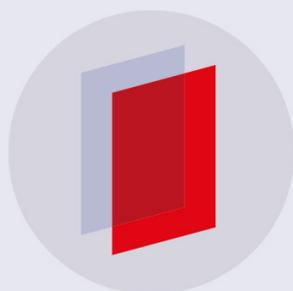**PAPER • OPEN ACCESS**

Codifference can detect ergodicity breaking and non-Gaussianity

To cite this article: Jakub Izak *et al* 2019 *New J. Phys.* **21** 053008

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Codifference can detect ergodicity breaking and non-Gaussianity

Jakub Ślęzak[1,2], Ralf Metzler[3] and Marcin Magdziarz[2]

1   Department of Physics, Bar Ilan University, Israel
2   Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Poland
3   Institute of Physics and Astronomy, Potsdam University, Germany

E-mail: rmetzler@uni-potsdam.de

Abstract

We show that the codifference is a useful tool in studying the ergodicity breaking and non-Gaussianity properties of stochastic time series. While the codifference is a measure of dependence that was previously studied mainly in the context of stable processes, we here extend its range of applicability to random-parameter and diffusing-diffusivity models which are important in contemporary physics, biology and financial engineering. We prove that the codifference detects forms of dependence and ergodicity breaking which are not visible from analysing the covariance and correlation functions. We also discuss a related measure of dispersion, which is a nonlinear analogue of the mean squared displacement.

1. Introduction

1.1. Statistical measures in modelling of diffusion

The analysis of stochastic systems has three important and partially distinct aspects: models, properties and estimation. These roughly correspond to physical, mathematical and statistical aspects of research. Modelling is concerned with explaining the nature of a system according to the underlying theory (e.g. 'the particle undergoes Brownian motion, because it rapidly exchanges momenta with the molecules of liquid'). The analysis of statistical properties (also called 'measures')[4] relates these models with observable quantities ('Brownian motion has a linear mean squared displacement'). By using suitable estimators we link these parameters to the experimental data ('the mean squared displacement can be efficiently estimated by an arithmetic average over squared displacements').

This work is motivated by our conviction that the choice of statistical measures is too small for contemporary needs, as the scope and number of models increased considerably [1]. The classical models based on the Langevin equation [2], the generalised Langevin equation [3, 4], as; well as short- [5] and long- [6, 7] memory random walks were complemented by motions on fractals [8], motions in complex energy landscapes [9], random walks in random environments [10, 11], random walks with correlated steps and waiting times [12–16] and Lévy walks [17], spatially heterogeneous diffusion processes [18], diffusing-diffusivity [19] and more. Distinguishing between different models from this wide class is of course crucially dependent on the physical understanding of the system, but this requirement does not lessen the importance of empirical verification based on various measures and corresponding estimators. From an experimental point of view the large range of different stochastic processes is called for by ever more detailed insights garnered in highly complex environments such as living biological cells or membranes, for instance, by single particle tracking of individual sub-micron tracers of even fluorescently labelled single molecules [20–22].

Traditionally, in the study of diffusion phenomena, the three most basic and popular statistical measures in use are: the mean as a measure of location, the mean squared displacement (MSD) as a measure of dispersion and the covariance as a measure of dependence, respectively

---

[4] Strictly speaking, these are sufficiently regular functionals acting on the space of observations. In quantum mechanics each such linear functional corresponds to an observable. In statistical mechanics a similar rôle is fulfilled by $\mathbb{E}[f(X)]$ for bounded continuous functions f. In statistics linearity is usually not required and various measures have the form $g(\mathbb{E}[f(X)])$. This is the case also in the present work.

$$\mu_X(t) \coloneqq \mathbb{E}[X_t], \quad \delta_X^2(t) \coloneqq \mathbb{E}[(X_t - \mu_X(t))^2], \quad r_X(t) \coloneqq \mathbb{E}[(X_{s+t} - \mu_X(s+t))(X_s - \mu_X(s))]. \quad (1.1)$$

Other, alternative choices of measures could be, for example: the median for the location [23], entropy [24] or quantile ranges [23] for the dispersion, the rank correlation [23, 25] or the mutual information [26] for the dependence.

The covariance as defined above should not depend on the choice of s, which is true for stationary processes (the term 'non-ageing' is also in use). We will assume stationarity whenever we will be studying memory. In practical applications this condition is fulfilled by many types of confined motions or increments of free diffusions. The more general non-stationary case will be only briefly mentioned in equations (3.2) and (3.4). Many of the arguments presented here could be further extended to non-stationary models, but it would require a case-by-case study. Conversely, measures of dispersion and location are interesting mostly for non-stationary (ageing) processes, otherwise they are constant, and the discussed cases will fit into that category.

The present range of typically employed measures, which could be effectively used for studying diffusion is indeed quite limited, and the need of a wider range of methods has been acknowledged for many years. Various papers proposed, e.g. studying higher order moments and ratios of moments [27], running maximum [28], p-variation [29], or time averages and ensemble averages of time averages [30]. A prominent example of the last kind of measure is, e.g. the ergodicity breaking parameter [18, 31–33]. Recently also single-trajectory power spectral methods were proposed [34, 35]. These techniques are steadily gaining public recognition, but often the range of their application is still narrow. Moreover, a large part of this important research has a limitation of studying properties 'not very different' from the second order on. For example, any power function x^α for $\alpha > 1$ has a similar behaviour to x^2 (i.e. it is an increasing, convex function) and parameters based on it are usually not far away from the classical ones[5]. They all emphasise highly the tails of the distribution, and any change of distributions for large values of observations has a larger influence than for the small ones. This connection is very helpful in making comparisons, but the important part of the total information is lost and could be extracted using more distinct measures.

1.2. Overview of the codifference

Our main subject of interest, the codifference, is an example for a measure different from those based on moments. It was initially proposed as a tool to measure the dependence for α-stable processes, for which the second moment is infinite [42, 37–41]. However, in many systems the divergence of the second moment is not an expected physical property, which limits the range of possible applications of stable processes. It was already noticed, e.g. in [43] that the codifference may be useful for both models with or without finite second moment. In our present work we study the applications of the codifference for a class of models based on Gaussian distributions, which we call conditionally Gaussian processes; as we will demonstrate many useful and widely used models fit into this category.

The definition of the codifference which we will use is as follows: for any stationary process X it is given by the formula

$$\tau_X^\theta(t) \coloneqq \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{i\theta(X_{s+t} - X_s)}]}{\mathbb{E}[e^{i\theta X_{s+t}}]\mathbb{E}[e^{-i\theta X_s}]}. \quad (1.2)$$

The sample codifference is introduced in a standard way, by replacing the three ensemble averages $\mathbb{E}[\cdot]$ in the above expression by arithmetic averages $\frac{1}{n}\sum_{j=1}^n(\cdot)$. Similarly, one can consider a time-averaged codifference. For all symmetric distributions the considered averages should be real-valued, so in most of the practical applications one can average over $\cos(\theta(\cdot))$ instead of $\exp(i\theta(\cdot))$; this was used for the Monte Carlo simulations which will be presented further on.

Note that the so-called generalised codifference has X_{s+t} and X_s multiplied by θ_1 and θ_2 respectively and contains even more information [37]. In the context of models that we will consider this additional flexibility does not seem to be meaningful and so the cost of complicating our formulae would be unreasonable.

Conversely, the basic formula for the codifference in the classical book of Samorodnitsky and Taqqu [37] is similar to ours, but with $\theta = 1$. In the mathematical study of stable process this is sufficient, but in more broad physical applications introducing an arbitrary dimensional constant equal to unity is not desirable. In our choice of definition the codifference has the unit of X^2 due to the introduction of $1/\theta^2$. This factor makes the codifference comparable to the covariance, and allows us to show them on the same plots. When this is not important the factor $1/\theta^2$ can be omitted. There exists an even more simplified object, the dynamical functional [44], which is just the numerator minus the denominator from (1.2) with $\theta = 1$; it is used to study ergodicity breaking [30, 45].

---

[5] This similarity is what causes the 'strong anomalous diffusion' property, for which the power-law dependency $\mathbb{E}|X_t|^q \propto t^{qv(q)}$ is observed for non-constant function v [36].

Instead of moments such as the covariance, the codifference depends on sines and cosines of θX_{s+t} and θX_s. Expanding these functions into Taylor series around zero up to the two first terms and using the fact that for stationary process $\mathbb{E}[X_t] = $ const. shows that the codifference agrees with the covariance for distributions concentrated around the origin. The most essential difference is that the codifference measures mainly the dependence determined by the bulk of the probability density in contrast to the covariance, which puts much larger emphasis on the tails. This is caused by the cancellation of highly oscillatory terms in the tails of the PDF as stated by the Riemann–Lebesgue lemma, which is in contrast to the huge influence of the tails in the covariance caused by the quadratic factor in the probabilistic integral $\mathbb{E}[X_s X_{s+t}]$.

Because of the presence of two highly nonlinear transformations: sine/cosine and logarithm, definition (1.2) may initially not seem very intuitive. It becomes more natural if we interpret it as a conveniently transformed Fourier transform of the distribution (that is, the probabilistic characteristic function). In the full, multidimensional form, the characteristic function contains all information about the dependence. Moreover for Gaussian variables it has the very simple form $\exp(-(\theta\sigma)^2/2)$, so it seems reasonable to use it as a dependence measure for models related to the Gaussian distribution. Still, it is not obvious that the codifference behaves as we would require from a memory function. Fortunately, simple arguments show that this is the case.

(a) When $X_{s+t} = X_s$ (the case of total positive dependence) the codifference is a positive constant $\tau_X^\theta(t) = \tau_X^\theta(0) > 0$. If the values X_{s+t} and X_s become independent, the codifference converges to 0. Both facts are immediate consequences of the definition together with

$$\mathbb{E}[e^{i\theta X_{s+t}}]\mathbb{E}[e^{-i\theta X_s}] = |\mathbb{E}[e^{i\theta X_s}]|^2 < 1 \tag{1.3}$$

and, for X_{s+t} independent of X_s,

$$\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}] = \mathbb{E}[e^{i\theta X_{s+t}}]\mathbb{E}[e^{-i\theta X_s}]. \tag{1.4}$$

(b) If the process is a sum of independent components $X_t = Y_t + Z_t$ then the respective codifferences are additive

$$\tau_X^\theta(t) = \ln\frac{\mathbb{E}[e^{i\theta(Y_{s+t}-Y_s)}e^{i\theta(Z_{s+t}-Z_s)}]}{\mathbb{E}[e^{i\theta Y_{s+t}}e^{i\theta Z_{s+t}}]\mathbb{E}[e^{-i\theta Y_s}e^{-i\theta Z_s}]} = \ln\frac{\mathbb{E}[e^{i\theta(Y_{s+t}-Y_s)}]\mathbb{E}[e^{i\theta(Z_{s+t}-Z_s)}]}{\mathbb{E}[e^{i\theta Y_{s+t}}]\mathbb{E}[e^{i\theta Z_{s+t}}]\mathbb{E}[e^{-i\theta Y_s}]\mathbb{E}[e^{-i\theta Z_s}]}$$
$$= \tau_Y^\theta(t) + \tau_Z^\theta(t). \tag{1.5}$$

This property is important in common applications, where the observed process usually is at least to some degree disturbed by noise, which can most often be assumed to be additive and independent of the basic motion.

(c) If $\mathbb{E}[X_t^2] < \infty$, the covariance can be viewed as a limit of the codifference,

$$\lim_{\theta\to 0}\tau_X^\theta(t) = r_X(t), \tag{1.6}$$

which stems from expanding the complex exponents in definition (1.2) into a Taylor series up to the second term and noting that we obtained the logarithm of expression $(1 + \theta^2 r_X(t) + o(\theta^2))^{\theta^{-2}}$. It is then justified to treat the codifference as a generalisation of the covariance.

(d) For a Gaussian process the codifference equals the covariance for any θ

$$\tau_X^\theta(t) = r_X(t), \tag{1.7}$$

which follows immediately from a short calculation, see equation (3.6). Therefore comparing the codifference and the covariance can be used to measure non-Gaussianity.

One intuitive property, that the codifference does not have, is symmetry. Considering two variables we fix the first one and negate the second one ($x \mapsto -x$), and we expect the strength of dependence to be the same but for the sign to change. This is the case for the covariance, but not for the codifference, which is by design nonlinear. Even in the borderline case $X_{s+t} = -X_s$ we do not have a guarantee that $\tau_X^\theta(t) < 0$, counterexamples can be given even for the otherwise well-behaved class of processes considered later. It is actually possible to remove this sometimes inconvenient property by introducing the symmetrised codifference

$$\widetilde{\tau}_X^\theta(t) := \frac{1}{2\theta^2}\ln\frac{\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}]}{\mathbb{E}[e^{i\theta(X_{s+t}+X_s)}]} \tag{1.8}$$

which for all symmetric distributions changes sign with respect to reflection, $X_{s+t} \mapsto -X_{s+t}$. This quantity can be useful if one wants to compare the strength of positive and negative dependencies, but there is a cost: the

symmetrised codifference is 'linear enough' to ignore many types of nonlinear ergodicity breaking, similarly to the covariance, see equation (3.12). For this reason further on we will use the non-symmetrised codifference and study systems with a positive type of dependence, at least in some suitable limit, such as $t \to \infty$.

Note that if the codifference is a generalisation of the covariance, one should reasonably expect that there exists a generalisation of the MSD defined in a similar spirit. Indeed, let us consider the formula

$$\zeta_X^\theta(t) := -\frac{2}{\theta^2} \ln \mathbb{E}[e^{i\theta(X_t - \mu_X(t))}]. \tag{1.9}$$

This quantity may seem trivial, because studying the distribution in Fourier space is a classical method of basic probability theory. But, the distinguishing part of this definition is that the result is treated primarily as a function of time and it is conveniently transformed, so that it can be interpreted as a measure of dispersion with the same unit as X^2. Up to a rescaling it can be considered a cumulant generating function calculated at imaginary argument, but such a quantity does not seem to have an established name in the literature, so we will call it by the straightforward term 'log characteristic function', in short LCF. It is clear that in analogy to the features of the codifference, the LCF measures mainly the spread of the bulk of the probability and is much less influenced by the distribution's tails than the MSD. As before, the first factor, here $2/\theta^2$, is optional and only needed when one wants to compare the LCF to the MSD.

The LCF is indeed a reasonable measure of dispersion, as shown by the following properties.

(a) For independent Y_t, Z_t and $X_t = Y_t + Z_t$,

$$\zeta_X^\theta(t) = \zeta_Y^\theta(t) + \zeta_Z^\theta(t). \tag{1.10}$$

(b) For any Gaussian process the LCF equals the MSD,

$$\zeta_X^\theta(t) = \delta_X^2(t). \tag{1.11}$$

(c) As we stretch the probability density of X_t, the LCF diverges, that is,

$$\lim_{c \to \infty} \zeta_{cX}^\theta(t) = -\frac{2}{\theta^2} \ln \lim_{c \to \infty} \mathbb{E}[e^{i\theta c X_t}] = -\frac{2}{\theta^2} \ln 0^+ = \infty. \tag{1.12}$$

The first two facts are analogues of the corresponding properties of the codifference which allow one to trace the influence of the noise and detect non-Gaussianity. The point (c) is just the Riemann–Lebesgue lemma in disguise: it corresponds to the intuition that the rescaled process should have a larger spread. It should be mentioned that in general the LCF can be negative or complex valued, which is highly undesirable. However, for the considered models, which are based on internal Gaussian dynamics, this will never be the case, as proved in proposition 2.

Decomposing any process with independent increments into a sum of its jumps shows that in this case $\zeta_X^\theta(t)$ is a linear function. In particular, this holds of Lévy flights [37]. It also holds for continuous time random walks with exponential waiting times [5], for which

$$\zeta_X^\theta(t) = \frac{1 - \mathbb{E}[e^{i\theta J}]}{2\theta^2 \mathbb{E}[T]} t, \tag{1.13}$$

where J is one jump and T is one waiting time of diffusion X. The dependence on T is the same as for the MSD,

$$\delta_X^2(t) = \frac{\mathbb{E}[J]}{\mathbb{E}[T]} t, \tag{1.14}$$

only the scaling depending on J's distribution changes from nonlinear to linear.

The LCF can also be used for finite- or infinite-variance models which are 'anomalous' in some sense. A basic example is fractional Lévy stable motion L_α^H [46]. It is stable and self-similar which implies that

$$\zeta_{L_\alpha^H}^\theta(t) = C_\theta t^{\alpha H}, \tag{1.15}$$

for some constant C_θ, which depends on the chosen normalisation. This formula agrees with the intuition that a measure of the spread in this case should behave like a power law. Somewhat surprisingly, the situation is different for continuous time random walks with power-law waiting times, which are used to model subdiffusion. Such processes after rescaling converge to subordinated Brownian motion $B(S_\alpha(t))$, for which the LCF can be calculated directly, using the well-known properties of the inverse α-stable subordinator S_α [47],

$$\zeta^{\theta}_{B(S_\alpha)}(t) = -\frac{2}{\theta^2} \ln E_\alpha\left(-\frac{\theta^2}{2}t^\alpha\right),$$ (1.16)

where E_α is the Mittag-Leffler function [48]. This function approaches infinity like a logarithm; the exact asymptotic is shown in equation (2.3). The difference between these two models of anomalous diffusion is that L_α^H is self-similar, so its PDF spreads in the uniform manner, whereas for $B(S_\alpha)$ the bulk is much more constrained than the tails.

After this brief discussion about the general properties of the codifference and related notions, we will study its behaviour in more detail for models based on random parameters of motion and for models based on random and time-varying diffusion coefficient. The next section 2 provides a general physical overview and concrete examples useful for the modelling. The third and the last section 3 is dedicated to presenting mathematical results and calculation techniques. The paper is written such that, if the reader prefers, the physical and mathematical sections 2 and 3 can be read independently.

2. Modelling

2.1. Gaussian diffusion governed by random parameters

One of the core concepts behind ergodicity and ergodicity breaking is the idea of looking at information contained in a single trajectory. We speak about ergodicity if the data that can possibly be gained analysing one, sufficiently long, series of observations, is the same as if one analyses all possible trajectories in the ensemble [49]. Conversely, if this amount of information is smaller, we speak about ergodicity breaking. In other words, there is some information contained in a given trajectory, and using only a single trajectory we omit the amount contained in the rest. This is sometimes also rephrased as confinement in the phase space, but this language must be used carefully as the said space has a subtle structure[6].

From a different perspective, modelling based on the information content often leads to an intuitive description, because the differences between trajectories often stem from differences between diffusing particles and differences between their local surroundings. Both may occur, e.g. in biological systems. The latter case requires the additional assumption that the inhomogeneity present in the surroundings varies on a length scale of the mean distance between trajectories, but does not vary much at the scale of the trajectories themselves. That is, distinct trajectories have distinct surroundings, but each particle is sufficiently localised so that the state of the medium around it does note change significantly. This is reasonable for example when the particles are trapped or the measurement time is sufficiently short—compare, e.g. the absolute spread of the traced particles in [33].

In any case, this information can be parametrised, which leads to the so-called hierarchical or multilevel modelling [51], which in the context of physics is also called 'superstatistics' (a short term for 'superposition of statistics') [52]. Deterministic parameters of the basic model become random on an additional statistical layer.

2.1.1. Random diffusion coefficient

For diffusion the simplest example of an hierarchical model is the motion with a random diffusion coefficient, the situation when different trajectories depict movements with varying average mobilities. A typical model of such observations is the grey Brownian motion [53–55]

$$B_{2H,\beta}(t) = \sqrt{D_\beta}B_H(t).$$ (2.1)

Here B_H is fractional Brownian motion [56] and the diffusion coefficient D_β is an independent random variable with the so-called β M-Wright distribution [57]. The moments of grey Brownian motion are the same as those of fractional Brownian motion up to a multiplicative constant, therefore the MSD still grows as t^{2H} and the process models anomalous diffusion. Nevertheless, a straightforward calculation yields that the LCF can be expressed using the Mittag-Leffler function,

$$\zeta^{\theta}_{B_{2H,\beta}}(t) = -\frac{2}{\theta^2} \ln E_\beta\left(-\frac{\theta^2}{2}t^{2H}\right) \sim \frac{1}{\Gamma(\beta + 1)}t^{2H}, \quad t \to 0^+,$$ (2.2)

which also yields

$$\zeta^{\theta}_{B_{2H,\beta}}(t) \in \frac{4H}{\theta^2} \ln t + \frac{2}{\theta^2} \ln\left(\frac{\theta^2\Gamma(1-\beta)}{2}\right) + o(1), \quad \beta \neq 1, t \to \infty.$$ (2.3)

Here the asymptotic '$+o(1)$' is pointwise, which is stronger than the asymptotic proportionality '$\sim$'; in the sense of '$\sim$' the term $4H/\theta^2 \ln t$ is dominating and the logarithmic behaviour clearly distinguishes the LCF from the power-law MSD at long times. This crossover behaviour can be used to distinguish grey Brownian motion

---

[6] Even for classical Brownian motion it is the infinitely dimensional Wiener space [50].

from fractional Brownian motion (case $\beta = 1$ [53]) and diffusing-diffusivity model (equations (2.27) and (2.28)). The very slow log increase of the LCF is not surprising: because the diffusion constant is random, but fixed and it constrains the relaxation of the probability density—it is detected by the LCF, but ignored by the MSD; for a more general result see proposition 7(d).

Grey Brownian motion models free, unconfined movements and is therefore not stationary. Still, the codifference can be used for its increments $\Delta B_{2H,\beta}(t) := B_{2H,\beta}(t + \Delta t) - B_{2H,\beta}(t)$. The calculation is again not hard and yields

$$\tau^{\theta}_{\Delta B_{2H,\beta}}(t) = \frac{1}{\theta^2} \ln \frac{E_\beta(-\theta^2(\Delta t^{2H} - (|t + \Delta t|^{2H} + |t - \Delta t|^{2H})/2))}{E_\beta(-\theta^2 \Delta t^{2H}/2)}. \tag{2.4}$$

The covariance decays to zero like a power law t^{2H-1}, but the function above decays to the non-zero constant

$$\tau^{\theta}_{\Delta B_{2H,\beta}}(\infty) = \frac{1}{\theta^2} \ln \frac{E_\beta(-\theta^2 \Delta t^{2H})}{E_\beta(-\theta^2 \Delta t^{2H}/2)}. \tag{2.5}$$

This means that there is some degree of dependence left even at $t = \infty$ which the covariance does not detect, but the codifference does. Indeed, it can be interpreted as a joint dependency on the trajectory-wise fixed but random diffusion coefficient D_β.

The above simple example shows that the codifference does not directly detect non-ergodicity, it rather detects dependence. The notion of mixing is useful to describe this idea. It is a property which states that the future evolution of the process after a long delay becomes independent of its past values. Formally speaking, the process is mixing when, if we calculate some statistic in some finite time interval starting at s, and later on any other statistic starting at $s + t$, these two must become independent as $t \to \infty$ [58]. Therefore, analysing the codifference, which measures the dependence between $\exp(-i\theta X_s)$ and $\exp(i\theta X_{s+t})$, allows one to exclude mixing, i.e. to indicate the presence of a non-vanishing dependence. The latter means that the motion is constrained in phase space, which in turn implies ergodicity breaking[7].

Thus, for a very large class of systems one does not need to study time-averages to detect non-ergodicity. It is sufficient to find a proper memory function which will indicate non-mixing. As we demonstrate the covariance fails in this role for the considered models, but the codifference works.

These detecting capabilities of the codifference work under quite general circumstances. If we observe any ensemble of mixing, zero mean Gaussian trajectories, the covariance will converge to zero. This happens because for Gaussian process, mixing is equivalent to a decay of the covariance [58, 59], and the mixture of decaying covariance functions is decaying. But, the ensemble of trajectories as a whole will not be ergodic, which will not be detected by the covariance. Let $\mathcal{C}$ is some parametrisation of this mixture, then the conditional average $D = \mathbb{E}[X_t^2|\mathcal{C}]$ be the resulting, possibly random, conditional variance. We call it D because if the data X corresponds to the velocity or increments of displacements, it will be proportional to the diffusion coefficient. Under these assumptions the codifference converges to the constant

$$\tau^{\theta}_X(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D}]}{\mathbb{E}[e^{-\theta^2 D/2}]^2} \geqslant 0, \tag{2.6}$$

as proven in proposition 5. This quantity is related to the coefficient of variation defined as the standard deviation divided by the mean [23]. Denoting it by $CV[X]$, the formula above can be expressed as $\theta^{-2} \ln(CV[\exp(-\theta^2 D/2)]^2 + 1)$ which is an increasing function of $CV[\exp(-\theta^2 D/2)]$ and asymptotically quadratic for small CV. The coefficient of variation is a measure of dispersion, hence so is $\tau^{\theta}_X(\infty)$ which reflects the randomness of D. This behaviour is also equivalent to detecting a residual dependence and the resulting non-mixing/non-ergodicity.

Outside of the useful limit $t = \infty$ not much can be said about the properties of the codifference in such a wide and general class. The situation changes if we consider a more specific model. The idea behind grey Brownian motion and many works about superstatistics [52] is that the trajectories differ mainly by the diffusion coefficient, other properties are not significantly distinct. A simple model of such a system can be written as

$$X_t = \sqrt{D}\, Y_t. \tag{2.7}$$

We assume that the process Y describes the joint form of dependence common for all trajectories. We consider a Gaussian Y, which for grey Brownian motion would be fractional Brownian motion. Another reasonable choice would be, e.g. a solution of the Langevin equation. In this case, as long as Y is stationary (i.e. for free diffusion we consider increments or the velocity process), the covariance is

---

[7] The remaining class of processes which are ergodic but non-mixing is complicated and those do not seem to appear in applications. For a mathematically constructed example of such a process and the discussion see [59].

Table 1. Formulae for the codifference and the LCF corresponding to common models of D: gamma, one-sided stable, Gaussian and uniform.

Law of D	Codifference $\tau_X^\theta(t)$	LCF $\zeta_X^\theta(t)$
$\mathcal{G}(\alpha, \beta)$	$\dfrac{\alpha}{\theta^2} \ln \dfrac{(1 + \theta^2/(2\beta))^2}{1 + \theta^2(1 - r_Y(t))/\beta}$	$\dfrac{2\alpha}{\theta^2} \ln\left(\dfrac{\theta^2}{2\beta}\delta_Y^2(t) + 1\right)$
$\mathcal{S}(\alpha, c)$	$c^\alpha \theta^{2\alpha-2}(2^{1-\alpha} - (1 - r_Y(t))^\alpha)$	$2^{1-\alpha} c^\alpha \theta^{2\alpha-2}(\delta_Y^2(t))^\alpha$
$\mathcal{N}(\mu, \sigma^2)$	$\mu r_Y(t) + \dfrac{(\theta\sigma)^2}{2}(1 - r_Y(t))^2 - \mu - \left(\dfrac{\theta^3}{8}\sigma^2 - \dfrac{\theta}{2}\mu\right)^2$	$\mu \delta_Y^2(t) - \dfrac{(\theta\sigma)^2}{4}(\delta_Y^2(t))^2$
$\mathcal{U}(a, b)$	$a r_Y(t) + \dfrac{1}{\theta^2} \ln\left(\dfrac{\theta^2(b-a)}{4(1 - r_Y(t))}\dfrac{1 - e^{-\theta^2(b-a)(1-r_Y(t))}}{(1 - e^{-\theta^2(b-a)/2})^2}\right)$	$a\delta_Y^2(t) - \dfrac{2}{\theta^2} \ln\left(\dfrac{2(1 - e^{-\theta^2(b-a)\delta_Y^2(t)/2})}{\theta^2\delta_Y^2(t)(b-a)}\right)$

Figure 1. Codifference τ and covariance r of the process $X_t = \sqrt{D}\,Y_t$ with $D \stackrel{d}{=} \mathcal{G}(1, 1)$ and $r_Y(t) = \cos(t)\exp(-t)$, as given by table 1. Various properties of the codifference are visible: for $\theta \to 0$ it converges to the covariance; the codifference and the covariance increase and decay in the same intervals; at $t = 0$ the codifference is smaller than the covariance; as $t \to \infty$ the codifference converges to a θ-dependent value $\tau_X^\theta(\infty)$ which is a functional of the law of D; the type of asymptotic of $r_X(t)$ and $\tau_X^\theta(t) - \tau_X^\theta(\infty)$ is the same (here: exponential decay). The derivations are presented in proposition 6.

$$r_X(t) = \mathbb{E}[D]\, r_Y(t), \tag{2.8}$$

of course as long as $\mathbb{E}[D] < \infty$. If the process Y has sufficiently long memory, $r_Y(t) \approx 0$ in the considered time scale, also $r_X(t) \approx 0$. The covariance does not detect the additional dependence introduced by random D.

At the same time the codifference can be expressed as a function of the covariance of Y, precisely as

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D(1 - r_Y(t))}]}{\mathbb{E}[e^{-\theta^2 D/2}]^2} \tag{2.9}$$

for any D, no matter if $\mathbb{E}[D] < \infty$. It clearly converges to the constant (2.6) as $r_Y(t) \to 0$ and detects the additional nonlinear dependence.

For a general, possibly non-stationary Y with $\mathbb{E}[Y_t^2] = \delta_Y^2(t)$, the representation of the LCF is

$$\zeta_X^\theta(t) = -\frac{2}{\theta^2} \ln \mathbb{E}[e^{-\theta^2 D \delta_Y^2(t)/2}]. \tag{2.10}$$

Given some model of D these formulae can be made completely explicit, examples are given in table 1. The first example is the gamma distribution $D \stackrel{d}{=} \mathcal{G}(\alpha, \beta)$ in which the coefficient α describes the power-law behaviour of the PDF near 0 and β is the rate of exponential decay of the tails (the specific case $\mathcal{G}(1, \beta)$ is the exponential distribution); it models common types of experiments in which the distribution of diffusion coefficients resembles a bump concentrated around some finite constant and high values of D become exponentially less probable. This case is also illustrated in figure 1.

Diffusion coefficients with a heavy-tailed distribution result in a motion that itself exhibits heavy tails of the PDF, a phenomenon actively investigated in transport, finance, turbulence and many other systems [6, 60, 61]. A classical model of this case is the one-sided α-stable subordinator $\mathcal{S}(\alpha, c)$, determined by its Laplace transform $\exp(-(cs)^\alpha)$. The resulting type of process was thoroughly studied in the literature concerned with stable distributions [37]. This process is called sub-Gaussian, which is arguably a confusing term. In this case the process X has no second moment, therefore attempts to estimate its covariance will lead to a diverging result. This is visible in the formulae for the codifference and the LCF, which diverge as $\theta \to 0$. But, for any $\theta > 0$ the

codifference and the LCF are finite and can be estimated in a standard way, and from the result if one wishes the covariance and the MSD of Y can be reconstructed.

For a distribution concentrated around its mean value one can use Gaussian $\mathcal{N}(\mu, \sigma^2)$ or uniform $\mathcal{U}(a, b)$ distributions, however the former is only a valid model for $\sigma \ll \mu$, when the probability that $D < 0$ can be neglected.

Even if the precise model of D is not known, quite a lot can be said about the behaviour of the codifference. In proposition 6 we show that:

(a) the codifference is a monotonic function of the covariance. If one increases, the second one also increases, the same goes for decreases;

(b) if $\mathbb{E}[D] < \infty$ the codifference is smaller than the covariance for strong positive correlation, but larger for weak or negative correlations;

(c) the approach to the value $\tau_X^\theta(\infty)$ has the same asymptotic as the decay of the covariance

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) \sim \frac{\mathbb{E}[De^{-\theta^2 D}]}{\mathbb{E}[e^{-\theta^2 D}]} r_Y(t), \quad t \to \infty, \tag{2.11}$$

assuming $r_Y(t) \to 0$, which is a typical case.

These are all desirable properties: the memory structure of the internal process Y is reflected in a straightforward manner by the codifference. For small values of the covariance their relation is even linear, as stated in (c), and the proportionality constant is finite for any distribution of D, due to the truncating factor $\exp(-\theta^2 D)$.

Another property is that the codifference depends additively on D. Precisely speaking, if we decompose $D = D' + D''$ for some independent D' and D'', the codifference also decomposes for

$$\tau_X^\theta(t) = \tau_{X'}^\theta(t) + \tau_{X''}^\theta(t), \tag{2.12}$$

where X' and X'' are processes with diffusion coefficients D' and D' respectively. Therefore subtracting the codifferences estimated from different samples may be used to analyse different sources of diffusivity. The derivation is given in proposition 6.

Analogous features can also be checked for the LCF (proposition 7), which can also be decomposed for $D = D' + D''$ and is a monotonic function of the MSD, but is always smaller than the MSD, therefore detecting the additional constraints of the motion introduced by a random D.

At the end of the discussion about random diffusion coefficients we note that the behaviour of the codifference near $t = 0$ can also give valuable information. In proposition 8 we prove that for a typical case when $\mathbb{E}[D] < \infty$ its asymptotic reflects that of the covariance. However, if $\mathbb{E}[D] = \infty$ and D has power tails, corresponding to the presence of high-volatility trajectories, the asymptotic of the codifference has an additional power law. As for Gaussian processes the behaviour of the covariance near $t = 0$ is determined by their fractal dimension [62], chapter 8.8, the same is true for the codifference, which can be applied also for processes with no moments.

2.1.2. Random memory decay rate

Another interesting type of models are ensembles of particles for which the time dependence may vary from trajectory to trajectory. The simplest model of a time-varying dependency is the exponential decay $\exp(-t\Lambda)$, which is the covariance of Ornstein–Uhlenbeck process [63]. It models many kinds of linear relaxation disturbed by additive noise. It was also studied as a model of the additive measurement noise itself [64, 65]. In the hierarchical model the decay rate Λ may be random. The covariance of the resulting mixture of Ornstein–Uhlenbeck type trajectories was studied in [66] in the context of a randomly parametrised Langevin equation.

The coefficient Λ has a different physical interpretation depending on the details of the studied phenomenon. For the velocity of a Brownian particle it is proportional to the friction coefficient and its randomness is related to local changes of the viscosity and/or different shapes of the diffusing particles [67]; in this system the fluctuation-dissipation relation also links the scaling to the temperature. For trapped particles Λ is proportional to the stiffness of the confining harmonic potential (the prominent example being optical tweezers [21, 68]), therefore the randomness of Λ is equivalent to an ensemble of traps with varying sizes, which are proportional to Λ^{-1}.

Another case worth mentioning is that of viscoelastic anomalous diffusion [69], for which the velocity (or increments) have power-law dependence $\propto t^{2H-1}$. This function can be expressed as $\exp(-\ln(t)(1 - 2H))$. Therefore it is enough to replace t with $\ln t$ and the results further on will also follow for the ensemble of power-

law memory trajectories characterised by random parameter $(1 - 2H)$. It is worth to note that the variability of the of the Hurst index H seems to be more of a rule than an exception for biological systems [70–72].

We do not want to make the discussion overly technical, so below we will analyse only the case of deterministic scaling and random decay rate, $r_X(t|\mathcal{C}) = \sigma^2 \exp(-t\Lambda)$. Results for more general $Df(\Lambda)\exp(-t\Lambda)$ are presented in propositions 10, 11 and 12, which prove that the randomness of the scaling is not essential for most of the properties discussed below. We also note that sometimes one can remove the random scaling and normalise the trajectories using the estimate of scaling obtained from the Birkhoff ergodic theorem [58],

$$r_X(0|\mathcal{C}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t \ X(t)^2. \tag{2.13}$$

However, this procedure requires having access to sufficiently long trajectories.

A particular property of ensembles with fixed scaling is that any marginal distribution is Gaussian, i.e. all variables X_t have Gaussian distribution with variance σ^2. But the codifference can be found to be

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln \mathbb{E}[e^{(\theta\sigma)^2 e^{-t\Lambda}}], \tag{2.14}$$

and because it does not equal the covariance, the process as a whole is not Gaussian. The codifference indicates the presence of subtle non-Gaussianity of the memory structure. This formula can also be used to derive useful bounds between the codifference and the covariance, see proposition 9.

Expanding in a Taylor series the exponent from (2.14) leads to

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln\left(1 + \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} \mathbb{E}[e^{-kt\Lambda}]\right) \sim \frac{1}{\theta^2} \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} \mathbb{E}[e^{-kt\Lambda}], \quad t \to \infty. \tag{2.15}$$

Note that $\sigma^2\mathbb{E}[e^{-kt\Lambda}] = r_X(kt)$, so the result is a type of average over the values $r_X(kt)$. When the distribution of Λ is not sufficiently concentrated near 0 and the covariance decays fast (strictly speaking is rapidly varying [73, 74]), the term $k = 1$ dominates the $t \to \infty$ asymptotic. This is the case, e.g. for the one-sided stable variable $\Lambda \overset{d}{=} \mathcal{S}(\alpha, c)$ for which

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln\left(1 + \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} e^{-(ckt)^\alpha}\right) \sim \sigma^2 e^{-(ct)^\alpha}, \quad t \to \infty \tag{2.16}$$

that is, we observe a stretched exponential type of dependence.

When Λ is more concentrated around 0 the situation differs. A basic example would again be the gamma distribution $\Lambda \overset{d}{=} \mathcal{G}(\alpha, \beta)$, for which

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln\left(1 + \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} \frac{1}{(1 + kt/\beta)^\alpha}\right). \tag{2.17}$$

When $\alpha = 1$ (i.e. Λ has an exponential distribution) the above can also be written using the incomplete gamma function. For any α all terms in the sum decay like $t^{-\alpha}$ and they are comparable. Because of this, the codifference also decays with the same power law, but the proportionality constant is non-trivial,

$$\tau_X^\theta(t) \sim \frac{1}{\theta^2} \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} \frac{1}{\left(\frac{1}{t} + k/\beta\right)^\alpha} t^{-\alpha} \sim \frac{1}{\theta^2\beta^\alpha} \sum_{k=1}^\infty \frac{(\theta\sigma)^{2k}}{k!} \frac{1}{k^\alpha} t^{-\alpha}, \quad t \to \infty. \tag{2.18}$$

It is not surprising that this behaviour is not specific to a gamma distribution and can be observed for any Λ with power-law PDF near 0^+, see proposition 10. Similarly, if the PDF of Λ decays fast near 0^+, the codifference also decays fast. All these properties are analogous to those of the covariance [66], so here they can be used interchangeably or simultaneously, as a mean to obtain stronger statistical verification.

They are also similar in that both do not detect the non-ergodicity, more precisely the non-mixing, of this system. As was already demonstrated for the covariance it is a common occurrence resulting from its linearity. The codifference fails, because it does measure only a reduced form of mixing. For the process to be mixing it means that any two sets of multiple disjoint measurements must become asymptotically independent, i.e. the vectors $[X_{s_1}, X_{s_2}, \ldots, X_{s_n}]$ and $[X_{s_1+t}, X_{s_2+t}, \ldots, X_{s_n+t}]$ have to become independent as $t \to \infty$. The codifference (and for that matter also the covariance) measures only the dependence between two values X_s and X_{s+t}.

For a process with a random decay rate these are asymptotically independent and the one-point distributions are relaxing. Therefore, in order to detect non-ergodicity, we need to analyse the dependence between at least three values. A practical choice is to use four values divided into two pairs $[X_s, X_{s+\Delta t}]$ and $[X_{s+t}, X_{s+\Delta t+t}]$. The values in the first pair are correlated as $e^{-\Delta t\Lambda}$ trajectory-wise, analogously for the values of the second pair. This property of both pairs is fixed and random, i.e. it is a constant of motion which can be detected. Probably the
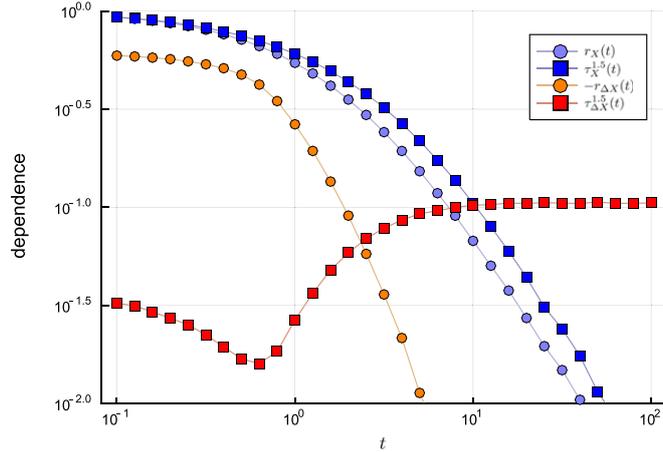
Figure 2. Estimated codifference τ and covariance r estimated from the process with random decay rate $\Lambda \overset{d}{=} \mathcal{G}(3/2, 1/2)$ and $\Delta t = 1$. In the presented domain the covariance $r_{\Delta X}$ was negative, so we plotted the negated value. One can observe the predicted power law decays $\propto t^{-3/2}$ and $\propto t^{-3/2-2}$ (equation (3.62)); the codifference of increments detects the non-ergodicity by converging to a constant $\tau_{\Delta X}^{1.5}(\infty) \approx 0.105$, which fits perfectly equation (2.20). The value $\theta = 1.5$ was chosen to best illustrate the interesting properties; for smaller θ the codifference τ_X^θ becomes closer to the covariance r_X, for larger θ the codifference $\tau_{\Delta X}^\theta$ converges faster to the $t \to \infty$ limit. To present smooth curves in the whole presented range we used a large 10^7 sample; the general shape of the presented functions is already visible for samples around 10^4; a significant difference between $r_{\Delta X}$ and $\tau_{\Delta X}$ is observed using even a few hundred trajectories. Examples for smaller sample sizes are presented in figure A1.

simplest method to achieve this is to calculate increments

$$\Delta X_t := X_{t+\Delta t} - X_t \tag{2.19}$$

and study the codifference of those. A short calculation given in proposition 11 shows that this method indeed works and

$$\tau_{\Delta X}^\theta(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-2(\theta\sigma)^2(1-e^{-\Delta t\Lambda})}]}{\mathbb{E}[e^{-(\theta\sigma)^2(1-e^{-\Delta t\Lambda})}]^2} \geqslant 0. \tag{2.20}$$

The result depends on Λ in a complex manner, but it can be easily estimated numerically. We can also use the fact that for small Δt the conditional covariance of increments is

$$r_{\Delta X}(t|\Lambda, D) \in 2\Delta t^2 \sigma^2 \Lambda^2 e^{-t\Lambda} + \mathcal{O}(\Delta t^4) \tag{2.21}$$

and normalise the process, $\Delta \widetilde{X}_t := \Delta X_t / \sqrt{\Delta t}$. The result then simplifies and becomes independent of Δt,

$$\lim_{\Delta t \to 0^+} \tau_{\Delta \widetilde{X}}^\theta(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-2(\theta\sigma)^2\Lambda}]}{\mathbb{E}[e^{-(\theta\sigma)^2\Lambda}]^2}. \tag{2.22}$$
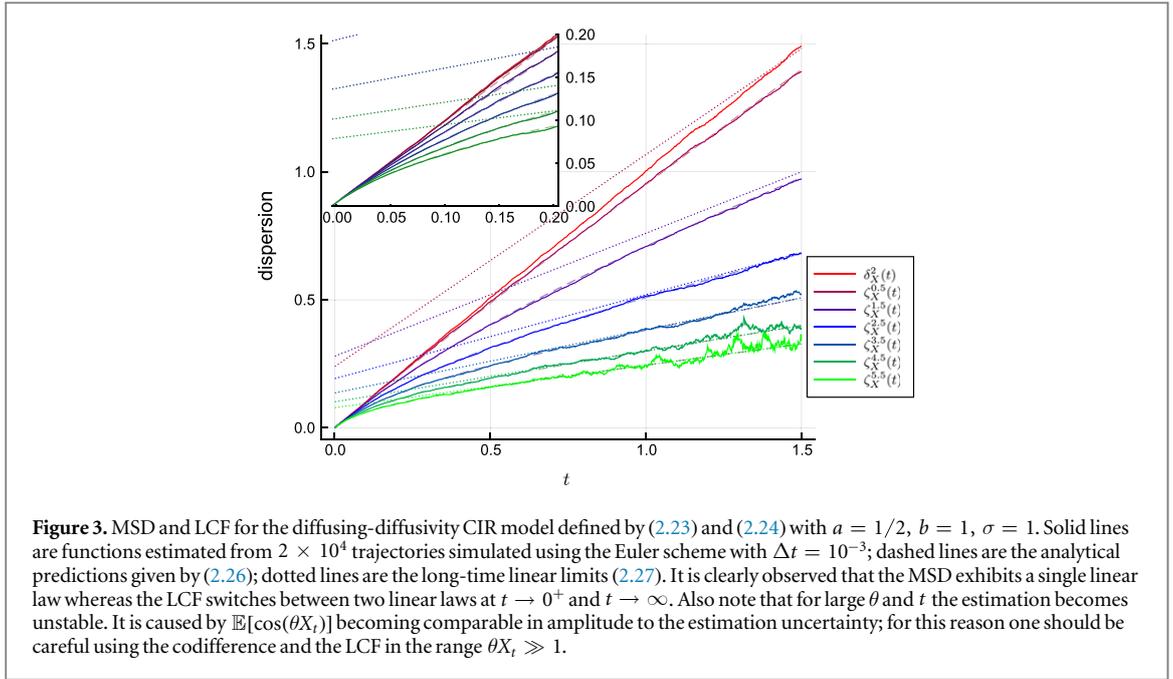
We stress here that this method cannot be applied using the covariance, which, calculated from increments, decays to 0 and does not detect this specific memory structure. Its decay is even quicker than for the original process and proportional to the power law decay $t^{-\alpha-2}$ [66]. Intuitively speaking, the decay rate is quicker by a factor t^{-2}, because the scale of ΔX depends on Λ as Λ^2 and the trajectories with stronger correlation have smaller amplitude and add less to the average. This property has its analogy for the codifference, for which $\tau_{\Delta X}^\theta(t) - \tau_{\Delta X}^\theta(\infty)$ also decays like $t^{-\alpha-2}$ (see proposition 12 for a more general result). This time the faster decay rate actually helps in detecting ergodicity breaking, making the limit $\tau_{\Delta X}^\theta(\infty)$ visible even at short times. The numerical illustration of the discussed behaviour is shown in figure 2.

2.2. Diffusing-diffusivity

In the preceding sections we considered models which were non-Gaussian and non-ergodic. For non-Gaussian but ergodic models the codifference can also be a useful measure of dependence. In particular we show that it can be successfully used to analyse diffusing-diffusivity models. We now assume that the increments of X_t are Brownian fluctuations, but rescaled by a time-dependent random diffusivity D_t,

$$dX_t = \sqrt{D_t} dB_t. \tag{2.23}$$

This is a generalisation of the random parameter model, for which $D_t = \text{const}$. Because we modified the dynamical equation by replacing the previously constant parameter with a stochastic process, models of this class are sometimes called 'doubly stochastic' [75]. Before application in physics, they were extensively used in

Figure 3. MSD and LCF for the diffusing-diffusivity CIR model defined by (2.23) and (2.24) with $a = 1/2$, $b = 1$, $\sigma = 1$. Solid lines are functions estimated from 2×10^4 trajectories simulated using the Euler scheme with $\Delta t = 10^{-3}$; dashed lines are the analytical predictions given by (2.26); dotted lines are the long-time linear limits (2.27). It is clearly observed that the MSD exhibits a single linear law whereas the LCF switches between two linear laws at $t \to 0^+$ and $t \to \infty$. Also note that for large θ and t the estimation becomes unstable. It is caused by $\mathbb{E}[\cos(\theta X_t)]$ becoming comparable in amplitude to the estimation uncertainty; for this reason one should be careful using the codifference and the LCF in the range $\theta X_t \gg 1$.

financial engineering, where it is natural to assume that parameters of the market, such as the volatility, vary in time. In 1985 Cox *et al* [76] proposed a model of interest rate (now commonly named CIR), which describes a non-negative stochastic process with linear mean-reverting property. In 2012 Chubynsky and Slater independently proposed a special case of the CIR process as a model of non-Gaussian diffusion [19, 77]. This led the way to a wider range of models based on fluctuating diffusivity coefficient with a short time memory [78–82]. The evolution of the diffusion coefficient in the CIR model is defined by the stochastic equation

$$\mathrm{d}D_t = a(b - D_t)\mathrm{d}t + \sigma\sqrt{D_t}\,\mathrm{d}B_t, \tag{2.24}$$

where $a > 0$ describes the speed of return to the mean $b > 0$, and $\sigma > 0$ regulates the amplitude of the fluctuations. In this equation as $D_t \to 0$ the term $a(b - D_t)\mathrm{d}t \approx ab\mathrm{d}t > 0$ starts to dominate the fluctuations with the mean-squared amplitude $\mathbb{E}[(\sqrt{D_t}\,\mathrm{d}B_t)^2] = D_t\mathrm{d}t$, consequently $\mathrm{d}D_t > 0$ which causes the motion to stay positive. We assume that the system evolved for a long time before the start of the measurement and has reached the stationary gamma distribution $D_0 \overset{d}{=} \mathcal{G}(2ab/\sigma^2, 2a/\sigma^2)$ [83]. Because of the non-Gaussianity the LCF function should differ from the MSD. Conditioning by D_t, it can be expressed by the formula

$$\zeta_X^\theta(t) = -\frac{2}{\theta^2}\ln\mathbb{E}\left[\exp\left(-\frac{\theta^2}{2}\int_0^t \mathrm{d}s\ D_s\right)\right]. \tag{2.25}$$

Expanding the above in powers of θ^2 shows that again $\zeta_X^\theta(t) \to \delta_X^2(t)$ as $\theta \to 0$.

The average in (2.25) appears in the calculation of the expected price of zero-coupon bond and was calculated in the initial paper of Cox *et al* [76], who derived the differential equation which it fulfils and then solved it; a more general result is also available in [83]. The calculation was performed for the case when D_0 is fixed and deterministic, however their result can be easily extended for stationary D by averaging over the equilibrium $\mathcal{G}(2ab/\sigma^2, 2a/\sigma^2)$ distribution of D_0. Then the formula for the LCF reads

$$\zeta_X^\theta(t) = \frac{4ab}{(\theta\sigma)^2}\ln\left(\left(\frac{1}{2} + \frac{(\theta\sigma)^2}{4\gamma_\theta a} + \frac{a}{2\gamma_\theta}\right)e^{(\gamma_\theta - a)t/2} + \left(\frac{1}{2} - \frac{(\theta\sigma)^2}{4\gamma_\theta a} - \frac{a}{2\gamma_\theta}\right)e^{-(\gamma_\theta + a)t/2}\right) \tag{2.26}$$
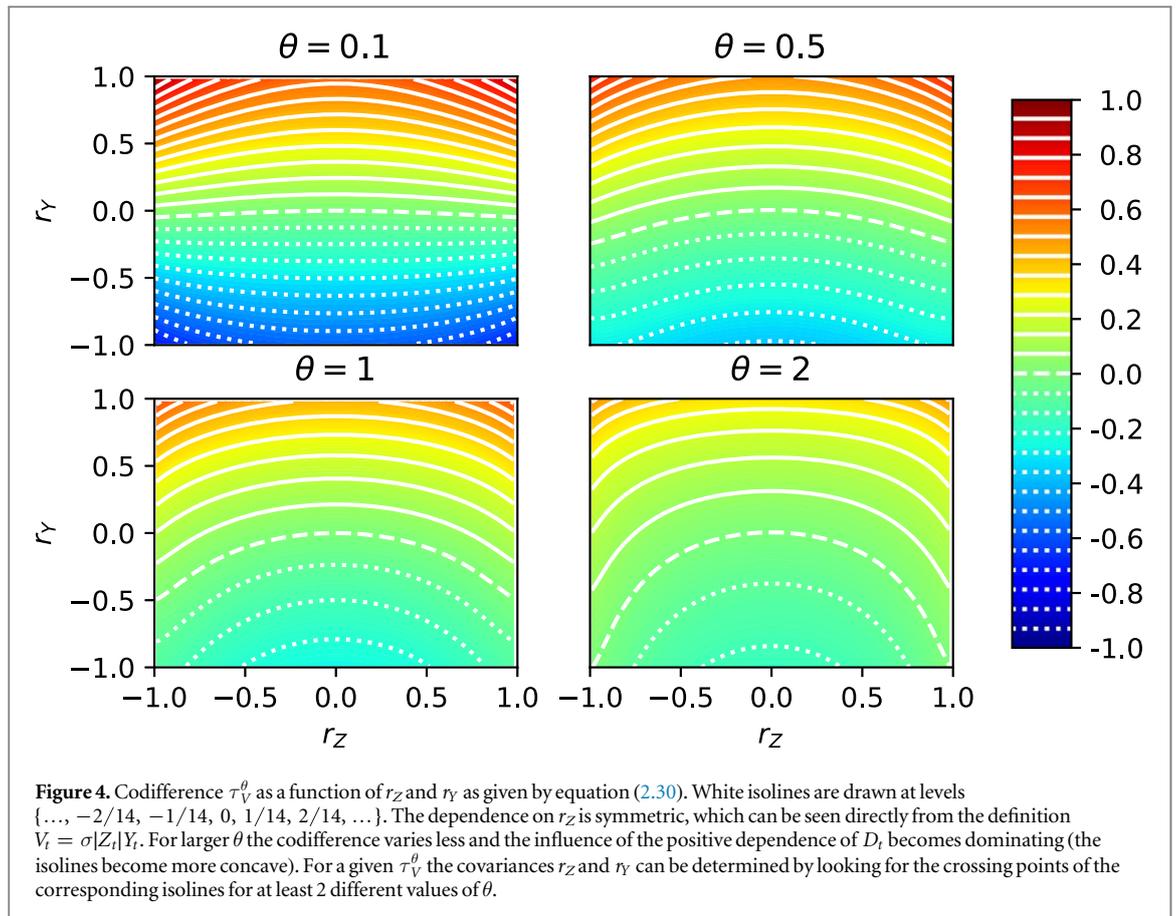
with $\gamma_\theta = \sqrt{a^2 + (\theta\sigma)^2}$. From that a brief calculation proves that the motion is Fickian for long times

$$\zeta_X^\theta(t) \in \frac{2ab}{(\theta\sigma)^2}(\gamma_\theta - a)t + \frac{4ab}{(\theta\sigma)^2}\ln\left(\frac{1}{2} + \frac{(\theta\sigma)^2}{4\gamma_\theta a} + \frac{a}{2\gamma_\theta}\right) + o(1), \quad t \to \infty, \tag{2.27}$$

and also for short time, albeit with a diffusion scale agreeing with the MSD

$$\zeta_X^\theta(t) \sim bt, \quad t \to 0^+, \tag{2.28}$$

which should come as no surprise. For an illustration of these formulae see figure 3, where we present results of Monte Carlo simulations compared to the theoretical predictions. See also the crossover behaviour of the MSD in the random diffusivity model in [81].

Figure 4. Codifference τ_V^θ as a function of r_Z and r_Y as given by equation (2.30). White isolines are drawn at levels $\{\ldots, -2/14, -1/14, 0, 1/14, 2/14, \ldots\}$. The dependence on r_Z is symmetric, which can be seen directly from the definition $V_t = \sigma|Z_t|Y_t$. For larger θ the codifference varies less and the influence of the positive dependence of D_t becomes dominating (the isolines become more concave). For a given τ_V^θ the covariances r_Z and r_Y can be determined by looking for the crossing points of the corresponding isolines for at least 2 different values of θ.

If we want to analyse the codifference of the CIR model, it would be required to study the memory of the velocity $V_t = \sqrt{D_t}\,\mathrm{d}B_t/\mathrm{d}t$. But the white noise $\mathrm{d}B_t/\mathrm{d}t$ is not well-defined in a classical sense. It can be interpreted as a distribution which leads to a similar redefinition of the covariance, the familiar Dirac delta. The codifference is, however, nonlinear and this approach fails. The solution is to consider only the well-defined velocity processes $V_t = \sqrt{D_t}\,Y_t$ with Y_t being some classical process which models the velocity as being undisturbed by the fluctuations of the diffusivity. The behaviour of the white noise can be studied if we consider t large enough such that $r_Y(t) = 0$ strictly or approximately. It is natural to assume that Y_t is Gaussian, while choosing the model of D_t is more subtle.

The CIR process for $ab \in \mathbb{N}$, can be proved to be a sum of squared independent Ornstein–Uhlenbeck processes, which follows directly from writing the stochastic differential equation of such a sum [83]. Thus, a natural generalisation is to consider D_t being a square of a Gaussian process [80, 81]. We will assume that the velocity can be decomposed as

$$V_t = \sigma|Z_t|Y_t, \tag{2.29}$$

where both Z_t and Y_t are Gaussian with variance one. In this model we have ample freedom in describing a wide range of memory types, because any covariance r_Z and r_Y can be used. By choosing r_Y we model the internal dynamics, if $r_Y(t) = 0$ in the considered time scale we arrive back at (2.23); by choosing r_Z we model the memory structure of D_t: exponential, power law, oscillating, etc. The one-dimensional distributions are more rigged, as we limit ourselves to D_t having the PDF of a square Gaussian, that is χ_1^2 distribution (a special case of the gamma distribution). A rather technical derivation (proposition 13) then shows that the exact form of the codifference is

$$\tau_V^\theta(t) = \frac{1}{\theta^2}\ln\left(\frac{\sqrt{1 - r_Z(t)^2}}{\pi}\frac{1 + (\theta\sigma)^2}{(\theta\sigma)^2(1 - r_Z(t)^2) + 1}\sum_{\rho\in\{\rho_+,\rho_-\}}\frac{\frac{\pi}{2} + \arctan\left(\frac{\rho}{\sqrt{1 - \rho^2}}\right)}{\sqrt{1 - \rho^2}}\right), \tag{2.30}$$

where

$$\rho_\pm := \frac{(\theta\sigma)^2(1 - r_Z(t)^2)r_Y(t) \pm r_Z(t)}{(\theta\sigma)^2(1 - r_Z(t)^2) + 1}.$$

This formula looks complicated, but is composed only of elementary functions. It is illustrated in figure 4, were we plotted the codifference τ_V^θ as a function of r_Z and r_Y for four different θs. Having calculated the codifference

for at least two θs, one can solve the system of equations resulting from (2.30) and calculate r_Z, r_Y. This procedure may be considered simpler than using the covariance r_Z, which requires calculating the average of $|Z_s Z_{s+t}|$ given by a hard-to-evaluate integral. The covariance r_V can also be obtained from taking the limit $\theta \to 0$ of the codifference.

More importantly, when $r_Y(t) = 0$ the codifference is clearly non-zero, so it detects the dependence introduced by $D_t = Z_t^2$. Its asymptotic for small $r_Z(t)$ (e.g. at long times) in this case is the simple relation

$$\tau_V^\theta(t) \sim \frac{\sigma^2}{2(1 + (\theta\sigma)^2)^2} r_Z(t)^2, \quad r_Z(t) \to 0. \tag{2.31}$$

Thus the codifference detects the memory structure of the time-varying diffusion coefficient $D_t = Z_t^2$ even in the regime $r_Y(t) = 0$ in which the covariance $r_V(t)$ is zero and does not contain any important information. This is also true when $r_Z(t) = 0$ but $r_Y(t) \neq 0$, this time the codifference is asymptotically proportional to $r_Y(t)$; the proportionality constant depends only on the one-dimensional distributions of D, the exact form of the dynamics does not matter, see proposition 14.

For some systems different models of D_t may be more suitable. When D_t is strongly concentrated around its mean value a possible choice is a simple Gaussian centred around some b, $V_t = (\sigma Z_t + b)Y_t$. This model permits the unphysical situation when $D_t < 0$, but when $\sigma \ll b$ the probability of this event is negligible. In this case an elementary formula for the codifference also can be given (see (3.78)) and again even for $r_Y(t) = 0$ the internal dependence of D_t is still detected, this time with asymptotic

$$\tau_V^\theta(t) \sim \frac{b^2}{(1 + (\theta\sigma)^2)^2} r_Z(t), \quad r_Z(t) \to 0. \tag{2.32}$$

2.3. Discussion

The aim of this work was to provide the theoretical background for using the codifference as a dependence measure suited for the study of various non-Gaussian and ergodicity breaking models. This goal was achieved in few steps. First we proved that the codifference has intuitive properties that one would expect from a reasonable memory function, such as additivity, positivity for the case of complete dependence and being null for the case of independence. Second, we showed that it can be calculated using fairly straightforward methods for typical random parameters and diffusing-diffusivity models, which represent a significant extension of the previously established results for stable and infinitely divisible processes. Finally, we analysed how the codifference detects forms of dependence and ergodicity breaking which cannot be easily studied using solely covariance-based methods.

We also showed one example of non-detected ergodicity breaking, the case of a Langevin equation with a random return rate. In this case we offer an easy fix: the codifference works well for the increments of this process. We note that within this paper we did not analyse ergodicity breaking caused by ageing. In principle, the codifference should work, but the analytical analysis will be challenging for many of these phenomena.

In addition to the codifference, we also discussed a related quantity, the logarithm of the characteristic function (LCF), which was interpreted as a measure of dispersion. Our contribution is an extension of the Fourier methods and a distinct view based on ideas previously developed only for heavy tailed α-stable distributions. The codifference is also very closely related to the theory of the dynamical functional, which was already successfully used for real data, and should be considered a part of the same framework.

The cost of using this technique is that linearity is a powerful analytical tool, especially for complicated models, and a significant part of this strength is lost when using the codifference. The more complicated defining formula also may make its form more complicated (e.g. see table 1). However, it is a clear application of the characteristic function which does not seem to be commonly acknowledged and the Fourier-based techniques by themselves are widely used by the scientific community. Thus, it has an advantage, offering a wide choice of established analytical methods and estimation techniques. In some cases (e.g. (2.30)) the codifference has a simpler form than the covariance.

We believe that the most important example that was considered was also the simplest: deterministic motion with its scale (diffusion coefficient) varying from trajectory to trajectory. The observed asymptotical behaviour of the codifference contains a lot of useful information and lays the foundation for possible future applications in more complex and realistic models, some of which we discussed. At the same time we stress that even this initial, highly simplified model is being commonly used, especially in biophysical systems.

We are confident that the obtained results are interesting in their own right, but we also promote their additional value by indicating the limitations of the methodology based on the MSD and the covariance. Both are, without a doubt, essential parts of the scientific language related to diffusion and complex phenomena, but their limitations are becoming more and more evident, as contemporary research starts to concentrate around

non-Gaussian systems with complicated memory structure; the change is stimulated by increasing experimental evidence. These complex and nonlinear phenomena require new complex and nonlinear methods.

3. Derivations

3.1. Basic definitions and properties

All processes considered in this work can be labelled as 'conditionally Gaussian'. In practical applications these processes are Gaussian locally, in the temporal or spatial sense. The formal definition is more general.

Definition 1. We call a process conditionally Gaussian when any of its finite-dimensional distributions is a Gaussian distribution under some conditioning by σ-algebra $\mathcal{C}$. That is, any finite dimensional distribution $\boldsymbol{X} := [X_{t_1}, \ldots, X_{t_n}]$ can be written as

$$\boldsymbol{X} = A\boldsymbol{Y} + \boldsymbol{\mu}, \tag{3.1}$$

where A and $\boldsymbol{\mu}$ are a $\mathcal{C}$-measurable $n \times n$ random matrix and an n-dimensional random vector. Both may depend on $t_1, \ldots, t_n$. The vector $\boldsymbol{Y}$ is i.i.d. $\mathcal{N}(0, 1)$ and is independent of A and $\boldsymbol{\mu}$.

If $\boldsymbol{\mu} = 0$ for any $t_1, \ldots, t_n$ we call a process conditionally centred Gaussian. Further on we will consider only this class. Similarly, we call a process conditionally stationary Gaussian, if the distribution of A and $\boldsymbol{\mu}$ does not depend on time translation $t_1, \ldots, t_n \mapsto t_1 + t, \ldots, t_n + t$.

Proposition 1. *The distribution of a conditionally Gaussian process is completely determined by the knowledge of $\mathcal{C}$, the conditional mean and the conditional covariance*

$$\mu_X(t|\mathcal{C}) = \mathbb{E}[X_t|\mathcal{C}], \quad r_X(s, t|\mathcal{C}) := \mathbb{E}[X_s X_t|\mathcal{C}]. \tag{3.2}$$

The process is conditionally centred if and only if $\mu_X(t|\mathcal{C}) = 0$. The process is conditionally stationary if and only if $\mu_X(t|\mathcal{C}) = $ const. and $r_X(s, t|\mathcal{C})$ is a function of $t - s$, denoted $r_X(t - s|\mathcal{C})$.

Proof. This is a direct consequence of the equality

$$\mathrm{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \mathbb{E}[\mathrm{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n|\mathcal{C})]. \tag{3.3}$$

The conditional probability on the right is a Gaussian integral and a function of $\mu_X(t|\mathcal{C})$ and $r_X(s, t|\mathcal{C})$. The representation of conditionally centred and stationary processes are just a reflection of the analogical representations for Gaussian processes. □

Definition 2. We define the codifference function as

$$\tau_X^\theta(s, t) := \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(X_t - X_s)}]}{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_t}]\mathbb{E}[\mathrm{e}^{-\mathrm{i}\theta X_s}]}. \tag{3.4}$$

For stationary process it is a function of $t - s$, which we denote as $\tau_X^\theta(t)$, similarly as for the covariance, see also equation (1.2).

Additionally, we define the log characteristic function (LCF) as

$$\zeta_X^\theta(t) := -\frac{2}{\theta^2} \ln \mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(X_t - \mu_X(t))}]. \tag{3.5}$$

All expected values in the above definitions are finite, but they may be complex and the denominator may be 0. This is however not the case in the class of processes considered herein.

Proposition 2. *For any conditionally centred Gaussian process the codifference and the LCF are well-defined real-valued functions.*

Proof. The Gaussian function centred at 0 is positive-definite. The mixture of positive-definite functions is positive-definite. Therefore all expected values in definition 2 are real numbers larger than 0 and less or equal 1. The logarithms are therefore real. □

We also note that for conditionally centred Gaussian processes a reduced formula for the codifference is available,

$$\tau_X^\theta(s,\,t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(X_t - X_s)}|\mathcal{C}]]}{\mathbb{E}[\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_t}|\mathcal{C}]]\mathbb{E}[\mathbb{E}[\mathrm{e}^{-\mathrm{i}\theta X_s}|\mathcal{C}]]}$$

$$= \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathrm{e}^{-\theta^2(r_X(t,t|\mathcal{C}) - 2r_X(s,t|\mathcal{C}) + r_X(s,s|\mathcal{C}))/2}]}{\mathbb{E}[\mathrm{e}^{-\theta^2 r_X(t,t|\mathcal{C})/2}]\mathbb{E}[\mathrm{e}^{-\theta^2 r_X(s,s|\mathcal{C})/2}]}, \tag{3.6}$$

which is very useful for calculations. For non-centred process the additional term

$$+ \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(\mu_X(t|\mathcal{C}) - \mu_X(s|\mathcal{C}))}]}{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta\mu_X(t|\mathcal{C})}]\mathbb{E}[\mathrm{e}^{-\mathrm{i}\theta\mu_X(s|\mathcal{C})}]} \tag{3.7}$$

appears. Here all averages are finite, but they can generally be complex values, moreover in particular cases the averages in the denominator can be 0. This strongly suggests the codifference should be used carefully in this case (the same applies to the LCF).

Additionally, representation (3.6) yields another desirable property of the codifference:

Proposition 3. *For a conditionally centred Gaussian process with positive covariance $r_X(s,\,t|\mathcal{C})$ the codifference $\tau_X^\theta(s,\,t)$ is also positive, a negative conditional covariance implies negative codifference.*

If the support of $r_X(s,\,t|\mathcal{C})$ is on both positive and negative half-axes, the sign of the codifference may vary, but it is worth noting that with $r_X(t,\,t|\mathcal{C})$ and $r_X(s,\,s|\mathcal{C})$ fixed, it depends monotonically on $r_X(s,\,t|\mathcal{C})$, so if the conditional covariance is smaller in the sense of stochastic dominance, the codifference will also be smaller.

Now, a simple fact follows only from the expansion $\ln(x) \in x - 1 + o(x)$ as $x \to 1$.

Proposition 4. *For any stationary process X with asymptotically independent values*

$$\tau_X^\theta(t) \sim \frac{1}{\theta^2}\left(\frac{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(X_{s+t} - X_s)}]}{\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_{s+t}}]\mathbb{E}[\mathrm{e}^{-\mathrm{i}\theta X_s}]} - 1 \right), \quad t \to \infty. \tag{3.8}$$

Proof. We assume that X_{s+t} and X_s are asymptotically independent as $t \to \infty$ (note that this property is not sufficient to imply that X is mixing). Therefore

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta(X_{s+t} - X_s)}] \overset{t \to \infty}{\longrightarrow} \mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_{s+t}}]\mathbb{E}[\mathrm{e}^{-\mathrm{i}\theta X_s}], \tag{3.9}$$

and the ratio of expected values under the logarithm converges to 1 so we can use the expansion $\ln(x) \approx x - 1$. ☐

This simple fact is a prototype for the later results, which describe cases when it is possible to remove the nonlinear logarithmic function if the process can be somehow decomposed as a transformation of some weakly dependent variables.

If the process X does not have asymptotically independent values the non-linearity cannot be removed at $t \to \infty$, but if it is an ensemble of such processes (i.e. the conditioned process is mixing), it can be shown that the codifference converges to a positive constant, non-linearly dependent on the law of D.

3.2. Random parameter models

Proposition 5. *If the process X is an ensemble of mixing stationary centred Gaussian processes, then, denoting $D = \mathbb{E}[X_t^2|\mathcal{C}]$,*

$$\tau_X^\theta(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathrm{e}^{-\theta^2 D}]}{\mathbb{E}[\mathrm{e}^{-\theta^2 D/2}]^2} \geqslant 0 \tag{3.10}$$

and equal 0 only for deterministic D.

Proof. The calculation is simple. Because $r_X(t|\mathcal{C}) \leqslant D$ almost surely the random variable $\mathrm{e}^{\theta^2(r_X(t|\mathcal{C}) - D)}$ is positive and bounded by 1 for every t. We can commute the limit with the logarithm and the averaging, getting

$$\lim_{t \to \infty} \tau_X^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}\left[\lim_{t \to \infty} \mathrm{e}^{\theta^2(r_X(t|\mathcal{C}) - D)}\right]}{\mathbb{E}[\mathrm{e}^{-\theta^2 D/2}]} = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[\mathrm{e}^{-\theta^2 D}]}{\mathbb{E}[\mathrm{e}^{-\theta^2 D/2}]^2}. \tag{3.11}$$

The non-negativity of the above stems from Jensen's inequality applied to the function $x \mapsto x^2$ and the variable $\mathrm{e}^{-\theta^2 D/2}$. ☐

Remark. A similar calculation repeated for symmetrised codifference (1.8) shows that it does not exhibit this behaviour. Under the same assumptions

$$\widetilde{\tau}_X^\theta(\infty) = \lim_{t\to\infty} \frac{1}{2\theta^2} \ln \frac{\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}]}{\mathbb{E}[e^{i\theta(X_{s+t}+X_s)}]} = \frac{1}{2\theta^2} \ln \frac{\mathbb{E}\left[\lim_{t\to\infty} e^{\theta^2(r_X(t|\mathcal{C})-D)}\right]}{\mathbb{E}\left[\lim_{t\to\infty} e^{\theta^2(-r_X(t|\mathcal{C})-D)}\right]}.$$

$$= \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D}]}{\mathbb{E}[e^{-\theta^2 D}]} = 0, \tag{3.12}$$

i.e. it cannot detect this form of residual dependence and ergodicity breaking.

Proposition 6. *Let the process X have the form*

$$X_t = \sqrt{D}\, Y_t, \tag{3.13}$$

where Y is a stationary Gaussian process, $\mathbb{E}[Y_t^2] = 1$*, and* $D > 0$ *is a random variable independent of Y. Then the codifference has the form*

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D(1-r_Y(t))}]}{\mathbb{E}[e^{-\theta^2 D/2}]^2}. \tag{3.14}$$

(a) *It is additive with respect to D, that is if* $D = D' + D''$ *for independent D' and D'', then*

$$\tau_X^\theta(t) = \tau_{X'}^\theta(t) + \tau_{X''}^\theta(t) \tag{3.15}$$

where $X_t' = \sqrt{D'}\, Y_t$ *and* $X_t'' = \sqrt{D''}\, Y_t$*.*

(b) *It is an increasing function of the covariance* $r_Y(t)$*, which is smaller than* $r_X(t)$ *for* $r_Y(t)$ *close to 1 and larger than* $r_X(t)$ *when the latter is close to 0. If* $\mathbb{E}[D] < \infty$ *the difference* $\tau_X^\theta(t) - r_X(t)$ *decreases as a function of* $r_Y(t)$*.*

(c) *For any mixing Y the difference* $\tau_X^\theta(t) - \tau_X^\theta(\infty)$ *exhibits the same type of asymptotic as the covariance* $r_Y(t)$*, that is*

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) \sim \frac{\mathbb{E}[De^{-\theta^2 D}]}{\mathbb{E}[e^{-\theta^2 D}]} r_Y(t), \quad t \to \infty. \tag{3.16}$$

Proof. Let us start from writing the conditional covariance,

$$\mathbb{E}[X_t^2|D] = D, \quad \mathbb{E}[(X_{s+t}-X_s)^2|D] = 2D(1-r_Y(t)), \tag{3.17}$$

which implies that

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D(1-r_Y(t))}]}{\mathbb{E}[e^{-\theta^2 D/2}]^2}. \tag{3.18}$$

If we substitute $D = D' + D''$ both numerator and denominator factorise as products of independent random variables. The formula

$$\tau_X^\theta(t) = \tau_{X'}^\theta(t) + \tau_{X''}^\theta(t) \tag{3.19}$$

follows.

In point (b) the monotonic dependence is a consequence of the fact that only the numerator of the fraction in (3.14) depends on $r_Y(t)$. It is a Laplace transform of the variable *D* calculated at the point $\theta^2(1 - r_Y(t))$, it decreases as the argument increases, so it is an increasing function of $r_Y(t)$. This dependence is continuous. When $r_X(t) = 0$, e.g. always for $t = 0$ formula (3.14) simplifies and we can apply Jensen's inequality,

$$\tau_X^\theta(0) = -\frac{2}{\theta^2} \ln \mathbb{E}[e^{-\theta^2 D/2}] \leqslant -\frac{2}{\theta^2}\mathbb{E}[\ln e^{-\theta^2 D/2}] = \mathbb{E}[D] = r_X(0). \tag{3.20}$$

For $r_Y(t)$ close to 0 we can use proposition 5 to determine that the codifference is positive. For the last property listed in b), let us write the difference $\tau_X^\theta(t) - r_X(t)$ as a function of $r = r_X(t)$,

$$f(r) := \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D(1-r)}]}{\mathbb{E}[e^{-\theta^2 D/2}]^2} - \mathbb{E}[D]r = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2(D-\mathbb{E}[D])(1-r)}]}{\mathbb{E}[e^{-\theta^2(D-\mathbb{E}[D])/2}]^2}. \tag{3.21}$$

Using the majorised convergence theorem, the derivative of the numerator exists and determines the sign of f'. Denoting $F_r := \theta^2(D - \mathbb{E}[D])(1 - r)$ we have

$$f'(r) \propto \frac{1}{1-r}\mathbb{E}[F_r e^{-F_r}] = \frac{1}{1-r}\mathbb{E}[F_r(e^{-F_r}-1)] \leqslant 0, \tag{3.22}$$

where we used the fact that $\mathbb{E}[F_r] = 0$ and $x(e^{-x}-1) \leqslant 0$.

For (c) consider $\tau_X(t) - \tau_X(\infty)$ and use the expansion $\ln(x) \approx x - 1$

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) = \frac{1}{\theta^2}\ln\frac{\mathbb{E}[e^{-\theta^2 D(1-r_Y(t))}]}{\mathbb{E}[e^{-\theta^2 D}]} \sim \frac{1}{\theta^2}\left(\frac{\mathbb{E}[e^{-\theta^2 D(1-r_Y(t))}]}{\mathbb{E}[e^{-\theta^2 D}]} - 1\right), \quad t \to \infty. \tag{3.23}$$

Now we can rearrange the right side of the above equation and get

$$\lim_{t\to\infty}\frac{\tau_X^\theta(t) - \tau_X^\theta(\infty)}{r_Y(t)} = \lim_{t\to\infty}\frac{\mathbb{E}\left[\frac{e^{\theta^2 D r_Y(t)}-1}{\theta^2 r_Y(t)}e^{-\theta^2 D}\right]}{\mathbb{E}[e^{-\theta^2 D}]} = \lim_{x\to 0}\frac{\mathbb{E}\left[\frac{e^{Dx}-1}{x}e^{-\theta^2 D}\right]}{\mathbb{E}[e^{-\theta^2 D}]} = \frac{\mathbb{E}[D e^{-\theta^2 D}]}{\mathbb{E}[e^{-\theta^2 D}]}. \tag{3.24}$$

$\square$

The analogues of (a) and (b) also hold for the LCF, the derivation is very similar as in proposition 6 so we only state the result.

Proposition 7. *Let the process X have the form*

$$X_t = \sqrt{D}\,Y_t, \tag{3.25}$$

where Y is a centred Gaussian process and $D > 0$ is a random variable independent of Y.

Then the LCF has the form

$$\zeta_X^\theta(t) = -\frac{2}{\theta^2}\ln\mathbb{E}[e^{-\theta^2 D \delta_Y^2(t)/2}] \tag{3.26}$$

and:

(a) *if $\mathbb{E}[D] < \infty$ then*

$$\zeta_X^\theta(t) \sim \delta_X^2(t), \quad t \to 0^+. \tag{3.27}$$

(b) *It is additive with respect to* D, *that is if $D = D' + D''$ for independent D' and D'', then*

$$\zeta_X^\theta(t) = \zeta_{X'}^\theta(t) + \zeta_{X''}^\theta(t) \tag{3.28}$$

where $X_t' = \sqrt{D'}\,Y_t$ and $X_t'' = \sqrt{D''}\,Y_t$.

(c) *It is an increasing function of the MSD $\delta_Y^2(t)$.*

(d) *For $\mathbb{E}[D] < \infty$ the difference $\delta_X^2(t) - \zeta_X^\theta(t)$ is non-negative and increases as $\delta_X^2(t)$ increases.*

The asymptotic of the codifference near zero depends on the tail behaviour of p_D and can be used to study it. This statement is clarified by the following result.

Proposition 8. *If the stationary Gaussian process Y is mean-square continuous and $X_t = \sqrt{D}\,Y_t$, then*

(a) *for $\mathbb{E}[D] < \infty$*

$$\tau_X^\theta(0) - \tau_X^\theta(t) \sim \mathbb{E}[D](1 - r_Y(t)), \quad t \to 0^+ \tag{3.29}$$

and

$$\zeta_X^\theta(t) \sim \mathbb{E}[D]\delta_Y^2(t), \quad t \to 0^+. \tag{3.30}$$

(b) *If*

$$p_D(d) \sim \frac{L(d)}{d^{1+\rho}}, \quad 0 < \rho < 1, \quad d \to \infty \tag{3.31}$$

for some slowly varying function L, then

$$\tau_X^\theta(0) - \tau_X^\theta(t) \sim \theta^{2\rho-2}\frac{\Gamma(1-\rho)}{\rho}L(\theta^{-2}(1-r_Y(t))^{-1})(1-r_Y(t))^\rho, \quad t \to 0^+. \tag{3.32}$$

Proof. For a mean-square continuous Y the covariance r_Y is a continuous function. The codifference is also continuous and $\ln(x) \approx x - 1$ implies that

$$\tau_X^\theta(0) - \tau_X^\theta(t) = -\frac{1}{\theta^2}\ln \mathbb{E}[e^{-\theta^2 D(1-r_Y(t))}] \sim \frac{1}{\theta^2}(1 - \mathbb{E}[e^{-D(1-r_Y(t))}]), \quad t \to 0^+. \tag{3.33}$$

Because

$$\lim_{t\to 0^+}\mathbb{E}\left[\frac{1 - e^{-\theta^2 D(1-r_Y(t))}}{\theta^2(1-r_Y(t))}\right] = \lim_{x\to 0^+}\mathbb{E}\left[\frac{1 - e^{-Dx}}{x}\right] = \mathbb{E}[D]. \tag{3.34}$$

The derivation for ζ_X^θ is similar. For point b) we write the asymptotic of $\tau_X^\theta(0) - \tau_X^\theta(t)$ as the integral

$$\tau_X^\theta(0) - \tau_X^\theta(t) \sim \frac{1}{\theta^2}\int_0^\infty \mathrm{d}d\,(1 - e^{-\theta^2 d(1-r_Y(t))})p_D(d) \tag{3.35}$$

and simplify the ratio under investigation

$$\frac{\tau_X^\theta(0) - \tau_X^\theta(t)}{L(\theta^{-2}(1-r_Y(t))^{-1})(1-r_Y(t))^\rho} \sim \theta^{2\rho-2}\lim_{x\to 0^+}\int_0^\infty \mathrm{d}d\,(1 - e^{-xd})\frac{1}{L(x^{-1})x^\rho}p_D(d)$$

$$= \theta^{2\rho-2}\int_0^\infty \mathrm{d}d\,\frac{1 - e^{-d}}{d}d^{-\rho}\lim_{x\to 0^+}\left(\frac{d}{x}\right)^{\rho+1}\frac{1}{L(x^{-1})}p_D\left(\frac{d}{x}\right)$$

$$= \theta^{2\rho-2}\int_0^\infty \mathrm{d}d\,\frac{1 - e^{-d}}{d}d^{-\rho} = -\theta^{2\rho-2}\Gamma(-\rho) = \theta^{2\rho-2}\frac{\Gamma(1-\rho)}{\rho}. \tag{3.36}$$

$\square$

Now, let us move our attention from a random D to the class of processes, for which the shape of the covariance function varies from trajectory to trajectory:

Proposition 9. *For a mixture of stationary Gaussian processes with fixed non-random scale $D = \sigma^2$*

$$\tau_X^\theta(t) = \frac{1}{\theta^2}\ln \mathbb{E}[e^{\theta^2 r_X(t|\mathcal{C})}]. \tag{3.37}$$

The above formula also implies that

$$r_X(t) \leqslant \tau_X^\theta(t) \leqslant \frac{1}{\theta^2}((\theta\sigma)^{-2}\sinh(\theta^2\sigma^2)r_X(t) + \cosh(\theta^2\sigma^2) - 1). \tag{3.38}$$

Proof. Assumption of a fixed variance means that $\mathbb{E}[X(t)^2|\mathcal{C}] = \sigma^2$ for some deterministic σ^2. Using the conditional expectancy it follows that

$$\tau_X^\theta(t) = \frac{1}{\theta^2}\ln \frac{\mathbb{E}[\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}|\mathcal{C}]]}{\mathbb{E}[\mathbb{E}[e^{i\theta X_{s+t}}|\mathcal{C}]]\mathbb{E}[\mathbb{E}[e^{-iX_s}|\mathcal{C}]]} = \frac{1}{\theta^2}\ln \frac{\mathbb{E}[e^{-\theta^2(\sigma^2 - r_X(t|\mathcal{C}))}]}{(e^{-\theta^2\sigma^2/2})^2}$$

$$= \frac{1}{\theta^2}\ln \mathbb{E}[e^{\theta^2 r_X(t|\mathcal{C})}]. \tag{3.39}$$

Now the left inequality is just Jensen's inequality applied to the function $\ln$. The right inequality follows from two approximations: the first is $\ln x \leqslant x - 1$, the second is $\exp(x) \leqslant L^{-1}\sinh(L)x + \cosh(L)$ for $-L \leqslant x \leqslant L$.

$\square$

For the exponentially decaying conditional covariance stronger results are available:

Proposition 10. *For a mixture of stationary centred Gaussian processes with conditional covariance $r_X(t|\Lambda, D) = De^{-t\Lambda}$, with Λ and D independent, we observe the following asymptotic properties.*

(a) *Power law behaviour: if $p_\Lambda(\lambda) \sim L(\lambda)\lambda^{\alpha-1}$, $\lambda \to 0^+$ for slowly varying L, then*

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) \sim C_{\alpha,\theta}\frac{L(t^{-1})}{t^\alpha}, \tag{3.40}$$

where the constant $C_{\alpha,\theta}$ is

$$C_{\alpha,\theta} = \frac{\Gamma(\alpha)}{\theta^2 \mathbb{E}[e^{-\theta^2 D}]^2} \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \frac{1}{k^{\alpha}}. \tag{3.41}$$

(b) *Quick decay behaviour: if* $p_{\Lambda}(\lambda) \in \mathcal{O}(\lambda^{\infty})$, $\lambda \to 0^+$ *then*

$$\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty) \in \mathcal{O}(t^{-\infty}), \quad t \to \infty. \tag{3.42}$$

(c) *Truncation: if* $\Lambda = \lambda_0 + \widetilde{\Lambda}$ *for deterministic* $\lambda_0 > 0$ *then*

$$\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty) \leqslant e^{-\lambda_0 t}(\tau_{\widetilde{X}}^{\theta}(t) - \tau_{\widetilde{X}}^{\theta}(\infty)), \tag{3.43}$$

where $\widetilde{X}$ *is a solution of the Langevin equation with viscosity* $\widetilde{\Lambda}$ *and the same D.*

Proof. For (a) first we apply the expansion $\ln(x) \approx x - 1$ to $\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty)$

$$\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D(1-e^{-t\Lambda})}]}{\mathbb{E}[e^{-\theta^2 D}]} \sim \frac{1}{\theta^2}\left(\frac{\mathbb{E}[e^{-\theta^2 D(1-e^{-t\Lambda})}]}{\mathbb{E}[e^{-\theta^2 D}]} - 1\right), \quad t \to \infty. \tag{3.44}$$

Therefore

$$\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty) \sim \frac{1}{\theta^2 \mathbb{E}[e^{-\theta^2 D}]} \mathbb{E}[e^{-\theta^2 D}(e^{\theta^2 D e^{-t\Lambda}} - 1)]$$

$$= \frac{1}{\theta^2 \mathbb{E}[e^{-\theta^2 D}]} \mathbb{E}\left[e^{-\theta^2 D} \sum_{k=1}^{\infty} \theta^{2k} \frac{D^k}{k!} e^{-kt\Lambda}\right]$$

$$= \frac{1}{\theta^2 \mathbb{E}[e^{-\theta^2 D}]} \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \mathbb{E}[e^{-kt\Lambda}], \quad t \to \infty. \tag{3.45}$$

Note that the sum within consists of positive terms, so the commutation of expectation and sum is justified. Now, knowing the asymptotic $p_{\Lambda}(\lambda) \sim \lambda^{\alpha-1}$, $\lambda \to 0^+$ we can apply the Tauberian theorem

$$\frac{t^{\alpha}}{L(t^{-1})} \mathbb{E}[e^{-kt\Lambda}] \overset{t \to \infty}{\longrightarrow} \Gamma(\alpha) \frac{1}{k^{\alpha}}. \tag{3.46}$$

The sum (3.45) consists of positive terms, so let us study its asymptotic

$$\frac{t^{\alpha}}{L(t^{-1})} \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \mathbb{E}[e^{-kt\Lambda}] = \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \frac{t^{\alpha}}{L(t^{-1})} \mathbb{E}[e^{-kt\Lambda}]$$

$$\overset{t \to \infty}{\longrightarrow} \Gamma(\alpha) \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \frac{1}{k^{\alpha}}, \tag{3.47}$$

where the commutation of taking the limit and the sum is justified by the inequality

$$t^{\alpha} \mathbb{E}[e^{-kt\Lambda}] \leqslant t^{\alpha} \mathbb{E}[e^{-t\Lambda}]. \tag{3.48}$$

The right term is convergent with respect to t, therefore it is bounded, so the left term is uniformly bounded with respect to k and we can use the dominated convergence theorem.

Note that the resulting sum is also bounded with respect to α,

$$\sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \frac{1}{k^{\alpha}} < \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!}$$

$$= \mathbb{E}\left[\sum_{k=0}^{\infty} \theta^{2k} \frac{D^k}{k!} e^{-\theta^2 D}\right] - \mathbb{E}[e^{-\theta^2 D}] = 1 - \mathbb{E}[e^{-\theta^2 D}]. \tag{3.49}$$

This concludes the derivation of (a). Now let us prove (b). We fix integer $N > 0$ and then make the estimation

$$\mathbb{E}[e^{-kt\Lambda}]t^N = \int_0^{\infty} \mathrm{d}\lambda \, p_{\Lambda}(\lambda) e^{-kt\lambda} t^N = k^{N+1} \int_0^{\infty} \mathrm{d}\lambda \, p_{\Lambda}\left(\frac{\lambda}{kt}\right) e^{-\lambda} \lambda^{N+1} \left(\frac{\lambda}{kt}\right)^{-N-1}$$

$$\overset{t \to \infty}{\longrightarrow} 0 \tag{3.50}$$

to obtain

$$\lim_{t \to \infty} (\tau_X^{\theta}(t) - \tau_X^{\theta}(\infty))t^N = \lim_{t \to \infty} \frac{1}{\theta^2} \sum_{k=1}^{\infty} \theta^{2k} \frac{\mathbb{E}[D^k e^{-\theta^2 D}]}{k!} \mathbb{E}[e^{-kt\Lambda}]t^N = 0. \tag{3.51}$$

The limit follows because it is a convergent sum of positive terms.

In order to prove the last point (c) notice that $e^{-t\lambda_0} < 1$ and $e^{t\lambda_0} > 1$ so $x \mapsto x^{e^{-t\lambda_0}}$ is a concave function and $e^{-\theta^2 D e^{t\lambda_0}} \leqslant e^{-\theta^2 D}$. Therefore

$$
\begin{aligned}
\tau_X^\theta(t) - \tau_X^\theta(\infty) &= \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D} e^{\theta^2 D e^{-t\Lambda} e^{-\lambda_0 t}}]}{\mathbb{E}[e^{-\theta^2 D}]} = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[(e^{-\theta^2 D e^{t\lambda_0}} e^{\theta^2 D e^{-t\Lambda}})^{e^{-t\lambda_0}}]}{\mathbb{E}[e^{-\theta^2 D}]} \\
&\leqslant \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D e^{t\lambda_0}} e^{\theta^2 D e^{-t\Lambda}}]^{e^{-t\lambda_0}}}{\mathbb{E}[e^{-\theta^2 D}]} = e^{-t\lambda_0} \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D e^{t\lambda_0}} e^{\theta^2 D e^{-t\Lambda}}]}{\mathbb{E}[e^{-\theta^2 D}]} \\
&\leqslant e^{-t\lambda_0} \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 D} e^{\theta^2 D e^{-t\Lambda}}]}{\mathbb{E}[e^{-\theta^2 D}]} = e^{-t\lambda_0}(\tau_{\widetilde{X}}^\theta(t) - \tau_{\widetilde{X}}^\theta(\infty)).
\end{aligned} \tag{3.52}
$$

$\square$

In the next proposition we will study the properties of the increment process

$$
\Delta X_t := X_{t+\Delta t} - X_t \tag{3.53}
$$

and use it to detect non-ergodicity.

Proposition 11. *Considering the same process as in proposition 10, the codifference of its increments ΔX_t converges to a constant*

$$
\lim_{t \to \infty} \tau_{\Delta X}^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{2\theta^2 D(e^{-\Delta t \Lambda} - 1)}]}{\mathbb{E}[e^{\theta^2 D(e^{-\Delta t \Lambda} - 1)}]^2} \geqslant 0, \tag{3.54}
$$

which equals 0 only when both D and Λ are deterministic. After suitable rescaling $\Delta \widetilde{X}_t := \Delta X_t / \sqrt{\Delta t}$ the limit becomes independent of Δt,

$$
\lim_{\Delta t \to 0^+} \lim_{t \to \infty} \tau_{\Delta \widetilde{X}}^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-2\theta^2 D \Lambda}]}{\mathbb{E}[e^{-\theta^2 D \Lambda}]^2}. \tag{3.55}
$$

Proof. The reasoning is similar to the one shown in the proof of proposition 6 (b). The increment process ΔX_t is a stationary process, which is conditionally Gaussian. We can calculate its conditional variance

$$
\mathbb{E}[\Delta X_t^2 | \Lambda, D] = 2D(1 - e^{-\Delta t \Lambda}) \tag{3.56}
$$

and the variance of the difference

$$
\mathbb{E}[(\Delta X_{s+t} - \Delta X_s)^2 | \Lambda, D] = 4D\left(1 - e^{-\Delta t \Lambda} - e^{-t\Lambda}\left(1 - \frac{e^{-\Delta t \Lambda} + e^{\Delta t \Lambda}}{2}\right)\right)
$$
$$
\xrightarrow{t \to \infty} 4D(1 - e^{-\Delta t \Lambda}). \tag{3.57}
$$

The limit of the codifference is

$$
\lim_{t \to \infty} \tau_{\Delta X}^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{2\theta^2 D(e^{-\Delta t \Lambda} - 1)}]}{\mathbb{E}[e^{\theta^2 D(e^{-\Delta t \Lambda} - 1)}]^2}. \tag{3.58}
$$

Applying Jensen's inequality to the variable $e^{2\theta^2 D e^{-\Delta t \Lambda}}$ and the function $x \mapsto x^2$ yields the inequality.

For the rescaled process it is straightforward to calculate that

$$
\lim_{\Delta t \to 0^+} \lim_{t \to \infty} \tau_{\Delta \widetilde{X}}^\theta(t) = \lim_{\Delta t \to 0^+} \frac{1}{\theta^2} \ln \frac{\mathbb{E}\left[\exp\left(2\theta^2 D \frac{e^{-\Delta t \Lambda} - 1}{\Delta t}\right)\right]}{\mathbb{E}\left[\exp\left(\theta^2 D \frac{e^{-\Delta t \Lambda} - 1}{\Delta t}\right)\right]^2} = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-2\theta^2 D \Lambda}]}{\mathbb{E}[e^{-\theta^2 D \Lambda}]^2}. \tag{3.59}
$$

$\square$

The last considered class of covariance functions is $Df(\Lambda)\exp(-t\Lambda)$. The increment process from proposition 11 fits this class with $f(\Lambda) = 1 - \exp(-\Delta t \Lambda)$, higher order increments and other similar transformations correspond to more complex f, but their behaviour at 0^+ can be easily traced. Note that the proposition below is not a straightforward generalisation of proposition 10. The statements and methods of the derivation below are similar, but the assumptions do not coincide, because the introduction of the scaling $f(\Lambda)$ with a power law at 0 was made at the cost of adding the strong requirement about the fast decay of tails of D, $\mathbb{E}[\exp(\theta^2 D)] < \infty$:

Proposition 12. *Let us consider the stationary, conditionally Gaussian process characterised by the conditional covariance*

$$r_X(t|\Lambda, D) = Df(\Lambda)e^{-t\Lambda}. \tag{3.60}$$

Now, let us assume that D *and* Λ *are independent,* $\mathbb{E}[e^{\theta^2 D}] < \infty$ *and the PDF of* Λ *has the form*

$$p_\Lambda(\lambda) \sim L(\lambda)\lambda^{\alpha-1}, \quad f(\lambda) \sim \lambda^\gamma, \quad \lambda \to 0^+; \quad \alpha, \gamma > 0, \tag{3.61}$$

for slowly varying function L. *Then, for this class of processes*

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) \sim \frac{\Gamma(\alpha + \gamma)\mathbb{E}[D]}{\mathbb{E}[e^{-\theta^2 Df(\Lambda)}]} L(t^{-1})t^{-\alpha-\gamma}. \tag{3.62}$$

Proof. We start from the formula

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2 Df(\Lambda)(1-e^{-t\Lambda})}]}{\mathbb{E}[e^{-\theta^2 Df(\Lambda)}]} \tag{3.63}$$

which has the asymptotic

$$\tau_X^\theta(t) - \tau_X^\theta(\infty) \sim \frac{1}{\theta^2}\left(\frac{\mathbb{E}[e^{-\theta^2 Df(\Lambda)(1-e^{-t\Lambda})}]}{\mathbb{E}[e^{-\theta^2 Df(\Lambda)}]} - 1\right)$$

$$= \frac{1}{\theta^2 \mathbb{E}[e^{-\theta^2 Df(\Lambda)}]}\mathbb{E}[e^{-\theta^2 Df(\Lambda)}(e^{\theta^2 Df(\Lambda)e^{-t\Lambda}} - 1)]. \tag{3.64}$$

We thus need to study the tail behaviour of

$$\mathbb{E}\left[e^{-\theta^2 Df(\Lambda)}\sum_{k=1}^\infty \frac{\theta^{2k}}{k!}D^k f(\Lambda)^k e^{-kt\Lambda}\right] = \sum_{k=1}^\infty \frac{\theta^{2k}}{k!}\mathbb{E}[D^k\mathbb{E}[f(\Lambda)^k e^{-\theta^2 Df(\Lambda)}e^{-kt\Lambda}|D]]. \tag{3.65}$$

We will analyse it using a bottom-up approach and start from considering the long time asymptotic of the conditional expected value $\mathbb{E}[\cdot|D]$ for one term,

$$\frac{t^{\alpha+k\gamma}}{L(1/t)}\mathbb{E}[f(\Lambda)^k e^{-\theta^2 Df(\Lambda)}e^{-kt\Lambda}|D] = \frac{t^{\alpha+k\gamma}}{L(1/t)}\int_0^\infty d\lambda\, f(\lambda)^k e^{-\theta^2 Df(\lambda)}p_\Lambda(\lambda)e^{-kt\lambda}$$

$$= \frac{t^{\alpha+k\gamma-1}}{L(1/t)}\frac{1}{k}\int_0^\infty d\lambda\, f\left(\frac{\lambda}{kt}\right)^k e^{-\theta^2 Df\left(\frac{\lambda}{kt}\right)}p_\Lambda\left(\frac{\lambda}{kt}\right)e^{-\lambda}$$

$$= \frac{1}{k^{\alpha+k\gamma}}\int_0^\infty d\lambda\, \left(\frac{f\left(\frac{\lambda}{kt}\right)}{\left(\frac{\lambda}{kt}\right)^\gamma}\right)^k e^{-\theta^2 Df\left(\frac{\lambda}{kt}\right)}\frac{p_\Lambda\left(\frac{\lambda}{kt}\right)}{L(1/t)\left(\frac{\lambda}{kt}\right)^{\alpha-1}}\lambda^{\alpha+k\gamma-1}e^{-\lambda}$$

$$\xrightarrow{t\to\infty} \frac{1}{k^{\alpha+k\gamma}}\int_0^\infty d\lambda\, \lambda^{\alpha+k\gamma-1}e^{-\lambda} = \frac{\Gamma(\alpha + k\gamma)}{k^{\alpha+k\gamma}}. \tag{3.66}$$

Now take $\delta > 0$ such that $f(\lambda) < 1$ for all $0 \leqslant \lambda < \delta$ and $\epsilon > 0$ such that $L(1/t) > t^{-1/2}$ for sufficiently large t

$$\frac{D^k}{k!}\frac{t^{\alpha+\gamma}}{L(1/t)}f(\Lambda)^k e^{-\theta^2 Df(\Lambda)}e^{-kt\Lambda} \leqslant \frac{D^k}{k!}\frac{t^{\alpha+\gamma}}{L(1/t)}(f(\Lambda)e^{-t\Lambda}1_{\Lambda<\delta} + f(\Lambda)^k e^{-\theta^2 Df(\Lambda)}e^{-kt\delta}1_{\Lambda\geqslant\delta})$$

$$\leqslant \frac{D^k}{k!}\frac{t^{\alpha+\gamma}}{L(1/t)}f(\Lambda)e^{-t\Lambda} + \frac{k^k}{k!}e^{-k}t^{\alpha+\gamma+1/2}e^{-kt\delta}, \tag{3.67}$$

where we additionally used the inequality $x^k e^{-x} \leqslant k^k e^{-k}$. Now, for the left term above observe that

$$\frac{t^{\alpha+\gamma}}{L(1/t)}\mathbb{E}[f(\Lambda)e^{-t\Lambda}] \xrightarrow{t\to\infty} \Gamma(\alpha), \tag{3.68}$$

so it is bounded with respect to t by some constant, let it be c_1,

$$\sum_{k=1}^\infty \mathbb{E}\left[\frac{\theta^{2k}D^k}{k!}\frac{t^{\alpha+\gamma}}{L(1/t)}f(\Lambda)e^{-t\Lambda}\right] \leqslant c_1 \sum_{k=1}^\infty \frac{\theta^{2k}\mathbb{E}[D^k]}{k!} = c_1\mathbb{E}[e^{\theta^2 D}]. \tag{3.69}$$

And for the right term, the Stirling formula shows that

$$\frac{k^k}{k!}e^{-k} \sim \sqrt{2\pi}k^{-1/2}, \quad k \to \infty. \tag{3.70}$$

Moreover straightforward calculation yields

$$t^{\alpha+\gamma+1/2}e^{-kt\delta} \leqslant c_2 k^{-\alpha-\gamma-1/2}, \tag{3.71}$$

so the whole series behaves like $k^{-1-\alpha-\gamma}$ and is summable.

Therefore, we have shown that we can use the dominated convergence theorem with respect to series (3.65) multiplied by $t^{\alpha+\gamma}/L(1/t)$. According to (3.66) the term $k = 1$ converges to $\mathbb{E}[D]\Gamma(\alpha + \gamma)$ and all terms $k > 1$ decay like $t^{-k\gamma}$. Only the first term remains in the limit $t \to \infty$ and

$$\frac{t^{\alpha+\gamma}}{L(1/t)} \sum_{k=1}^{\infty} \frac{\theta^{2k}}{k!} \mathbb{E}[D^k \mathbb{E}[f(\Lambda)^k e^{-\theta^2 Df(\Lambda)} e^{-kt\Lambda}|D]] \stackrel{t\to\infty}{\longrightarrow} \theta^2 \mathbb{E}[D]\Gamma(\alpha + \gamma). \tag{3.72}$$

$\square$

Remark. Propositions 10 and 12 above can be generalised by replacing t by $g(t)$ in the formula for the covariance, the only requirement is that $g(t) \to \infty$ as $t \to \infty$. This allows one to consider some more general types of the dependence, e.g. the power-law t^{-2H} corresponds to $\Lambda = 2H$ and $g(t) = \ln(t)$.

3.3. Diffusing diffusivity

Proposition 13. *Let us assume that Y and Z are centred stationary Gaussian processes. Without loss of generality we assume $\mathbb{E}[Y_t^2] = \mathbb{E}[Z_t^2] = 1$. Let X be given by*

(a)

$$X_t = (\sigma Z_t + d) Y_t, \tag{3.73}$$

(b)

$$X_t = \sigma |Z_t| Y_t; \tag{3.74}$$

with deterministic σ, $d > 0$. Then the codifference of X is given by elementary formulae, as given at the end of corresponding derivations in equations (3.78) and (3.82).

Proof. We begin by conditioning over Z_t, the averages then become Gaussian averages rescaled by values Z_t. Next we calculate the denominator in the codifference

$$\mathbb{E}[e^{i\theta X_{s+t}}]\mathbb{E}[e^{-i\theta X_s}] = \mathbb{E}[e^{-\theta^2(\sigma Z_s+d)^2/2}]^2 = \frac{1}{1 + (\theta\sigma)^2} \exp\left(-\theta^2 \frac{d^2}{1 + (\theta\sigma)^2}\right). \tag{3.75}$$

The last equality corresponds to calculating a Gaussian integral, which can also be interpreted as a Laplace transform of the distribution $\chi^2(1)$. The numerator is more complicated,

$$\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}] = \mathbb{E}[e^{-\theta^2((\sigma Z_{s+t}+d)^2 + (\sigma Z_s+d)^2 - 2r_Y(t)(\sigma Z_{s+t}+d)(\sigma Z_s+d))/2}]. \tag{3.76}$$

The above expectation can be calculated if we decompose $[Z_{s+t}, Z_s] \stackrel{d}{=} [A_+ + A_-, A_+ - A_-]$ where A_+, A_- are independent Gaussian variables, whose variances can be found to be $\mathbb{E}[A_\pm^2] = (1 \pm r_Z(t))/2$. After substitution the exponent in (3.76) factorises into

$$\mathbb{E}[e^{i\theta(X_{s+t}-X_s)}] = \mathbb{E}[e^{-(\theta\sigma)^2(1-r_Y(t))A_+^2 - 2(\theta\sigma)^2(1-r_Y(t))A_+ d}]$$
$$\times \mathbb{E}[e^{-(\theta\sigma)^2(1+r_Y(t))A_-^2}]e^{-\theta^2(1-r_Y(t))d^2}. \tag{3.77}$$

Both obtained terms are Gaussian integrals which can be easily evaluated. Taking both together and calculating the logarithm we obtain

$$\tau_X^\theta(t) = \frac{(1 - r_Y(t))^2((\theta\sigma)^2 + r_Z(t))}{1 + (1 - r_Y(t))((\theta\sigma)^2 + r_Z(t))}d^2 + \frac{d^2}{1 + (\theta\sigma)^2} - (1 - r_Y(t))d^2 + \frac{1}{\theta^2}\ln(1 + (\theta\sigma)^2)$$
$$- \frac{1}{2\theta^2}\ln((1 + ((\theta\sigma)^2 + r_Z(t))(1 - r_Y(t)))(1 + ((\theta\sigma)^2 - r_Z(t))(1 + r_Y(t)))). \tag{3.78}$$

For (b) the denominator is simple and yields $(1 + (\theta\sigma)^2)^{-1}$. The numerator can be expressed as

$$\mathbb{E}[e^{i\theta\sigma(|Z_{s+t}|Y_{s+t} - |Z_s|Y_s)}] = \mathbb{E}\left[\exp\left(-\frac{(\theta\sigma)^2}{2}(Z_1^2 + Z_2^2 - 2r_Y(t)|Z_1 Z_2|)\right)\right]. \tag{3.79}$$

Using the formula for the two-dimensional density of $[Z_{s+t}, Z_s]$, the term under the logarithm in the formula for the codifference can be expressed as an integral over the function

$$C \exp\left(-\frac{(\theta\sigma)^2}{2}(z_1^2 + z_2^2 - 2r_Y(t)|z_1 z_2|)\right) \exp\left(-\frac{1}{2(1 - r_Z(t)^2)}(z_1^2 + z_2^2 - 2r_Z(t)z_1 z_1)\right)$$
$$= C \exp\left(-sz_1^2 - sz_2^2 + 2s\rho_{\text{sgn}(z_1 z_2)}|z_1 z_2|\right), \tag{3.80}$$

where

$$C := \frac{1 + (\theta\sigma)^2}{2\pi\sqrt{1 - r_Z(t)^2}}, \quad s := \frac{1}{2}\left((\theta\sigma)^2 + \frac{1}{1 - r_Z(t)^2}\right),$$
$$\rho_{\pm} := \frac{(\theta\sigma)^2(1 - r_Z(t)^2)r_Y(t) \pm r_Z(t)}{(\theta\sigma)^2(1 - r_Z(t)^2) + 1}. \tag{3.81}$$

The integration over $\mathbb{R}^2$ of (3.80) can be changed to an integration over $\mathbb{R}_+^2$

$$I = 2C \int_0^\infty dz_1 \int_0^\infty dz_2 \ e^{-sz_1^2 - sz_2^2} \sum_{\rho \in \{\rho_+, \rho_-\}} e^{2s\rho z_1 z_2}$$
$$= \frac{2C}{s} \int_0^\infty dz_1 \int_0^\infty dz_2 \ e^{-z_1^2 - z_2^2} \sum_{\rho \in \{\rho_+, \rho_-\}} e^{2\rho z_1 z_2}$$
$$= \frac{\sqrt{\pi}\,C}{s} \sum_{\rho \in \{\rho_+, \rho_-\}} \int_0^\infty dz_1 \ e^{-(1-\rho)z_1^2} \operatorname{erfc}(-\rho z_1)$$
$$= \frac{\sqrt{1 - r_Z(t)^2}}{\pi} \frac{1 + (\theta\sigma)^2}{(\theta\sigma)^2(1 - r_Z(t)^2) + 1} \sum_{\rho \in \{\rho_+, \rho_-\}} \frac{\frac{\pi}{2} + \arctan\left(\frac{\rho}{\sqrt{1 - \rho^2}}\right)}{\sqrt{1 - \rho^2}}. \tag{3.82}$$

Now, the codifference is $\tau_X^\theta(t) = \theta^{-2}\ln I$. □

When $r_Y(t) = 0$ the above formulae simplify significantly and simple asymptotic can be derived by direct computation, see equations (2.31) and (2.32). The case $r_Z(t) = 0$ also leads to a simplification and can be considered in a more general setting.

Proposition 14. *If Y_t is a stationary Gaussian process, $\mathbb{E}[Y_t^2] = 1$ and for large enough t values D_s and D_{s+t} are i.i.d. and independent of Y, then for $X_t = \sqrt{D_t}\,Y_t$*

$$\tau_X^\theta(t) \sim \frac{\mathbb{E}[\sqrt{D}\,e^{-\theta^2 D/2}]^2}{\mathbb{E}[e^{-\theta^2 D/2}]^2} r_Y(t), \quad r_Y(t) \to 0, \tag{3.83}$$

where D has the same distribution as D_s or D_{s+t}.

Proof. We take t large enough so that we can represent the values of X as $X_s = \sqrt{D_1}\,Y_s$ and $X_{s+t} = \sqrt{D_2}\,Y_{s+t}$ for i.i.d. D_1 and D_2. Using a conditioning on D_1, D_2 the codifference can be expressed as

$$\tau_X^\theta(t) = \frac{1}{\theta^2} \ln \frac{\mathbb{E}[e^{-\theta^2(D_1 + D_2 - 2r_Y(t)\sqrt{D_1 D_2})/2}]}{\mathbb{E}[e^{-\theta^2(D_1 + D_2)/2}]}$$
$$= \frac{1}{\theta^2} \ln\left(\frac{\mathbb{E}[e^{-\theta^2(D_1 + D_2)/2}(e^{\theta^2 r_Y(t)\sqrt{D_1 D_2}} - 1)]}{\mathbb{E}[e^{-\theta^2(D_1 + D_2)/2}]} + 1\right). \tag{3.84}$$

Now we consider the numerator in the above, divide it by $r_Y(t)$ and, using dominated convergence as in previous propositions,

$$\mathbb{E}\left[e^{-\theta^2(D_1 + D_2)/2} \frac{e^{\theta^2 r_Y(t)\sqrt{D_1 D_2}} - 1}{r_Y(t)}\right] \xrightarrow{r_Y(t) \to 0} \theta^2 \mathbb{E}[\sqrt{D_1 D_2}\,e^{-\theta^2(D_1 + D_2)/2}]. \tag{3.85}$$

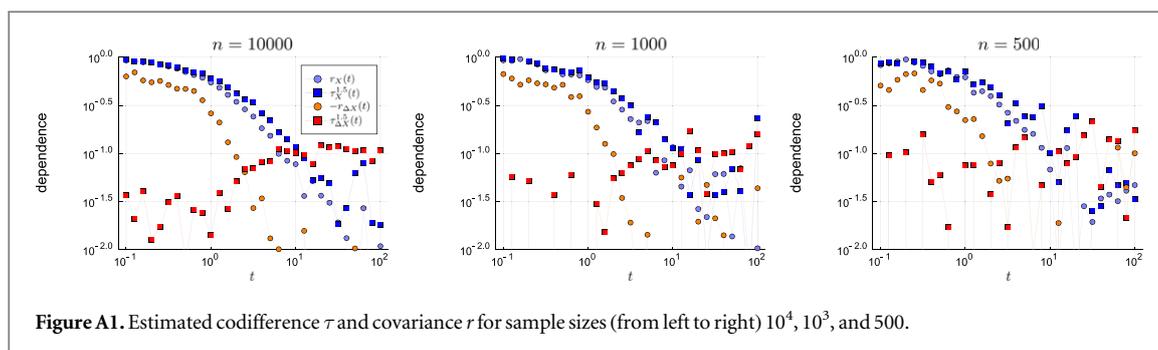The result follows. □

Acknowledgments

Figure A1. Estimated codifference τ and covariance r for sample sizes (from left to right) 10^4, 10^3, and 500.

Appendix. Sample size dependence of codifference and covariance

In supplement to figure 2 we show in figure A1 that even for smaller sample sizes such as 10^4, 10^3, and 500 significant differences between the covariance and codifference of increments are visible.

ORCID iDs

Ralf Metzler ☉ https://orcid.org/0000-0002-6013-7020

References

[1] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking *Phys. Chem. Chem. Phys.* **16** 24128
[2] Coffey W T, Kalmykov Y P and Waldron J T 1996 *The Langevin Equation* (Singapore: Word Scientific)
[3] Lutz E 2001 Fractional Langevin equation *Phys. Rev.* E **64** 051106
[4] Kou S C 2008 Stochastic modelling in nanoscale physics: subdiffusion within proteins *Ann. Appl. Stat.* **2** 501
[5] Codling E A, Plank M J and Benhamou S 2008 Random walk models in biology *J. R. Soc. Interface* **5** 813
[6] Metzler R and Klafter J 2004 The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics *J. Phys. A: Math. Gen.* **37** R161
[7] Schulz J H P, Barkai E and Metzler R 2014 Aging renewal theory and application to random walks *Phys. Rev.* X **4** 011028
[8] O'Shaughnessy B and Procaccia I 1985 Diffusion on fractals *Phys. Rev.* A **32** 3073
[9] Camacho C J, Weng Z, Vajda S and DeLisi C 1999 Free energy landscapes of encounter complexes in protein–protein association *Biophy. J.* **76** 1166
[10] Comtet A and Dean D S 1998 Exact results on Sinai's diffusion *J. Phys. A: Math. Gen.* **31** 8595
[11] Bouchaud J and Georges A 1990 Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications *Phys. Rep.* **195** 127
[12] Renshaw E and Henderson R 1981 The correlated random walk *J. App. Prob.* **18** 403
[13] Bovet P and Benhamou S 1988 Spatial analysis of animal's movements using a correlated random walk model *J. Theor. Biol.* **131** 419
[14] Tejedor V and Metzler R 2010 Anomalous diffusion in correlated continuous time random walks *J. Phys.* A **43** 082002
[15] Magdziarz M, Metzler R, Szczotka W and Zebrowski P 2012 Correlated continuous time random walks in external force fields *Phys. Rev.* E **85** 051103
[16] Schulz J H P, Chechkin A V and Metzler R 2013 Correlated continuous-time random walks: combining scale-invariance with long-range memory for spatial and temporal dynamics *J. Phys. A: Math. Theor.* **46** 475001
[17] Zaburdaev V, Denisov S and Klafter J 2015 Lévy walks *Rev. Mod. Phys.* **87** 483
[18] Cherstvy A G, Chechkin A V and Metzler R 2013 Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes *New J. Phys.* **15** 083039
[19] Chubynsky M V and Slater G W 2014 Diffusing diffusivity: a model for anomalous, yet Brownian, diffusion *Phys. Rev. Lett.* **113** 098302
[20] Höfling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* **76** 046602
[21] Nørregaard K, Metzler R, Ritter C M, Berg-Sørensen K and Oddershede L B 2017 Manipulation and motion of organelles and single molecules in living cells *Chem. Rev.* **117** 4342
[22] Metzler R, Jeon J-H and Cherstvy A G 2016 Non-Brownian diffusion in lipid membranes: experiments and simulations *Biochim. Biophys. Acta* **1858** 2451
[23] Everitt B and Skrondal A 2010 *The Cambridge Dictionary of Statistics* (Cambridge: Cambridge University Press)
[24] Downarowicz T 2007 Entropy *Scholarpedia* **2** 3901
[25] Kendall M D 1970 *Rank Correlation Methods* (London: Griffin)
[26] Latham P E and Roudi Y 2009 Mutual information *Scholarpedia* **4** 1658
[27] de Oliveira M M and Dickman R 2006 Moment ratios for the pair-contact process with diffusion *Phys. Rev.* E **74** 011124
[28] Tejedor V, Bénichou O, Voituriez R, Jungmann R, Simmel F, Selhuber-Unkel C, Oddershede L B and Metzler R 2010 Quantitative analysis of single particle trajectories: mean maximal excursion method *Biophys. J.* **98** 1364
[29] Magdziarz M and Klafter J 2010 Detecting origins of subdiffusion: p-variation test for confined systems *Phys. Rev.* E **82** 011129
[30] Weron A, Burnecki K, Akin E J, Solé L, Balcerek M, Tamkun M M and Krapf D 2017 Ergodicity breaking on the neuronal surface emerges from random switching between diffusive states *Sci. Rep.* **7**
[31] Metzler R 2015 Weak ergodicity breaking and ageing in anomalous diffusion *Int. J. Mod. Phys. Conf. Ser.* **36** 1560007
[32] He Y, Burov S, Metzler R and Barkai E 2008 Random time-scale invariant diffusion and transport coefficients *Phys. Rev. Lett.* **101** 058101

[33] Jeon J-H, Tejedor V, Burov S, Barkai E, Selhuber-Unkel C, Berg-Sørensen K, Oddershede L and Metzler R 2011 *In vivo* anomalous diffusion and weak ergodicity breaking of lipid granules *Phys. Rev. Lett.* **106** 048103

[34] Krapf D, Marinari E, Metzler R, Oshanin G, Squarcini A and Xu X 2018 Power spectral density of a single Brownian trajectory: what one can and cannot learn from it *New J. Phys.* **20** 023029

[35] Krapf D, Lukat N, Marinari E, Metzler R, Oshanin G, Selhuber-Unkel C, Squarcini A, Stadler L, Weiss M and Xu X 2019 Spectral content of a single non-Brownian trajectory *Phys. Rev.* X **9** 011019

[36] Castiglione P, Mazzino A, Muratore-Ginanneschi P and Vulpiani A 1999 On strong anomalous diffusion *Physica* D **134** 75

[37] Samorodnitsky G and Taqqu M S 1994 *Stable Non-Gaussian Random Processes* (London: Chapman and Hall)

[38] Kokoszka P S and Taqqu M S 1995 Fractional ARIMA with stable innovations *Stoch. Proc. Appl.* **60** 19

[39] Kokoszka P S and Taqqu M S 1996 Infinite variance stable moving averages with long memory *J. Econom.* **73** 79

[40] Magdziarz M 2007 Short and long memory fractional Ornstein–Uhlenbeck α-stable processes *Stoch. Models* **23** 451

[41] Magdziarz M 2007 Fractional langevin equation with α-stable noise. a link to fractional ARIMA time series *Stud. Math.* **181** 47

[42] Burnecki K, Klafter J, Magdziarz M and Weron A 2008 From solar flare time series to fractional dynamics *Physica* A **387** 1077

[43] Wyłomańska A, Chechkin A, Gajda J and Sokolov I 2015 Codifference as a practical tool to measure interdependence *Physica* A **421** 412

[44] Magdziarz M and Weron A 2011 Anomalous diffusion: testing ergodicity breaking in experimental data *Phys. Rev.* E **84** 051138

[45] Loch-Olszewska H and Szwabiński J 2018 Detection of ϵ-ergodicity breaking in experimental data-a study of the dynamical functional sensibility *J. Chem. Phys.* **148** 204105

[46] Weron A, Burnecki K, Mercik S and Weron K 2005 Complete description of all self-similar models driven by Lévy stable noise *Phys. Rev.* E **71** 016113

[47] Magdziarz M 2009 Stochastic representation of subdiffusion processes with time-dependent drift *Stoch. Proc. Appl.* **119** 3238

[48] Haubold H, Mathai A and Saxena R 2011 Mittag-Leffler functions and their applications *J. Appl. Math.* **2011** 298628

[49] Lebowitz J L and Penrose O 1973 Modern ergodic theory *Phys. Today* **26** 23

[50] Janson S 1997 *Gaussian Hilbert Spaces, Cambridge Tracts in Mathematics* (Cambridge: Cambridge University Press)

[51] Goldstein H 2011 *Multilevel Statistical Models* (London: Wiley)

[52] Beck C and Cohen E G 2003 Superstatistics *Physica* A **322** 267

[53] Schneider W 1990 *Grey Noise* (Singapore: World Scientific)

[54] Mainardi F 1996 Fundamental solutions for the fractional diffusion-wave equation *Appl. Math. Lett.* **9** 23

[55] da Silva J L and Erraoui M 2015 Grey Brownian motion local time: existence and weak-approximation *Stochastics* **87** 347

[56] Mandelbrot B B and van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422

[57] Pagnini G 2013 The M-Wright function as a generalization of the Gaussian density for fractional diffusion processes *Frac Calc. Appl. Anal.* **16** 436

[58] Cornfel I, Fomin S and Sinai Y 1982 *Ergodic Theory* (Heidelberg: Springer)

[59] Ślęzak J 2017 Asymptotic behaviour of time averages for non-ergodic Gaussian processes *Ann. Phys.* **383** 285

[60] Klafter J, Shlesinger M F and Zumofen G 1996 Beyond Brownian motion *Phys. Today* **49** 33

[61] Barthelemy P, Bertolotti J and Wiersma D S 2008 A Lévy flight for light *Nature* **453** 495

[62] Adler R J 2007 *Random Fields and Geometry* (Berlin: Springer)

[63] Uhlenbeck G E and Ornstein L S 1930 On the theory of the Brownian motion *Phys. Rev.* **36** 823

[64] Bryson J A E and Henrikson L J 1968 Estimation using sampled data containing sequentially correlated noise *J. Spacecraft Rockets* **5** 662

[65] Jeon J-H, Barkai E and Metzler R 2013 Noisy continuous time random walks *J. Chem. Phys.* **139** 121916

[66] Ślęzak J, Metzler R and Magdziarz M 2018 Superstatistical generalised Langevin equation: non-Gaussian viscoelastic anomalous diffusion *New J. Phys.* **20** 023026

[67] Zwanzig R 2001 *Nonequilibrium Statistical Mechanics* (Oxford: Oxford University Press)

[68] Ashkin A 1970 Acceleration and trapping of particles by radiation pressure *Phys. Rev. Lett.* **24** 156

[69] Goychuk I 2012 Viscoelastic subdiffusion: generalized Langevin equation approach *Adv. Chem. Phys.* **150** 187

[70] Sungkaworn T, Jobin M-L, Burnecki K, Weron A, Lohse M J and Calebiro D 2017 Single-molecule imaging reveals receptor-G protein interactions at cell surface hot spots *Nature* **550** 543

[71] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 Bayesian nested sampling analysis of single particle tracking data: maximum likelihood model selection applied to stochastic diffusivity data *Phys. Chem. Chem. Phys.* **20** 29018

[72] Cherstvy A G, Thapa S, Wagner C E and Metzler R Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels *Soft Matter* accepted (https://doi.org/10.1039/C8SM02096E)

[73] de Haan L 1970 On regular variation and its applications to the weak convergence of sample extremes *Mathematical Centre Tracts* vol 32 (Amsterdam: Mathematisch Centrum)

[74] Mikosch T 1999 Regular variation, subexponentiality and their applications in probability theory *Report Eurandom* **vol 99013** (Eindhoven: Eurandom)

[75] Tjøstheim D 1986 Some doubly stochastic time series models *J. Time Ser. Anal.* **7** 51

[76] Cox J C, Ingersoll J E and Ross S A 1985 A theory of the term structure of interest rates *Econometrica* **53** 385

[77] Jain R and Sebastian K L 2017 Diffusing diffusivity: a new derivation and comparison with simulations *J. Chem. Sci.* **129** 929

[78] Jain R and Sebastian K L 2016 Diffusion in a crowded, rearranging environment *J. Phys. Chem.* B **120** 3988

[79] Tyagi N and Cherayil B J 2017 Non-Gaussian Brownian diffusion in dynamically disordered thermal environments *J. Phys. Chem.* B **121** 7204

[80] Chechkin A V, Seno F, Metzler R and Sokolov I M 2017 Brownian yet non-Gaussian diffusion: From superstatistics to subordination of diffusing diffusivities *Phys. Rev.* X **7** 021002

[81] Sposini V, Chechkin A, Seno F, Pagnini G and Metzler R 2018 Random diffusivity from stochastic equations: comparison of two models for Brownian yet non-Gaussian diffusion *New J. Phys.* **20** 043044

[82] Lanoiselée Y, Moutal N and Grebenkov D S 2018 Diffusion-limited ractions in dynamic heterogeneous media *Nat. Commun.* **9** 4398

[83] Jeanblanc M, Yor M and Chesney M 2009 *Mathematical Methods for Financial Markets* (Berlin: Springer)