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

Hans Kersting1,∗, Antonio Orvieto2, Frank Proske3 and Aurelien Lucchi4

1 INRIA, Ecole Normale Supérieure, PSL Research University, Paris, France
2 Department of Computer Science, ETH Zürich, Switzerland
3 Department of Mathematics, University of Oslo, Norway
4 Department of Mathematics & Computer Science, University of Basel, Switzerland

E-mail: h.p.kersting@gmail.com

Received 26 August 2022; revised 17 January 2023
Accepted for publication 17 March 2023
Published 4 May 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 θ ⩾ 0 and a diffusion coefficient weighted by σ > 0. In the absence of drift (i.e. θ = 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 ≫ 0. We prove that, asymptotically for d → ∞, the OUP takes (on average) no longer to exit than BM. In other words, the mean-reverting drift of the OUP (scaled by θ ⩾ 0) has asymptotically no effect on its MFET. This finding might be surprising because, for small d ∈ 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.

Keywords: Ornstein–Uhlenbeck process, Brownian motion, first-exit time, asymptotics, high-dimensional limit

(Some figures may appear in colour only in the online journal)
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:

$$
\text{d}X_t = -\theta X_t \text{d}t + \sigma \text{d}B_t,
$$

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

$$
\text{cov}(X_{t_1}, X_{t_2}) = \delta_{ij} \frac{\sigma^2}{2\theta} \left(e^{-\theta|t_1-t_2|} - e^{-\theta(t_1+t_2)}\right),
$$

where $\delta_{ij}$ is the Kronecker delta and $X_{t_1}, X_{t_2}$ are the independent components of $X_t = (X_{t,1}, \ldots, X_{t,d})$.

The OUP has been extensively studied (Borodin and Salminen 2002, part II, section 7), sometimes with the parameter $\lambda = \frac{\theta}{2\sigma^2} \in \mathbb{R}$ instead of $\theta$. 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. 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, Rohlfis et al 2014), and single-cell differentiation for multiple cells (Matsumoto and Kiyuru 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 equation (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
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 $\theta$ impacts the asymptotic rate (see section 5 for a discussion). Also, see figure 2, where the mean asymptotics are depicted for the same parameters of $L$, $\sigma$, and $\theta$.

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 figure 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 theoretical results and experiments (figure 2) will confirm this.

1.1. 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).
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 \(\lambda\) (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 equations (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 \((\theta = 0)\).

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

1.2. 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 equation (1). We are concerned with the first exit time

\[
\tau_L := \inf_{t \geq 0} \{ \|X_t\| = L \}.
\]

We can see that this stopping time only depends on \(X_t\) via \(\rho_t := \|X_t\| = (\sum_{i=1}^d X_{i,t}^2)^{1/2}\). The stochastic process \(\rho_t\) 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 equation (1). Then, $\rho_t^2 = \sum_{i=1}^{d} X_{t,i}^2$ and $\rho_t = (\sum_{i=1}^{d} X_{t,i}^2)^{1/2}$ follow the SDEs

$$\begin{align*}
    d\rho_t^2 &= (\sigma^2 d - 2\theta \rho_t^2) \, dt + 2\sqrt{\rho_t^2} \, dB_t \\
    \text{and} \quad d\rho_t &= \left[ \frac{(d-1)\sigma^2}{2\rho_t} - \theta \rho_t \right] dt + \sigma dB_t,
\end{align*}$$

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 $X_t = [X_{1,t}, \ldots, X_{d,t}]$ are independent OUPs of the form

$$dX_{i,t} = -\theta X_{i,t} \, dt + \sigma dB_{i,t}. \quad (6)$$

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

$$\begin{align*}
    d\rho_t^2 &= 2X_t^T dX_t + \frac{1}{2} \text{tr}(2I_d [dX_t dX_t^T]) \\
    &= 2X_t^T dX_t + \sigma^2 dt.
\end{align*}$$

The remaining term to compute is

$$X_t^T dX_t \overset{(6)}{=} -\theta \rho_t^2 dt + \sigma \sum_{i=1}^{d} X_{t,i} dB_{i,t} = -\theta \rho_t^2 dt + \sigma \sqrt{\rho_t^2} \, dY_t, \quad (9)$$

where

$$Y_t := \sum_{i=1}^{d} \int_0^t \frac{X_{s,i}}{\sqrt{\rho_s^2}} \, dB_{i,s}^{(i)}. \quad (10)$$

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 equation (9) into equation (8) yields equation (4).

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

Now, we have derived the SDE for the rOUP in equation (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
\[ \text{d}X_t = a(X_t, t) \text{d}t + b(X_t, t) \text{d}B_t, \]
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
\[ L^*_t u(x, t) = \sum_{i=1}^{n} a_i(t) \frac{\partial u(x, t)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} [b_i(X_t, t) b_j(X_t, t)]_{ij} \frac{\partial^2 u(x, t)}{\partial x^i \partial x^j}. \]

For any bounded domain \( D \subset \mathbb{R}^d \), we define the first exit time from this domain:
\[ \pi_D := \inf \{ t > 0 \mid x(t) \notin D \}. \]
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 \( \pi_D \).

**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
\[ \frac{\partial u(x, t)}{\partial t} + L^*_t u(x, t) = -1, \text{ for } x \in D \text{ and } t \in \mathbb{R}, \quad u(x, t) = 0, \text{ for } x \in \partial D, \]
has a unique bounded solution \( u(x, t) \). Then, the mean first passage time \( \mathbb{E}^* \pi_D \) is finite and takes the form
\[ \mathbb{E}^* \pi_D = u(x, 0). \]

**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 equation (11) are independent of \( t \), the solution \( u \) of equation (14) is also independent of \( t \) and therefore solves the ODE
\[ L^*_t u(x) = -1, \quad \text{for } x \in D, \quad u(x) = 0, \text{ for } x \in \partial D. \]
In particular, \( \mathbb{E}^* \pi_D = u(x) \).

**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 \( \mathbb{E}^* \tau_L \).

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 \( \lambda := \frac{a}{\sigma^2} \). Then, for all \( \sigma > 0 \) and \( \lambda \in \mathbb{R} \), we have that
\[ \mathbb{E}^* \tau_L = \frac{2}{\sigma^2} \int_x^{L} z^{1-d} e^{\lambda z} \left[ \int_0^z r^{d-1} e^{-\lambda r} \text{d}r \right] \text{d}z. \]
If \( \lambda > 0 \), then this formula further simplifies to

\[
\mathbb{E}^x \tau_L = \frac{1}{\lambda^2 \sigma^2} \int_x^L z^{1-d} e^{\lambda z^2} \gamma(d/2, \lambda z^2) \, dz,
\]

(18)

where the function \( \gamma \) denotes the upper incomplete Gamma function

\[
\gamma(n, y) := \int_0^y t^{n-1} \exp(-t) \, dt.
\]

(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 equation (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 (2017) derives our formula (17) in his equation (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 \( \lambda \) in their notation). (Note that they use hypergeometric functions to express the integral in equation (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 equation (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 equation (18) is easily derived from equation (17) by a change of variable (see proof), it will be essential to prove the scaling for \( d \to \infty \) below. Equation (18) is thus an essential part of our analysis.

**Proof of Theorem 4.** We first prove equation (17) and then equation (18). The proof of equation (17) is split into three parts. First, we fix \( x > 0 \) and show that the right-hand side of equation (17) solves the Andronov–Vitt–Pontryagin BVP, equation (16), associated with the radial Ornstein–Uhlenbeck process \( \rho_t \), with \( \rho_0 = x \). Second, we show that the solution to this BVP is unique. (Together, the first two steps imply that equation (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 \( \rho_t = (\sum_{i=1}^d X_{t-}^i)^{1/2} \), we have \( \tau_L = \inf \{ t > 0 : \rho_t = L \} \). Hence, by the SDE (5) of \( \rho_t \), the associated Andronov–Vitt–Pontryagin BVP (16) reads

\[
\left[ (d-1) \sigma^2 / 2x - 6x \right] u'(x) + \frac{\sigma^2}{2} u''(x) = -1, \quad \text{with } u(L) = 0, \tag{18}
\]

or equivalently

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

(Note that we here used that \( \tau_{R_t(0)} = \tau_L \) by construction in equation (13), so that we could apply the Andronov–Vitt–Pontryagin formula.) We will now show that the right-hand side of equation (17) solves the above equation, that is
\( u(x) := \frac{2}{\sigma^2} \int_x^L e^{-1/d} e^{-\lambda x^2} \left[ \int_0^t e^{-1/d} e^{-\lambda x^2} \, dt \right] \, dz. \)  

(21)

Its derivatives are

\[ u'(x) = -\frac{2}{\sigma^2} x^{1/d} e^{\lambda x^2} I(x), \]

(22)

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

(23)

where

\[ I(x) := \left[ \int_0^x e^{1/d} e^{-\lambda x^2} \, dt \right]. \]

(24)

Insertion of these formulas yields equation (20):

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

(26)

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

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

(27)

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

\[
L(\tilde{x}) := \sup_{u \neq v \in \mathbb{R}^M} \frac{\|f(\tilde{x}, u) - f(\tilde{x}, v)\|}{\|u - v\|} = \min \left\{ 1, \left| 2\lambda \tilde{x} - (d-1)/\tilde{x} \right| \right\} \leq |2\lambda \tilde{x} - (d-1)/\tilde{x}| \\
\leq 2\lambda L + (d-1)/L < \infty, \quad \forall \tilde{x} \in [x, L].
\]

(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 \( x > 0 \):

\[
\int_x^L L(\tilde{x}) \, d\tilde{x} \leq (L - x) \cdot (2\lambda L + (d-1)/x) < \infty.
\]

(29)

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

**Third step:** To provide the missing case \( x = 0 \), we first observe that \( \lim_{x \to 0} \mathbb{E}_0^0 \tau_1 = 0 — 

because otherwise the radial Ornstein–Uhlenbeck \( \rho_t \) 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

\[
\mathbb{E}_0^0 \tau_L = \mathbb{E}_0^0 \tau_l + \mathbb{E}_0^l \tau_L. 
\]

(30)

Hence, in the limit \( l \to 0 \), we indeed obtain \( \mathbb{E}_0^0 \tau_L = u(0) \) which concludes the proof of equation (17). Note that all above steps worked for all \( \lambda = \frac{\rho}{\sigma^2} \in \mathbb{R} \), and thus equation (17) holds for all \( \lambda \in \mathbb{R} \).
To prove the missing equation (18) for \( \lambda > 0 \), we observe by a change of variable, using the substitution function \( \phi(t) := \sqrt{\frac{t}{\lambda}} \) with derivative \( \phi'(t) = \sqrt{\frac{4}{\lambda} t} \), that

\[
\hat{z} = \phi(\lambda z^2)
\]

\[
\int_0^{\epsilon = \phi(\lambda z^2)} t^{d-1} \exp(-\lambda t^2)dt = \int_0^{\lambda z^2} \sqrt{4\lambda t} \frac{e^{\epsilon^2}}{\epsilon} \exp(-\epsilon)dt,
\]

\[
= \frac{1}{2} \lambda^{-d/2} \int_0^{\lambda z^2} t^{d-1} \exp(-t)dt.
\]

Insertion of the above formula into equation (17) yields equation (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 \( \gamma \) from equation (19), the inequalities

\[
\frac{x^a}{a} \exp\left(-\frac{ax}{a+1}\right) \leq \gamma(a,x) \leq \frac{x^a}{a(a+1)} (1 + a \exp(-x))
\]

are valid for all \( a > 0 \).

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

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

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

(33)

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

(34)

and lower bounds

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

(35)

\[
\geq \left[ \frac{L^2 - x^2}{\sigma^2} \right] d^{-1}.
\]

(36)

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

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

\[
\gamma(d/2, \lambda z^2) \leq \frac{4\lambda^{d/2} \epsilon^{d/2}}{d(d+2)} \left[ 1 + \frac{d}{2} \exp(-\lambda z^2) \right]
\]

(37)

\[
\leq \frac{4\lambda^{d/2} \epsilon^{d/2}}{d(d+2)} \cdot \frac{2 + d}{2} = \frac{2\lambda^{d/2} \epsilon^{d/2}}{d}.
\]

(38)
Insertion of the inequality (37) into equation (18) yields
\[ E^x \tau_L \leq \frac{4}{\sigma^2 d(d + 2)} \int_x^L z \exp \left( \lambda z^2 \right) \left[ 1 + \exp(\lambda z^2) \right] \, dz, \tag{39} \]
which is equation (33). For the missing upper bound (34), we insert the less sharp inequality (38) into equation (18):
\[ E^x \tau_L \leq \frac{2}{\sigma^2} \int_x^L z \exp \left( \lambda z^2 \right) \, dz = \frac{1}{\theta} \left[ \exp(\lambda L^2) - \exp(\lambda x^2) \right] d^{-1}, \tag{40} \]
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
\[ \gamma(d/2, \lambda z^2) \geq \frac{2}{d} (\lambda z^2)^{d/2} \exp \left( - \frac{d/2}{d/2 + 1} \lambda z^2 \right) \geq \frac{2}{d} (\lambda z^2)^{d/2} \exp \left( - \lambda z^2 \right). \tag{41} \]

Now, for the inequalities in equations (35) and (36), we insert the first and second inequality from equation (41) into equation (18), respectively. For the first inequality, we thereby obtain
\[ E^x \tau_L \geq \frac{2}{\sigma^2 d} \int_x^L \exp \left( \frac{2}{d + 2} \lambda z^2 \right) \, dz = \frac{2}{\sigma^2 d} \left[ \frac{d + 2}{4\lambda} \exp \left( \frac{2\lambda}{d + 2} \right) \right]^L_{z=x} \tag{42} \]
\[ = \frac{1 + \frac{3}{2\lambda \sigma^2} \exp \left( \frac{2\lambda}{d + 2} L^2 \right) - \exp \left( \frac{2\lambda}{d + 2} x^2 \right) \right]. \tag{43} \]

For the second inequality, we analogously obtain
\[ E^x \tau_L \geq \frac{2}{\sigma^2 d} \int_x^L z \, dz = \left[ \frac{L^2 - x^2}{\sigma^2} \right] d^{-1}. \tag{44} \]
As for the upper bounds, the ordering of inequalities in equation (41) implies that the lower bound in equation (44) is even lower than the one in equation (43).

\[ \square \]

**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
\[ E^x \tau_L \sim \left[ \frac{L^2 - x^2}{\sigma^2} \right] d^{-1}, \quad \forall \theta \geq 0, \tag{45} \]
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. equation (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 equation (45).)

**Proof.** From equation (33), we know that
\[ E^x \tau_L \leq \frac{2}{\lambda \sigma^2 d(d + 2)} \left[ \exp(\lambda L^2) - \exp(\lambda x^2) \right] + \frac{1}{\sigma^2(d + 2)} \left[ L^2 - x^2 \right], \tag{46} \]
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

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

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

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

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

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

which is equivalent to the desired equation (45).

\[\square\]

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 $\theta$ does not impact $\mathbb{E}^\tau_L$ in high dimensions.

5.1. 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 $\theta$ (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 $\mathbb{E}^\tau_L$ goes to zero, and that (ii) $\mathbb{E}^\tau_L$ is for any $\theta > 0$ asymptotically equivalent to $\mathbb{E}^\tau_L$ 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 $\tau_L$ will converge to zero (in probability) for all choices of $\theta$ (or in fact, for any $d$-dimensional stochastic process with i.i.d. components). This can be seen from equation (3), where $\tau_L$ was defined as $\tau_L = \inf\{t > 0 : \|X_t\|_2^2 = \sum_{i=1}^d X_{t,i}^2 = L^2\}$. Since the processes $X_{t,i}$ are i.i.d. the sum $\sum_{i=1}^d X_{t,i}^2$ will likely reach $L^2$ faster for a high value of $d$. Thus, each value of the cumulative distribution function of $\tau_L$, 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 $\theta$, as $d \to \infty$. Next, we will explain why this is the case.
5.2. For \( \rho_t \in [0,L] \), the dynamics of the rOU P is asymptotically independent of \( \theta \), as \( d \to \infty \)

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

\[
\sigma^2 d - 2\theta \rho_t^d \sim \sigma^2 d, \quad \text{for } d \to \infty, \tag{50}
\]

because \( \sigma^2 \), \( \theta \) 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 \( \rho_t^d \) with \( \theta = 0 \) (Pitman and Winkel 2018, equation (1.1)). Hence, equation (50) shows that the coefficients of the SDE of \( \rho_t^d \), equation (4), is asymptotically independent of \( \theta \) on \( \rho_t \in [0,L] \).

The left subplot of figure 3 visualizes this effect. With this in mind, it is unsurprising that the MFET \( \mathbb{E}^d \tau_L \) becomes independent of \( \theta \), as \( d \to \infty \).

5.3. For \( \rho_t \in \mathbb{R}_{>0} \), the dynamics of rOUP still depend on \( \theta \)

For unbounded \( \rho_t \) the situation is however different. If \( \rho_t^2 \) can take arbitrarily large values in the left-hand side of equation (50), then the asymptotics do not hold. This is demonstrated on the right subplot of figure 3. In fact, as long as \( L \) grows at least like \( O(\sqrt{d}) \), we could still conclude from equation (50) that the \( d \to \infty \) asymptotics of \( \mathbb{E}^d \tau_L \) will depend on \( \theta \).

Therefore, it would be incorrect to conclude that corollary 8) implies that the \( d \)-dim. squared rOUP \( \rho_t^2 \) 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 \( \rho_t^2 \) 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, equation (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, equation (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 figure 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 figure 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.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgment

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.

Hans Kersting thanks the European Research Council for support through ERC grant 724063. Frank Proske thanks the Norway-Ukrainian Cooperation in Mathematical Education, project number CPEA-LT-2016/10139 and Norges forskningsråd 274410. Aurelien Lucchi acknowledges the financial support of the Swiss National Foundation, SNF, grant No 207392.

ORCID iD

Hans Kersting @ https://orcid.org/0000-0002-2782-868X
References

Borodin A N and Salminen P 2002 Handbook of Brownian Motion - Facts and Formulae (Probability and Its Applications) 2nd edn (Basel, Switzerland: Birkhäuser)

Butler M A and King A A 2004 Phylogenetic comparative analysis: A modeling approach for adaptive evolution Am. Nat. 164 683–95

Ciołek G, Marushkevych D and Podolskij M 2020 On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein–Uhlenbeck model Electron. J. Stat. 14 4395–420

Faugeras O, Touboul J and Cessac B 2004 A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs Front. Comput. Neurosci. 3 1

Graczyk P and Jakubowski T 2008 Exit times and Poisson kernels of the Ornstein–Uhlenbeck diffusion Stoch. Models 24 314–37

Graversen S-E and Peskir G 2000 Maximal inequalities for the Ornstein-Uhlenbeck process Proc. Am. Math. Soc. 128 3035–41

Grebenkov D S 2014 First exit times of harmonically trapped particles: a didactic review J. Phys. A: Math. Theor. 48 013001

Jia C and Zhao G 2020 Moderate maximal inequalities for the Ornstein-Uhlenbeck process Proc. Am. Math. Soc. 148 3607–15

Lucchi A, Proske F, Orvieto A, Bach F and Kersting H 2022 On the theoretical properties of noise correlation in stochastic optimization Advances in Neural Information Processing Systems

Matsumoto H and Kiyu H 2016 SCOUP: a probabilistic model based on the Ornstein–Uhlenbeck process to analyze single-cell expression data during differentiation BMC Bioinform. 17 232

Neuman E 2013 Inequalities and bounds for the incomplete Gamma function Results Math. 63 1209–14

Nguyen T H, Simsekli U, Gurbuzbalaban M and Richard G 2019 First exit time analysis of stochastic gradient descent under heavy-tailed gradient noise Advances in Neural Information Processing Systems

Øksendal B 2003 Stochastic Differential Equations: An Introduction With Applications 6th edn (Berlin: Springer)

Pitman J and Winkel M 2018 Squared Bessel processes of positive and negative dimension embedded in Brownian local times Electron. Commun. Probab. 23 1–13

Riccio L M and Sacerdote L 1979 The Ornstein-Uhlenbeck process as a model for neuronal activity. I. Mean and variance of the firing time Biol. Cybern. 35 1–9

Rohls R V, Harrigan P and Nielsen R 2014 Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation Mol. Biol. Evol. 31 201–11

Schuss Z 2010 Theory and Applications of Stochastic Processes: An Analytical Approach (Applied Mathematical Sciences vol 170) (New York: Springer)

Teschl G 2012 Ordinary Differential Equations and Dynamical Systems (Graduate Studies in Mathematics vol 40) (Providence, RI: American Mathematical Society)