}$$

Suppose also the energy centered at u_* is dissipative under the drift h, so for some $\lambda, M > 0$,

$$\langle u - u_*, h(u)\rangle \leq -\lambda\|u - u_*\|^2 + M_\lambda. \tag{3.2}$$

Then $(b, h) \in \mathcal{A}_m$.

Corollary 3.7. *Suppose $b(u) = 0$ when $\|u - u^*\| \leq \delta$, and b has polynomial growth for some n*

$$b(u) \leq D\|u - u_*\|^n.$$

Suppose the energy centered at u_* is dissipative under the drift h, so (3.2) holds. Then $(b, h) \in \mathcal{A}_m$ for all m such that $2^{m+1} \geq n + 2$. In other words, X_t will have an exponential like tail.

Proof. We just need to verify (3.1) for some C. Simply note that when $\|u - u_*\| \geq \delta$

$$\begin{aligned} b(u) \leq D\|u - u_*\|^n &= D\delta^n\|(u - u_*)/\delta\|^n \\ &\leq D\delta^n\|(u - u_*)/\delta\|^{2^{m+1}-2} = D\delta^{n+2-2^{m+1}}\|u - u_*\|^{2^{m+1}-2}. \end{aligned}$$

And when $\|u - u_*\| < \delta$, $b(u) = 0$, so (3.1) holds automatically. $\square$

4 Sub-Gaussian tails from strictly positive dampings

The following analysis is rather standard. But since it is short and we want to be self-contained, we provide the details rather than finding a reference.

Theorem 4.1. *Suppose $b(u_t) \geq b_0$ for a $b_0 > 0$, then if $\alpha < b_0$,*

$$\limsup_{t\to\infty} \mathbb{E}\exp(\alpha\|X_t\|^2) < \infty.$$

This leads to

$$\limsup_{p\to\infty} \limsup_{t\to\infty} \frac{\mathbb{E}\|X_t\|^{2p}}{p\log p} \leq 1.$$

If u_t dissipates the energy, that is

$$\langle h(u), u\rangle \leq -\lambda\|u\|^2 + M_\lambda,$$

then if $\lambda < a$,

$$\limsup_{t\to\infty} \mathbb{E}\exp(\alpha\|u_t\|^2) < \infty.$$

Proof. Let $\mathcal{E}(x) = \exp(\alpha\|x\|^2)$, with $\alpha < b_0$. Apply the generator to it,

$$\begin{aligned}\mathcal{L}\mathcal{E} &= -2\alpha\|x\|^2 b(u)\mathcal{E} + \sigma_x^2(\alpha d_X + 2\alpha^2\|x\|^2)\mathcal{E} \\ &\leq (-\delta\|x\|^2 + \sigma_x^2 d_X)\alpha\mathcal{E},\end{aligned} \tag{4.1}$$

where $\delta = 2b_0 - 2\alpha$. When $\delta\|x\|^2 \leq (1+d_X)\sigma_x^2$, $\mathcal{E}(x) \leq \exp((1+d_X)\alpha\sigma_x^2/\delta)$, otherwise $\mathcal{L}\mathcal{E}_t \leq -\alpha\sigma_x^2\mathcal{E}_t$. Therefore

$$\mathcal{L}\mathcal{E} \leq -\alpha\sigma_x^2\mathcal{E} + \alpha\sigma_x^2\exp((1+d_X)\alpha\sigma_x^2/\delta)$$

So Dynkin's formula and Gronwall's inequality immediately gives us

$$\mathbb{E}\mathcal{E}(X_t) \leq \exp(-\alpha\sigma_x^2 t)\mathbb{E}\mathcal{E}(X_0) + \exp((1+d_X)\alpha\sigma_x^2/\delta) < \infty.$$

Finally note that by Taylor expansion of $\exp(\alpha\|x\|^2)$,

$$\|x\|^{2p} \leq p!\alpha^{-p}\exp(\alpha\|x\|^2)$$

Apply an estimate of $\log k$ as in (A.1),

$$\log\mathbb{E}\|X_t\|^{2p} \leq \sum_{k=1}^{p}\log k - p\log\alpha + \log\mathbb{E}\exp(\alpha\|X_t\|^2) = p\log p + O(p).$$

A similar analysis applies to u_t as well. Let $\mathcal{E}(u) = \exp(\alpha\|u\|^2)$. Apply the generator to it,

$$\begin{aligned}\mathcal{L}\mathcal{E} &= -2\alpha\langle h(u), u\rangle\mathcal{E} + (\alpha d_u + 2\alpha^2\|u\|^2)\mathcal{E} \\ &\leq (-2(\lambda-\alpha)\|u\|^2 + d_u + 2M_\lambda)\alpha\mathcal{E}.\end{aligned}$$

The follow up analysis is much the same as after (4.1). □

5 General conditional Gaussian system

The X_t part in a general multivariate conditional Gaussian system (1.2) can be written as

$$dX_t = -B(u_t)X_t + \Sigma_X dW_t,$$

where u_t is the same as in system (1.1). This formulation is different from (1.1), since B is matrix-valued. In this section, we show how to build moment bounds for (1.2) by building surrogate damping rates as in (1.1).

To begin, we decompose $B(u_t)$ as

$$B(u_t) = \sum_{i=1}^{n} b_i(u_t)B_i, \tag{5.1}$$

for some matrixs B_i and functions b_i. This decomposition always exists, since

$$B(u_t) = \sum_{j,k=1}^{d_X} [B(u_t)]_{j,k}E_{j,k},$$

where $E_{j,k}$ is the matrix with all components being zero, except the (j,k)-th component being 1. Yet this choice of decomposition can sometimes be sub-optimal.

With decomposition (5.1), let $N_i \subseteq \{1,\dots,d_X\}$ includes all indices that B_i involves, so if $j \notin N_i$ then $[B_i]_{j,k} = 0$. There are constants m_i and M_i such that

$$m_i I_{N_i} \preceq \frac{1}{2}(B_i + B_i^T) \preceq M_i I_{N_i}.$$

Here I_{N_i} is the diagonal matrix with (j,j)-th term being one if and only if $j \in N_i$, and with two symmetric matrices A and B, $A \preceq B$ indicates that $B - A$ is positive semidefinite. One easy choice of N_i can be $N_i = \{1,\dots,d_X\}$ for all i, then m_i and M_i are simply the minimum and maximum eigenvalues of $\frac{1}{2}(B_i + B_i^T)$.

Consider the following two scalar value damping functions where $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$.

$$\bar{b}(u) = \bigwedge_{j=1}^{d} \sum_{i:j\in N_i} M_i b_i \wedge m_i b_i, \quad \underline{b}(u) = \bigvee_{j=1}^{d} \sum_{i:j\in N_i} M_i b_i \vee m_i b_i. \tag{5.2}$$

In the special case when $N_i = \{1,\cdots,d_X\}$ for all i, the formulation can be simplified

$$\bar{b}(u) = \sum_{i=1}^{n} M_i b_i \wedge m_i b_i, \quad \underline{b}(u) = \sum_{i=1}^{n} M_i b_i \vee m_i b_i.$$

Then we have the following lemma

Lemma 5.1. *When applying the generator to the approximated moment $\mathcal{E}_q(x)$ in Lemma 1.2, for any fixed $\delta > 0$, there is a constant C_δ,*

$$-(q\underline{b} + \delta|\underline{b}| + \delta)\mathcal{E}_q(x) - C_\delta(1 + |\underline{b}|) \le \mathcal{L}\mathcal{E}_q(x) \le -(q\bar{b} - \delta|\bar{b}| - \delta)\mathcal{E}_q(x) + C_\delta(1 + |\bar{b}|).$$

Consequentially, the role of $b(u)$ in the dyadic model (1.1) can be replaced by $\bar{b}$ and $\underline{b}$, when finding upper and lower bounds. So following the proof of Theorem 2.3 and 3.1, and 4.1, we have the following corollaries:

Corollary 5.2. *Assume u_t is asymptotically contractive, $\bar{b}$ is Lipschitz and $\langle \pi, \bar{b}\rangle > 0$, then*

$$\limsup_{t\to\infty} \mathbb{E}\|X_t\|^p < \infty \quad \text{for some } p > 0.$$

If further more $\bar{b} \geq 0$, then tail of $\|X_t\|$ is sub-exponential. If $\bar{b} \geq b_0 > 0$, then the tail of $\|X_t\|$ is sub-Gaussian.

Corollary 5.3. *Assume u_t is asymptotically contractive, $\underline{b}$ and h follow Assumption 2.1.*

1. *If $\underline{b}(u_*) < 0$ for some u_*, then for some $p > 0$, $\limsup_{t\to\infty} \mathbb{E}\|X_t\|^p = \infty$.*
2. *If $\underline{b}(u) = 0$ for u close to u_*, the energy centered at u_* is dissipative as in (3.2), $\underline{b} \leq D\|u - u_*\|^n$ for some D and n, then $\|X_t\|$ have a tail heavier than exponential.*

6 Conclusion

Extreme events are happening more often due to the global climate change, and they can induce heavy economic losses. The capability to analyze and predict them is crucial for our society. Mathematically, they appear as strong anomalies in time series and form heavy tails in the histograms. They are typically associated with stochastic instability caused by hidden unresolved processes. Such instability can be modeled by stochastic dampings in conditional Gaussian models. This has been justified by extensive numerical experiments, while there is little theoretical understanding. This can be problematic, since extreme events can be difficult to simulate.

This paper closes this gap by creating a theoretical framework, in which the tail density of conditional Gaussian models can be rigorously determined. Theorem 2.3 shows that if the stochastic damping takes negative values, the tail is polynomial. Theorem 3.1 shows that if the stochastic damping is nonnegative but takes value zero, the tail is between exponential and exponential. These results can be generalized to multivariate conditional Gaussian systems (1.2), as long as certain surrogate damping rates follow the conditions in Theorems 2.3 and 3.1. Moreover, we can apply the same framework to obtain a large deviation bound for long time averaging processes. This is shown in Corollary 2.8.

Acknowledgement

The authors Prof. Ramon van Handel and Nan Chen for the discussion of certain parts of this paper. This research of A. J. M. is partially supported by the Office of Naval Research through MURI N00014-16-1-2161 and DARPA through W911NF-15-1-0636. This research of X. T. T. is supported by the National University of Singapore grant R-146-000-226-133.

A Appendix

We allocate most the technical verifications in this appendix.

A.1 Some useful tools

Proof of Lemma 1.2. The upper bound $\mathcal{E}_p$ is trivial. The lower bound comes from Hölder's inequality $\|x\|^{p+2}+1 \geq \|x\|^p$, so

$$2\mathcal{E}_p - 1 = \frac{2\|x\|^{p+2}+\|x\|^2+1}{\|x\|^2+1} \geq \frac{\|x\|^{p+2}+\|x\|^p}{\|x\|^2+1} = \|x\|^p.$$

The gradient and Hessian of $\mathcal{E}_p$ can be computed directly:

$$\nabla_x \mathcal{E}_p = \frac{p\|x\|^p x}{1+\|x\|^2} + \frac{2\|x\|^p x}{(1+\|x\|^2)^2}.$$

$$\nabla_x^2 \mathcal{E}_p = \frac{p^2\|x\|^{p-2}xx^t}{1+\|x\|^2} + \frac{p\|x\|^p}{1+\|x\|^2} - \frac{2p\|x\|^p xx^t}{(1+\|x\|^2)^2} + \frac{2p\|x\|^{p-2}xx^t}{(1+\|x\|^2)^2} + \frac{2\|x\|^p}{(1+\|x\|^2)^2} - \frac{8\|x\|^p xx^t}{(1+\|x\|^2)^3}.$$

Since

$$\mathcal{L}\mathcal{E}_p(x) = \langle \nabla_x \mathcal{E}_p(x), b(u)x\rangle + \frac{1}{2}\mathrm{tr}(\nabla_x^2 \mathcal{E}_p(x)).$$

Note that

$$\langle \nabla_x \mathcal{E}_p(x), b(u)x\rangle = b(u)\left(\frac{p\|x\|^{p+2}}{1+\|x\|^2} + \frac{2\|x\|^{p+2}}{(1+\|x\|^2)^2}\right),$$

by Young's inequality, there is a constant C_δ

$$-(pb(u)+\tfrac{1}{2}\delta)\mathcal{E}_p(x) - \tfrac{1}{2}C_\delta|b(u)| \leq -\langle \nabla_x \mathcal{E}_p(x), b(u)x\rangle \leq -(pb(u)-\tfrac{1}{2}\delta)\mathcal{E}_p(x) + \tfrac{1}{2}C_\delta|b(u)|.$$

Lastly, by Young's inequality, we can further increase C_δ so that

$$|\mathrm{tr}(\nabla_x^2 \mathcal{E}_p(x))| = p^2 O((1+\|x\|)^{p-1}) \leq \delta\mathcal{E}_p(x) + C_\delta.$$

In combination, we have reached our claim. □

Lemma A.1. *Consider an OU process $u_t = -\Gamma u_t + dB_t$, $u_0 = u$, where the real parts of the eigenvalues of Γ are all strictly positive, then u_t is asymptotically contractive.*

Proof. Consider $v_t = -\Gamma v_t + dB_t$, and $v_0 = v$, where B_t is the same as the one in the SDE of u_t. Then the distribution of v_t is P_t^v. Yet

$$d(u_t - v_t) = -\Gamma(u_t - v_t)dt \quad \Rightarrow \quad \|u_t - v_t\| \leq \|\exp(-\Gamma t)\|\|u - v\|.$$

In other words $d(P_t^u, P_t^v) \leq \|\exp(-\Gamma t)\|\|u-v\|$. Because all eigenvalues of Γ have positive real parts, so u_t is asymptotically contractive. □

Lemma A.2. *Suppose u_t is asymptotically contractive. For any Lipschitz ψ with at most polynomial growth, then*

$$\theta(u) = \int_0^\infty (\mathbb{E}^u\psi(u_t) - \langle \pi, \psi\rangle)dt,$$

is well defined, while $\mathcal{L}\theta = \psi - \langle \pi, \psi\rangle$.

Proof. Let u_t' be an independent copy of the SDE $du_t' = h(u_t') + dB_t'$, while $u_0' \sim \pi$. Then $u_t' \sim \pi$ by invariance. So

$$\mathbb{E}^u\psi(u_t) - \langle \pi, \psi\rangle = \mathbb{E}\psi(u_t) - \psi(u_t').$$

By the asymptotic contractiveness, we have

$$|\mathbb{E}\psi(u_t) - \psi(u_t')| \le \|\psi\|_{Lip} d(P_t^u, P_t^\pi) \le C_\gamma e^{-\gamma t}\|\psi\|_{Lip} d(\delta_u, \pi).$$

Therefore ψ is well defined. Next, note that

$$\theta(u) = \int_0^\infty dt \int dz p_t^u(z)(\psi(z) - \langle \pi, \psi\rangle),$$

where $p_t^u(z)$ is the density of P_t^u. Apply Fubini's theorem, we have:

$$\mathcal{L}\theta(u) = \int dz(\psi(z) - \langle \pi, \psi\rangle) \int_0^\infty dt\mathcal{L}p_t^u(z).$$

Moreover by the Kolmogorov backward equation,

$$-\frac{\partial}{\partial t} p_t^u(z) = \mathcal{L}p_t^u(z).$$

Thus by $P_\infty^u = \pi$,

$$\mathcal{L}\theta(u) = \int dz(\psi(z) - \langle \pi, \psi\rangle)(p_0^u(z) - p_\infty^u(z)) = \psi(u) - \langle \pi, \psi\rangle.$$

□

A.2 Polynomial tails

Proof of Lemma 2.5. First we derive the upper bound. Consider the temporally inflated version of V, $\widetilde{V}(x, u, t) := e^{\rho t}V(x, u)$, then by Lemma 2.4

$$\mathcal{L}\widetilde{V}(x, u, t) = e^{\rho t}\mathcal{L}V(x, u) + \rho e^{\rho t}V(x, u) \le C_\delta(1 + |b(u)|)e^{\rho t} f(u).$$

Thus by Dynkin's formula:

$$\mathbb{E}\widetilde{V}(X_t, u_t, t) = \mathbb{E}\widetilde{V}(X_0, u_0, 0) + \mathbb{E}\int_0^t \mathcal{L}\widetilde{V}(X_s, u_s, s)ds \le \mathbb{E}V(X_0, u_0) + C_\delta \int_0^t e^{\rho s}\mathbb{E}(1 + |b(u_s)|)f(u_s)ds.$$

By our assumption on f,

$$\mathbb{E}f(u_t)\mathcal{E}_q(X_t) = \mathbb{E}V(X_t, u_t) = e^{-\rho t}\mathbb{E}\widetilde{V}(X_t, u_t, t),$$

is bounded by $C_\delta M_0/\rho$ when $t \to \infty$. In order to remove the $f(u_t)$ inside the expectation, we apply Hölder's inequality. For any $p < q$, when $t \to \infty$,

$$\mathbb{E}\|X_t\|^p \le \mathbb{E}[2\mathcal{E}_q(X_t)]^{\frac{p}{q}} \le 2\left[\mathbb{E}f(u_t)\mathcal{E}_q(X_t)\right]^{\frac{p}{q}}\left[\mathbb{E}f(u_t)^{\frac{-q}{q-p}}\right]^{\frac{q-p}{q}} \le \frac{2}{\rho}C_\delta M_0 M_{\frac{q-p}{q}}.$$

To prove the converse direction, it is almost identical. Let $\widetilde{U}(x, u, t) := e^{-\rho t}U(x, u)$, then

$$\mathcal{L}\widetilde{U}(x, u, t) \ge -C_\delta e^{-\rho t}g(u).$$

Thus by Dynkin's formula:

$$\mathbb{E}U(X_t, u_t) = e^{\rho t}\mathbb{E}\widetilde{U}(X_t, u_t, t) \ge e^{\rho t}\mathbb{E}U(X_0, t_0) - C_\delta\int_0^t e^{\rho(t-s)}\mathbb{E}g(u_s)(1 + |b(u_s)|)ds.$$

So if $\mathbb{E}U(X_0, t_0) > C_\delta M_0/\rho$, then $\mathbb{E}U(X_t, u_t) \to \infty$ as $t \to \infty$.

We can generalize this using the Markov property:

$$\begin{aligned}\mathbb{E}U(X_{t_0+t}, u_{t_0+t}) &= \mathbb{E}\mathbb{E}^{X_{t_0}, u_{t_0}}U(X_t, u_t) \\ &\ge \mathbb{E}1_{U(X_{t_0}, u_{t_0}) > k_v\rho+\delta}\mathbb{E}^{X_{t_0}, r_{t_0}}U(X_t, u_t)\end{aligned}$$

which goes to ∞ as $t \to \infty$ if $\mathbb{P}(U(X_{t_0}, u_{t_0}) > k_v\rho + \delta) > 0$. Yet when $\sigma_x > 0$, system 1.1 is controllable, so given any ϵ-ball $\mathcal{B}$ centered at any point (x', u'), $\mathbb{P}(X_{t_0}, u_{t_0} \in \mathcal{B}) > 0$ [?]. Then since $U = \mathcal{E}_q g$, so $\mathbb{P}(U(X_{t_0}, u_{t_0}) > k_v\rho + \delta) > 0$.

Lastly, we note that for $p > q$

$$\mathbb{E}U(X_t, u_t) = \mathbb{E}g(u_t)\mathcal{E}_q(X_t) \le \left[\mathbb{E}g(u_t)^{-\frac{p}{p-q}}\right]^{\frac{p-q}{p}}\left[\mathbb{E}\mathcal{E}_q(u_t)^{\frac{p}{q}}\right]^{\frac{q}{p}} \le 2M_{\frac{p-q}{p}}(\mathbb{E}\|X_t\|^p + 1).$$

This gives us $\mathbb{E}\|X_t\|^p \to \infty$. $\square$

Proof of Theorem 2.3. By Lemmas 2.6 and 2.7, there are functions g and f that satisfy (2.3) and (2.2), while $\mathcal{E}_q$ in Lemma 1.2 satisfies the conditions of Lemma 2.4. So Lemmas 2.4 and 2.5 apply, which provide us the claim about the moments.

We just need to show the ergodicity part. According to the arguments in [?], we only to construct a Lyapunov function for (X_t, u_t). Note that by Lipschitz condition and Lemma 2.7, both b and θ have at most linear growth. Therefore in Lemma 2.4, for a certain constant C_3

$$(1 + |b(u)|)f(u) \le C_3\exp(C_3\|u\|).$$

If $\langle h(u), u\rangle \le -\lambda\|u\|^2 + M_\lambda$ for some $\lambda, M_\lambda > 0$.

Let $\mathcal{E}(u) = \exp(\alpha\|u\|^2)$, with an $\alpha < \frac{1}{2}\lambda$. Apply the generator,

$$\mathcal{L}\mathcal{E} = 2\alpha\langle h(u), u\rangle\mathcal{E} + (\alpha + 2\alpha^2\|u\|^2)\mathcal{E} \le (-2\epsilon\|u\|^2 + 1 + 2M_\lambda)\alpha\mathcal{E}.$$

Here $\epsilon = \lambda - \frac{1}{2}\alpha > 0$. When $\epsilon\|u\|^2 \leq 2\sigma_x^2 + 4M_2$, $\mathcal{E}(x) \leq \exp(2\alpha(1+2M_\lambda)/\delta)$, otherwise $\mathcal{L}\mathcal{E}_t \leq -\delta\mathcal{E}_t$. Therefore for some constant M_4,

$$\mathcal{L}\mathcal{E} \leq -\epsilon\mathcal{E} + \alpha(1+2M_\lambda)\exp(2\alpha(1+2M_\lambda)/\delta) =: -\epsilon\mathcal{E} + M_4.$$

We can find another constant M_5, so that

$$(1+|b(u)|)f(u) \leq C_3\exp(C_3\|u\|) \leq \frac{1}{2}\epsilon M_5 \exp(\alpha\|u\|^2).$$

Let $\widetilde{V}(x,u) = f(u)\mathcal{E}_q(x) + M_5\mathcal{E}(u)$, then by Lemma 2.4,

$$\begin{aligned}\mathcal{L}\widetilde{V}(x,u) &\leq -\rho f(u)\mathcal{E}_q(x) + C_\delta(1+|b(u)|)f(u) - M_5\epsilon\mathcal{E} + M_4M_5\\ &\leq -\min\{\rho, \tfrac{1}{2}\epsilon\}\widetilde{V}(x,u) + M_4M_5.\end{aligned}$$

This qualifies $\widetilde{V}$ as a Lyapunov function for the process (X_t, u_t). □

A.3 Exponential tails

Proof of Proposition 3.2. Given a vector $\alpha \in \mathbb{R}^d$, let $\mathcal{E}_t = \exp\langle\alpha, X_t\rangle$, apply the chain rule (1.5), we find

$$\mathcal{L}\mathcal{E}_t = -\langle\alpha, X_t\rangle b(u_t)\mathcal{E}_t + \frac{1}{2}\mathcal{E}_t\|\sigma_x\alpha\|^2.$$

Consider the function $H(x) = \exp(x)(\|\alpha\| - x)$. Its derivative is $\dot{H}(x) = \exp(x)(\|\alpha\| - x - 1)$, so $H(x)$ reaches its maximum $\exp(\|\alpha\|-1)$ at $x = \|\alpha\| - 1$. Therefore

$$H(\langle\alpha, X_t\rangle) = \mathcal{E}_t(\|\alpha\| - \langle\alpha, X_t\rangle) \leq \exp(\|\alpha\| - 1),$$

and by $b(u_t) \geq 0$

$$-b(u_t)\langle\alpha, X_t\rangle\mathcal{E}_t \leq \exp(\|\alpha\|-1)b(u_t) - \|\alpha\|\mathcal{E}_t b(u_t).$$

So

$$\mathcal{L}\mathcal{E}_t \leq (\tfrac{1}{2}\|\sigma_x\alpha\|^2 - \|\alpha\|b(u_t))\mathcal{E}_t + \exp(\|\alpha\|-1)b(u_t).$$

Next, we apply the generator to $f = \exp(\|\alpha\|\theta)$, by the chain rule (1.5) is,

$$\mathcal{L}f = (\|\alpha\|(b - \langle\pi, b\rangle) + \tfrac{1}{2}\|\alpha\|^2\|\nabla_u\theta\|^2)f.$$

We apply Lemma 2.4 first part to $\mathcal{E}_t$ and $f = \exp(\|\alpha\|\theta)$,

$$\mathcal{L}\mathcal{E}_t f(u_t) \leq (\tfrac{1}{2}\|\alpha\|^2\|\nabla_u\theta\|^2 + \tfrac{1}{2}\sigma_x^2\|\alpha\|^2 - \|\alpha\|\langle\pi, b\rangle)\mathcal{E}_t f(u_t) + \exp(\|\alpha\|-1)b(u_t)f(u_t).$$

Since $\langle\pi, b\rangle > 0$, and $\|\nabla_u\theta\|$ is bounded, so if

$$\rho_\alpha = -(\tfrac{1}{2}\|\alpha\|^2\|\nabla_u\theta\|^2 + \tfrac{1}{2}\sigma_x^2\|\alpha\|^2 - \|\alpha\|\langle\pi, b\rangle) > 0.$$

Then by Gronwall's inequality,

$$\mathbb{E}\mathcal{E}_t\exp(\|\alpha\|\theta(u_t)) \leq e^{-\rho_\alpha t}\mathbb{E}\mathcal{E}_0\exp(\|\alpha\|\theta(u_0)) + \rho_\alpha^{-1}\int_0^t e^{-\rho_\alpha(t-s)}\mathbb{E}\exp(\|\alpha\|\theta(u_s))ds < \infty.$$

Finally by Hölder's inequality, for any $\rho < 1$

$$\mathbb{E}\exp(\rho\langle\alpha, X_t\rangle) \leq [\mathbb{E}\mathcal{E}_t \exp(\|\alpha\|\theta(u_t))]^\rho [\mathbb{E}\exp(-(1-\rho)^{-1}\|\alpha\|\theta(u_t))]^{1-\rho} < \infty.$$

To get our claim in the proposition, one simply let $\alpha = \rho^{-1}\beta$, with a proper $\rho < 1$ so that $\rho_\alpha > 0$.

For the last claim, note that if $\|\cdot\|_\infty$ denote the l_∞ norm, then

$$\mathbb{E}\exp(a\|X_t\|) \leq \mathbb{E}\exp(a\sqrt{d_X}\|X_t\|_\infty) \leq \sum_{i=1}^{d_X} \mathbb{E}\exp(a\sqrt{d_X}\langle e_i, X_t\rangle) + \exp(-a\sqrt{d_X}\langle e_i, X_t\rangle).$$

Here e_i is the i-th standard Euclidean basis vector, so $\langle e_i, X_t\rangle$ is the i-th component of X_t. So for sufficiently small α, $\limsup_{t\to\infty} \mathbb{E}\exp(a\|X_t\|) < \infty$.

Finally note that by Taylor expansion of $\exp(a\|x\|)$,

$$\|x\|^{2p} \leq (2p)! a^{-2p} \exp(a\|x\|).$$

So we have our claim since

$$\log \mathbb{E}\|X_t\|^{2p} \leq \sum_{k=1}^{2p} \log k - 2p\log a + \log \mathbb{E}\exp(a\|X_t\|) \leq 2p\log p + O(p).$$

□

Proof of Lemma 3.4. Consider the the following process

$$U_{p,t} = \exp\left(\int_0^t (-p^{\frac{1}{2^m}}\mathcal{L}g_m(u_s) - p\Gamma(g_1)(u_s))ds + G_p(u_t)\right).$$

Apply the generator, we find that

$$\begin{aligned}\mathcal{L}U_{p,t} &= (\mathcal{L}G_p + \Gamma(G_p) - p^{\frac{1}{2^m}}\mathcal{L}g_m - p\Gamma(g_1))U_{p,t} \\ &= \left(\sum_{k=1}^{m-1} p^{\frac{1}{2^k}}(\mathcal{L}g_k + \Gamma(g_{k+1})) + \sum_{k\neq j} p^{\frac{1}{2^j}+\frac{1}{2^k}}\Gamma(g_j, g_k)\right) U_{p,t} \geq 0.\end{aligned}$$

By Dynkin's formula, $U_{p,t}$ is a submartingale

$$\mathbb{E}U_{p,t} \geq \mathbb{E}\exp(G_p(u_0)).$$

Moreover, note that

$$\begin{aligned}\exp\left(-p\int_0^t b(u_s)ds + G_p(u_t)\right) &\geq \exp\left(-p\int_0^t \Gamma(g_1)(u_s)ds + G_p(u_t)\right) \\ &= U_{p,t}\exp(p^{\frac{1}{2^m}}\mathcal{L}g_m(u_s)ds) \geq \exp(-p^{\frac{1}{2^m}}Mt)U_{p,t}.\end{aligned}$$

And by Cauchy Schwartz

$$\mathbb{E}\exp\left(-2p\int_0^t b(u_s)ds\right)\mathbb{E}\exp(2G_p(u_t)) \geq \left(\mathbb{E}\exp\left(-p\int_0^t b(u_s)ds + G_p(u_t)\right)\right)^2.$$

As a consequence

$$\mathbb{E}\exp\left(-2p\int_0^t b(u_s)ds\right) \geq \frac{\exp(-2p^{\frac{1}{2m}}Mt)\left(\mathbb{E}\exp(G_p(u_0))\right)^2}{\mathbb{E}\exp(2G_p(u_t))}.$$

Then by Jensen's inequality, $\mathbb{E}\exp(G_p(u_0)) \geq \exp(\mathbb{E}G_p(u_0)) \geq \exp(-\sqrt{p}M_1)$, moreover $\mathbb{E}\exp(2G_p(u_t)) \leq \exp(2\sqrt{p}M_0)$. Therefore

$$\mathbb{E}\exp\left(-2p\int_0^t b(u_s)ds\right) \geq \exp(-2p^{\frac{1}{2m}}Mt - 2\sqrt{p}M_0 - 2\sqrt{p}M_1).$$

$\square$

Proof of Theorem 3.5. We will only look at integer p. For non-integer p, one can get similar bounds using Hölder's inequality. By the Duhamel's formula, we can write

$$X_t = A_{0,t}X_0 + \sigma_x\int_0^t A_{s,t}dW_s, \quad A_{s,t} := \exp\left(-\int_s^t b(u_r)dr\right).$$

Conditioned on the realization of u_s, $A_{0,t}X_0$ and $\int_0^t A_{s,t}dW_s$ are independent, so the conditional expectation of $\|X_t\|^{2p}$ will be larger than the conditional expectation of $\|\sigma_x\int_0^t A_{s,t}dW_s\|^{2p}$. So without lost of generality, we can assume $X_0 = 0$.

Next, conditioned on the realization of the u_s process, X_t has a Gaussian distribution with mean zero and variance $\int_0^t A_{s,t}^2 ds$. Therefore

$$\mathbb{E}\|X_t\|^{2p} = \mathbb{E}Z^{2p}\mathbb{E}\left(\int_0^t A_{s,t}^2 ds\right)^p,$$

where Z is $\mathcal{N}(0,1)$. Note that

$$\mathbb{E}\left(\int_0^t A_{s,t}^2 ds\right)^p = \int_{s_1,\dots,s_p=0}^t \mathbb{E}A_{s_1,t}^2\cdots A_{s_p,t}^2 ds_1\cdots ds_p,$$

and because $A_{r,t}^2 = A_{r,s}^2A_{s,t}^2 \leq A_{s,t}^2$ for any $r \leq s \leq t$, therefore

$$\mathbb{E}\left(\int_0^t A_{s,t}^2 ds\right)^p \geq \int_{s_1,\dots,s_p=0}^t \mathbb{E}A_{s_*,t}^{2p} ds_1\cdots ds_p,$$

where $s_* = \min\{s_1,\dots,s_p\}$. Applying Lemma 3.4 with a time shift of s, we find that

$$\mathbb{E}A_{s_*,t}^{2p} \geq \exp(-2p^{\frac{1}{2m}}M(t-s_*) - 2\sqrt{p}M_0 - 2\sqrt{p}M_1).$$

Notice the volume inside $\{(s_1,\dots,s_p) : s_i \leq t\}$ corresponds to $s_* \geq s$ is $(t-s)^p$, so by a change of variable,

$$\begin{aligned}\mathbb{E}\left(\int_0^t A_{s,t}^2 ds\right)^p &\geq \exp(-2\sqrt{p}M_0 - 2\sqrt{p}M_1)\int_0^t \exp(-2p^{\frac{1}{2m}}M(t-s))p(t-s)^{p-1}ds \\ &= p\exp(-2\sqrt{p}M_0 - 2\sqrt{p}M_1)\int_{s=0}^t \exp(-p^{\frac{1}{2m}}Ms)s^{p-1}ds.\end{aligned}$$

By Lemma A.3 in below,

$$\log \int_{s=0}^{\infty} \exp(-p^{\frac{1}{2^m}} M s) s^{p-1} ds = \left(1 - \frac{1}{2^m}\right) p \log p + O(p).$$

Using the inequality above, we find that

$$\log \mathbb{E}\|X_t\|^{2p} \geq \log \mathbb{E} Z^{2p} + \left(1 - \frac{1}{2^m}\right) p \log p + O(p) = \left(2 - \frac{1}{2^m}\right) p \log p + O(p).$$

The $\log \mathbb{E} Z^{2p} = p \log p + O(p)$ can be obtained by standard Gaussian moment formula or using Lemma A.3. □

Lemma A.3. *Fixed any $r > 0$, then with any sequence $c_p = O(p)$, the following holds*

$$\log \left(\int_0^{\infty} \exp(-c_p x^r) x^p dx \right) = \frac{p}{r} \log \frac{p}{c_p} + O(p).$$

Here a term is $O(p)$, if this term is bounded by $[M^{-1}p, Mp]$ for a constant M independent of p.

Proof. Let $y = c_p x^r$, we apply the change of variable, the integral can be written as

$$\int_0^{\infty} \exp(-c_p x^r) x^p dx = \frac{1}{r} c_p^{-\frac{p+1}{r}} \int_0^{\infty} \exp(-y) y^{\frac{p+1}{r} - 1} dy.$$

Let $q = \frac{p+1}{r} - 1$, and denote

$$M_q := \int_0^{\infty} \exp(-y) y^q dy.$$

Using integration by part, we find that

$$M_q = q \int_0^{\infty} \exp(-y) y^{q-1} dy = q M_{q-1} = \cdots = \left(\prod_{k=0}^{\lfloor q \rfloor - 1} (q - k) \right) M_{q - \lfloor q \rfloor}.$$

Since $q - \lfloor q \rfloor \in [0, 1)$, M_q is bounded by constants from both below and above. Moreover, note that

$$\log \left(\prod_{k=0}^{\lfloor q \rfloor - 1} (q - k) \right) = \sum_{k=0}^{\lfloor q \rfloor - 1} \log(q - k).$$

While clearly for $z \in [n, n+1]$ and $n \geq 1$,

$$\frac{1}{\sqrt{2}} \int_n^{n+1} \log u du \leq \log z \leq \sqrt{2} \int_n^{n+1} \log u du. \tag{A.1}$$

Combining these inequalities, and returning to the formulation of M_q, we can conclude that

$$\log M_q = \int_1^q \log u + O(q) = q \log q + O(q) = \frac{p}{r} \log p + O(p).$$

Then our claim holds as long as

$$\log \left(c_p^{-\frac{p+1}{r}} \right) = -\frac{p}{r} \log c_p + O(p).$$

Our assumption on c_p guarantees this. □

Proof of Lemma 3.6. Without loss of generality, we will assume u_* is the origin. We will use

$$g_1(u) = -M_1\|u\|^{2^m}, \quad M_1 := \frac{\sqrt{C}}{2^m}.$$

It is easy to see that

$$\nabla_u g_1 = -2^m M_1 \|u\|^{2^m-2} u, \quad \nabla_u^2 g_1 = -2^m M_1 \|u\|^{2^m-2} I_d - 2^m(2^m-2) M_1 \|u\|^{2^m-4} uu^T. \tag{A.2}$$

The level-1 constraint is met because

$$\Gamma(g_1) = 2^{m+1} M_1^2 \|u\|^{2^{m+1}-2} \geq b(u).$$

Next, we notice that,

$$\begin{aligned}\mathcal{L}g_1 &= -2^m M_1 \|u\|^{2^m-2}\langle u, h(u)\rangle + \frac{1}{2}\mathrm{tr}(\nabla_u^2 g_1) \\ &= -2^m M_1 \|u\|^{2^m-2}\langle u, h(u)\rangle - \frac{1}{2}2^m(2^m-1)M_1\|u\|^{2^m-2} \\ &\geq \lambda 2^m M_1 \|u\|^{2^m} - 2^{m-1}(2^m + 2M_\lambda) M_1 \|u\|^{2^m-2}. \end{aligned} \tag{A.3}$$

So if we let

$$C_1 = 2^{m-1}(2^m + 2M_\lambda)M_1, \quad g_2(u) = -M_2\|u\|^{2^{m-1}}, \quad M_2 := \sqrt{\frac{C_1}{2^m}}.$$

Then the gradient and Hessian of g_2 are similar to (A.2). In particular, it solves the level-2 constraint with g_1 since,

$$\Gamma(g_2) \geq 2^m M_2^2 \|u\|^{2^m-2} = C_1 \|u\|^{2^m-2}.$$

And a similar lower bound as (A.3) holds for $\mathcal{L}g_2$ with a $C_2 > 0$:

$$\mathcal{L}g_2 \geq \lambda 2^{m-1} M_1 \|u\|^{2^{m-1}} - C_2 \|u\|^{2^{m-1}-2}.$$

Clearly we can iterate this construction, and obtain a series of $g_k(u) = -M_k\|u\|^{2^{m+1-k}}, k = 1, \cdots, m$, wile

$$\mathcal{L}g_{k-1} + \Gamma(g_k) \geq 0.$$

One the other hand, we can verify that for $g_m(u) = -M_m\|u\|^2$,

$$\mathcal{L}g_m = -2M_m\langle u, h(u)\rangle - M_m \geq 2\lambda M_m \|u\|^2 - 2M_m M - M_m \geq -2M_m M - M_m.$$

Finally, we check conditions (5) and (6) in Definition 3.3. The first part of (5) holds as $G_p \leq 0$, and the second part holds since the power of p in G_p is at most $\frac{1}{2}$.

As for the alignment condition (6), note that

$$\nabla_u g_k = -2^{m+1-k} M_k \|u\|^{2^{m-1-k}} u,$$

so $\Gamma(g_j, g_k) \geq 0$. □

Proof of Theorem 3.1. By Lemma A.2, we know that θ is well defined and Lipschitz, so $\Gamma(\theta) = \|\nabla_u \theta\|^2$ is bounded. This indicates that $\exp(C\theta(u)) \leq \exp(CM\|u\| + CM)$ for a constant M. Then by Theorem 4.1, and the fact the energy is dissipative, we have $\mathbb{E}\exp(CM\|u_t\| + CM)$ is bounded uniformly in time. Therefore, we can use Proposition 3.2 to find the upper tail.

For the other direction, Corollary 3.7 indicates that $(b, h) \in \mathcal{A}_m$ for any m. So Proposition 3.5 indicates the lower bound. □

A.4 General conditional Gaussian models

Proof of Lemma 5.1. Simply note that

$$\begin{aligned}\langle B(u)x, x\rangle &= \frac{1}{2}\langle (B(u)+B(u)^T)x, x\rangle \\ &= \frac{1}{2}\sum_{i=1}^n \langle b_i(u)(B_i + B_i^T)x, x\rangle \\ &\le \frac{1}{2}\sum_{i=1}^n b_i M_i \vee b_i m_i \langle x I_{N_i}, x\rangle \\ &\le \sum_{i=1}^n b_i M_i \vee b_i m_i \sum_{j\in N_i} x_j^2 \le \underline{b}(u)\|x\|^2.\end{aligned}$$

Use the derivatives derived in Lemma 1.2 for $\mathcal{E}_q$, apply Young's inequality, we find that for any fixed $\delta > 0$, there is a C_δ

$$\begin{aligned}\mathcal{L}\mathcal{E}_q(x) &= -\left(\frac{q\|x\|^q}{1+\|x\|^2} + \frac{2\|x\|^q}{(1+\|x\|^2)^2}\right)\langle B(u)x, x\rangle + \frac{1}{2}\mathrm{tr}(\Sigma_X \nabla^2 \mathcal{E}_q(x)\Sigma_X^T) \\ &\ge -q\underline{b}(u)\mathcal{E}_q(x) - \underline{b}(u)O((1+\|x\|)^{q-2}) - O((1+\|x\|)^{q-1}) \\ &\ge -(q\underline{b} + \delta|\underline{b}| + \delta)\mathcal{E}_q(x) - C_\delta(1+|\underline{b}|).\end{aligned}$$

The converse direction comes in very similarly. $\square$

References

[1] Y Bakhtin and T Hurth. Invariant densities for dynamical systems with random switching. *Nonlinearity*, 25(10):2937–2952, 2012.

[2] Y Bakhtin, T Hurth, and J C Mattingly. Regularity of invariant densities for 1D systems with random switching. *Nonlinearity*, 28(11):3755–3787, 2015.

[3] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Grundlehren der mathematischen Wissenschaften. Springer, 2013.

[4] J B Bardet, H Guérin, and F Malrieu. Long time behavior of diffusions with markov switching. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7:151– 170, 2010.

[5] M. Branicki, B. Gershgorin, and A. J. Majda. Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters. *J. Comput. Phys.*, 231(4):1462–1498, 2012.

[6] B. Castaing, G Guanratne, F Heslot, L Kadanoff, A Libchaber, S Thomae, X Wu, S Zaleski, and G Zanetti. Scaling of hard thermal turbulence in rayleigh-bénard convection. *J. Fluid Mech.*, 204(1):1–30, 1989.

[7] L. Chen and R. C. Dalang. Moments and growth indices for the nonlinear stochastic heat equation with rough conditions. *Annals of Probability*, 2015.

[8] L. Chen, D. Khoshnevisan, and K. Kim. A boundedness trichotmoy for the stochastic heat equation. *Ann. Inst. H. Poincaré Probab. Statist.*, 53(4):1991–2004, 2017.

[9] N. Chen, D. Giannakis, R. Herbei, and A. J. Majda. An MCMC algorithm for parameter estimation in signals with hidden intermittent instability. *SIAM/ASA J. Uncertainty Quantification*, 2(1):647–669, 2014.

[10] N. Chen and A. J. Majda. Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions. *J. Comput. Phys.*, 354:242–268, 2018.

[11] N. Chen, A. J. Majda, and D. Giannakis. Predicting the cloud patterns of the MaddenJulian oscillation through a loworder nonlinear stochastic model. *Geophysical Res. Lett.*, 41(15):5612–5619, 2014.

[12] N. Chen, A. J. Majda, and X. T. Tong. Rigorous analysis for efficient statistically accurate algorithms for solving fokker-planck equations in large dimensions. arXiv:1709.05585.

[13] N Chen, A J Majda, and X T Tong. Information barriers for noisy lagrangian tracers in filtering random incompressible flows. *Nonlinearity*, 27:2133–2163, 2014.

[14] N. Chen, A. J. Majda, and X. T. Tong. Noisy lagrangian tracers for filtering random rotating compressible flows. *J. Non. Sci.*, 25(3):451–488, 2014.

[15] B Cloez and M Hairer. Exponential ergodicity for markov processes with random switching. *Bernoulli*, 21(1):505–536, 2015.

[16] D. R. Easterling, J. L. Evans, P. Y. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje. Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society*, 81(3):417–425, 2000.

[17] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.

[18] A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov. arXiv:1606.06012.

[19] B. Eichengreen, A. Mody, M. Nedeljkovic, and L. Sarno. How the subprime crisis went global: evidence from bank credit default swap spreads. *ournal of International Money and Finance*, 31(5):1299–1318, 2012.

[20] M. Farazmand and T. Sapsis. Reduced-order prediction of rogue waves in two dimensional water waves. *J. Comput. Phys.*, 340:418–434, 2017.

[21] B. Gershgorin, J. Harlim, and A. J. Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229:1–31, 2010.

[22] B Gershgorin and A J Majda. Filtering a statistically exactly solvable test model for turbulent tracers from partial observations. *J. Comput. Phys*, 230:1602–1638, 2011.

[23] J P Gollub, J Clarke, M Gharib, B Lane, and O N Mesquita. Fluctuations and transport in a stirred fluid with a mean gradient. *Phys. Rev. Lett.*, 67:3507–3510, 1991.

[24] Y. Gu and W. Xu. Moments of 2d parabolic Anderson model. https://arxiv.org/abs/1702.07026.

[25] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numerical Analysis*, 40(3):1041–1063, 2006.

[26] R. Huser and A. C. Davison. Space-time modelling of extreme events. *J. Royal Statistical Society: Series B*, 76(2):439–461, 2014.

[27] B Khouider, J A Biello, and A J Majda. A stochastic multicloud model for tropical convection. *Comm. Math. Sci.*, 8(1):187–216, 2010.

[28] B Khouider, A J Majda, and S Stechmann. Climate science in the tropics: Waves, vortices, and pdes. *Nonlinearity*, 26(R1-R68), 2013.

[29] Y. Lee, A. J. Majda, and D. Qi. Stochastic superparameterization and multiscale filtering of turbulent tracers. *SIAM Multiscale Model. Simul.*, 14(1), 2016.

[30] R S Liptser and A N Shiryaev. *Statistics of random processes. I, II,*, volume 5 of *Applications of Mathematics*. Springer-Verlag, 2001.

[31] A J Majda, C Franzke, and B Khouider. An applied mathematics perspective on stochastic modelling for climate. *Phil. Trans. Roy. Soc.*, 336(1875):2427–2453, 2008.

[32] A J Majda and B Gershgorin. Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum, and intermittency. *Phil. Trans. Roy. Soc.*, 371(1982), 2013.

[33] A J Majda and J Harlim. *Filtering complex turbulent systems*. Cambridge University Press, Cambridge, UK, 2012.

[34] A J Majda, J Harlim, and B Gershgorin. Mathematical strategies for filtering turbulent dynamical systems. *Discrete and Continuous Dynamical Systems*, 27:441–486, 2010.

[35] A J Majda and S Stechmann. The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci.*, 106(21):8417–8422, 2009.

[36] A J Majda and S Stechmann. The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci.*, 106:8417–8422, 2009.

[37] A. J. Majda and X. T. Tong. Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity*, 28(11), 2015.

[38] A. J. Majda and X. T. Tong. Moment bounds and geometric ergodicity of diffusions with random switching and unbounded transition rates. *Research in the mathematical sciences*, 3(41), 2016.

[39] A J Majda and X Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, Cambridge, UK, 2006.

[40] J Neelin, B Lintner, B Tian, Q Li, L Zhang, Chahine M Patra, P, and S Stechmann. Long tails in deep columns of natural and anthropogenic tropospheric tracers. *Geophysical Res. Lett.*, 37:L05804, 2010.

[41] J. Shao. Strong solutions and strong feller properties for regime-switching diffusion processes in an infinite state space. *SIAM J. Control Optim*, 53(4):2462–2479, 2015.

[42] R. S. Tsay. *Analysis of financial time series*. John Wiley & Sons, 2005.