Probability distribution of the time-averaged mean-square displacement of a Gaussian process

Denis S. Grebenkov*

Laboratoire de Physique de la Matière Condensée (UMR 7643), CNRS, Ecole Polytechnique, F-91128 Palaiseau, France, Laboratoire Poncelet (UMI 2615), CNRS, Independent University of Moscow, Bolshoy Vlasyevskiy Pereulok 11, 119002 Moscow, Russia, and Chebyshev Laboratory, Saint Petersburg State University, 14th line of Vasil'evskiy Ostrov 29, Saint Petersburg, Russia

(Received 19 June 2011; published 21 September 2011)

We study the probability distribution of the time-averaged mean-square displacement of a discrete Gaussian process. An empirical approximation for the probability density is suggested and numerically validated for fractional Brownian motion. The optimality of quadratic forms for inferring dynamical and microrheological quantities from individual random trajectories is discussed, with emphasis on a reliable interpretation of single-particle tracking experiments.

DOI: 10.1103/PhysRevE.84.031124 PACS number(s): 02.50.−r, 05.60.−k, 05.10.−a, 02.70.Rr

I. INTRODUCTION

Single-particle tracking (SPT) techniques that survey an individual trajectory of a tracer have become the standard tools for studying dynamics in complex or viscoelastic media, notably in living cells [1–7]. By considering each acquired trajectory as a random realization of an unknown stochastic process, one aims to infer dynamical quantities of this process (e.g., diffusion coefficient or subdiffusion scaling exponent) or microrheological quantities of the medium (e.g., stiffness [6], compliance and dynamic moduli [8,9], or viscosities [10]). For this purpose, various quadratic forms (or functionals) of the trajectory have been introduced: mean-square displacement (MSD), squared root-mean square of the position, velocity autocorrelation function, power spectral density, etc. Since a limited number (from few to few hundred) of random trajectories are typically available, statistical "ensemble" averaging of these forms may be problematic, especially for tracers that move in spatially heterogeneous and time-evaluating media such as living cells. An appropriate time averaging is then used to reduce fluctuations of a quantity of interest (e.g., MSD). However, even if the data sample is long, the quadratic form of a random trajectory is still a random variable. The knowledge of its probability distribution is therefore of primary importance for a reliable interpretation of SPT experiments.

General quadratic forms of Gaussian (or normal) random variables were intensively studied in mathematical statistics (see Ref. [11] and references therein). Since the first approximation by a chi-squared distribution suggested by Rice in 1944 [12], several rigorous series representations of the probability density $p(z)$ of a quadratic form have been proposed. One can mention the representation as a mixture of chi-squared distributions [13–16], a power series expansion [17,18], and a series expansion over Laguerre polynomials [19–21] (see Appendix A for a brief summary). Quite surprisingly, this mathematical theory seems to remain unknown in the biophysical community. The major contribution is attributed to Qian and co-workers, who showed for Brownian motion that the time-averaged MSD with the lag time 1 has a Gamma (or chi-squared) distribution [22]. The distribution of diffusion coefficients was also studied via Monte Carlo simulations by Saxton [23,24].

In this paper, we employ the mathematical theory of Gaussian quadratic forms for studying the probability distribution of the time-averaged MSD of a discrete Gaussian process. We propose an empirical approximation of the probability density $p(z)$ by the generalized Gamma distribution:

$$\psi(z) \propto z^{\nu-1} \exp\left(-\frac{a}{z} - \frac{z}{b}\right),$$

with three fitting parameters a, b, and ν, which can be found through their relation to three moments of the quadratic form. The function $\psi(z)$ captures the essential features of the probability density $p(z)$, namely, the exponential decay at large z and very sharp decay at intermediate z. In turn, the approximation fails in the region of small z, which, however, turns out to be irrelevant for most practical applications. We check numerically the accuracy of this approximation for the time-averaged MSD of fractional Brownian motion.

It is worth stressing the difference between our approximation and the aforementioned mathematical approaches. When an infinite series representation of the probability density is truncated, such an approximation can be made as accurate as needed by taking a large enough number of terms. However, if this number is too large, such representations are of limited interest for practical applications (see Appendix A). In turn, our approach starts from a convenient explicit function $\psi(z)$, which seems to approximate the probability density $p(z)$. Although the accuracy of this empirical approximation cannot be improved, its advantages are the functional simplicity and the need for knowing only three moments of the quadratic form. These advantages are attractive for the analysis of SPT experiments.

The paper is organized as follows. In Sec. II A, we start by recalling the theoretical basis for studying quadratic forms of discrete Gaussian processes. After that, we discuss in Sec. II B the optimality of quadratic forms for inferring model parameters from individual random trajectories. An empirical approximation for $p(z)$ by the generalized Gamma distribution is then proposed. In Sec. III we investigate the quadratic form related to the time-averaged MSD, which is the most often way for analyzing random trajectories of SPT experiments. The probability density $p(z)$ of the time-averaged MSD of

*denis.grebenkov@polytechnique.edu

fractional Brownian motion with different Hurst exponents is computed numerically and compared to the empirical approximation. The paper ends by conclusions, in which the main results are summarized and their potential applications are discussed.

II. THEORETICAL BASIS

A. Classical approach to quadratic forms

In this subsection, we remind readers of classical tools for studying quadratic forms of Gaussian random variables (e.g., see Refs. [11,25]). We consider a discrete Gaussian process $\mathbf{x} = \{x_1, \ldots, x_N\} \in \mathbb{R}^N$ with mean $\mathbf{x}^0 \in \mathbb{R}^N$ and the covariance matrix $\mathbf{C}_{n_1 n_2} = \langle x_{n_1} x_{n_2} \rangle$, where $\langle \cdots \rangle$ denotes the expectation with respect to the Gaussian probability density of $\mathbf{x}$:

$$P_N(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}^0)^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{x}^0)\right]. \tag{1}$$

A given matrix $\mathbf{M} \in \mathbb{R}^N \times \mathbb{R}^N$ defines the quadratic form:

$$\chi = \tfrac{1}{2}(\mathbf{x}^T \mathbf{M} \mathbf{x}).$$

We are interested in computing explicitly the probability distribution of this random variable. Its characteristic function $\phi(k) = \langle e^{ik\chi} \rangle$ is easily found by regrouping two quadratic forms and computing Gaussian integrals:

$$\phi(k) = \frac{1}{\sqrt{\det(\mathbf{I} - ik\mathbf{MC})}} \times \exp\left\{-\frac{1}{2}\mathbf{x}^{0T}\left[\mathbf{C}^{-1} - \mathbf{C}^{-1}(\mathbf{C}^{-1} - ik\mathbf{M})^{-1}\mathbf{C}^{-1}\right]\mathbf{x}^0\right\}. \tag{2}$$

Throughout the paper, we focus on centered Gaussian processes (with $\mathbf{x}^0 = 0$), for which

$$\phi(k) = \frac{1}{\sqrt{\det(\mathbf{I} - ik\mathbf{MC})}}. \tag{3}$$

Note that $\phi(k)$ is a complex-valued function.

Using the Schur factorization of the square matrix $\mathbf{MC} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ (where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Lambda}$ is an upper triangular matrix [26]), one can represent the determinant as

$$\det(\mathbf{I} - ik\mathbf{MC}) = \prod_{q=1}^{\tilde{N}} (1 - ik\lambda_q),$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_{\tilde{N}}$ denote all *nonzero* diagonal elements of the Schur matrix $\mathbf{\Lambda}$ (indexed in the descending order), and $\tilde{N} \leqslant N$ is the number of these elements. When the Schur matrix is diagonal, λ_q are the nonzero *eigenvalues* of the matrix $\mathbf{MC}$, and we use this term throughout the paper. One obtains therefore

$$\phi(k) = \prod_{q=1}^{\tilde{N}} (1 - ik\lambda_q)^{-\frac{1}{2}} = \exp\left[-\frac{1}{2}\sum_{q=1}^{\tilde{N}} \ln(1 - ik\lambda_q)\right]. \tag{4}$$

It is worth noting that $\phi(k)$ from this equation is also the characteristic function of another quadratic form

$$\chi' = \frac{1}{2}\sum_{q=1}^{\tilde{N}} \lambda_q y_q^2, \tag{5}$$

where y_q are independent Gaussian variables with mean zero and variance 1. This representation is called the canonical quadratic form, to which a given quadratic form $\mathbf{M}$ can be reduced by an orthogonal transformation. Here the eigenvalues λ_q stand as the weights of independent Gaussian variables. Although these weights can in general be positive or negative (see, e.g., Ref. [19]), we focus on nonnegative definite quadratic forms with $\lambda_q > 0$ (zero eigenvalues are allowed but not included in the set $\{\lambda_q\}$).

Developing the logarithm in Eq. (4) into a Taylor series, one gets

$$\phi(k) = \exp\left[\frac{1}{2}\sum_{m=1}^{\infty} \frac{(ik)^m}{m} \sum_{q=1}^{\tilde{N}} \lambda_q^m\right].$$

The logarithm of $\phi(k)$ is the generating function for the cumulant moments $\kappa_m \equiv \langle \chi^m \rangle_c$:

$$\ln \phi(k) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} \kappa_m.$$

Comparing with the previous relation, one finds then

$$\kappa_m = \frac{(m-1)!}{2}\sum_{q=1}^{\tilde{N}} \lambda_q^m = \frac{(m-1)!}{2}\mathrm{tr}[(\mathbf{MC})^m]. \tag{6}$$

For instance, $\kappa_1 = \frac{1}{2}\mathrm{tr}(\mathbf{MC})$ and $\kappa_2 = \frac{1}{2}\mathrm{tr}((\mathbf{MC})^2)$ are the mean and the variance of χ, respectively. The skewness and kurtosis are also expressed in terms of the cumulant moments as $\kappa_3/\kappa_2^{3/2}$ and κ_4/κ_2^2, respectively. Note that the moments $\mu_m \equiv \langle \chi^m \rangle$, which are generated by the characteristic function $\phi(k)$,

$$\phi(k) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \mu_m,$$

are easily expressed through the cumulant moments, e.g.,

$$\begin{aligned} \mu_1 &= \kappa_1, \\ \mu_2 &= \kappa_2 + \kappa_1^2, \\ \mu_3 &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3, \\ \mu_4 &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4. \end{aligned} \tag{7}$$

The negative-order moments $\mu_{-\alpha} \equiv \langle \chi^{-\alpha} \rangle$ can be computed using the identity

$$z^{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} ds s^{\alpha-1} e^{-sz} \qquad (\alpha > 0)$$

[here $\Gamma(\alpha)$ is the Gamma function], which yields

$$\mu_{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} ds s^{\alpha-1} \int_{\mathbb{R}^N} d\mathbf{x} P_N(\mathbf{x}) \exp\left[-\frac{s}{2}(\mathbf{x}^T \mathbf{M} \mathbf{x})\right],$$

from which

$$\mu_{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} ds s^{\alpha-1} \varphi(s) \qquad (\alpha > 0), \tag{8}$$

where $\varphi(s) \equiv \phi(is)$ is the Laplace transform of the probability density:

$$\varphi(s) \equiv \phi(is) = \langle e^{-s\chi} \rangle = \int_0^\infty dz e^{-sz} p(z)$$
$$= \frac{1}{\sqrt{\det(\mathbf{I} + s\mathbf{MC})}} = \prod_{q=1}^{\tilde{N}} (1 + s\lambda_q)^{-\frac{1}{2}}. \quad (9)$$

The moments $\mu_{-\alpha}$ appear thus as the Mellin transform of the function $\varphi(s)$. Using the identity

$$\ln z = \lim_{\alpha \to 0} \frac{z^\alpha - 1}{\alpha},$$

one gets a closed formula for the mean logarithm of χ:

$$\langle \ln(\chi/\chi_0) \rangle = \int_0^\infty \frac{ds}{s} [e^{-s} - \varphi(s/\chi_0)],$$

where χ_0 is an arbitrary normalization constant in units of χ.

Finally, the probability density $p(z)$ of the random variable χ is retrieved through the inverse Fourier transform of $\phi(k)$:

$$p(z) = \int_{-\infty}^\infty \frac{dk}{2\pi} e^{-ikz} \phi(k). \quad (10)$$

In practice, this computation can be rapidly performed by a fast Fourier transform.

In a similar way, one can compute the joint distribution of several quadratic forms and analyze correlations between them (see Appendix B).

B. "Best" and "worst" probability densities

Given the explicit form for the mean and variance, one may wonder what is the "best strategy" to infer information from a single trajectory measurement? By "the best strategy" we mean such a choice of the matrix $\mathbf{M}$ that would result in the smallest variance. Evidently, the minimal variance is achieved for a trivial choice $\mathbf{M} = 0$. One needs therefore to add a reasonable constraint that would eliminate this trivial solution. A natural choice is to keep the mean value κ_1 to be a constant.

More generally, one may ask the following question: How can one choose the quadratic form (i.e., the matrix $\mathbf{M}$) in order to minimize a chosen cumulant moment κ_m under the constraint that the mean value κ_1 is fixed? Using Eq. (6), one has to minimize the function

$$f(\mathbf{M}) = \frac{(m-1)!}{2} \mathrm{tr}((\mathbf{MC})^m) + \alpha \left[\frac{1}{2} \mathrm{tr}(\mathbf{MC}) - \kappa_1 \right]$$

over all non-negative definite matrices $\mathbf{M}$, where α is a Lagrange multiplier. Writing the traces in terms of the eigenvalues λ_q, one gets another function to be minimized:

$$f(\lambda_1, \ldots, \lambda_N) = \frac{(m-1)!}{2} \sum_{q=1}^N \lambda_q^m + \alpha \left(\frac{1}{2} \sum_{q=1}^N \lambda_q - \kappa_1 \right)$$

over all *non-negative* λ_1,..., λ_N (here N is the length of the data sample). Equating the derivative $\partial f/\partial \lambda_q$ to 0 for each $q = 1, \ldots, N$, one easily checks that this function is minimized when all λ_q are equal to each other, i.e.,

$$\lambda_1 = \cdots = \lambda_N = \lambda = 2\kappa_1/N$$

(the last equality allows one to satisfy the condition of the fixed mean value). Such a matrix $\mathbf{\Lambda}$ is proportional to the identity matrix, from which

$$\mathbf{M}_{\mathrm{opt}} = (2\kappa_1/N)\mathbf{C}^{-1}. \quad (11)$$

Note that this optimal quadratic form minimizes at once all cumulant moments (except for κ_1 which was fixed).

In this "best" situation, the characteristic function is reduced to

$$\phi(k) = (1 - ik\lambda)^{-N/2},$$

from which one deduces the Gamma distribution:

$$p(z) = \frac{z^{N/2-1}}{\Gamma(N/2)\lambda^{N/2}} e^{-z/\lambda}, \quad (12)$$

whose moments are $\mu_k = \lambda^k \Gamma(N/2 + k)/\Gamma(N/2)$. In particular, the mean and variance are $\kappa_1 = \lambda N/2$ and $\kappa_2 = \lambda^2 N/2 = 2\kappa_1^2/N$. Note that the canonical form (5) is reduced to $2\chi/\lambda = y_1^2 + \cdots + y_N^2$, which is known as the chi-squared random variable with N degrees of freedom.

On the other hand, the largest cumulant moments κ_m (always with a fixed mean value κ_1) are obtained when all eigenvalues λ_q are zero except one: $\lambda_1 = 2\kappa_1$ and $\lambda_q = 0$ for $q = 2, 3, \ldots, N$. The "worst" probability density is once again the Gamma distribution (12), with $N = 1$ and $\lambda = 2\kappa_1$:

$$p(z) = \frac{e^{-z/\lambda}}{\sqrt{\pi \lambda z}}. \quad (13)$$

This density does not have even a maximum, and its variance is $\kappa_2 = 2\kappa_1^2$.

We conclude that whatever the quadratic form (with fixed κ_1) is, the variance of χ lies between $2\kappa_1^2/N$ and $2\kappa_1^2$. One gets therefore the exact lower and upper bounds for the variance (similarly, both bounds can be found for any cumulant moment).

In practice, the use of the optimal quadratic form $\mathbf{M}_{\mathrm{opt}}$ may be problematic because the dynamics of tracers (i.e., the covariance matrix and even the Gaussian character of motion) is not known and has to be inferred from the acquired trajectories. In this situation, one starts by assuming an appropriate model for the dynamics (e.g., restricted or hindered diffusion, fractional Brownian motion, generalized Langevin equation, or continuous-time random walks). For Gaussian models, one needs to assume a specific form of the covariance matrix [e.g., Eqs. (25), (27) below], which may contain some unknown physical parameters to be inferred (e.g., diffusion coefficient or Hurst exponent). After that, numerical simulations can be used to study the dependence of $\mathbf{M}_{\mathrm{opt}}$ on these parameters, aiming to determine a nearly optimal quadratic form for a given range of physical parameters.

C. Asymptotic behavior of the probability density

The asymptotic behavior at large z can be obtained by deforming the path of integration in Eq. (10) and by noting that most of the contribution arises from the region around

$1/(i\lambda_1)$. If the largest eigenvalue λ_1 is $m \geqslant 1$ times degenerate, one gets [27]

$$p(z) \simeq \frac{z^{m/2-1}e^{-z/\lambda_1}}{\Gamma(m/2)\lambda_1^{m/2}} \prod_{q=m+1}^{\tilde{N}} (1-\lambda_q/\lambda_1)^{-1/2} \qquad (z \gg \lambda_1). \tag{14}$$

This is the Gamma distribution in Eq. (12) with $N = m$ and $\lambda = \lambda_1$ multiplied by an additional factor that accounts for the marginal contribution from the remaining eigenvalues.

The analysis of the probability density at small z is more subtle. We first consider the behavior of the Laplace transform $\varphi(s)$ when $s \to \infty$. When $s \gg 1/\lambda_{\tilde{N}}$ ($\lambda_{\tilde{N}}$ being the smallest nonzero eigenvalue), one can replace each $1 + s\lambda_q$ by $s\lambda_q$ in Eq. (9), from which the asymptotic behavior of $p(z)$ is

$$p(z) \simeq \frac{z^{\tilde{N}/2-1}}{\Gamma(\tilde{N}/2)\sqrt{\lambda_1 \cdot \cdots \cdot \lambda_{\tilde{N}}}} \qquad (z \ll \lambda_{\tilde{N}}). \tag{15}$$

When $\tilde{N}$ is relatively large (in practice, $\tilde{N} = 10$ is enough), the probability density $p(z)$ for $z < \lambda_{\tilde{N}}$ is negligible, and Eq. (15) becomes of a limited interest. One needs therefore to describe the intermediate region of z between $\lambda_{\tilde{N}}$ and λ_1.

If the eigenvalues λ_q are far from each other, $\lambda_1 \gg \lambda_2 \gg \cdots \gg \lambda_{\tilde{N}}$, for s lying between certain $1/\lambda_n$ and $1/\lambda_{n+1}$, one can approximate $1 + s\lambda_q \approx s\lambda_q$ for $q \leqslant n$ and $1 + s\lambda_q \approx 1$ for $q > n$:

$$\varphi(s) = \prod_{q=1}^{\tilde{N}} (1+s\lambda_q)^{-1/2} \approx s^{-n/2} \prod_{q=1}^{n} \lambda_q^{-1/2},$$

from which for $\lambda_{n+1} \ll z \ll \lambda_n$,

$$p(z) \sim \frac{z^{n(z)/2-1}}{\Gamma(n(z)/2)\sqrt{\lambda_1 \cdot \cdots \cdot \lambda_{n(z)}}}, \tag{16}$$

where we wrote $n(z)$ to stress that the index n depends on z. More formally, $n(z)$ can be defined as a counting function of the eigenvalues λ_q, $n(z) = \max\{q : \lambda_q \geqslant z\}$, so that $n(\lambda_q) = q$.

In practice, the computation of the counting function $n(z)$ requires the knowledge of all the eigenvalues λ_q. If these values are known (or can be numerically computed for a given covariance matrix), the direct Fourier inversion of the characteristic function $\phi(k)$ would result in a faster and much more accurate computation of the probability density $p(z)$. Nevertheless, the qualitative relation (16) illustrates that the decay of the probability density $p(z)$ for the *intermediate* values of z (between $\lambda_{\tilde{N}}$ and λ_1) may be *faster* than a power law, due to the dependence $n(z)$.

D. Generalized Gamma distribution

Since the "best" and "worst" probability densities are the Gamma distributions (12) and (13), the probability density for any quadratic form is somehow "comprised" between them. At the same time, the asymptotic relations (14) and (15) indicate that the Gamma distribution (12) is not "flexible" enough to approximate $p(z)$ in general. In fact, both asymptotic behaviors of the Gamma distribution at small and large z are determined by the same scale λ, while the probability density $p(z)$ has multiple scales ranging from $\lambda_{\tilde{N}}$ to λ_1.

Looking at the qualitative relation (16), it is tempting to include a faster decay in the Gamma distribution through a generalized Gamma distribution

$$\psi(z) = \frac{z^{\nu-1}}{2(ab)^{\nu/2}K_\nu\left(2\sqrt{a/b}\right)} \exp\left(-\frac{a}{z} - \frac{z}{b}\right) \qquad (z > 0) \tag{17}$$

[and $\psi(z) = 0$ for $z \leqslant 0$], where $K_\nu(z)$ is the modified Bessel function of the second kind. This distribution, which is characterized by three real-valued parameters $a \geqslant 0$, $b > 0$ and ν, is reduced to the Gamma distribution in the limit $a \to 0$. For $a > 0$, the exponential $e^{-a/z}$ determines the asymptotic behavior at $z \to 0$ and ensures that all derivatives of $\psi(z)$ vanish at $z = 0$. This behavior could be expected for the probability density $p(z)$ in the limit $N \to \infty$.

We emphasize that the choice of the function $\psi(z)$ is empirical. Moreover, Eq. (17) is expected to approximate the probability density $p(z)$ *only* for intermediate and large values of z. At small z, Eq. (17) evidently fails to approximate $p(z)$ which follows a power-law decay according to Eq. (15). However, as we already mentioned, the range of small z is irrelevant for most practical applications.

The Fourier transform of Eq. (17) is [28]

$$\phi(k) = (1-ikb)^{-\nu/2} \frac{K_\nu[2\sqrt{(a/b)(1-ikb)}]}{K_\nu(2\sqrt{a/b})}, \tag{18}$$

while the moments of the above distribution are

$$\mu_k \equiv \langle \chi^k \rangle = (ab)^{k/2} \frac{K_{\nu+k}(2\sqrt{a/b})}{K_\nu(2\sqrt{a/b})}. \tag{19}$$

Using the recurrence relations

$$K_{\alpha+1}(z) = K_{\alpha-1}(z) + \frac{2\alpha}{z} K_\alpha(z),$$

one can further simplify the expressions for the moments:

$$\begin{aligned}
\mu_1 &= \sqrt{ab}\,\eta, \\
\mu_2 &= (ab)\left[1 + \frac{2(1+\nu)}{c}\eta\right], \\
\mu_3 &= (ab)^{3/2}\left\{\frac{2(1+\nu)}{c} + \left[1 + \frac{4(2+\nu)(1+\nu)}{c^2}\right]\eta\right\}, \\
\mu_4 &= (ab)^2\left\{1 + \frac{4(3+\nu)(2+\nu)}{c^2} \right. \\
&\quad \left. + \frac{8(2+\nu)}{c}\left[1 + \frac{(3+\nu)(1+\nu)}{c^2}\right]\eta\right\}, \ldots,
\end{aligned}$$

where $c = 2\sqrt{a/b}$ and $\eta \equiv K_{\nu+1}(c)/K_\nu(c)$. These four moments allow one to determine the mean, variance, skewness, and kurtosis of the probability density $\psi(z)$.

The most probable value of χ is obtained by equating the derivative $\partial\psi/\partial z$ to 0, from which one gets a quadratic equation on the most probable value with the single positive solution:

$$\chi_{\text{most}} = \frac{\sqrt{(1-\nu)^2b^2 + 4ab} - (1-\nu)b}{2}. \tag{20}$$

Note that the probability density $\psi(z)$ in Eq. (17) has the following symmetry: If χ is distributed according to $\psi(z)$,

then $\chi' = 1/\chi$ is distributed according to $\psi'(z)$ with $a' = 1/b$, $b' = 1/a$, and $\nu' = -\nu$.

E. Reconstruction by three moments

The knowledge of three moments of the probability density $\psi(z)$ allows one to retrieve the parameters a, b, and ν. Using Eq. (19), one can represent the ratios μ_2/μ_1^2 and μ_3/μ_1^3 as the system of two nonlinear equations:

$$\begin{aligned}\frac{\mu_2}{\mu_1^2} &= \frac{K_{\nu+2}(c)K_\nu(c)}{K_{\nu+1}^2(c)},\\ \frac{\mu_3}{\mu_1^3} &= \frac{K_{\nu+3}(c)K_\nu^2(c)}{K_{\nu+1}^3(c)}.\end{aligned} \tag{21}$$

If the moments in the left-hand side are known, these two equations can be solved numerically to extract ν and c, from which one gets a and b as

$$b = 2\mu_1\frac{K_{\nu+1}(c)}{cK_\nu(c)}, \qquad a = \frac{bc^2}{4}.$$

These relations can be used to find the parameters a, b, and ν and thus constructing an approximation $\psi(z)$ of the probability density $p(z)$ *directly* from the three moments μ_1, μ_2, and μ_3 of the quadratic form. Either for a given class of Gaussian processes, or directly from experimental data, the computation of the moments is much easier than that of the whole probability distribution. For instance, for a large class of stochastic processes governed by a generalized Langevin equation, the first two moments of the time-averaged MSD were recently computed [29]. The computation can be extended to the third (and higher) moments that would allow the application of the above reconstruction scheme.

III. TIME-AVERAGED MEAN-SQUARE DISPLACEMENT

In this section, we focus on the time-averaged MSD, which is often used for the analysis of individual trajectories:

$$\chi = \frac{1}{N-n}\sum_{k=1}^{N-n}(x_{k+n}-x_k)^2 = \frac{1}{2}(\mathbf{x}^T\mathbf{M}\mathbf{x}), \tag{22}$$

where n is the lag time, and

$$\mathbf{M}_{n_1n_2} = \frac{2}{N-n}(m_{n_1}\delta_{n_1,n_2} - \delta_{|n_1-n_2|,n}),$$

with

$$m_k = \begin{cases}\begin{cases}1 & k \leqslant n \quad \text{or} \quad k \geqslant N-n\\ 2 & n < k < N-n\end{cases} & n \leqslant N/2,\\ \begin{cases}1 & k \leqslant N-n \quad \text{or} \quad k \geqslant n\\ 0 & N-n < k < n\end{cases} & n > N/2.\end{cases}$$

$\mathbf{M}$ is a three-diagonal matrix in which two subdiagonals are shifted by n steps from the main diagonal.

1. Case $n = N-1$

In this case, $\chi = (x_N - x_1)^2$ and the matrix $\mathbf{MC}$ has the only nonzero eigenvalue:

$$\lambda_1 = 2(\mathbf{C}_{1,1} - \mathbf{C}_{N,1} - \mathbf{C}_{1,N} + \mathbf{C}_{N,N}).$$

As shown in Sec. II B, the probability density is given by Eq. (13), which has the highest variance. This is not surprising as all data points except x_1 and x_N are thrown out by the quadratic form (22) with $n = N-1$.

2. Case $n = N-2$

For $n = N-2$, there are two nonzero eigenvalues of the matrix $\mathbf{MC}$, which can be found by solving the determinant equation:

$$\lambda_{1,2} = \frac{u_1+u_2 \pm \sqrt{(u_1-u_2)^2+4u_{12}^2}}{2},$$

where

$$\begin{aligned}u_1 &= \mathbf{C}_{1,1} - \mathbf{C}_{N-1,1} - \mathbf{C}_{1,N-1} + \mathbf{C}_{N-1,N-1},\\ u_2 &= \mathbf{C}_{2,2} - \mathbf{C}_{N,2} - \mathbf{C}_{2,N} + \mathbf{C}_{N,N},\\ u_{12} &= \mathbf{C}_{1,2} - \mathbf{C}_{N-1,2} - \mathbf{C}_{1,N} + \mathbf{C}_{N-1,N}.\end{aligned} \tag{23}$$

In this case the probability density can be explicitly obtained through the inverse Laplace transform of $\varphi(s)$ as [30] (this result was reported in Ref. [31])

$$p(z) = \frac{1}{\sqrt{\lambda_1\lambda_2}}\exp\left(-\frac{z}{2}\left[\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right]\right)I_0\left(\frac{z}{2}\left|\frac{1}{\lambda_2}-\frac{1}{\lambda_1}\right|\right), \tag{24}$$

where $I_0(z)$ is the modified Bessel function of the first kind. The asymptotic behavior of $p(z)$ is

$$\begin{aligned}p(z) &\simeq \frac{e^{-z/\lambda_1}}{\sqrt{\pi(\lambda_1-\lambda_2)z}}\left[1+O(z^{-1})\right] \quad (z\to\infty),\\ p(z) &\simeq \frac{1}{\sqrt{\lambda_1\lambda_2}} + O(z) \quad (z\to 0),\end{aligned}$$

in agreement with the general Eqs. (14) and (15).

3. General case

For any n, the matrix $\mathbf{MC}$ has exactly $\tilde{N} = N-n$ nonzero eigenvalues λ_q, as shown in Appendix C. Since the quadratic form (22) with smaller n includes a larger number of points x_n, it is expected to have a smaller variance, as shown below.

A. Brownian motion

We start the analysis by considering a one-dimensional off-lattice random walk, for which the position of a particle after n steps is

$$x_n = \sum_{j=1}^{n}\sigma_j.$$

The random displacements σ_j are independent Gaussian variables with mean zero and variance 1 (note that a given variance can be straightforwardly implemented through a dimensional analysis). This is a discrete version of Brownian motion (with the diffusion coefficient equal to 1/2), for which the covariance matrix is simply

$$\mathbf{C}_{n_1n_2} = \langle x_{n_1}x_{n_2}\rangle = \min\{n_1,n_2\}. \tag{25}$$

It is worth noting that in practice, the position measurements are affected by experimental noises. In other words, the "true"

position x_n is superimposed with the noise ϵ_n (i.e., localization uncertainty) in the acquired position $\tilde{x}_n = x_n + \epsilon_n$ [32]. In a first approximation, $\{\epsilon_k\}$ can be treated as independent Gaussian random variables with mean zero and variances ε_k^2. In this case, the acquired trajectory $(\tilde{x}_1, \ldots, \tilde{x}_N)$ is again a Gaussian process with mean zero and covariance matrix $\tilde{\mathbf{C}} = \mathbf{C} + \mathrm{diag}(\varepsilon_1^2, \ldots, \varepsilon_N^2)$, where the second term is the diagonal matrix with entries ε_k^2. Using the covariance matrix $\tilde{\mathbf{C}}$, one can study the influence of localization errors onto measured diffusion coefficients and search for the lag time that would optimally reduce these errors. Since this effect was already investigated in Ref. [32], the localization uncertainties are not included in the following analysis.

Given the explicit structures of the matrices $\mathbf{C}$ and $\mathbf{M}$, the moments of the quadratic form χ can be found according to Eq. (6). The first moment is simply n, while the second moment was derived by Qian *et al*. [22]:

$$\kappa_1 = n, \qquad \kappa_2 = n^2 F_{n,N}, \tag{26}$$

where

$$F_{n,N} = \begin{cases} \frac{4n^2N - 5n^3 + 2N - n}{6n(N-n)^2} & (n \leqslant N/2), \\ \frac{(N-n)^3 - 4n(N-n)^2 + 6n^2(N-n) + 5n - N}{6n^2(N-n)} & (n \geqslant N/2). \end{cases}$$

The prefactor $F_{n,N}$ is an increasing function of n, which ranges from $1/(N-1)$ at $n = 1$ to 1 at $n = N-1$.

For the special case $n = N-1$, the only nonzero eigenvalue is $\lambda_1 = 2(N-1)$, and the probability density $p(z)$ is given by Eq. (13). For $n = N-2$, one computes $u_1 = u_2 = (N-2)$ and $u_{12} = (N-3)$ in Eqs. (23) so that $\lambda_1 = (2N-5)$, $\lambda_2 = 1$, and $p(z)$ is given by Eq. (24).

The inverse of the matrix $\mathbf{C}$ is proportional to a three-diagonal matrix, which resembles the Laplacian matrix:

$$\mathbf{C}^{-1} = \begin{pmatrix} 2 & -1 & 0 & \ldots & 0 & 0 \\ -1 & 2 & -1 & \ldots & 0 & 0 \\ 0 & -1 & 2 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 2 & -1 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}$$

(the only difference appears in the last diagonal element, which is equal to 1, not 2). The optimal quadratic form is then expressed through $\mathbf{C}^{-1}$ according to Eq. (11). This is equivalent to the choice of the following quadratic form:

$$\chi_{\mathrm{opt}} = \frac{1}{N} \sum_{n=1}^{N} (x_n - x_{n-1})^2$$

(with $x_0 = 0$). One sees that the "best" quadratic form is (almost) the time-averaged MSD with the lag time $n = 1$ [in comparison to Eq. (22), there is an additional term x_1^2 and the prefactor $1/N$ instead of $1/(N-1)$]. This result agrees with the general Cramér-Rao lower bound, which for Brownian motion is achieved by the time-averaged MSD estimator [33,34].

1. *The eigenvalues of the matrix* MC

Since the probability distribution of a quadratic form is completely defined through the eigenvalues of the matrix **MC**, we start the analysis by considering this spectrum. Figure 1(a) shows the eigenvalues λ_q for the time-averaged MSD of Brownian motion, with $N = 200$ and six lag times $n = 1, 10, 50, 100, 150, 190$. Since the mean value κ_1 increases with the lag time, it is convenient to renormalize the eigenvalues by $2\kappa_1 = \lambda_1 + \cdots + \lambda_N = 2n$. The ratio $\lambda_q/(2\kappa_1)$ can thus be seen as a relative weight of the eigenvalue λ_q.

In agreement with the theoretical result, one clearly sees that there are $\tilde{N} = N - n$ nonzero eigenvalues. For $n = 1$, the time-averaged MSD was shown to minimize the variance of the probability density $p(z)$ (with a fixed mean). In this optimal case, all the eigenvalues λ_q provide identical contributions (dotted horizontal line). The larger the lag time n is, the higher the contribution of the first eigenvalue λ_1 is, while the contributions of other eigenvalues progressively vanish. This behavior is better illustrated on Fig. 1(b), which shows the three largest eigenvalues as functions of the lag time n. In the ultimate case $n = 199$, there is only one nonzero eigenvalue λ_1 that results in the highest variance of $p(z)$.

Since discrete Gaussian processes are often considered as approximations for the continuous ones, one can study the behavior of the spectrum in the limit N going to infinity. We observed numerically that doubling both N and

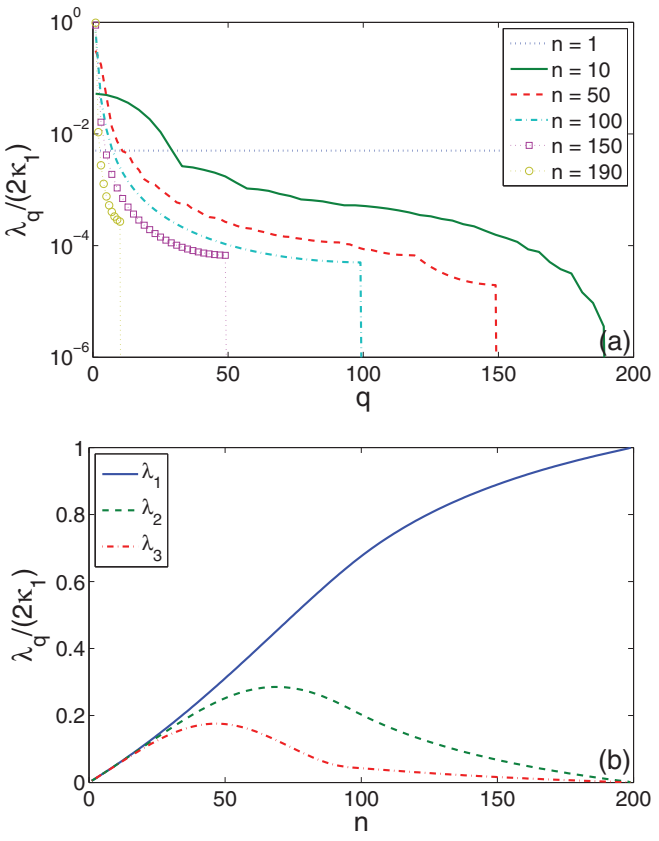

FIG. 1. (Color online) The eigenvalues λ_q of the matrix **MC** for the time-averaged MSD of Brownian motion, with $N = 200$. The vertical axis is normalized by $2\kappa_1 = \lambda_1 + \cdots + \lambda_N = 2n$ so that $\lambda_q/(2\kappa_1)$ can be seen as a relative weight of each eigenvalue. (a) λ_q as a function of q for six lag times $n = 1, 10, 50, 100, 150, 190$. (b) Three largest eigenvalues $\lambda_1, \lambda_2, \lambda_3$ as functions of the lag time n. For large n, the first eigenvalue λ_1 provides the overwhelming contribution.

n results in approximately twofold larger eigenvalues. After renormalization, one may then expect to get a well-defined limit for $N \to \infty$. We also checked numerically that for large q, the eigenvalues λ_q decrease as $\propto 1/q^2$. Further analysis is beyond the scope of this paper.

2. *Probability density*

Figure 2 shows the probability density $p(z)$ (lines) computed numerically through the inverse Fourier transform (10). We consider the time-averaged MSD of Brownian motion, with $N = 200$ and three lag times $n = 50, 100, 150$. Since the mean value κ_1 increases with n, we renormalize the horizontal axis and put z/κ_1.

First of all, one can see that the most probable value [i.e., the position of the maximum of the probability density $p(z)$ shown by vertical dotted lines] is always smaller than the mean value (which is equal to 1 due to normalization by κ_1). This is an important observation for interpreting SPT data. For instance, the diffusion coefficient inferred from the time-averaged MSD of a single random trajectory would most probably underestimate the mean (true) value.

For larger n, the probability density $p(z)$ is broader, and the most probable value is still smaller. Table I presents the reduced variance κ_2/κ_1^2, as well as the skewness $\kappa_3/\kappa_2^{3/2}$ and the kurtosis κ_4/κ_2^2 of the distribution (see the data for $H = 0.5$). All these characteristics increase with the lag time n.

As expected, the use of larger lag times (i.e., a smaller averaging window in the time-averaged MSD) degrades the accuracy of estimations because of a larger variance. From this perspective, the smallest lag time $n = 1$ is evidently the best choice. However, other criteria for choosing n may be relevant in practice. For instance, the unavoidable presence of localization uncertainty in the position detection results in larger lag times n, which are needed for a more stable estimation of the diffusion coefficient [32].

Figure 2 also presents two approximations of the probability density $p(z)$ by the generalized Gamma distribution $\psi(z)$. The first approximation (shown by circles, squares, and triangles for $n = 50, 100, 150$, respectively) was obtained by fitting the numerically computed $p(z)$ to Eq. (17). Three parameters a, b, and ν, which correspond to the best fit, are given in Table I (note that the fit and its parameters depend on the chosen interval of z values). One can see that the function $\psi(z)$ is indeed an accurate approximation of $p(z)$ for the shown range of z, for which the probability density $p(z)$ is not too small. This result justifies our empirical choice for the generalized Gamma distribution.

The second approximation (shown by pluses, crosses, and asterisks for $n = 50, 100, 150$, respectively) was constructed from three moments of the probability density $p(z)$, as discussed in Sec. II E. In this case, we first compute the three cumulant moments by Eq. (6), retrieve from them the moments μ_1, μ_2, and μ_3 according to Eqs. (7), and solve numerically the system of two nonlinear Eqs. (21) to get the parameters a, b, and ν. In contrast to the first approximation, which uses the whole probability density $p(z)$, the second approximation is obtained uniquely from the first three moments, without fitting. As a consequence, one expects lower accuracy for

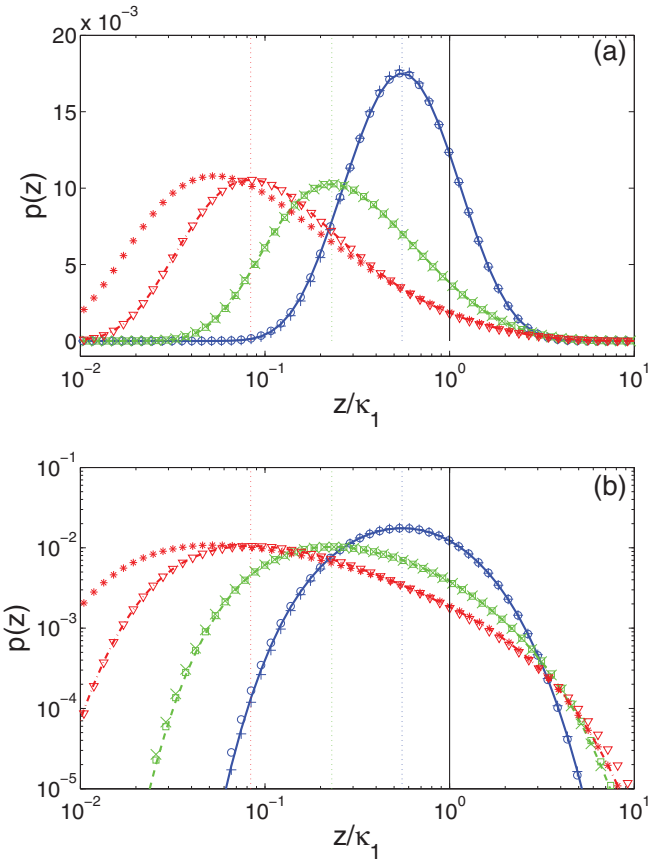

FIG. 2. (Color online) The probability density $p(z)$ (lines) and its approximations (symbols) by the generalized Gamma distribution, for the time-averaged MSD of Brownian motion ($H = 0.5$), with $N = 200$ and three lag times: $n = 50$ (blue solid curve, circles and pluses), $n = 100$ (green dashes line, squares and crosses), and $n = 150$ (red dash-dotted line, triangles and asterisks), shown at semilogarithmic (a) and logarithmic (b) scales. The horizontal axis is normalized by the mean value $\kappa_1 = n$ (the solid vertical line points on the normalized mean value at 1). Circles, squares, and triangles show the first approximation by fitting $p(z)$ to Eq. (17). Three dotted vertical lines indicate the most probable value of this approximation according to Eq. (20). Pluses, crosses, and asterisks indicate the second approximation obtained by computing three moments, as discussed in Sec. II E.

the second approximation. As clearly seen on Fig. 2, the second approximation remains accurate for $n = 50$ and $n = 100$, while for $n = 150$, there is a significant deviation at small z.

The reason for this deviation is clear. For large n, the contribution of the first eigenvalue λ_1 dominates the others and essentially determines the positive-order moments in Eq. (6). The parameters a, b, and ν which are computed from the moments μ_1, μ_2, and μ_3, are thus mainly determined by λ_1, while the information about the other eigenvalues is partially discarded. As a consequence, the second approximation works well for large z but may fail at small z. The quality of the second approximation at small z could potentially be improved by using the negative-order moments $\mu_{-\alpha}$ to retrieve the parameters a, b, ν. This option can be explored in the future.

TABLE I. Characteristics of the probability densities $p(z)$ shown in Figs. 2, 4, and 5: the mean value κ_1, the reduced variance κ_2/κ_1^2, the skewness $\kappa_3/\kappa_2^{3/2}$, the kurtosis κ_4/κ_2^2, and three parameters a, b, and ν of the best fit to the generalized Gamma distribution (17).

H	n	κ_1	κ_2/κ_1^2	$\kappa_3/\kappa_2^{3/2}$	κ_4/κ_2^2	a	b	ν
0.25	50	7.071	0.210	1.398	3.141	22.21	5.004	−2.359
	100	10.00	0.516	2.233	8.050	17.08	237.2	−2.462
	150	12.25	0.988	2.713	11.29	9.563	10 000	−1.533
0.50	50	50.00	0.408	1.614	4.164	26.98	27.14	1.042
	100	100.0	1.000	2.533	10.08	23.61	150.8	0.125
	150	150.0	1.593	2.804	11.85	12.14	408.7	0.063
0.75	50	353.6	0.752	2.305	8.620	63.11	242.6	1.027
	100	1000	1.438	2.746	11.49	45.09	1214	0.563
	150	1837	1.857	2.825	11.98	18.03	2748	0.471

B. Fractional Brownian motion

The above analysis of the time-averaged MSD for Brownian motion can now be extended to more sophisticated processes. Various types of motion in living cells, biological tissues, and mineral samples were observed and analyzed: restricted, obstructed, and hindered diffusion, directed motion, anomalous diffusion, diffusion through traps, etc. [1,7,23,35–37]. In this paper, we consider a (discrete) fractional Brownian motion (fBm), for which the role of strong correlations between steps can be investigated. A fBm with the Hurst exponent $0 < H < 1$ was introduced by Kolmogorov and later by Mandelbrot and van Ness [38,39]. This is a Gaussian process that starts from 0 and has mean zero and the covariance matrix, which in the discrete case can be written as

$$\mathbf{C}_{n_1 n_2} = \tfrac{1}{2}\left(n_1^{2H} + n_2^{2H} - |n_1 - n_2|^{2H}\right). \qquad (27)$$

When $H = 1/2$, Eq. (27) is reduced to Eq. (25) for Brownian motion. The fBm is persistent (has positive correlations between steps) or antipersistent (has negative correlations between steps) for $H > 1/2$ and $H < 1/2$, respectively. The mean value of the time-averaged MSD of fBm, $\kappa_1 = n^{2H}$, may exhibit superdiffusive ($2H > 1$) and subdiffusive ($2H < 1$) character.

In what follows, we present the eigenvalues and the probability density for the time-averaged MSD of fBm with $H = 0.25$ and 0.75 and compare them with the previous results for Brownian motion ($H = 0.5$).

1. *The eigenvalues of the matrix* **MC**

Figure 3 shows the eigenvalues λ_q of the matrix **MC** for the time-averaged MSD of persistent ($H = 0.75$) and antipersistent ($H = 0.25$) fractional Brownian motion. As previously, we take $N = 200$ and six lag times, $n = 1, 10, 50, 100, 150, 190$. Although the curves look similar to those of Fig. 1 for Brownian motion ($H = 0.5$), several differences can be noticed.

(1) The ratio between the largest and the smallest (nonzero) eigenvalues, $\lambda_1/\lambda_{\tilde{N}}$, is much bigger for $H = 0.75$, yielding a broader intermediate region $\lambda_{\tilde{N}} \leqslant z \leqslant \lambda_1$ and a broader probability density $p(z)$. One may therefore expect a higher variance κ_2 of the time-averaged MSD for persistent fBm, as confirmed by the data from Table I. In fact, positive correlations between steps can be qualitatively interpreted as a smaller number of effectively "independent measurements" in the time-averaged MSD that results in a higher variance.

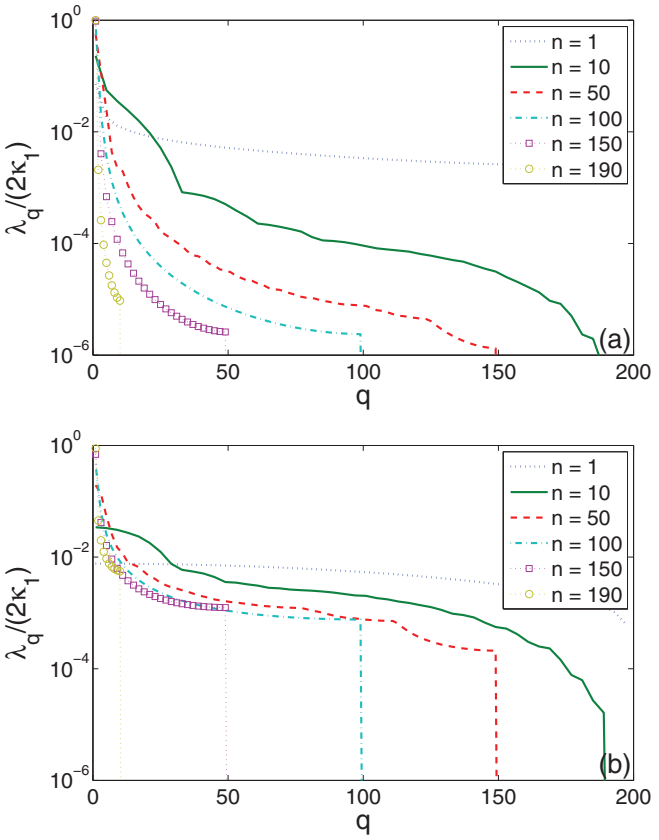

FIG. 3. (Color online) The eigenvalues λ_q of the matrix **MC** for the time-averaged MSD of fBm, with $N = 200$, six lag times $n = 1, 10, 50, 100, 150, 190$, and $H = 0.75$ (a) and $H = 0.25$ (b). The eigenvalues λ_q are normalized by $2\kappa_1 = 2n^{2H}$.

(2) On the opposite, the antipersistence of fBm with $H = 0.25$ leads to a smaller variance as compared to Brownian motion (see Table I). The intermediate region $\lambda_{\tilde{N}} \leqslant z \leqslant \lambda_1$ and the probability density $p(z)$ are narrower.

(3) In contrast to Fig. 1, the eigenvalues λ_q for $n = 1$ are not identical. This means that the time-averaged MSD with $n = 1$ is not optimal, while the probability distribution is not Gamma distribution, as it was for Brownian motion ($H = 0.5$). Interestingly, the curves λ_q for $n = 1$ [dotted lines in Figs. 3(a) and 3(b)] for $H = 0.25$ and 0.75 exhibit different features. For antipersistent fBm ($H = 0.25$), the eigenvalues λ_q with $1 \leqslant q \leqslant 100$ remain almost equal to each other and slightly decrease for $q > 100$. In turn, when $H = 0.75$, there is a relatively fast decrease for small q and slow decrease for larger q. But the behavior at small q (in particular, λ_1) essentially determines the properties of the probability density $p(z)$. From this point of view, the properties of the time-averaged MSD for the antipersistent motion are somewhat closer to those for Brownian motion.

2. *Probability density*

Figures 4 and 5 present the probability density $p(z)$ (lines) obtained numerically through the inverse Fourier transform in

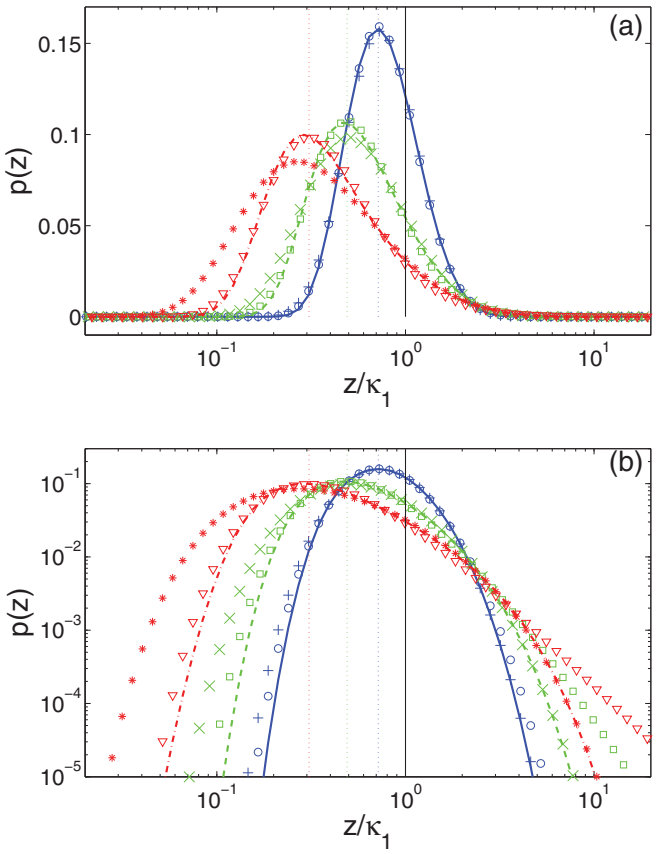

FIG. 4. (Color online) The probability density $p(z)$ (lines) and its approximations (symbols) by a generalized Gamma distribution, for the time-averaged MSD of fBm, with $H = 0.25$, $N = 200$ and three lag times: $n = 50$ (blue solid curve, circles and pluses), $n = 100$ (green dashes line, squares and crosses), and $n = 150$ (red dash-dotted line, triangles and asterisks), shown at semilogarithmic (a) and logarithmic (b) scales. The horizontal axis is normalized by the mean value $\kappa_1 = n^{2H}$ (the solid vertical line points on the normalized mean value at 1). Circles, squares, and triangles show the first approximation by fitting $p(z)$ to Eq. (17). Three dotted vertical lines indicate the most probable value of this approximation according to Eq. (20). Pluses, crosses, and asterisks indicate the second approximation obtained by computing three moments, as discussed in Sec. II E.

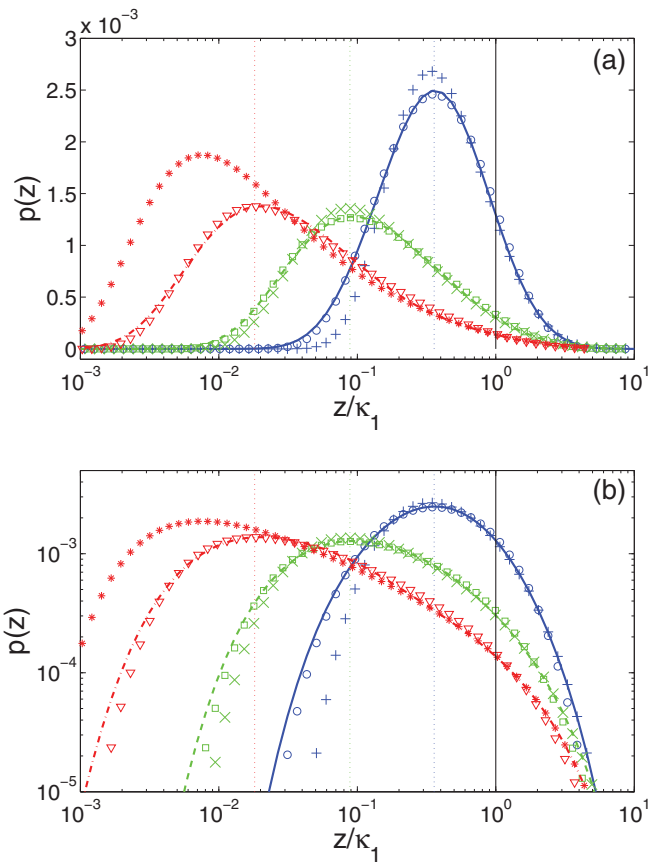

FIG. 5. (Color online) The probability density $p(z)$ (lines) and its approximations (symbols) by a generalized Gamma distribution, for the time-averaged MSD of fBm, with $H = 0.75$, $N = 200$ and three lag times: $n = 50$ (blue solid curve, circles and pluses), $n = 100$ (green dashes line, squares and crosses), and $n = 150$ (red dash-dotted line, triangles and asterisks), shown at semilogarithmic (a) and logarithmic (b) scales. The horizontal axis is normalized by the mean value $\kappa_1 = n^{2H}$ (the solid vertical line points on the normalized mean value at 1). Circles, squares, and triangles show the first approximation by fitting $p(z)$ to Eq. (17). Three dotted vertical lines indicate the most probable value of this approximation according to Eq. (20). Pluses, crosses, and asterisks indicate the second approximation obtained by computing three moments, as discussed in Sec. II E.

Eq. (10), for $H = 0.25$ and 0.75, respectively. As previously, we take $N = 200$ and three lag times $n = 50,100,150$. As discussed earlier, the probability density is narrower for $H = 0.25$ and broader for $H = 0.75$. The most probable value is closer to the mean value (i.e., 1) for $H = 0.25$ and farther from 1 for $H = 0.75$ (e.g., the most probable value in the case $n = 150$ is approximately 50 times smaller than the mean value).

Two approximations by the generalized Gamma distribution are shown by symbols. The first approximation obtained by fitting the probability density $p(z)$ to Eq. (17) remains accurate, although small deviations can be distinguished at small z and larger deviations at large z. Its parameters a, b, and ν are given in Table I. In turn, the second approximation, which is based on the three moments, is more accurate at large z and less accurate at small z, as discussed in Sec. III A.

3. *Variance, skewness, and kurtosis*

Finally, the reduced variance κ_2/κ_1^2, the skewness $\kappa_3/\kappa_2^{3/2}$ and the kurtosis κ_4/κ_2^2 of the time-averaged MSD are plotted as functions of the lag time n in Fig. 6. All these characteristics increase with n as the number of data points is reduced. However, the increase is slower for small H because of the antipersistent character of displacements. In turn, for $H = 0.9$, the skewness and kurtosis get the largest values and slowly vary with n.

IV. CONCLUSION

The knowledge of the probability distribution of time-averaged quadratic forms is of primary importance for a reliable inference of the dynamical and microrheological quantities from individual random trajectories of tracers. In

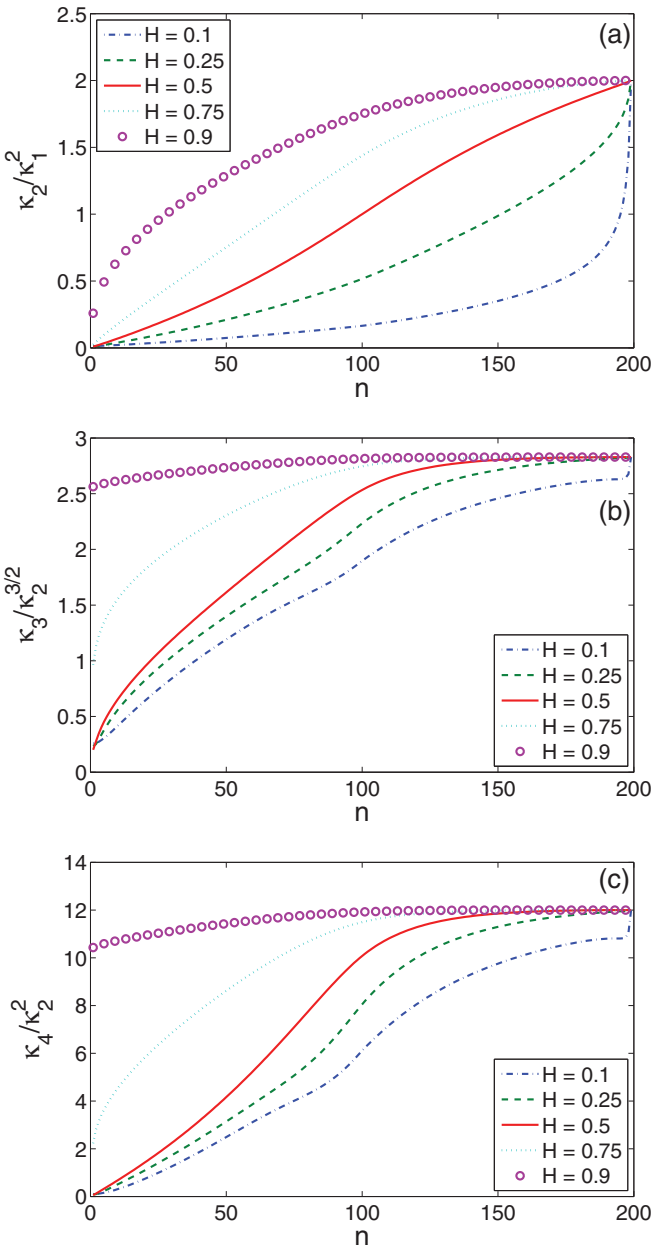

FIG. 6. (Color online) The reduced variance κ_2/κ_1^2 (a), the skewness $\kappa_3/\kappa_2^{3/2}$ (b), and the kurtosis κ_4/κ_2^2 (c) of the time-averaged MSD as functions of the lag time n, with different H.

this paper we employed the classical mathematical theory in order to study the quadratic form associated to the time-averaged mean-square displacement which is often used in practice. Although rigorous infinite series representations of the probability density $p(z)$ of quadratic forms are available, their practical implementation is problematic for relatively long data samples (Appendix A). Our main theoretical result is an empirical approximation of $p(z)$ by a generalized Gamma distribution (17). The accuracy of this approximation was checked numerically for normal, persistent and antipersistent fractional Brownian motion. We showed that for the whole range of relevant values z [for which $p(z)$ is not too small], Eq. (17) provides an accurate fit for the probability density $p(z)$ for moderate lag times. Moreover, three moments of the quadratic form are enough for constructing this approximation.

When an experimental histogram of time-averaged MSD is available from the analysis of SPT data (e.g., see Ref. [40]), its fitting to the generalized Gamma distribution (17) may allow one to characterize the dynamics of tracers through three parameters, a, b, and ν. Further work is needed for providing a clearer biophysical interpretation of these parameters. Note that Eq. (17) significantly extends the Gamma distribution by Qian *et al.*, which was applicable only for normal Brownian motion and unit lag time [22]. In particular, we have shown that even for the lag time 1, the Gamma distribution is not valid for fractional Brownian motion with the Hurst exponent $H \neq 1/2$ (Fig. 3).

Rigorous asymptotic relations for the probability density $p(z)$ and its explicit approximation by (17) clarify the properties of the time-averaged MSD. For instance, a seemingly parabolic shape of the probability density $p(z)$ in Fig. 2(b) could suggest approximating $p(z)$ by a log-normal distribution:

$$p_{\rm LN}(z) = \frac{1}{z\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln z - \mu)^2}{2\sigma^2}\right],$$

where $\mu = \ln(\kappa_1/\sqrt{1+\kappa_2/\kappa_1^2})$ and $\sigma^2 = \ln(1+\kappa_2/\kappa_1^2)$. However, a more careful analysis reveals limitations of a parabolic fit to the curves in Figs. 2(b), 4(b), and 5(b), especially for large lag times. In turn, the generalized Gamma distribution (17), which is functionally as simple as the log-normal one, is a much more accurate approximation.

It is worth stressing that Hoffman *et al.* reported the empirical distribution of the time-averaged MSD, which was close to log-normal [41]. This finding was attributed to tracer-to-tracer and cell-to-cell variability. At the same time, the probabilistic nature of the time-averaged MSD would also result in nearly log-normal variability of empirical values. It is therefore important to quantify the stochastic uncertainty of the time-averaged MSD in order to distinguish these two sources of variability.

The spatial heterogeneity of living cells (e.g., spatially varying viscosity and shear moduli) was reported in the literature (see Refs. [40,41] and references therein). If the observation time is long enough, a tracer may explore various regions of the cell and thus experience different impacts from the surrounding medium. For instance, the tracer displacements may be faster or slower depending on the local viscosity, while the presence of obstacles (such as plasma membrane or large organelles) would hinder the motion. A theoretical analysis of single-particle trajectories in such a heterogeneous medium is a very difficult problem. Moreover, even if the motion of the tracer is local (e.g., a tracer trapped by an optical tweezer), the probability distribution of the time-averaged MSD may depend on the position of the tracer inside the cell. Although these issues are important for an accurate analysis of the experimental data, they go beyond the scope of the paper, in which the medium was considered as homogeneous.

Since the spatial distribution of heterogeneities and obstacles is often not available, one resorts to the paradigm of "effective diffusion models," in which the medium is homogenized, while the heterogeneities are effectively taken

into account in the diffusive model. In the simplest case, the presence of obstacles or spatially varying viscosity heterogeneities can be expressed through an average (or apparent) diffusion coefficient. In more sophisticated cases, the influence of heterogeneities may be modeled either through long-time trapping (continuous-time random walks) or as memory correlations between displacements (fractional Brownian motion or generalized Langevin equation). Although the paradigm of homogenized medium is simplified, it is very broadly used in the biophysical community.

We also addressed the problem of choosing the optimal quadratic form. Under the constraint of a fixed mean value κ_1, the smallest variance $\kappa_2 = 2\kappa_1^2/N$ is achieved by using $\mathbf{M} \propto \mathbf{C}^{-1}$, for which all the eigenvalues λ_q are identical. In this case the probability density $p(z)$ is the Gamma distribution (12). In turn, the largest variance $\kappa_2 = 2\kappa_1^2$ is provided when all the eigenvalues are zero except one. This probability density is again given by the Gamma distribution.

In the future, it would be interesting to improve the accuracy of the empirical approximation at small z by using other moments of the quadratic form for retrieving the parameters a, b, and ν (Sec. II E). The study of other quadratic forms (e.g., squared root-mean square of the position, velocity autocorrelation function, power density spectrum) may also be helpful. Finally, a more elaborate statistical analysis of the acquired data has to be developed for a better estimation of the mean quantities from several empirical observations. In particular, the difference between the mean value and the most probable value has to be taken into account for a reliable interpretation of single-particle tracking experiments.

ACKNOWLEDGMENTS

The author acknowledges B.-T. Nguyen for providing an elegant proof in Appendix C.

APPENDIX A: SERIES REPRESENTATIONS OF THE PROBABILITY DENSITY

We sketch three series representations of the probability density $p(z)$ following a general approach by Kotz and coworkers [42,43].

1. Power series expansion

Pachares and later Shah and Khatri suggested the power series expansion, which can be written as [17,18]

$$p(z) = \sum_{k=0}^{\infty} c_k^P \frac{(-1)^k}{\Gamma(\tilde{N}/2+k)} z^{\tilde{N}/2+k-1}, \tag{A1}$$

where the coefficients c_k^P are given by recurrent relations:

$$c_0^P = \frac{1}{\sqrt{\lambda_1 \cdots \lambda_{\tilde{N}}}}, \quad c_k^P = \frac{1}{k}\sum_{j=0}^{k-1} d_{k-j}^P c_j^P \quad (k \geqslant 1),$$

and

$$d_k^P = \frac{1}{2}\sum_{q=1}^{\tilde{N}} \lambda_q^{-k}.$$

2. Expansion over Laguerre polynomials

In a series of papers, Gurland developed the expansion over Laguerre polynomials [19–21]:

$$p(z) = g_{\tilde{N}}(z/\beta) \sum_{k=0}^{\infty} c_k^L \frac{k!\Gamma(\tilde{N}/2)}{\beta\Gamma(\tilde{N}/2+k)} L_k^{(\tilde{N}/2-1)}[z/(2\beta)], \tag{A2}$$

where $g_n(z)$ is the central chi-squared density with n degrees of freedom,

$$g_n(z) = \frac{(z/2)^{n/2-1}e^{-z/2}}{2\Gamma(n/2)},$$

and $L_k^{(a)}(x)$ are the generalized Laguerre polynomials defined by Rodrigues's formula,

$$L_k^{(a)}(x) = \frac{1}{k!} e^x x^{-a} \frac{d^k}{dx^k}(e^{-x}x^{k+a}) \quad (a > 1).$$

The coefficients c_k^L are given by recurrent relations:

$$c_0^L = 1, \quad c_k^L = \frac{1}{k}\sum_{j=0}^{k-1} d_{k-j}^L c_j^L \quad (k \geqslant 1),$$

with

$$d_k^L = \frac{1}{2}\sum_{q=1}^{\tilde{N}} [1-\lambda_q/(2\beta)]^k = \frac{1}{2}\mathrm{tr}\{[\mathbf{I} - \mathbf{MC}/(2\beta)]^k\}.$$

Finally, $\beta > \lambda_1/4$ is a positive constant that ensures the uniform convergence of the series [e.g., $\beta = (\lambda_1 + \lambda_{\tilde{N}})/4$] [42]. The advantage of this expansion is that the coefficients c_k^L can be found through the traces, without computing the eigenvalues λ_q.

3. Chi-squared expansion

Robbins and later Ruben provided a representation in terms of chi-squared distributions [13–16]:

$$p(z) = \sum_{k=0}^{\infty} c_k^C \frac{1}{\beta} g_{\tilde{N}+2k}(z/\beta), \tag{A3}$$

where the coefficients c_k^C are given by recurrent relations:

$$c_0^C = \prod_{q=1}^{\tilde{N}} (2\beta/\lambda_q)^{1/2}, \quad c_k^C = \frac{1}{k}\sum_{j=0}^{k-1} d_{k-j}^C c_j^C \quad (k \geqslant 1),$$

with

$$d_k^C = \frac{1}{2}\sum_{q=1}^{\tilde{N}} (1-2\beta/\lambda_q)^k, \quad \beta = \left(\prod_{q=1}^{\tilde{N}} \lambda_q/2\right)^{1/\tilde{N}}.$$

4. Comparison

All three series expansions were shown to be uniformly convergent, and the accuracy of the truncated series approximation of the probability density $p(z)$ was investigated (see Ref. [42] and references therein). Since these approximations are well understood, it is tempting to employ them in the study of the time-averaged MSD. Figure 7 compares the probability

density $p(z)$ (found numerically through the inverse Fourier transform of $\phi(k)$ computed for $N = 10$ and $n = 3$), its best fit to Eq. (17), and the above approximations truncated at $K = 10$ (11 terms). Several comments are in order:

(1) The power-law expansion (A1) and the chi-squared expansion (A3) are very accurate for $z < 0.5$. In turn, these approximations drastically deviate from $p(z)$ for $z > 0.5$. Note that the truncation at $K = 100$ (not shown) shifts the "threshold" from 0.5 to 3, but deviations for $z > 3$ are still extreme. In practice, these two approximations are of limited interest (at least, for this example).

(2) The expansion (A2) over Laguerre polynomials is a better approximation. Although it slightly underestimates $p(z)$ for small z and exhibits oscillations at large z, it correctly reproduces the overall behavior. A larger truncation order ($K = 20$, not shown) further improves its quality.

(3) The generalized Gamma distribution $\psi(z)$ provides the best fit for the region of relevant values of z [see Fig. 7(a)]. However, deviations at small z are clearly seen at logarithmic scale [Fig. 7(b)]. We emphasize again that the accuracy of this empirical approximation cannot be improved, as the "best" set of parameters a, b, and ν has already been chosen.

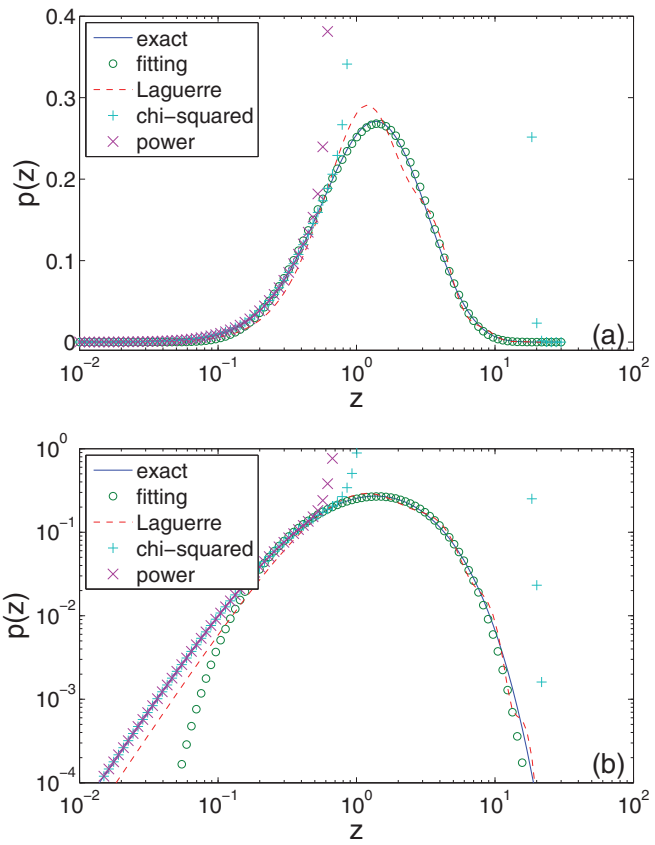

FIG. 7. (Color online) The probability density $p(z)$ (solid line) of the time-averaged MSD of Brownian motion, with $N = 10$ and $n = 3$, shown at semilogarithmic (a) and logarithmic (b) scales. For comparison, we present the best fit to Eq. (17) (circles), the expansion (A2) over Laguerre polynomials (dashed line), the expansion (A3) over chi-squared distributions (pluses), and the power series expansion (A1) (crosses). These three expansions are truncated to $K = 10$ (11 terms).

The above example with $N = 10$ and $n = 3$ is of limited interest for practical applications. We also tried to apply the above approximations for the case $N = 200$ and $n = 50$, which was studied in this paper, but the results were not satisfactory. Although further improvements may be interesting, they are beyond the scope of the paper.

APPENDIX B: JOINT DISTRIBUTIONS

Extending the analysis of Sec. II A, one can study joint distributions of several quadratic forms $\chi_1,\ldots,\chi_\ell$: $\chi_j = \frac{1}{2}(\mathbf{x}^T\mathbf{M}_j\mathbf{x})$, $j = 1,\ldots,\ell$, with given matrices $\mathbf{M}_j$. The multiple characteristic function is

$$\phi(k_1,\ldots,k_\ell) = \int_{\mathbb{R}^N} d\mathbf{x} P_N(\mathbf{x}) \exp\left[\frac{i}{2}\sum_{j=1}^{\ell} k_j(\mathbf{x}^T\mathbf{M}_j\mathbf{x})\right].$$

Replacing $k\mathbf{M}$ in Eq. (2) by $\sum_j k_j\mathbf{M}_j$, one gets

$$\phi(k_1,\ldots,k_\ell) = \frac{1}{\sqrt{\det\left[\mathbf{I} - i\left(\sum_{j=1}^{\ell} k_j\mathbf{M}_j\right)\mathbf{C}\right]}} \times \exp\left\{-\frac{1}{2}\mathbf{x}^{0T}\left[\mathbf{C}^{-1} - \mathbf{C}^{-1}\left(\mathbf{C}^{-1} - i\sum_{j=1}^{\ell} k_j\mathbf{M}_j\right)^{-1}\mathbf{C}^{-1}\right]\mathbf{x}^0\right\}.$$

For centered Gaussian processes ($\mathbf{x}^0 = 0$), one gets

$$\phi(k_1,\ldots,k_\ell) = \frac{1}{\sqrt{\det\left[\mathbf{I} - i\left(\sum_{j=1}^{\ell} k_j\mathbf{M}_j\right)\mathbf{C}\right]}}.$$

As in Sec. II A, one can rewrite this function in term of traces:

$$\phi(k_1,\ldots,k_\ell) = \exp\left\{\frac{1}{2}\sum_{m=1}^{\infty}\frac{i^m}{m}\mathrm{tr}\left[\left(\sum_{j=1}^{\ell} k_j\mathbf{M}_j\mathbf{C}\right)^m\right]\right\}.$$

The joint cumulant moments are obtained by taking the respective derivatives:

$$\begin{aligned}\left\langle \chi_1^{m_1}\cdots\chi_\ell^{m_\ell}\right\rangle_c &= \frac{1}{i^m}\left[\frac{\partial^m \ln\phi(k_1,\ldots,k_\ell)}{\partial k_1^{m_1}\ldots\partial k_\ell^{m_\ell}}\right]_{k_1=\ldots=k_\ell=0} \\ &= \frac{1}{2m}\left\{\frac{\partial^m}{\partial k_1^{m_1}\ldots\partial k_\ell^{m_\ell}}\mathrm{tr}\left[\left(\sum_{j=1}^{\ell} k_j\mathbf{M}_j\mathbf{C}\right)^m\right]\right\}_{k_1=\ldots=k_\ell=0},\end{aligned}$$

where $m = m_1 + \cdots + m_\ell$. For instance, the covariance between two quadratic forms χ_1 and χ_2 is

$$\langle\chi_1\chi_2\rangle_c = \tfrac{1}{2}\mathrm{tr}(\mathbf{M}_1\mathbf{C}\mathbf{M}_2\mathbf{C}).$$

This simple result can be used to analyze the correlation ρ between the values of the time-averaged MSD with different lag times n_1 and n_2:

$$\rho = \frac{\langle\chi_1\chi_2\rangle_c}{\sqrt{\langle\chi_1^2\rangle_c\langle\chi_2^2\rangle_c}}.$$

As an example, we take $n_1 = 1$ and $n_2 = n$, for which the correlation ρ as a function of n is shown on Fig. 8 for fBm with

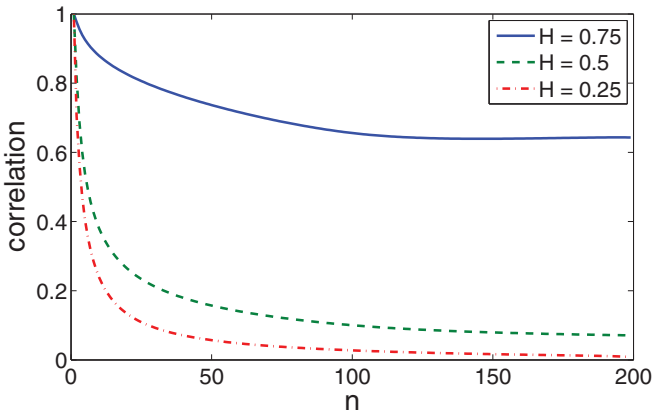

FIG. 8. (Color online) Correlation ρ between the time-averaged MSD with the lag time 1 (χ_1) and the time-averaged MSD with the lag time n (χ_2), for persistent ($H = 0.75$), normal ($H = 0.5$), and antipersistent ($H = 0.25$) fBm.

$H = 0.25, 0.5, 0.75$. As expected, the negative correlations between steps for antipersistent fBm ($H < 0.5$) result in smaller correlations between two time-averaged MSDs.

APPENDIX C: THE NUMBER OF NONZERO EIGENVALUES

We sketch the proof by B.-T. Nguyen for the fact that the matrix **MC** for the time-averaged MSD with the lag time n has exactly n zero eigenvalues. The proof is based on two lemmas.

Lemma 1. Let **M** be a symmetric matrix and **C** is an invertible matrix. Then the number n of zero eigenvalues of **MC** is equal to the number m of zero eigenvalues of **M**.

Proof. Let $\mathbf{u}_1,...,\mathbf{u}_n$ be all the eigenvectors of the matrix **MC** with zero eigenvalue, i.e., $\mathbf{MCu}_k = 0$ for $k = 1, \ldots, n$. Then $\mathbf{Cu}_1,...,\mathbf{Cu}_n$ are also the eigenvectors of the matrix **M** with zero eigenvalue. Moreover, these vectors are linearly independent. In fact, if these vectors were linearly dependent, then there would exist real coefficients a_1, ..., a_n such that

$$a_1\mathbf{Cu}_1 + \cdots + a_n\mathbf{Cu}_n = \mathbf{C}(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n) = 0;$$

i.e., $a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n$ would be an eigenvector of the matrix **C** with zero eigenvalue. This is not possible because the matrix **C** is invertible. One concludes that the matrix **M** has at least n zero eigenvalues, i.e., $m \geqslant n$.

Inversely, let $\mathbf{v}_1,...,\mathbf{v}_m$ be all the eigenvectors of the matrix **M** with zero eigenvalue, i.e., $\mathbf{Mv}_k = 0$ for $k = 1, \ldots, m$. Since **C** is invertible, $\mathbf{C}^{-1}\mathbf{v}_1,...,\mathbf{C}^{-1}\mathbf{v}_m$ are also the eigenvectors of the matrix **MC** with zero eigenvalue. Since these vectors are linearly independent, the matrix **MC** has at least m zero eigenvalues, i.e., $n \geqslant m$. One concludes that $m = n$.

Lemma 2. The matrix **M** for the time-averaged MSD with the lag time n has exactly n zero eigenvalues.

Proof. Let $\mathbf{u} = (u_1, \ldots, u_N)^T$ be an eigenvector of the matrix **M** with zero eigenvalue, i.e., $\mathbf{Mu} = 0$. Multiplying it by **u**, one gets

$$\frac{1}{2}(\mathbf{u}^T\mathbf{Mu}) = \frac{1}{N-n}\sum_{k=1}^{N-n}(u_{k+n} - u_k)^2 = 0,$$

from which $u_{k+n} = u_k$ for all $k = 1, \ldots, N-n$. This means that any eigenvector of **M** with zero eigenvalue can be represented as

$$\mathbf{u} = (u_1, u_2, \ldots, u_n, u_1, u_2, \ldots, u_n, \ldots, u_1, \ldots, u_r)^T,$$

where r is the remainder of the division of N by n. The linear space of such vectors has n dimensions so that the matrix **M** has n zero eigenvalues.

Corollary. Since **M** is symmetric and any covariance matrix **C** is invertible, one can apply the above lemmas for proving that the number of zero eigenvalues of **MC** is equal to n or, equivalently, $\tilde{N} = N - n$.

[1] M. J. Saxton and K. Jacobson, Annu. Rev. Biophys. Biomol. Struct. **26**, 373 (1997).
[2] I. M. Tolić-Norrelykke, E.-L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sorensen, Phys. Rev. Lett. **93**, 078102 (2004).
[3] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).
[4] D. Arcizet, B. Meier, E. Sackmann, J. O. Rädler, and D. Heinrich, Phys. Rev. Lett. **101**, 248103 (2008).
[5] C. Wilhelm, Phys. Rev. Lett. **101**, 028101 (2008).
[6] D. Wirtz, Annu. Rev. Biophys. **38**, 301 (2009).
[7] R. Metzler, V. Tejedor, J.-H. Jeon, Y. He, W. H. Deng, S. Burov, and E. Barkai, Acta Phys. Pol. B **40**, 1315 (2009).
[8] R. M. L. Evans, M. Tassieri, D. Auhl, and T. A. Waigh, Phys. Rev. E **80**, 012501 (2009).
[9] H. Lee, J. M. Ferrer, F. Nakamura, M. J. Lang, and R. D. Kamm, Acta Biomaterialia **6**, 1207 (2010).
[10] C. Guzman, Appl. Phys. Lett. **93**, 184102 (2008).
[11] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, Vols. 1 and 2 (John Wiley & Sons, New York, 1970).
[12] S. O. Rice, Bell Syst. Tech. J. **23**, 282 (1944); **24**, 46 (1945).
[13] H. Ruben, Ann. Math. Stat. **33**, 542 (1962).
[14] H. Ruben, Ann. Math. Stat. **34**, 1582 (1963).
[15] H. Robbins, Ann. Math. Stat. **19**, 266 (1948).
[16] H. Robbins and E. J. G. Pitman, Ann. Math. Stat. **20**, 552 (1949).
[17] J. Pachares, Ann. Math. Stat. **26**, 128 (1955).
[18] B. K. Shah and C. G. Khatri, Ann. Math. Stat. **32**, 883 (1961); *ibid*. **34**, 673 (1963).
[19] J. Gurland, Ann. Math. Stat. **24**, 416 (1953).
[20] J. Gurland, Ann. Math. Stat. **26**, 122 (1955); *ibid*. **33**, 813 (1962).
[21] J. Gurland, Ind. J. Stat. **17**, 37 (1956).
[22] H. Qian, M. P. Sheetz, and E. L. Elson, Biophys. J. **60**, 910 (1991).
[23] M. J. Saxton, Biophys. J. **64**, 1766 (1993).
[24] M. J. Saxton, Biophys. J. **72**, 1744 (1997).
[25] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
[26] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed. (SIAM, Philadelphia, 1997).

[27] S. O. Rice, SIAM J. Sci. Stat. Comput. **1**, 438 (1980).
[28] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York, 1980).
[29] D. S. Grebenkov, Phys. Rev. E **83**, 061117 (2011).
[30] A. P. Prudnikov, Yu. A. Brychkov and O. I. Marichev, *Integrals and Series, Vol. 5: Inverse Laplace Transforms* (Gordon and Breach, New York, 1992).
[31] A. Grad and H. Solomon, Ann. Math. Stat. **26**, 464 (1955).
[32] X. Michalet, Phys. Rev. E **82**, 041914 (2010).
[33] G. Voisinne, A. Alexandrou, and J.-B. Masson Biophys. J. **98**, 596 (2010).
[34] A. J. Berglund, Phys. Rev. E **82**, 011917 (2010).
[35] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).
[36] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[37] D. S. Grebenkov, Rev. Mod. Phys. **79**, 1077 (2007).
[38] A. N. Kolmogorov, C. R. Doklady Acad. Sci. URSS (N.S.) **26**, 115 (1940).
[39] B. Mandelbrot and J. W. van Ness, SIAM Rev. **10**, 422 (1968).
[40] M. H. G. Duits, Y. Li, S. A. Vanapalli, and F. Mugele, Phys. Rev. E **79**, 051910 (2009).
[41] B. D. Hoffman, G. Massiera, K. M. Van Citters, and J. C. Crocker, Proc. Natl. Acad. Sci. USA **103**, 10259 (2006).
[42] S. Kotz, N. L. Johnson, and D. W. Boyd, Ann. Math. Stat. **38**, 823 (1967).
[43] S. Kotz, N. L. Johnson, and D. W. Boyd, Ann. Math. Stat. **38**, 838 (1967).