Mean first exit times of Ornstein–Uhlenbeck processes in high-dimensional spaces

Hans Kersting
Inria, Ecole Normale Supérieure
PSL Research University, Paris, France.
hans.kersting@inria.fr

Antonio Orvieto
Department of Computer Science
ETH Zürich, Zürich, Switzerland.
antonio.orvieto@inf.ethz.ch

Frank Proske
Department of Mathematics
University of Oslo, Oslo, Norway.
proske@math.uio.no

Aurelien Lucchi
Dpt. of Mathematics & Computer Science
University of Basel, Basel, Switzerland.
aurelien.lucchi@unibas.ch

January 18, 2023

Abstract

The d-dimensional Ornstein–Uhlenbeck process (OUP) describes the trajectory of a particle in a d-dimensional, spherically symmetric, quadratic potential. The OUP is composed of a drift term weighted by a constant $\theta \geq 0$ and a diffusion coefficient weighted by $\sigma > 0$. In the absence of drift (i.e. $\theta = 0$), the OUP simply becomes a standard Brownian motion (BM). This paper is concerned with estimating the mean first-exit time (MFET) of the OUP from a ball of finite radius L for large $d \gg 0$. We prove that, asymptotically for $d \to \infty$, the OUP takes (on average) no longer to exit than BM. In other words, the mean-reverting drift of the OUP (scaled by $\theta \geq 0$) has asymptotically no effect on its MFET. This finding might be surprising because, for small $d \in \mathbb{N}$, the OUP exit time is significantly larger than BM by a margin that depends on θ. As it allows for the drift to be ignored, it might simplify the analysis of high-dimensional exit-time problems in numerous areas.

Finally, our short proof for the non-asymptotic MFET of OUP, using the Andronov–Vitt–Pontryagin formula, might be of independent interest.

1 Introduction

The d-dimensional Ornstein–Uhlenbeck process (OUP) is one of the most important stochastic processes. It can be defined as a d-dimensional Brownian motion (aka. Wiener process) where a drift is added which is proportional to the displacement from its mean. It is usually defined as the solution of the following SDE:

$$\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \sigma \mathrm{d}B_t, \tag{1}$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$ are the drift and diffusion parameters and B_t is a d-dimensional Brownian motion (BM). Alternatively, if $X_0 = 0$, it can be defined as the unique d-dimensional zero-mean Gaussian process with the covariance function

$$\mathrm{cov}(X_{i,s}, X_{j,t}) = \delta_{ij} \cdot \frac{\sigma^2}{2\theta} \cdot \left(e^{-\theta|t-s|} - e^{-\theta(t+s)} \right), \tag{2}$$

where δ_{ij} is the Kronecker delta and $X_{j,t}$ are the independent components of $X_t = (X_{1,t}, \ldots, X_{d,t})$.

The OUP has been extensively studied (Borodin and Salminen, 2002, Part II, Section 7), sometimes with the parameter $\lambda = \frac{\theta}{\sigma^2} \in \mathbb{R}$ instead of θ. Many authors define the OUP only for $\theta > 0$, while others allow for $\theta \in \mathbb{R}$. If $\theta \in \mathbb{R}$, one usually distinguishes between the positively recurrent case ($\theta > 0$), the Brownian motion case ($\theta = 0$), and the transient case ($\theta < 0$); see p. 137 in Borodin and Salminen (2002). Here, we will allow for $\theta \in \mathbb{R}$ and only restrict this range if mathematically necessary.

In this paper, we are interested in the mean *first exit time* (aka. first passage time) of a d-dimensional OUP from a ball of radius L, $B_L(0) = \{x \in \mathbb{R}^d : \|x\|_2 < L\}$. More specifically, we study the expectation $\mathbb{E}^x\tau_L$ of the stopping time $\tau_L = \inf_{t\geq 0}\{\|X_t\|_2 = L\}$, where the x in $\mathbb{E}^x\tau_L$ indicates that $X_0 \in \mathbb{R}^d$ is chosen such that a.s. $\|X_0\| = x$ for some $x \in [0, L]$. In particular, we want to understand the asymptotics of $\mathbb{E}^x\tau_L$ *in high dimensions*, i.e. as $d \to \infty$.

Such high-dimensional OUPs are important models in numerous applications ares. For instance in biophysical sciences, high-dimensional OUPs are used to model the membrane potential of multiple neurons (Ricciardi and Sacerdote, 1979; Faugeras et al., 2004), the phylogenetic dynamics of quantitative traits (Butler and King, 2004; Rohlfs et al., 2014), and single-cell differentiation for multiple cells (Matsumoto and Kiryu, 2016). In complex systems, many neurons, traits, or cells have to be modeled – leading to high-dimensional OUP models, whose exit times mark when a specific average variation over all components is reached.

High-dimensional OUPs are also important in other scientific fields. Grebenkov (2014, Section 3.4.) describes how MFETs matter in algorithmic trading, where the dimensionality depends on the number of different assets in a portfolio (which can get very large, e.g. for exchange-traded funds). In economics, such models can be used to model the wealth of trading agents in an economy (Ciołek et al., 2020).

In non-convex optimization, high-dimensional Langevin equations are used as continuous-time models for stochastic gradient descent methods; these models are Ornstein–Uhlenbeck processes in local minima (Li et al., 2017) whose dimensionality can go up to $d = 10^7$ in modern machine-learning problems. The first exit time of high-dimensional OUPs hence describe the time of escape from a spurious local minimum (Nguyen et al., 2019; Lucchi et al., 2022).

MFETs of the OUP have of course already been studied to some extent in the literature. We refer the reader to the review paper by Grebenkov (2014) for an overview and detailed literature review. Grebenkov proves a formula (see Eq. (75) in his paper) for the MFET of a d-dimensional OUP for all $d \in \mathbb{N}$. Grebenkov's proof makes use of the Fokker-Planck equation and the eigenfunctions of the corresponding backward Laplace operator. Previously, for the special case of $\sigma = 1$ and $\theta \geq 0$, Graczyk and Jakubowski (2008) proved the same formula in their Theorem 2.2 by deriving and solving a suitable boundary value problem (BVP); see their Theorem 2.1. In this paper, we will re-prove and extend the formulas by Grebenkov (2014) and Graczyk and Jakubowski (2008). We emphasize that our proof of Theorem 4 uses a strategy similar to Graczyk and Jakubowski (2008); we refer the reader to our discussion in Remark 5 for a comparison.

While the previous formulas are satisfactory for finite $d \in \mathbb{N}$, our reformulation in terms of the incomplete Gamma function enables us to get the asymptotic scaling for $d \to \infty$. We will find that, as $d \to \infty$, the MFET of OUP is asymptotically equivalent to that of BM. The initial motivation for this paper is illustrated Fig. 1: as d increases the first exit times of both BM and OUP converge to zero at a rate that does not seem to differ between BM and OUP. Our theroetical results and experiments (Fig. 2) will confirm this.

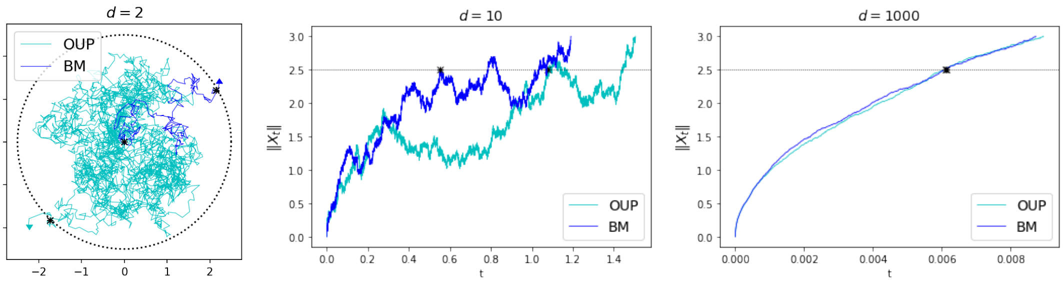

Figure 1: *Motivation of this paper*: the scaling in d of the first-exit time of OUP ($\theta > 0$) and BM ($\theta = 0$) from a ball of radius L. In each of the three plots, a d-dim. OUP and BM are run until they exit a ball of radius $L = 2.5$ (visualized by the dotted black line). The point of the first exits is marked by black stars. The three plots are for $d = 2$, $d = 10$ and $d = 1000$ respectively, with parameters $\sigma = 1.0$ and $\theta = 0.7$. On the left ($d = 2$), the entire trajectory is plotted in the 2-d plane, while in the other plots only $\|X_t\|$ is plotted against time. We can see for $d = 2$ how the trajectory of the OUP (i.e. the first-exit time) is much longer than of BM, due to the mean-reverting drift of the OUP. For $d = 10$, this effect is already less pronounced, but still significant. For $d = 1000$, the first exit times are almost the same. (For even larger d, the two lines becomes indistinguishable.) This highlights what we will show in Corollary 8, namely that (on average, for large d) OUP takes asymptotically no longer to exit than BM. Note that we can already observe here that the exit times go to zero as d grows, but it is a-priori not obvious how θ impacts the asymptotic rate (see Section 5 for a discussion). Also, see Fig. 2, where the mean asymptotics are depicted for the same parameters of L, σ, and θ.

Contributions In this paper, we

(i) reprove the known general formula for the MFET of a d-dimensional OUP (Theorem 4) by a shorter proof using the Andronov–Vitt–Pontryagin formula (see Remark 5 for the relation to prior work) and express this formula in terms of the incomplete Gamma function,

(ii) prove asymptotically sharp bounds for the MFET (Theorem 7), as $d \to \infty$, by relying on inequalities on the incomplete Gamma function from Neuman (2013), and

(iii) demonstrate that (perhaps surprisingly) the MFET of OUP is asymptotically equal to the one of Brownian motion (Corollary 8).

Note that, while point (iii) might be initially surprising, it is much less surprising after careful examination; see our Discussion in Section 5.

1.1 Structure of paper

The remainder of the paper is structured as follows. First, in Section 2, we explain how the MFET of the d-dimensional OUP can be considered as a MFET of the radial Ornstein–Uhlenbeck (rOUP) process. Second, in Section 3, we introduce the Andronov–Vitt–Pontryagin (AVP) formula and its corresponding boundary value problem for the autonomous case. Third, in Section 4, we use both the rOUP and AVP formula to derive an explicit expression for the MFET of OUP and reformulate it in terms of the incomplete Gamma function. Fourth, in Section 4.1, we prove asymptotically sharp bounds for MFET (as $d \to \infty$) by exploiting inequalities from Neuman

(2013) on the incomplete Gamma function. Fifth, in Corollary 8, we show that these bounds imply that, as $d \to \infty$, the OUP does (on average) take no longer than Brownian motion to exit balls of arbitrary radius. Sixth, in Section 5, we discuss the intuitive reasons and implications of our results. Finally, we summarize and conclude in Section 6.

2 The radial Ornstein–Uhlenbeck process

Let X_t denote a d-dimensional OUP process, as defined in Eq. (1). We are concerned with the first exit time

$$\tau_L := \inf_{t \geq 0} \{ \underbrace{\|X_t\|_2}_{=:\rho_t} = L \}. \tag{3}$$

We can see that this stopping time only depends on X_t via $\rho_t := \|X_t\|_2 = (\sum_{i=1}^d X_{i,t}^2)^{1/2}$. The stochastic process ρ_t is called the *radial Ornstein–Uhlenbeck process (rOUP)*. It (as well as its square) can be represented by the following Itô diffusions (Borodin and Salminen, 2002, Appendices 1.25 and 1.26).

Lemma 1. *Let $d \in \mathbb{N}$, $\theta \in \mathbb{R}$ and $\sigma > 0$. Consider the d-dimensional OUP X_t from Eq. (1). Then, $\rho_t^2 = \sum_{i=1}^d X_{i,t}^2$ and $\rho_t = (\sum_{i=1}^d X_{i,t}^2)^{1/2}$ follow the SDEs*

$$\mathrm{d}\rho_t^2 = (\sigma^2 d - 2\theta\rho_t^2)\,\mathrm{d}t + 2\sigma\sqrt{\rho_t^2}\,\mathrm{d}B_t \tag{4}$$

and

$$\mathrm{d}\rho_t = \left[\frac{(d-1)\sigma^2}{2\rho_t} - \theta\rho_t \right] \mathrm{d}t + \sigma\,\mathrm{d}B_t, \tag{5}$$

respectively. (Nb: Here B_t is not the original BM, but another one (see proof). We however use the same symbol to declutter the notation.)

Proof. First, note that the components of $X_t = [X_{1,t}, \dots, X_{d,t}]$ are independent OUPs of the form

$$\mathrm{d}X_{i,t} = -\theta X_{i,t} + \sigma\,\mathrm{d}B_{i,t}. \tag{6}$$

Then, by $\rho_t^2 = \sum_{i=1}^d X_{i,t}^2$, application of Itô's lemma yields

$$\begin{aligned} \mathrm{d}\rho_t^2 &= 2X_t^\intercal\,\mathrm{d}X_t + \frac{1}{2}\operatorname{tr}(2I_d\,\overbrace{[dX_t dX_t^\intercal]}^{=\sigma^2 I_d \mathrm{d}t}) & (7)\\ &= 2X_t^\intercal \mathrm{d}X_t + \sigma^2 d\,\mathrm{d}t. & (8) \end{aligned}$$

The remaining term to compute is

$$X_t^\intercal \mathrm{d}X_t \overset{(6)}{=} -\theta\rho_t^2\,\mathrm{d}t + \sigma\sum_{i=1}^d X_{t,i}\,\mathrm{d}B_t^{(i)} = -\theta\rho_t^2\,\mathrm{d}t + \sigma\sqrt{\rho_t^2}\,\mathrm{d}Y_t, \tag{9}$$

where

$$Y_t := \sum_{i=1}^d \int_0^t \frac{X_{i,s}}{\sqrt{\rho_t^2}}\,\mathrm{d}B_s^{(i)}. \tag{10}$$

But Y_t is a continuous martingale with quadratic variation $[Y,Y]_t = t$. This means, by Lévy's characterization of BM, that Y_t is another Brownian motion. To declutter notation, we denote Y_t by B_t, too. Now, insertion of Eq. (9) into Eq. (8) yields Eq. (4).

For the missing Eq. (5), we observe that $\rho_t = \sqrt{\rho_t^2}$. Then, application of Itô's lemma and insertion of Eq. (4) gives Eq. (5). □

Now, we have derived the SDE for the rOUP in Eq. (5). To exploit it, we will need to solve its Andronov–Vitt–Pontryagin formula which we will introduce next.

3 The Andronov–Vitt–Pontryagin formula

The Andronov–Vitt–Pontryagin (AVP) formula is a formula to compute the expected exit times of an Itô diffusion by use of its Kolmogorov backward operator.

Consider the Itô SDE

$$\mathrm{d}X_t = a(X_t,t)\,\mathrm{d}t + b(X_t,t)\,\mathrm{d}B_t, \tag{11}$$

where $a : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, $b : \mathbb{R}^d \times [0,T] \to \mathbb{R}^{d\times n}$, and B_t is an n-dimensional Brownian motion. The associated *backward Kolmogorov operator* is defined as

$$L_x^* u(x,t) = \sum_{i=1}^n a_i(t)\frac{\partial u(x,t)}{\partial x^i} + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n [b(X_t,t)b(X_t,t)^\intercal]_{ij}\frac{\partial^2 u(x,t)}{\partial x^i \partial x^j}. \tag{12}$$

For any bounded domain $D \subset \mathbb{R}^d$, we define the first exit time from this domain:

$$\pi_D := \inf\{t > 0 \mid x(t) \notin D\}. \tag{13}$$

Note that, if D is equal to a centered ball $B_L(0)$ of radius L, then $\pi_D = \tau_L$. The next theorem characterizes the expected value of π_D.

Theorem 2 (The Andronov–Vitt–Pontryagin formula)**.** *Let $D \subset \mathbb{R}^d$ be a bounded set, and assume that the following boundary value problem*

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} + L_x^* u(x,t) &= -1, \quad \text{for } x \in D \text{ and } t \in \mathbb{R}, \\ u(x,t) &= 0, \ \text{for } x \in \partial D, \end{aligned} \tag{14}$$

has a unique bounded solution $u(x,t)$. Then, the mean first passage time $\mathbb{E}^x \pi_D$ is finite and takes the form

$$\mathbb{E}^x \pi_D = u(x,0). \tag{15}$$

Proof. See Theorem 4.4.3. in Schuss (2010). □

Fortunately, in the autonomous case, the PDE (14) turns into an ODE which can be solved analytically in many cases.

Corollary 3 (The Andronov–Vitt–Pontryagin boundary value problem for the autonomous case)**.** *Under the assumptions of Theorem 2, if the coefficients a and b from Eq. (11) are independent of t, the solution u of Eq. (14) is also independent of t and therefore solves the ODE*

$$\begin{aligned} L_x^* u(x) &= -1, \quad \text{for } x \in D. \\ u(x) &= 0, \ \text{for } x \in \partial D, \end{aligned} \tag{16}$$

In particular, $\mathbb{E}^x \pi_D = u(x)$.

Proof. See Corollary 4.4.1. in Schuss (2010). □

We can now apply the above Corollary to the SDE (5) of the radial OUP. We will be able to solve the resulting ODE (16) and thereby obtain an exact formula for $\mathbb{E}^x\tau_L$.

4 Mean first exit time of the d-dimensional OUP

Theorem 4. *Let $L > 0$ and let $x \in [0, L]$ such that a.s. $\|X_0\| = x$. Denote $\lambda := \frac{\theta}{\sigma^2}$. Then, for all $\sigma > 0$ and $\lambda \in \mathbb{R}$, we have that*

$$\mathbb{E}^x\tau_L = \frac{2}{\sigma^2}\int_x^L z^{1-d}e^{\lambda z^2}\left[\int_0^z t^{d-1}e^{-\lambda t^2}\,\mathrm{d}t\right]\mathrm{d}z. \tag{17}$$

If $\lambda > 0$, then this formula further simplifies to

$$\mathbb{E}^x\tau_L = \frac{1}{\lambda^{\frac{d}{2}}\sigma^2}\int_x^L z^{1-d}e^{\lambda z^2}\gamma(d/2, \lambda z^2)\,\mathrm{d}z, \tag{18}$$

where the function γ denotes the upper incomplete Gamma function

$$\gamma(n, y) := \int_0^y t^{n-1}\exp(-t)\ \mathrm{d}t. \tag{19}$$

Remark 5 (Relation to prior work)**.** *Theorem 4 gives an exact representation of the desired MFET of the OUP for all $d \in \mathbb{N}$. Note that Theorem 4 is already* partially known. *While both the physics and mathematics literature have independently derived variations of Eq. (17), our reformulation with the incomplete Gamma function (18) is new. Importantly, this will enable us to derive novel lower and upper bounds on the MFET (see Theorem 7).*

In physics, Grebenkov (2014) derives our formula (17) in his Eq. (75), but in a different parametrization and with a different proof using the eigenfunctions of the backward Laplace operator.

In mathematics, Graczyk and Jakubowski (2008), in their Theorem 2.2, also prove our formula (17) with a similar strategy, but only for $\sigma = 1$ and $\theta \geq 0$ (called λ in their notation). (Note that they use hypergeometric functions to express the integral in Eq. (17).) Their proof strategy is similar to ours: our Corollary 3 corresponds to their Theorem 2.1, which they derive by different means than Schuss (2010) used for our Corollary 3. Accordingly, our BVP (20) appears on page 320 of Graczyk and Jakubowski (2008). Hence, one can think of our Eq. (17) as an extension of Theorem 2.1 in Graczyk and Jakubowski (2008) to all $\sigma > 0$ and all $\theta \in \mathbb{R}$, and of our proof as a shorter version of theirs.

While our Eq. (18) is easily derived from Eq. (17) by a change of variable (see proof), it will be essential to prove the scaling for $d \to \infty$ below. Eq. (18) is thus an essential part of our analysis.

Proof of Theorem 4. We first prove Eq. (17) and then Eq. (18). The proof of Eq. (17) is split into three parts. First, we fix $x > 0$ and show that the right-hand side of Eq. (17) solves the Andronov–Vitt–Pontryagin BVP, Eq. (16), associated with the radial Ornstein–Uhlenbeck process ρ_t, with $\rho_0 = x$. Second, we show that the solution to this BVP is unique. (Together, the first two steps imply that Eq. (17) holds for all $x > 0$.) Third, we will show that it also holds for $x = 0$ which will conclude the proof.

We now go through the proof step by step.

First step: By the definition of the radial OU process $\rho_t = (\sum_{i=1}^d X_{i,t}^2)^{1/2}$, we have $\tau_L = \inf\{t > 0 : \rho_t = L\}$. Hence, by the SDE (5) of ρ_t, the associated Andronov–Vitt–Pontryagin BVP (16) reads

$$\left[\frac{(d-1)\sigma^2}{2x} - \theta x\right] u'(x) + \frac{\sigma^2}{2} u''(x) = -1, \qquad \text{with } u(L) = 0,$$

or equivalently

$$u''(x) = \left[2\lambda x - \frac{d-1}{x}\right] u'(x) - \frac{2}{\sigma^2}, \qquad \text{with } u(L) = 0. \tag{20}$$

(Note that we here used that $\pi_{B_L(0)} = \tau_L$ by construction in Eq. (13), so that we could apply the Andronov–Vitt–Pontryagin formula.) We will now show that the right-hand side of Eq. (17) solves the above equation, that is

$$u(x) := \frac{2}{\sigma^2} \int_x^L z^{1-d} e^{\lambda z^2} \left[\int_0^z t^{d-1} e^{-\lambda} \, \mathrm{d}t\right] \mathrm{d}z. \tag{21}$$

Its derivatives are

$$u'(x) = -\frac{2}{\sigma^2} x^{1-d} e^{\lambda x^2} I(x), \tag{22}$$

$$u''(x) = -\frac{2}{\sigma^2} \left[(1-d) x^{-d} e^{\lambda x^2} + 2\lambda x^{2-d} e^{\lambda x^2}\right] I(x) - \frac{2}{\sigma^2}, \tag{23}$$

where

$$I(x) := \left[\int_0^x t^{d-1} e^{-\lambda t^2} \, \mathrm{d}t\right]. \tag{24}$$

Insertion of these formulas yields Eq. (20):

$$\left[2\lambda x - \frac{d-1}{x}\right] u'(x) - \frac{2}{\sigma^2} = \left[2\lambda x - \frac{d-1}{x}\right] \left[-\frac{2}{\sigma^2} x^{1-d} e^{\lambda x^2}\right] I(x) - \frac{2}{\sigma^2} \tag{25}$$

$$= -\frac{2}{\sigma^2} \left[(1-d) x^{-d} e^{\lambda x^2} + 2\lambda x^{2-d} e^{\lambda x^2}\right] I(x) - \frac{2}{\sigma^2} = u''(x). \tag{26}$$

Second step: To prove that u from Eq. (21) is the *unique* solution to the second-order ODE (20) for any $x > 0$, we recast it as a first-order ODE

$$\frac{\mathrm{d}}{\mathrm{d}x} \begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix} = \begin{bmatrix} u_2(x) \\ \left[2\lambda x - \frac{d-1}{x}\right] u_1(x) - \frac{2}{\sigma^2} \end{bmatrix} =: f\left(x, \begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix}\right). \tag{27}$$

Now, on the domain $\tilde{x} \in [x, L]$, the vector field $f(\cdot, [u_1, u_2]^\intercal)$, has Lipschitz uniformly bounded Lipschitz constants:

$$\begin{aligned} L(\tilde{x}) :&= \sup_{u \neq v \in \mathbb{R}^{2d}} \frac{\|f(\tilde{x}, u) - f(\tilde{x}, v)\|}{\|u - v\|} = \min\{1, |2\lambda\tilde{x} - (d-1)/x|\} \leq |2\lambda\tilde{x} - (d-1)/x| \\ &\leq 2\lambda L + (d-1)/x < \infty, \qquad \forall \tilde{x} \in [x, L]. \end{aligned} \tag{28}$$

Hence, the following global Lipschitz condition required for a global-Lipschitz version of Picard–Lindelöf's theorem (Teschl, 2012, Corollary 2.6.) is satisfied for any choice of $x > 0$:

$$\int_x^L L(\tilde{x}) \, \mathrm{d}\tilde{x} \leq (L - x) \cdot (2\lambda L + (d-1)/x) < \infty \tag{29}$$

Thus, this version of Picard–Lindelöf's theorem yields that $u(x)$ is the unique solution of Eq. (20), for any fixed choice of $x > 0$.

Third step: To provide the missing case $x = 0$, we first observe that $\lim_{l\to 0} \mathbb{E}^0 \tau_l = 0$ – because otherwise the radial Ornstein–Uhlenbeck ρ_t would stay at the origin for a positive time with positive probability, which it does not. Moreover, by the strong Markov property of X_t, we have for all $l \in (0, L]$ that

$$\mathbb{E}^0 \tau_L = \mathbb{E}^0 \tau_l + \underbrace{\mathbb{E}^l \tau_L}_{=u(l)} . \tag{30}$$

Hence, in the limit $l \to 0$, we indeed obtain $\mathbb{E}^0 \tau_L = u(0)$ which concludes the proof of Eq. (17). Note that all above steps worked for all $\lambda = \frac{\theta}{\sigma^2} \in \mathbb{R}$, and thus Eq. (17) holds for all $\lambda \in \mathbb{R}$.

To prove the missing Eq. (18) for $\lambda > 0$, we observe by a change of variable, using the substitution function $\phi(t) := \sqrt{\frac{t}{\lambda}}$ with derivative $\phi'(t) = \sqrt{4\lambda t}$, that

$$\begin{aligned} \int_0^{z=\phi(\lambda z^2)} t^{d-1} \exp(-\lambda t^2) \mathrm{d}t &= \int_0^{\lambda z^2} \sqrt{4\lambda t} \left(\frac{t}{\lambda}\right)^{\frac{d-1}{2}} \exp(-t) \mathrm{d}t \cdot \\ &= \frac{1}{2} \lambda^{-d/2} \underbrace{\int_0^{\lambda z^2} t^{\frac{d}{2}-1} \exp(-t)\, \mathrm{d}t}_{=\gamma\left(\frac{d}{2}, \lambda z^2\right)} . \end{aligned} \tag{31}$$

Insertion of the above formula into Eq. (17) yields Eq. (18). □

4.1 Scaling for $d \to \infty$

Equipped with our formula (18), we will derive the scaling for dimension $d \to \infty$ by use of the following existing bounds on the incomplete Gamma function.

Lemma 6 (Theorem 4.1. by Neuman (2013)). *For the lower incomplete Gamma function γ from Eq. (19), the inequalities*

$$\frac{x^a}{a} \exp\left(\frac{-ax}{a+1}\right) \le \gamma(a, x) \le \frac{x^a}{a(a+1)}(1 + a \exp(-x)) \tag{32}$$

are valid for all $a > 0$.

Now, we use the above result, to show the the scaling of $\mathbb{E}^x \tau_L$ in d. The result of the following theorem is verified numerically in Fig. 2.

Theorem 7. *Let $L > 0$ and let $x \in [0, L]$ such that a.s. $\|X_0\| = x$. Denote $\lambda := \frac{\theta}{\sigma^2}$. Let $\theta > 0$ (or equivalently $\lambda = \frac{\theta}{\sigma^2} > 0$). Then, we have the following upper bounds*

$$\mathbb{E}^x \tau_L \le \frac{2}{\sigma^2 d(d+2)} \left(\frac{1}{\lambda} \left[\exp(\lambda L^2) - \exp(\lambda x^2)\right] + \frac{d}{2} \left[L^2 - x^2\right] \right) \tag{33}$$

$$\le \frac{1}{\theta} \left[\exp(\lambda L^2) - \exp(\lambda x^2)\right] d^{-1}, \tag{34}$$

and lower bounds

$$\mathbb{E}^x \tau_L \ge \frac{1 + \frac{2}{d}}{2\lambda\sigma^2} \left[\exp\left(\frac{2\lambda}{d+2} L^2\right) - \exp\left(\frac{2\lambda}{d+2} x^2\right) \right] \tag{35}$$

$$\ge \left[\frac{L^2 - x^2}{\sigma^2}\right] d^{-1}. \tag{36}$$

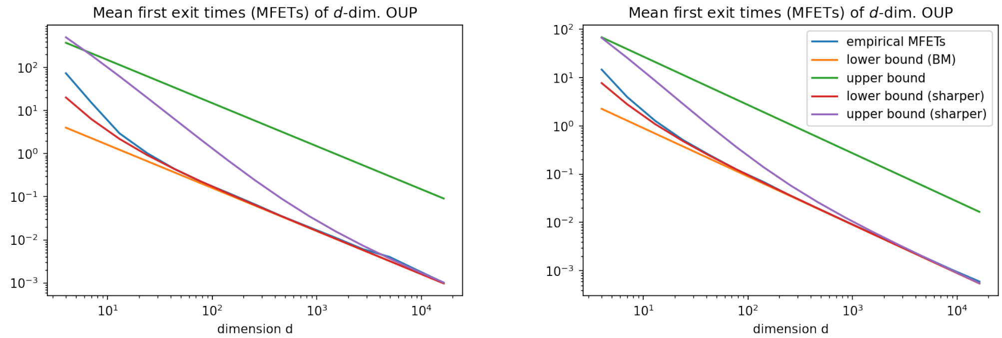

Figure 2: Log-log plot of the bounds of Theorem 7. Parameters are set as $(L, x, \sigma, \lambda) = (4.0, 0.0, 1.0, 0.5)$ (on the left plot) and as $(L, x, \sigma, \lambda) = (3.0, 0.0, 1.0, 0.7)$ (on the right plot); but other parameters give the same behavior. The BM lower bound is the smaller lower bound that is independent of λ (which is the exact MFET of BM). To compute the MFETs, 100 simulations with a step size of 0.001 (on the left) and of 0.0001 (on the right) were run for each choice of d. In both plots, the step size was chosen to reproduce the upper bounds (otherwise the point of exit can be missed by the discrete time steps); smaller choices also work. The bounds can be computed in closed form with our Eqs. (33)–(36). The bounds from Theorem 7 are verified by this plot. Remarkably, as we prove in Corollary 8, the asymptotics of OUP ($\theta > 0$) are equivalent to the BM lower bound ($\theta = 0$).

Proof. We first show the upper bounds and then the lower bounds.

Upper bounds: By the second inequality from Lemma 6, we have

$$\gamma(d/2, \lambda z^2) \le \frac{4\lambda^{d/2} z^d}{d(d+2)} \cdot \left[1 + \frac{d}{2}\exp(-\lambda z^2)\right] \tag{37}$$

$$\le \frac{4\lambda^{d/2} z^d}{d(d+2)} \cdot \frac{2+d}{2} = \frac{2\lambda^{d/2} z^d}{d}. \tag{38}$$

Insertion of the inequality (37) into Eq. (18) yields

$$\mathbb{E}^x \tau_L \le \frac{4}{\sigma^2 d(d+2)} \cdot \underbrace{\int_x^L z\exp\left(\lambda z^2\right)\left[1 + \exp(\lambda z^2)\right] \mathrm{d}z}_{=\int_x^L z\exp(\lambda z^2)\mathrm{d}z + \frac{d}{2}\int_x^L z\mathrm{d}z \;=\; \left[\frac{1}{2\lambda}\exp(\lambda z^2)\right]_{z=x}^{z=L} + \frac{d}{4}[L^2 - x^2]}, \tag{39}$$

which is Eq. (33). For the missing upper bound (34), we insert the less sharp inequality (38) into Eq. (18):

$$\mathbb{E}^x \tau_L \le \frac{2}{\sigma^2 d}\int_x^L z\exp\left(\lambda z^2\right)\mathrm{d}z = \frac{1}{\theta}\left[\exp(\lambda L^2) - \exp(\lambda x^2)\right] d^{-1}, \tag{40}$$

where we used that $\lambda = \theta/\sigma^2$. Since the inserted upper bound from inequality (37) is sharper than the one from (38), the first proved inequality (39) is lower than the second one (40).

Lower bounds: By the first inequality from Lemma 6, we have

$$\gamma(d/2, \lambda z^2) \ge \frac{2}{d}(\lambda z^2)^{d/2}\exp\left(-\frac{d/2}{d/2+1}\lambda z^2\right) \ge \frac{2}{d}(\lambda z^2)^{d/2}\exp\left(-\lambda z^2\right). \tag{41}$$

Now, for the inequalities in Eqs. (35) and (36), we insert the first and second inequality from Eq. (41) into Eq. (18), respectively. For the first inequality, we thereby obtain

$$\mathbb{E}^x \tau_L \;\geq\; \frac{2}{\sigma^2 d} \int_x^L \exp\left(\frac{2}{d+2}\lambda z^2\right) z \,\mathrm{d}z \;=\; \frac{2}{\sigma^2 d}\left[\frac{d+2}{4\lambda}\exp\left(\frac{2\lambda}{d+2}z^2\right)\right]_{z=x}^{L} \tag{42}$$

$$= \frac{1+\frac{2}{d}}{2\lambda\sigma^2}\left[\exp\left(\frac{2\lambda}{d+2}L^2\right) - \exp\left(\frac{2\lambda}{d+2}x^2\right)\right]. \tag{43}$$

For the second inequality, we analogously obtain

$$\mathbb{E}^x \tau_L \;\geq\; \frac{2}{\sigma^2 d}\int_x^L z \,\mathrm{d}z = \left[\frac{L^2 - x^2}{\sigma^2}\right] d^{-1}. \tag{44}$$

As for the upper bounds, the ordering of inequalities in Eq. (41) implies that the lower bound in Eq. (44) is even lower than the one in Eq. (43). □

Corollary 8. *Let $L > 0$ and let $x \in [0, L]$ such that a.s. $\|X_0\| = x$. Then, we have, for $d \to \infty$, that*

$$\mathbb{E}^x \tau_L \;\sim\; \left[\frac{L^2 - x^2}{\sigma^2}\right] d^{-1}, \qquad \forall\, \theta \geq 0, \tag{45}$$

i.e. the MFET of a d-dimensional radial Ornstein Uhlenbeck process with arbitrary $\theta > 0$ is asymptotically equivalent to the one of a d-dimensional Brownian motion with $\theta = 0$. (See e.g. Eq. (7.4.2) in Øksendal (2003) for a proof that the MFET of a d-dimensional Brownian motion is equal to the right-hand side of Eq. (45).)

Proof. From Eq. (33), we know that

$$\mathbb{E}^x \tau_L \;\leq\; \frac{2}{\lambda\sigma^2 d(d+2)}\left[\exp(\lambda L^2) - \exp(\lambda x^2)\right] + \frac{1}{\sigma^2(d+2)}\left[L^2 - x^2\right], \tag{46}$$

where the first term is of order $O(d^{-2})$ and the second of order $O(d^{-1})$. Thus, the first term is irrelevant for the asymptotics and we have

$$\limsup_{d\to\infty} \frac{\mathbb{E}^x \tau_L}{\sigma^2 d/(L^2 - x^2)} = \limsup_{d\to\infty} \frac{\mathbb{E}^x \tau_L}{\sigma^2 (d+2)/(L^2 - x^2)} \leq 1. \tag{47}$$

On the other hand, by Eq. (36), we have that

$$\liminf_{d\to\infty} \frac{\mathbb{E}^x \tau_L}{\sigma^2 d/(L^2 - x^2)} \geq 1. \tag{48}$$

This means that the liminf and the limsup coincide at 1, i.e.

$$\lim_{d\to\infty} \frac{\mathbb{E}^x \tau_L}{\sigma^2 d/(L^2 - x^2)} = 1, \tag{49}$$

which is equivalent to the desired Eq. (45). □

5 Discussion

Theorem 7 provides asymptotically tight bounds for the MFET of a d-dimensional OUP from a ball of finite radius $L > 0$, as $d \to \infty$. These bounds imply by virtue of Corollary 8 that, in high dimensions, the d-dimensional OUP takes (on average) no longer than a d-dimensional BM to exit a ball of radius L. While we provided a rigorous proof above, this section gives some intuition on whether this result is surprising and why the drift coefficient θ does not impact $\mathbb{E}^x \tau_L$ in high dimensions.

Is Corollary 8 surprising? (Short answer: At first, it might be; but a careful examination dispels the surprise.) At first sight, Corollary 8 may be surprising. After all, the mean-reverting drift of OUP (parametrized by $\theta > 0$) makes all the difference with the drift-less BM. In fact, due to the drift, the analytical properties of OUP and BM differ significantly; e.g., the OUP has a stationary distribution (unlike the BM) and its maximum grows in θ (Graversen and Peskir, 2000; Jia and Zhao, 2020). But, our result suggests that – for the MFET – the drift becomes irrelevant as $d \to \infty$. The following paragraphs contain an attempt to explain this.

First, we want to discern which part of our result might be surprising. We proved that (i) the MFET $\mathbb{E}^x \tau_L$ goes to zero, and that (ii) $\mathbb{E}^x \tau_L$ is for any $\theta > 0$ asymptotically equivalent to $\mathbb{E}^x \tau_L$ when $\theta = 0$ (Brownian motion case). We feel that, while point (i) is unsurprising, point (ii) may be surprising at first (before a closer examination on the drift below).

Regarding (i), it is clear that, as d grows, the first-exit time τ_L will converge to zero (in probability) for all choices of θ (or in fact, for any d-dimensional stochastic process with iid. components). This can be seen from Eq. (3), where τ_L was defined as $\tau_L = \inf\{t > 0 : \|X_t\|_2^2 = \sum_{i=1}^d X_{i,t}^2 = L^2\}$. Since the processes $X_{i,t}$ are iid., the sum $\sum_{i=1}^d X_{i,t}^2$ will likely reach L^2 faster for a high value of d. Thus, each value of the cumulative distribution function of τ_L, as well as its mean, will monotonously decrease to zero, as $d \to \infty$. But, from this, it does not follow how the asymptotic rates depend on drift coefficient θ, as $d \to \infty$. Next, we will explain why this is the case.

For $\rho_t \in [0, L]$, the dynamics of the rOUP is asymptotically independent of θ, as $d \to \infty$. From Eq. (3), we can rewrite the first exit time as $\tau_L = \sup_{t \geq 0}\{0 \leq \rho_s^2 \leq L^2, \forall s \in [0, t]\}$. This means that, for τ_L, only the dynamics of ρ_t^2 on the interval $\rho_t \in [0, L]$ matters. But the SDE of ρ_t^2, Eq. (4), only depends on θ through the SDE's drift coefficient $(\sigma^2 d - 2\theta\rho_t^2)$, and this drift coefficient becomes irrelevant for $d \to \infty$ on the bounded interval $\rho_t \in [0, L]$:

$$\sigma^2 d - 2\theta\rho_t^2 \;\sim\; \sigma^2 d, \qquad \text{for } d \to \infty, \tag{50}$$

because σ^2, θ and L are chosen as constants independent of d. Note that the right-hand side, $\sigma^2 d$, is indeed the drift of the d-dim. squared Bessel process, i.e. of the squared rOUP ρ_t^2 with $\theta = 0$ (Pitman and Winkel, 2018, Eq. (1.1)). Hence, Eq. (50) shows that the coefficients of the SDE of ρ_t^2, Eq. (4), is asymptotically independent of θ on $\rho_t \in [0, L]$. The left subplot of Fig. 3 visualizes this effect. With this in mind, it is unsurprising that the MFET $\mathbb{E}^x \tau_L$ becomes independent of θ, as $d \to \infty$.

For $\rho_t \in \mathbb{R}_{>0}$, the dynamics of rOUP still depend on θ. For unbounded ρ_t the situation is however different. If ρ_t^2 can take arbitrarily large values in the left-hand side of Eq. (50), then the asymptotics do not hold. This is demonstrated on the right subplot of Fig. 3. In fact, as long as L

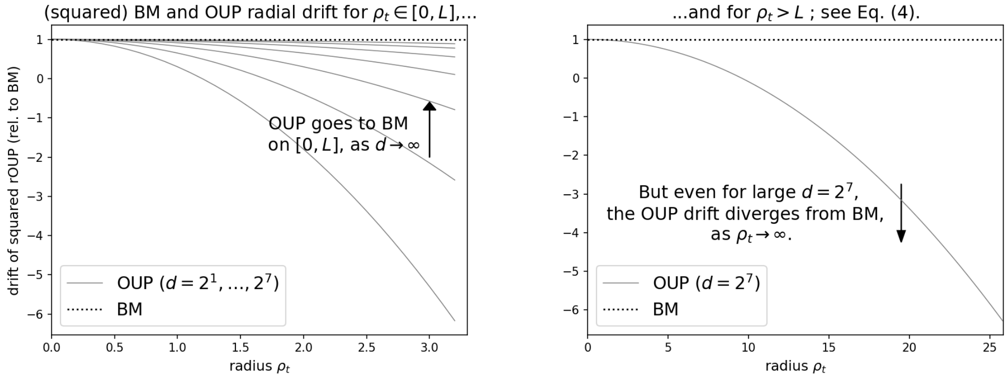

Figure 3: Simulations to give intuition for Corollary 8. Parameters are set as $L = 3.0$, $\sigma^2 = 1.0$, and $\theta = 0.7$. Both the *left* and the *right* plot show the drift of the squared radial OUP $\sigma^2 d - 2\theta$ divided by the squared BM drift $\sigma^2 d$, i.e. $\frac{\sigma^2 d - 2\theta}{\sigma^2 d}$, on different domains; see Eq. (4) for these drift terms. Hence, BM is a vertical line at $y = 1$ and the OUPs have lower values (in some cases negative, i.e. the drift pushes back to zero). The *left* plot shows these ratios for $d = 2^1, 2^2, ..., 2^7$ in gray, on $\rho_t \in [0, L]$. The k^{th} line from below belongs to $d = 2^k$. We can see that, as $d \to \infty$, the OUP drift approaches BM. However the *right* plot shows that, for $\rho_t > L$, even the highest OUP drift $d = 2^7$ still diverges to $-\infty$, as $\rho_t \to \infty$. In summary, the left plot explains why $\mathbb{E}^x \tau_L$ is asymptotically independent of θ, while the right plot highlights that the rOUP still depends on θ for large values of ρ_t. More details in main text.

grows at least like $\mathcal{O}(\sqrt{d})$, we could still conclude from Eq. (50) that the $d \to \infty$ asymptotics of $\mathbb{E}^x \tau_L$ will depend on θ.

Therefore, it would be incorrect to conclude that Corollary 8) implies that the d-dim. squared rOUP ρ_t^2 will resemble the d-dim. squared Bessel process on all of $\mathbb{R}_{>0}$, as $d \to \infty$. But, on $[0, L]$, the SDE of the squared rOUP ρ_t^2 is asymptotically equivalent to the squared Bessel process. (With this in mind, one can probably show more asymptotic similarities between these processes on compact domains.)

6 Conclusion

In the above material, we proved two new theorems and one corollary.

First, Theorem 4 gave two explicit formulas for the mean first exit time of a d-dimensional Ornstein–Uhlenbeck process from a ball of radius L. The first of these formulas, Eq. (17), coincides with the ones derived by prior work (Graczyk and Jakubowski, 2008; Grebenkov, 2014), but our proof is very short (see discussion in Remark 5) and leverages Andronov–Vitt–Pontryagin theory (Schuss, 2010). The second of these formulas, Eq. (18), is a novel reformulation in terms of the incomplete Gamma function.

Second, Theorem 7 exploits this reformulation by bounding the Gamma function as suggested by Neuman (2013). The resulting bounds are new and are verified numerically in Fig. 2. Since the upper and lower bounds are asymptotically equivalent, they have the (perhaps at first surprising) implication (Corollary 8) that, for $d \to \infty$, the d-dimensional OUP takes no longer than a d-dimensional Brownian motion to exit a ball of arbitrary radius L. Thus, for large d, the drift does not matter for the (mean) exit time of OUP. This might be surprising because the d-dimensional Brownian motion is just an OUP without drift, i.e. with $\theta = 0$, and one would expect that a larger drift back to zero leads to a slower exit (as is the case for small $d \in \mathbb{N}$). The simulations in Fig. 2 verify this asymptotic relation between OUP and Brownian motion. In Section 5, we then give intuition to disperse any initial surprise that readers might have experienced.

Our findings shed light on some unusual behavior of the OUP in high dimensions. We hope that our Corollary 8 will be applicable in the numerous research areas (biology, economics, machine learning, statistical mechanics, etc.), where high-dimensional OUPs are used for modeling.

Acknowledgements

We thank Ralf Metzler for his helpful comments on an early version of the manuscript. We thank the anonymous reviewers for their careful remarks which helped us to improve this paper.

References

Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian Motion - Facts and Formulae*. Probability and Its Applications. Birkhäuser Basel, 2nd edition, 2002.

Marguerite A. Butler and Aaron A. King. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American naturalist*, 164(6):683–695, 2004. URL https://doi.org/10.1086/426002.

Gabriela Ciołek, Dmytro Marushkevych, and Mark Podolskij. On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein–Uhlenbeck model. *Electronic Journal of Statistics*, 14(2):4395 – 4420, 2020. URL https://doi.org/10.1214/20-EJS1775.

Olivier Faugeras, Jonathan Touboul, and Bruno Cessac. A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Frontiers in Computational Neuroscience*, 3(2009), 2004. URL https://doi.org/10.3389/neuro.10.001.2009.

Piotr Graczyk and Tomasz Jakubowski. Exit times and Poisson kernels of the Ornstein–Uhlenbeck diffusion. *Stochastic Models*, 24(2):314–337, 2008. URL https://doi.org/10.1080/15326340802009337.

Svend-Erik Graversen and Goran Peskir. Maximal inequalities for the Ornstein-Uhlenbeck process. *Proceedings of the American Mathematical Society*, 128(10):3035–3041, 2000. URL https://www.ams.org/journals/proc/2000-128-10/S0002-9939-00-05345-4/S0002-9939-00-05345-4.pdf.

Denis S. Grebenkov. First exit times of harmonically trapped particles: a didactic review. *Journal of Physics A: Mathematical and Theoretical*, 48(1):013001, 2014. URL https://iopscience.iop.org/article/10.1088/1751-8113/48/1/013001.

Chen Jia and Guohuan Zhao. Moderate maximal inequalities for the Ornstein-Uhlenbeck process. *Proceedings of the American Mathematical Society*, 148(8):3607–3615, 2020. URL https://www.ams.org/journals/proc/2020-148-08/S0002-9939-2020-14804-6/home.html.

Qianxiao Li, Cheng Tai, and Weinan E. Stochastic modified equations and adaptive stochastic gradient algorithms. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2101–2110. PMLR, 2017. URL https://proceedings.mlr.press/v70/li17f.html.

Aurelien Lucchi, Frank Proske, Antonio Orvieto, Francis Bach, and Hans Kersting. On the theoretical properties of noise correlation in stochastic optimization. In *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=cNrglG_OAeu.

Hirotaka Matsumoto and Hisanori Kiryu. SCOUP: a probabilistic model based on the Ornstein–Uhlenbeck process to analyze single-cell expression data during differentiation. *BMC Bioinformatics*, 17(1):232, 2016. URL https://doi.org/10.1186/s12859-016-1109-3.

Edward Neuman. Inequalities and bounds for the incomplete Gamma function. *Results in Mathematics*, 63 (3):1209–1214, 2013. URL https://doi.org/10.1007/s00025-012-0263-9.

Thanh Huy Nguyen, Umut Simsekli, Mert Gurbuzbalaban, and Gaël Richard. First exit time analysis of stochastic gradient descent under heavy-tailed gradient noise. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper/2019/file/a97da629b098b75c294dffdc3e463904-Paper.pdf.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Berlin, Heidelberg, 6th edition, 2003.

Jim Pitman and Matthias Winkel. Squared Bessel processes of positive and negative dimension embedded in Brownian local times. *Electronic Communications in Probability*, 23:1–13, 2018. URL https://doi.org/10.1214/18-ECP174.

Luigi M. Ricciardi and Laura Sacerdote. The Ornstein-Uhlenbeck process as a model for neuronal activity. i. mean and variance of the firing time. *Biological cybernetics*, 35(1):1–9, 1979. URL https://doi.org/10.1007/BF01845839.

Rori V. Rohlfs, Patrick Harrigan, and Rasmus Nielsen. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Molecular Biology and Evolution*, 31:201–211, 2014. URL https://doi.org/10.1093/molbev/mst190.

Zeev Schuss. *Theory and Applications of Stochastic Processes: An Analytical Approach*, volume 170 of *Applied Mathematical Sciences*. Springer New York, 2010.

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012. URL https://www.mat.univie.ac.at/~gerald/ftp/book-ode/ode.pdf.