Published in final edited form as:

Phys Rev E Stat Nonlin Soft Matter Phys. 2012 June ; 85(6 Pt 1): 061916. doi:10.1103/PhysRevE.85.061916.

Optimal diffusion coefficient estimation in single-particle tracking

Xavier Michalet[1,*] and **Andrew J. Berglund**[2,†]

[1]Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095, USA

[2]Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

Abstract

Single-particle tracking is increasingly used to extract quantitative parameters on single molecules and their environment, while advances in spatial and temporal resolution of tracking techniques inspire new questions and avenues of investigation. Correspondingly, sophisticated analytical methods are constantly developed to obtain more refined information from measured trajectories. Here we point out some fundamental limitations of these approaches due to the finite length of trajectories, the presence of localization error, and motion blur, focusing on the simplest motion regime of free diffusion in an isotropic medium (Brownian motion). We show that two recently proposed algorithms approach the theoretical limit of diffusion coefficient uncertainty. We discuss the practical performance of the algorithms as well as some important implications of these results for single-particle tracking.

I. INTRODUCTION

Single-particle tracking (SPT) methods are well established and widely used for studying the microscopic behavior and interactions of individual molecules or microscopic objects in soft matter and biological environments [1,2] and are part of the growing toolbox available to researchers studying nanoparticle and single-molecule dynamics. A single particle can be localized with nanoscale accuracy through optical microscopy—a fact that has been extensively analyzed and applied in recent years in general nanoscale measurement methods and specifically for super-resolution microscopy [3–5], high-resolution colocalization studies [6,7], or single-particle and single-molecule tracking [8–10]. However, the consequences of finite localization accuracy on measurements of properties of moving particles have not yet received much attention. As it turns out, this can lead to very large errors on dynamic parameter estimation, preventing in some cases any quantitative conclusion on the particle motion. The purpose of this article is to compare and summarize new theoretical and numerical developments, allowing a precise assessment of these effects, as well as to provide simple guidelines to adjust experimental control parameters

*michalet@chem.ucla.edu
†ajberglund@gmail.com

(localization accuracy, exposure time, frame separation, and trajectory length) in order to ensure a desired precision on the measured parameters.

A basic dynamic characteristic of microscopic behavior is the diffusion coefficient D, which may be constant in an isotropic medium or vary spatially or temporally in complex environments [11,12]. For example, changes from slow to fast diffusion are sometimes observed in membrane proteins, and can in principle be characterized as changes in D [13–17]. Although not all motion regimes can be characterized by a single D value (a regime we will refer to as "free diffusion") or even a few different D values—for example, diffusion may be superimposed to drift (directed motion), confined within hard boundaries (confined motion), or taking place within a fractal domain (anomalous diffusion) [2]—free diffusion is one of the most common models used to interpret experimental data, in particular when only short trajectories are available. In these complex situations, it is of fundamental importance to understand how accurately a free diffusion coefficient can be estimated from a finite-length trajectory. In particular, this knowledge is necessary to assess the statistical significance of apparent differences in diffusion coefficients between segments of a single trajectory, between multiple trajectories observed under identical conditions or between experiments obtained under different conditions.

The previous discussion raises the question of how well D can be obtained from the analysis of a finite length single-particle trajectory, but a no less important aspect of this question is the influence of localization error on the measured D. Surprisingly, although the precision of noise-limited single-particle localization has been the focus of significant attention due to its importance in super-resolution imaging [18,19], the consequence of a finite localization precision on D estimation (and its worsening due to diffusion) has until recently been scarcely addressed theoretically [20–22].

In this paper we answer two fundamental theoretical questions regarding the optimal estimation of free diffusion coefficients from SPT data.

1. What is the best theoretically achievable precision in estimating D from a *single* trajectory knowing its length (number of positions N), localization precision, and camera integration time? Since a theoretical bound without a practical means of attaining it is of limited use, we also address a second, more practical question.
2. What estimation procedure extracts D from the data in a way that achieves such precision?

Answering the first question provides relations allowing optimizing experimental data acquisition parameters in order to achieve a desired level of precision on the measured D.

The answer to the second question provides tools to extract an optimal estimation of the diffusion coefficient from the resulting experimental trajectories.

This work extends and ties together our two recent independent publications addressing partial aspects of these questions [20,21]. On one hand, asymptotic (large N) expressions for the optimal precision (or Cramér-Rao lower bound) as well as a maximum likelihood

estimator (MLE) of D were obtained in Ref. [20]. On the other hand, an algorithm for optimized least-square fit (OLSF) of the familiar mean-square displacement curve, accounting for localization accuracy and trajectory length, was presented in Ref. [21]. Both results were incomplete in the following sense: the applicability of the asymptotic results of Ref. [20] to real data remained unknown (in particular, the question of how large must N be for the asymptotic results to apply), while the information theoretic optimality of the OLSF approach [21] remained to be addressed.

Here we give a complete theoretical analysis of diffusion coefficient estimation in SPT by first providing an exact Cramér-Rao lower bound formula for the diffusion coefficient uncertainty valid for all N, based on maximum likelihood estimation, thus answering question (1). We then briefly summarize the OLSF method and compare its theoretical performance with the Cramér-Rao lower bound, finding it to perform near optimally. We next compare numerical implementations of the MLE and OLSF approaches to the fundamental limit (Cramér-Rao bound) over a large simulated data set, verifying that both approach the theoretical performance. We conclude that (1) dynamic localization uncertainty σ sets a fundamental limit on the precision with which D can be determined, quantified by the Cramér-Rao lower bound; (2) both the MLE and OLSF provide a good approximation of the best possible diffusion coefficient estimate over a large portion of parameter space [thus answering question (2)]; and (3) in many experimental scenarios—including the estimation of short-time diffusion coefficients as might be encountered in complex or switching regimes—D cannot be measured precisely by any data analysis procedure. This, of course, should be of immediate practical interest, as it sets a practical limit to the claims that can be drawn from experimental data. We also hope that this work will motivate further studies of similar constraints existing in other diffusion or complex motion models.

This article is organized as follows. Section II briefly summarizes the maximum likelihood approach used in Ref. [20] and provides exact results for the Cramér-Rao lower bounds (CRLB) on fitted parameters (detailed in Appendix B), briefly discussing the general implications for parameter estimation in SPT. Section III provides a brief summary of the optimal least-square fit (OLSF) analysis of the mean-square displacement introduced in Ref. [21], while improvements introduced in this work are detailed in Appendix A. Section IV compares the results of OLSF with the CRLB, establishing that OLSF is nearly optimal. Section V then compares the performance of numerical implementations of the MLE and OLSF methods with the theoretical limits. Finally, Sec. VI summarizes our main results and discusses some applications.

II. MAXIMUM LIKELIHOOD ESTIMATION AND CRAMÉR-RAO LOWER BOUNDS ON THE VARIANCE OF D AND σ^2

In a typical SPT experiment, a trajectory is obtained by determining a particle's positions at equally spaced time intervals Δt, for instance using a series of N images (or frames) obtained by video microscopy [1,2]. The particle's position at each time point $i\Delta t$ is known with an uncertainty σ.

It is worth mentioning that the *dynamic localization error* σ involved in a tracking experiment is larger than the static localization error σ_0 involved in super resolution or high-resolution colocalization studies [18,19,23]. Indeed, as discussed in Ref. [21], the effect of diffusion is to expand the single-particle image and thus "spread out" the total signal over a larger area, therefore decreasing the signal-to-noise ratio and correspondingly, the localization accuracy. A simple model to take this effect into account was proposed in Ref. [21], resulting in the following approximation:

$$\sigma \sim \sigma_0 \left(1+\frac{Dt_E}{s_0^2}\right)^{1/2}, \quad (1)$$

where s_0 is the standard deviation of a Gaussian approximation of the microscope point-spread function and t_E is the camera exposure duration ($t_E \leq \Delta t$). From this expression, which involves the unknown parameter D, it is clear that σ should formally be considered as an unknown parameter of the model.

After SPT data have been collected, estimating model parameters such as D and σ from the trajectory can be accomplished using various methods [24–26]. Mean-square displacement (MSD) analysis is among the most popular one [2,24], but has a number of fundamental limitations, which make it wholly inappropriate for reliable parameter estimation, unless special care is taken, as discussed in a later section. An alternative approach to diffusion coefficient estimation based on maximum likelihood estimation was introduced in Ref. [15] and updated in Ref. [20], where it was noted that, in contrast with the original assumption of Ref. [15], SPT displacements $\{\Delta_i = \vec{r_{i+1}} - \vec{r_i}\}$ $(i = 1, \ldots, N-1)$ calculated from realistic data are correlated due to localization error and motion blur, resulting in a multivariate Gaussian distribution of observed displacements with nondiagonal covariance matrix Σ [20,27,28]. For the one-dimensional case:

$$p(\Delta) = \frac{1}{(2\pi)^{(N-1)/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\Delta^T \Sigma^{-1} \Delta\right). \quad (2)$$

In this expression Δ is the $(N-1)$-component column vector of observed displacements and Σ is the covariance matrix with $(N-1) \times (N-1)$ elements defined by [20]

$$\Sigma_{ij} = \begin{cases} 2D\Delta t(1+2x), & i=j, \\ -xD\Delta t, & i=j\pm 1, \\ 0, & \text{otherwise}, \end{cases} \quad (3)$$

where the reduced square localization error x is defined by

$$x = \frac{\sigma^2}{D\Delta t} - 2R, \quad (4)$$

and R is the *motion blur coefficient*, a dimensionless parameter specifying the duration and type of camera exposure used in the experiment [20,21]:

$$R=\frac{1}{\Delta t}\int_0^{\Delta t} S(t)[1-S(t)]dt. \quad (5)$$

In this expression, $S(t)$ is the illumination percentage occurring before time t (between the beginning and end of a frame, separated by Δt). R ranges from 0 to 1/4 depending on the camera exposure mode and duration. For instance, in the common experimental case of a uniform exposure of duration $t_E \leq \Delta t$, $R = 1/6(t_E/\Delta t)$ [setting $S(t) = t/t_E$ in Eq. (5) for $t < t_E$, $S(t) = 1$ otherwise] [15,20,21,29,30].

Using a standard assumption of statistical inference theory, the "best fit" parameters of the model are taken as those maximizing the probability distribution of the displacements, given the observed data. It is equivalent and in general simpler to maximize the log-likelihood, $L(\Delta) = \log[p(\Delta)]$, which can be done by diagonalizing the covariance matrix Σ. This computation can be done approximately (for large N) by a discrete Fourier transform, as discussed in Ref. [20]. From the diagonalized data, a numerical algorithm approximating the MLE for large N was given in Ref. [20], allowing an iterative search of the best fit parameters.

In Ref. [20] this approximation was used to compute the Fisher information matrix $\mathbf{I}$ in the two parameters D and σ^2, whose matrix elements $\mathbf{I}_{jk}$ are equal to

$$\mathbf{I}_{jk}=E\left[\frac{\partial L(\Delta)}{\partial \theta_j}\frac{\partial L(\Delta)}{\partial \theta_k}\right], \quad (6)$$

where θ_j designates either parameter D or parameter σ and E indicates the mean value of the quantity in parentheses. From this matrix, the *Cramér-Rao lower bounds* (CRLB) on the model parameter variances for large N were obtained as the inverse of the diagonal elements of the Fisher information matrix. The CRLB is an information theoretic quantity that sets a lower limit on the variances of all unbiased estimation procedures (not just the MLE [31]). In other words, no practical algorithm used to extract parameters D and σ from the data can expect to achieve a smaller uncertainty on these parameters than the lower limit set by the CRLB. Expressed in yet another way, any algorithm achieving an error on the fitted parameters that approaches the CRLB result is performing optimally from an information theory point of view.

In Ref. [20] the number of points N required to make these approximations valid was not discussed. Here we present exact results for the MLE and CRLB for any of N, allowing an exhaustive comparison of the MLE and OLSF algorithms performance with the CRLB. As shown in Appendix A, the CRLB on D depends only on the number of trajectory points N and the reduced square localization error x.

The CRLB[1] on σ^2 also depends on the motion blur coefficient R. From the expressions for the CRLB on the variance of D and σ^2 given in Appendix A, the relative uncertainties [ratio between the standard deviation (S) and the expected value of the parameter] satisfy

$$\frac{S(D)}{D} \geq \sqrt{\frac{2}{d(N-1)}}\left[1+\delta(x)^2\right]^{1/2}, \quad (7)$$

$$\frac{S(\sigma^2)}{\sigma^2} \geq \sqrt{\frac{2}{d(N-1)}}\left[1+\varsigma(x)^2\right]^{1/2}, \quad (8)$$

where d is the number of spatial dimensions of the trajectory and $\delta(x)$ [respectively, $\varsigma(x)$] is an increasing (respectively, decreasing) function of x, whose exact expressions as a function of x, N, and R are given in Appendix A.

Equations (7) and (8) completely specify how accurately D and σ^2 can be determined in SPT by an unbiased estimator. They are represented in Fig. 1, which shows contours of regions in the (x, N) plane within which each quantity can be estimated with a specified relative uncertainty ε.

It is worth noting that the dependence of the uncertainties on D and σ^2 with respect to the reduced square localization error x are anticorrelated: better precision on D is achieved when the reduced square localization error x is small, whereas the opposite behavior is observed for the localization error σ^2.

Equations (7) and (8) also express the fact that the standard deviation of both parameters vary as $[d(N-1)]^{-1/2}$, where $d(N-1)$ is the total number of one-dimensional displacements in the trajectory. In other words, an improvement of a factor of 2 of the precision on D requires 4 times more trajectory points (assuming the localization uncertainty remains constant). In particular, since both vary as $d^{-1/2}$, where d is the number of spatial dimensions, this means that from a statistical point of view it is not recommended to perform a separate MSD analysis for each dimension, as sometimes seen in the literature [32], unless there is a reason to expect some diffusion anisotropy. Section VI will discuss different limiting cases of these results and their application in practical situations.

III. OPTIMIZED LEAST-SQUARE FIT OF THE MSD

As mentioned above, one of the most popular approaches to SPT data analysis involves computing the so-called mean-square displacement (MSD) curve of the trajectory and fitting it with the appropriate dynamic model. Among different possible definitions of the MSD, we consider the most commonly used:

[1]To facilitate comparison of the CRLB results with those of the optimized least square fit approach presented in the next section, it is easier to consider the CRLB on the variance of σ^2 rather than of σ.

$$\overline{\rho}_n=\frac{1}{N-n}\sum_{i=1}^{N-n}(\overrightarrow{r}_{i+n}-\overrightarrow{r}_i)^2,\quad n=1,\ldots,N-1, \tag{9}$$

where $\vec{r_i} \in \{1, \ldots, N\}$ is the measured position of the particle in frame i. The MSD curve represents the average square distance traveled by the particle after different time lags $i\Delta t$. For an isotropic free diffusion regime in d dimensions, the theoretical MSD curve grows linearly with time, the slope of the curve being proportional to the diffusion coefficient D:

$$\rho(i\Delta t)=2d(\sigma^2-2RD\Delta t)+2dDi\Delta t. \tag{10}$$

Equation (10) can be normalized by the average single-frame mean square displacement $2dD\Delta t$ to obtain

$$\frac{\rho(i\Delta t)}{2dD\Delta t}=\left(\frac{\sigma^2}{D\Delta t}-2R\right)+i=x+i, \tag{11}$$

which identifies the reduced square localization error x defined in the previous section as the offset of the normalized MSD curve. When $R = 0$, a small value of x ($x < 1$) indicates a regime where the dynamic localization error σ is small compared to the mean displacement due to diffusion during a single frame duration, $\sqrt{2dD\Delta t}$.

Although a linear fit to the experimental MSD curve should provide a good estimate of D, the problem is complicated by the increased variance and correlated values of the MSD at large time lags, as noted originally [24,33]. This suggests using only the first few points of the MSD curve in order to reduce the influence of the large variance of the later MSD points. How few points to use has remained largely an empirical decision and varies from author to author. It was recently recognized that the MSD variance is also severely affected at short time lags by dynamic localization error σ and motion blur R [20,21] (the explicit expression for the MSD variance and covariance for arbitrary dimension d including dynamic localization error and motion blur is provided in Appendix C for reference). The increased variance of the MSD curve at short time lags is particularly noticeable when the reduced square localization error x is larger than 1, and results in poor MSD fits if only small time lags are used for the fit, as is often practiced. This suggested that there exists an optimal number $p_{\min}$ of MSD points to use (depending on the reduced parameter x as well as on the number of trajectory points N) in order to obtain the smallest uncertainty on the fitted parameter D.

This heuristic argument was confirmed by computing the expected error on both parameters (D and σ^2) as a function of number of fitting points p. As shown in Ref. [21], a minimal error on each fitted parameter can be achieved using an optimal number of MSD points $p_{\min}$, whose exact value depends on x and N. Arguably the most important conclusion of this work was the fact that using any other number of MSD points (say 10% of the total number,

or the first four points, or any other empirical choice) results in potentially much larger uncertainty on the fitted parameters. Although $p_{\min}$ can be computed without approximation as detailed in Ref. [21], it is more computationally efficient to use the approximate formulas for $p_{\min}(D)$ and $p_{\min}(\sigma^2)$ provided in Appendix B [Eqs. (B3) and (B4)]. The resulting error on each parameter D and σ^2 can then be computed using Eqs. (B11) and (B16).

The least-square fit of the unweighted MSD using the optimal number of points described above is what we will refer to in the following as the *optimized least square fit* (OLSF) of the MSD curve. Note that since this fit requires the *a priori* knowledge of x to compute the optimal number of points needed to fit D and σ^2, it is not directly applicable to experimental data, for which, by definition, D and σ are not known. However, an iterative search of the optimal number of fitting points can be devised, as briefly discussed in Ref. [21]. An improved version of this algorithm and results of extensive study of its performance are presented in Appendix B, and discussed in Sec. V.

IV. COMPARISON OF THE THEORETICAL OLSF RESULTSWITH THE CRAMÉR-RAO LOWER BOUNDS

One of the unanswered questions regarding the theoretical OLSF approach is in what sense the fitted OLSF parameters are optimal. An indication that they might be more than optima for the least-square fit approach only, is given by the fact that uncertainties on fitted parameters [D and σ^2, Eqs. (B11) and (B16)] appear numerically indistinguishable whether one used a weighted or an unweighted optimized least-square fit [21]. This is a nontrivial result, as both types of fit (weighted or unweighted) result in very different behavior of the parameter uncertainties as a function of the number of MSD points p used for the fit (see Ref. [21] for details). The fact that both fits yield the same minimum uncertainty on the parameters is thus a strong indication that this minimum is an absolute minimum. However, the only way to establish that the OLSF results are truly optimal in a information theoretic sense is by a direct comparison with the lowest statistical bounds on each parameter's uncertainty.

We thus compared the OLSF theoretical results for the uncertainties on D and σ^2 [Eqs. (B11) and (B16)] with the respective CRLB obtained in Sec. II.

As shown in Fig. 2 for $R = 1/6$ (camera integration during the whole frame duration) and $d = 1$, there is an excellent agreement between the two approaches over several orders of magnitude for x and N. Similar results hold for $R = 0$ (no integration, see Fig. 8 in Appendix A) and values in between.

Since the CRLB are by definition the theoretical limits of the MLE approach, these results indicate than *numerical* implementations of both MLE and OLSF methods should be capable of approaching these theoretical limits in experimental situations. This, however, can only be determined by investigating the performance of actual algorithms implementing these approaches, which we do in the next section.

V. COMPARISON OF ACTUAL MLE AND OLSF ALGORITHMS WITH ONE ANOTHER AND WITH THE CRLB

To test the practical performance of MLE and OLSF algorithms we performed a series of simulations over a large range of reduced square localization error x and number of trajectory points N. For each (x, N) pair we simulated 1000 trajectories and ran both the MLE and OLSF estimators (provided as a Matlab routine in Supplemental Material [34]) on each of them, and analyzed the resulting distribution of fitted D and σ^2 parameters.

Figures 3 and 4 summarize the results of our numerical study in the form of "success maps" for both algorithms. For each (x, N) pair we computed the fraction of trajectories whose fitted D (Fig. 3) or σ^2 (Fig. 4) was within 25% of the true value (the value used for the simulation), as a measure of the success rate of both approaches. We also overlaid smooth contours showing the expected success fraction (defined as above) for an optimal unbiased estimator that achieves the CRLB. Note that the choice of a 25% distance to the true value is arbitrary and does not change the fact that for any other choice the agreement between theory and algorithms is excellent.

From Figs. 3 and 4, which show a near perfect match between the success maps of both methods and the predicted success rates, it appears that both MLE and OLSF estimators achieve nearly optimal performance. That is, each estimator performs almost as well as an ideal unbiased estimator. In other words, no other unbiased estimator can significantly outperform the MLE or OLSF estimators presented here over the vast majority of the (x, N) space. A detailed comparison of the complete data set is available online as a hyperlinked PDF document [34] and is worth looking at to understand the difference between the situation in high success rate regions (green) and a low success rate regions (red). In the latter, the fitted parameters can be wildly off the exact result, casting doubt on the usefulness of even attempting to fit the data in these regions of the (x, N) plane.

In particular, it is important to realize that, although MLE and OLSF are nearly optimal, they cannot give reliable estimates of D for small numbers N of trajectory points ($N < 10$). Indeed, it is impossible for any estimator to obtain reliable estimates in this case, even when the reduced square localization error x is fairly modest ($x < 1$). This can be seen from Eq. (7) and Figs. 1–4, and suggests that a number of approaches using short trajectory segments to infer subtle changes of diffusion regimes may need to be revisited (e.g., Refs. [35,36]). We discuss this important point further in Sec. VI.

Although they approach optimality, MLE and OLSF do not give identical results. Indeed, in contrast to OLSF, MLE cannot return an unphysical negative value for D or σ^2. Although this seems initially to be a desirable attribute, it means that estimates of a parameter with a large variance (compared to the true value) tend to have an asymmetric distribution with positive values results "piling up" at $D = 0$ and $\sigma^2 = 0$ (see Supplemental Material [34] for examples). This results in a positive bias (overestimation of the mean), which is particularly marked for $x > 1$. OLSF on the contrary is not constrained to return positive values and thus results in symmetrically distributed, unbiased results. However, in the same regime of large parameter variance (compared to the true value), this can result in unphysical negative fitted

values. Each of these phenomena (biased results and unphysical estimates) are undesirable, but it is important to note that both issues appear in "red zones" of the parameter space (Figs. 3 and 4), where the CRLB is large compared to the mean, and even an optimal unbiased estimator performs rather poorly. In these regimes there is simply not enough information in the data to obtain high-quality estimates by any estimation procedure.

Another noteworthy difference between the two algorithms (implemented in the Matlab routine provided as Supplemental Information [34]) is their execution speed. A simple estimate of computation time as a function of the number of points in the trajectory shows an approximately linear behavior for MLE versus a quadratic dependence for the OLSF (Fig. S1). However, the OLSF algorithm appears to outperform the MLE algorithm for $N < 3000$ points. As the computer speed steadily improves, these differences might become irrelevant.

So far we only discussed results for cases where σ is unknown and must be estimated along with D. If σ is already known, the CRLB on D is improved and a smaller number of trajectory points are needed to obtain a given uncertainty (Fig. 5). It is common to attempt to measure this quantity using immobilized particles. However, σ depends on all aspects of the experiment including motion blur contributions not present for static particles [Eq. (1)]. We therefore consider that the most important case of practical interest is that in which σ is unknown, as discussed in the text.

VI. DISCUSSION

In the previous sections we have presented exact results for the MLE and OLSF approaches and established that the two corresponding numerical algorithms perform optimally in most practical situations. In this section we discuss some possible applications of these results for SPT analysis.

A. Estimating the uncertainty on *D*

A first application of these results consists, after using either one of these algorithms to obtain the best possible estimates of D and σ^2 from a given trajectory, of estimating the uncertainty on these parameters using the CRLB results [Eqs. (7) and (8)] *computed using the fitted parameters*. However, since Eqs. (7) and (8) depend on the reduced localization error x, which is very sensitive to the values of D and σ^2, a large relative uncertainty on either one or both fitted parameters may yield artificially small or large x value, resulting in erroneous estimates of their relative uncertainties. Although analytically computable using the results of Appendix A, the exact distribution of x values is complex and can moreover be bimodal (with positive or negative values) [37], making a general discussion very tedious.

It is possible, however, to discuss a few common experimental cases. For instance, in the presence of *negligible dynamic localization error* ($\sigma^2/D\Delta t < 1$) and assuming the exposure time t_E is equal to the frame duration Δt ($R = 1/6$), the results of Fig. 2 show that a better estimate of D than of σ^2 is expected. Supposing for instance that the true value of the right-hand side of Eq. (8) defining the minimum relative error on σ^2 is of the order of 1, we could potentially obtain a fitted value of σ_{fit}^2: (i) a few times larger than the true value [e.g., $\sigma_{fit}^2 \sim$

$\sigma^2 + V(\sigma^2)^{1/2} \sim \lambda\sigma^2$, with $\lambda = O(1)$] or (ii) a few orders of magnitude smaller than the true value [e.g., $\sigma_{fit}^2 \sim \sigma^2 - V(\sigma^2)^{1/2} \sim \sigma^2/\lambda$, with $\lambda = O(1)$].

In the first case (i) the value x_{fit} of x computed using the fitted values of D and σ^2 [Eq. (4)], $D_{fit} \sim D$, and $\sigma_{fit}^2 > \sigma^2$ would be such that $x_{fit} > x$, and would result in an overestimation of the error on D and an underestimation of the error on σ^2. Note however, that if $x_{fit} + 2R = \sigma_{fit}^2/D_{fit}\Delta t \ll 1$, the overestimation of the error on D will be negligible (Fig. 2). Indeed, in this regime, the relative error on D is approximately independent of x.

In the second case (ii) [e.g., $\sigma_{fit}^2 \sim \sigma^2 - V(\sigma^2)^{1/2}$], the severe underestimation of σ^2 could result in $x_{fit} + 2R \ll 1$, which, as we have just seen, would have little effect on the estimated error on D.

Inversely, in presence of *significant dynamic localization error* ($\sigma^2/D\Delta t > 1$), the opposite would be expected: larger error on D and small error on σ^2. Suppose the fitted D_{fit} is underestimated and a few orders of magnitude smaller than its true value [e.g., $D_{fit} \sim D - V(D)^{1/2} \sim D/\lambda$, with $\lambda = O(1)$]. Then x_{fit} would be much larger than its true value, and the inferred relative error on D much larger than it is really. If, on the contrary, D_{fit} is a few times larger than its true value [e.g., $D_{fit} \sim D + V(D)^{1/2} \sim \lambda D$, with $\lambda = O(1)$], $x_{fit} + 2R = \sigma_{fit}^2/D_{fit}\Delta t$ would be a few times smaller than its true value. If $x_{fit} + 2R \ll 1$, the effect on the estimated relative error on D would be negligible, whereas in the other cases, it would be difficult to determine how bad this estimate is.

In summary, it appears that the CRLB result for the relative error on D [Eq. (7), calculated using the fitted values] can be used reliably provided the fitted x value ($x_{fit} = \sigma_{fit}^2/D_{fit}\,\Delta t - 2R$) is smaller than 1, whereas no reliable error estimation can be made if $x_{fit} > 1$. The opposite conclusion holds for the relative error on σ^2 [Eq. (8)].

Naturally, if a large series of trajectories of *identical* particles can be studied, a reliable estimate of the *mean value* of the diffusion constant D may be obtained (law of large numbers), from which an estimate of σ could be made using Eq. (1) and an estimate of x using Eq. (4)). The resulting estimate of the relative uncertainty on D [Eq. (7)] could then be compared to the relative statistical standard deviation of D for self-consistency. Significant discrepancy between the two might indicate an underlying distribution of diffusion coefficients for the particles studied.

B. Adjusting experimental parameters to improve the precision on *D*

A second application of these results consists in optimizing SPT experimental parameters (N, Δt, R, and σ) to achieve a desired precision on the measured diffusion constant D (assuming one has an approximate idea of its order of magnitude). As shown in Appendix A and Fig. 6, when the desired relative uncertainty is smaller than 1, it is sufficient to use the asymptotic forms of Eqs. (A7) and (A8) for the CRLB. In this approximation, the relative uncertainty on D can be expressed as

$$\frac{S(D)}{D} \gtrsim \sqrt{\frac{2}{d(N-1)}}(1+2\sqrt{1+2x})^{1/2}, \quad (12)$$

showing that when N and x are large (compared to 1), $\delta(x)^2 \sim \sqrt{8x}$, and the relative uncertainty on D increases as $x^{1/4}$ (i.e., $\sigma^{1/2}$) but decreases as $N^{-1/2}$. In other words, when the reduced square localization error x is large, halving the dynamic localization error σ (for instance by increasing the brightness of the particle's PSF) allows halving the number of data points N needed to obtain a given precision on D.

When N is large and the reduced square localization error x is small, $\delta(x)^2 \sim 2$, indicating that the relative uncertainty on D varies as $\sim \sqrt{6}(dN)^{-1/2}$. For instance, in order to obtain a 10% precision on D in two dimensions, $N > 300$ trajectory points are necessary. A 30 points trajectory with negligible dynamic localization error would only allow at best a ~30% precision on D, while a 10 points trajectory only a ~54% precision. These figures would be increased by a factor ~2 for a reduced square localization error $x = 1$ and significantly more for larger x values (see also Fig. 5 for a visual guide). Note that none of these figures are recommendable goals since a 30% relative uncertainty translates into a three-sigma confidence range going approximately from zero to twice the actual value of the diffusion constant. In other words, the fitted value has potentially very little resemblance to the true value.

Clearly there is quite some flexibility in adjusting x or N in order to reduce the uncertainty on D. For instance, x can always be decreased by increasing the frame interval Δt or reducing the static localization error σ_0 [Eq. (1)]. The latter can for instance be obtained by increasing the particle brightness (number of collected photons). This, however, may reduce the lifetime of the particle due to premature photobleaching, thus limiting the number of observable trajectory points N, which is always detrimental to the precision of the measurement.

On the other hand, if one is interested in measuring D over short time scales (in order to detect changes in diffusion coefficient), the total duration of the measurement is constrained by the time scale of interest: $N\Delta t = T$, therefore increasing Δt would reduce N and vice versa.

We can quantify these different statements using the approximate expression for the dynamic localization error σ [Eq. (1)].

The reduced square localization error x can be expressed as

$$x = \frac{\sigma^2}{D\Delta t} - 2R \sim \frac{\sigma_0^2}{D\Delta t}\left(1 + \frac{Dt_E}{s_0^2}\right) - \frac{1}{3}\frac{t_E}{\Delta t}. \quad (13)$$

In the absence of pixilation, background, detector noise, and amplifier excess noise, the static localization error σ_0 itself can be approximated by [21]

$$\sigma_0 \sim \frac{s_0}{\sqrt{P}} \sim \frac{s_0}{\sqrt{Qt_E}}, \quad (14)$$

where P is the *total number of photons* in the PSF image and Q is the *average photon rate* emitted by the particle. From these two equations, it is straightforward to verify that x is minimized when $t_E = \Delta t$ ($R = 1/6$) and when Δt is maximized. In fact, according to Fig. 2, given a number of trajectory points N, a quasi-optimal precision on D can be obtained as soon as $x + 1/3 < 1$, which reads

$$P \gtrsim 1 + s_0^2/D\Delta t. \quad (15)$$

Using a PSF size $s_0 \sim 100$ nm and a typical membrane protein diffusion coefficient $D \sim 0.1$ $\mu m^2/s$, we see that the number of photons required to satisfy Eq. (15) is easily achieved in practice ($\Delta t = 100$ ms $\rightarrow P > 2$, $\Delta t = 1$ ms $\rightarrow P > 100$). For particles diffusing more slowly, accurate measurements of D will require longer integration times or brighter images ($D \sim 10^{-3}$ $\mu m^2/s$, $\Delta t = 100$ ms $\rightarrow P > 100$, $\Delta t = 10$ ms $\rightarrow P > 1000$).

So far we have not specified the precision desired for D. According to Fig. 2 it can be improved either by decreasing $x + 2R$ or by increasing the number of trajectory points N. If we are interested in measuring the diffusion coefficient rapidly with a relative precision ε, we can use the asymptotic behavior $\varepsilon \sim \sqrt{6}(dN)^{-1/2}$ provided the number of frames N is sufficiently large ($N > 10$) and $x < 1$. The latter condition can be ensured by adjusting Δt as discussed previously and therefore the minimum number of frames needed to achieve a precision ε will be given by $N > 6/d\varepsilon^2$. For $d = 2$ and $\varepsilon = 0.5$ (a 50% precision on D, which again is not a recommendable goal), we obtain $N > 12$. For a better precision (for instance, to distinguish between D and $D/2$), $\varepsilon = 0.1$ would require $N > 300$. Since the total duration of the measurement $T = N\Delta t$ might be constrained by other considerations, Δt might become relatively short, requiring stronger signal to keep x small.

VII. CONCLUSION

In conclusion, we show that both MLE and OLSF estimators achieve nearly optimal performance in estimating D and σ^2. Exact expressions for the CRLB on parameter uncertainties provided here can be used for determining when an experiment can reasonably be expected to give any useful information at all. It could also be used to determine *a priori*, based on experimental parameters such as integration time, frame rate, signal-to-noise ratio, maximum experiment duration, etc., what maximum precision can be expected for an estimate of the diffusion coefficient D. These results will also allow benchmarking other analysis methods with respect to the exact Cramér-Rao lower bound for various (x, N) values. In regimes where experimental data contains sufficient information to give a useful estimate, MLE and OLSF can be implemented to achieve near-optimal results with high

confidence. Extension of these results to more sophisticated dynamic models will obviously be of great practical interest but are beyond the scope of this study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by Grants No. NIH EB006353 and No. GM084327 (XM).

APPENDIX A

CRAMÉR-RAO LOWER BOUND ON THE UNCERTAINTY OF *D* AND σ^2

We briefly extend results from Ref. [20] to find exact, rather than approximate, expressions for the maximum likelihood estimator (MLE) and Cramér-Rao (CR) bounds for D and σ^2. For a one-dimensional ($d = 1$), N-element trajectory (X_1, …, X_N), we define the $(N-1)$-element displacement (column) vector to be $\Delta = [X_2 - X_1, X_3 - X_2, \ldots, X_N - X_{N-1}]^T$. For free diffusion with diffusion coefficient D and zero drift velocity, recorded at a frame interval Δt with noise σ in each frame, the $(N-1) \times (N-1)$ covariance matrix of Δ was found to be tridiagonal with matrix elements [20]:

$$\Sigma = \left\langle \Delta\Delta^T \right\rangle = 2D\Delta t \begin{pmatrix} 1+x & -x/2 & \cdots & 0 \\ -x/2 & & \ddots & \vdots \\ \vdots & \ddots & & -x/2 \\ 0 & \cdots & -x/2 & 1+x \end{pmatrix}, \quad \text{(A1)}$$

where x is the reduced square localization error [21] including a motion blur term [Eq. (5)].

The displacement covariance matrix is diagonal if and only if $x = 0$. R is the motion blur coefficient defined in Eq. (5). In Ref. [20] approximate expressions for the eigenvalues and eigenvectors of Σ were used to derive the MLE, Fisher information matrix, and Cramér-Rao lower bound on estimates of D and σ. Those results hold asymptotically for large N. Here we update those results with exact expressions. Let $\mathbf{v}^{(k)}$ and Ψ_k be the kth eigenvector and corresponding eigenvalue of Σ. These are found to be [38]

$$\mathbf{v}_j^{(k)} = \sqrt{\frac{2}{N}} \sin\left(\frac{jk\pi}{N}\right), \quad \text{(A2)}$$

$$\psi_k = 2D\Delta t \left\{1 + x\left[1 - \cos\left(\frac{\pi k}{N}\right)\right]\right\}, \quad \text{(A3)}$$

and satisfy $\Sigma \mathbf{v}^{(k)} = \Psi_k \mathbf{v}^{(k)}$ and $\mathbf{v}^{(k)} \cdot \mathbf{v}^{(l)} = \delta_{kl}$. Equation (A2) corresponds to the eigenvectors of the discrete sine transform (DST), for which a fast numerical implementation is known, similar to the fast Fourier transform [39]. In the terminology of stochastic processes, we have observed that the DST is the Karhunen-Loève transform [38] of the single-particle-tracking displacement process Δ. Consequently, the DST of Δ has a *diagonal* covariance matrix (its components are uncorrelated), and an *exact* numerical implementation of the MLE can be efficiently computed in this basis using Eq. (18) of Ref. [20] but setting $\tilde{\Delta} = DST(\Delta)$. The exact Fisher information matrix $\mathbf{I}$ in the two parameters D and σ^2 can also be found from the eigenvalues given above using the formula in Eq. (19) of Ref. [20]. Its inverse, which sets a lower bound on the covariance matrix of estimates of D and σ^2 is found after algebraic manipulation to be (in d dimensions)

$$[\mathbf{I}_{D\sigma^2}]^{-1} = \frac{2}{d(N-1)} \begin{pmatrix} D^2(1+\delta^2) & D\sigma^2(1-\delta\varsigma) \\ D\sigma^2(1-\delta\varsigma) & \sigma^4(1+\varsigma^2) \end{pmatrix}, \quad \text{(A4)}$$

where

$$\delta = \delta(x) = \frac{1 - \acute{O}\acute{r}_1}{\sqrt{\acute{O}\acute{r}_2 - \acute{O}\acute{r}_1^2}}, \quad \text{(A5)}$$

$$\varsigma = \varsigma(x) = \frac{\bar{y}_1 - 2R/(x+2R)}{\sqrt{\bar{y}_2 - \bar{y}_1^2}}, \quad \text{(A6)}$$

$$\bar{y}_1 = \frac{1}{N-1} \sum_{k=1}^{N-1} \left\{ 1 + x \left[1 - \cos\left(\frac{\pi k}{N} \right) \right] \right\}^{-1} \sim \frac{1}{\sqrt{1+2x}} (N \gg 1), \quad \text{(A7)}$$

$$\bar{y}_2 = \frac{1}{N-1} \sum_{k=1}^{N-1} \left\{ 1 + x \left[1 - \cos\left(\frac{\pi k}{N} \right) \right] \right\}^{-2} \sim \frac{1+x}{(1+2x)^{3/2}} (N \gg 1). \quad \text{(A8)}$$

These formulas can be used to compute the exact CRLB of Eqs. (7) and (8). The approximate results indicated by "~" arise from approximating the sum over N as a continuous integral. Plugging the asymptotic forms of Eqs. (A7) and (A8) into Eqs. (A5) and (A6), and in turn plugging these into Eq. (A4) gives the asymptotic form of the CRLB.

The asymptotic formulas can be used for a large range of (x, N) values as shown in Fig. 6, which represent curves of constant relative CRLB for D using either the exact formula (open symbols) or the approximate ones (plain curves). The asymptotic formulas result in an

underestimation of the minimum number of trajectory points N needed to obtain a given relative error ε on D only if ε > 1 (which corresponds to situations where D cannot be estimated reliably and therefore should be avoided).

APPENDIX B

OPTIMIZED LEAST-SQUARE FIT (OLSF) OF THE MSD

As discussed in detail in Ref. [21], for each value of x [Eq. (4)], there is an optimal number of MSD points $p_{\min}$ to use in order to obtain the "best" least-square fit parameters a (intercept) and b (slope) of the MSD, where

$$a=2d\sigma^2 - 4dRD\Delta t, \quad b=2dD \quad \text{(B1)}$$

in which d is the space dimension. The fitted parameters are "optimal" fits in the sense that their standard deviation (or uncertainty) is minimum among all fitted parameters obtained with a different number of MSD points p. Reference [21] only considered the blur-free case ($R = 0$) in the derivation of the fit parameters uncertainty and provided only a very short discussion of a practical algorithm to obtain $p_{\min}$. This Appendix provides additional details on the algorithm and extends the analysis to the general case $0 \le R \le 1/4$.

Optimal number of fitting points

Ref. [21] provided an approximate formula to estimate the optimal number of MSD points to use for an unweighted fit of a and b:

$$p_{\min}=E(2+2.3x^{0.52}), \quad \text{(B2)}$$

where x is given by Eq. (4). Note that if $x < 0$, Eq. (B2) should read $p_{\min} = 2$. This approximation is poor for large x values ($x \gg 1$) and $N < 1000$ values and does not reflect the fact that slightly different optimal number of points are needed to fit the slope b (i.e., D) or the intercept a (i.e., σ^2).[2]

Figure 7 (open symbols) represents the number of points to use to obtain the optimal intercept a [Fig. 7(a)] or slope b [Fig. 7(b)] using an unweighted LSF of the MSD curve, as a function of both x and N. These values were obtained by direct calculation of the relative fitted parameter uncertainties as a function of number of fitting points p and determination of the minimum location [21]. They are therefore *exact* results of the OLSF algorithm with prior information on x. The corresponding relative uncertainties on each parameter can be computed using Eqs. (B16) and (B11) [21].

To avoid computing the relative uncertainties for all p ($2 \le p < N$) in order to find the minimum uncertainty, an iterative search of the minimum can be used since the relative

[2]The values of $p_{\min}$ given here correspond to the fit of the MSD curve as defined by Eq. (9). Any other definition would require a new computation of the variance and covariance of the MSD, and consequently, of the theoretical uncertainty of the fitted parameters.

uncertainties are, in general, convex curves of p. Although this approach is very fast for modest values of trajectory points N, computation of the uncertainties can become slow for large N ($N > 1000$). Since this procedure needs to be repeated at each step of the iterative algorithm mentioned in Sec. III and described next, using a heuristic formula can significantly speed up the approach.

Plain curves in Fig. 7 represent heuristic formulas providing a good approximation of $p_{\min}$ for $x \in [10^{-2}, 10^{7}]$ and $N \leq 10^{4}$. The heuristic formulas represented on these graphs are

$$p_{\min}^{(a)}(x,N) = E\{f_a(x)L_a(N)/[f_a(x)^3 + L_a(N)^3]^{1/3}\},\ f_a(x) = 2 + 1.6x^{0.51},\quad L_a(N) = 3 + (4.5N^{0.4} - 8.5)^{1.2}. \tag{B3}$$

and

$$p_{\min}^{(b)}(x,N) = \min(L_b(N),\ E\{f_b(x)L_b(N)/[f_b(x)^3 + L_b(N)^3]^{1/3}\}),$$

$$f_b(x) = 2 + 1.35x^{0.6},\quad L_b(N) = 0.8 + 0.564N, \tag{B4}$$

$L_{a,b}(N)$ are the limits for $x \to \infty$ of $p_{\min}^{(a,b)}(x, N)$. Note that these formulas are *ad hoc* formulas and could be replaced by any other algebraic form fitting the exact results sufficiently well. In practice, the fitted a and b obtained using these approximate formulas are indistinguishable from those obtained using the exact results.

Determining the optimum number of fitting points without a priori knowledge

In the absence of any a priori knowledge on x, an iterative search of $p_{\min}^{(a,b)}(x,N)$ [yielding a series of $p_i(a)$ and $p_i(b)$ values for each parameter a and b] is needed. The algorithm proposed in Ref. [21] (modified here to account for the improved formulas (B3) and (B4)) starts with $p_1(a) = p_1(b) = N/10$ and proceed as follows:

1. Using the new p_i, fit the MSD curve and obtain new fit parameters a_i and b_i.
2. Using the new fit parameters a_i and b_i compute a new x_i.
3. Using the new x_i, compute new p_{i+1} using Eqs. (B3) and (B4).
4. Repeat steps 1–4 until a previously found set of p_i is repeated.
5. In the latter case, use the previous set of p_i.

In practice, this algorithm does not function well for very small or very large x, because in these cases, either the relative error on a or the relative error on b is very large (and the computed x is therefore extremely unreliable) if the optimum p are not used. These cases are however easily detected. When the relative error on a is very large, a will occasionally have

a negative value. In this case, the algorithm is stopped and a value of $x = 0$ is assumed (in other words, $p_{\min} = 2$ is used for both parameters).

When the relative error on b is large, b will occasionally have a negative value. In this case, the algorithm is stopped and a value of $x = \infty$ is assumed [in other words, $p_{\min}(a,b) = L_{a,b}(N)$ are used].

This simple algorithm is surprisingly successful and very rapid (only a few iterations are necessary).

Figure 8 shows the percentage of success as a function of x and N in determining the optimal $p_{\min}$ using simulated data described in the text. Success if defined here as computing a $p_{\min}$ with the previous algorithm, which, if used to perform the MSD least-square fit, would result in an error no more than 25% larger than the optimal error. The algorithm is successful in more than 50% of cases even in the most unfavorable regions of the (x, N) plane ($x \gg 1$ for D and $x \ll 1$ for σ^2).

Relative error on D and σ^2 in the presence of motion blur ($R > 0$)

This section expands the results of Ref. [21] in the presence of blur, limiting it to the case of unweighted linear fit. We explicitly introduce the number of spatial dimensions d in the final results.

The first $p \leq N-1$ points of the MSD curve are fitted by the equation

$$\rho(t) = a + bt, \quad \text{(B5)}$$

where a and b are related to the variable parameters σ^2 and D by Eq. (B1). The best fit values of an unweighted LSF limited to the first p points of the MSD are obtained by minimization of the χ square:

$$\chi^2(\sigma^2, D) = \sum_{i=1}^{p} (\overline{\rho}_i - a - bt_i)^2, \quad \text{(B6)}$$

with respect to σ^2 and D. Here we have used the notations of Ref. [21]:

$$\overline{\rho}_n = \frac{1}{N-n} \sum_{i=1}^{N-n} (\overrightarrow{r}_{i+n} - \overrightarrow{r}_i)^2,$$

$$n = 1, \ldots, N-1, \quad t_n = n\Delta t. \quad \text{(B7)}$$

After some straightforward algebra we obtain

$$2dD=\frac{II_{xy}-I_xI_y}{\Delta},$$

$$2d\sigma^2=\frac{I_{xx}I_y-I_xI_{xy}}{\Delta}+2R\Delta t\frac{II_{xy}-I_xI_y}{\Delta}, \quad \text{(B8)}$$

where we have introduced the intermediate quantities

$$I=\sum_{i=1}^{p}1, \quad I_x=\sum_{i=1}^{p}t_i, \quad I_{xx}=\sum_{i=1}^{p}t_i^2,$$

$$I_y=\sum_{i=1}^{p}\overline{\rho}_i, \quad I_{xy}=\sum_{i=1}^{p}t_i\overline{\rho}_i, \quad \text{(B9)}$$

and Δ is defined by

$$\Delta=II_{xx}-I_x^2. \quad \text{(B10)}$$

The results of Ref. [21] for the relative variance (Var) of the intercept *a* and slope *b* need to be multiplied by a factor 2/*d* for an arbitrary number of spatial dimensions *d*

$$\frac{\mathrm{Var}(a)}{a^2}=\frac{2}{d}\frac{1}{x^2}\left[\frac{6}{p(p-1)}\right]^2\left[\sum_{i=1}^{p}\left(\frac{2p+1}{3}-i\right)^2 f(i,N,x)^{-1}+2\sum_{i=1}^{p}\sum_{j<i}\left(\frac{2p+1}{3}-i\right)\left(\frac{2p+1}{3}-j\right)g(i,j,N,x)\right]$$

$$\frac{\mathrm{Var}(b)}{b^2}=\frac{2}{d}\left[\frac{12}{p(p^2-1)}\right]^2\left[\sum_{i=1}^{p}\left(\frac{p+1}{2}-i\right)^2 f(i,N,x)^{-1}+2\sum_{i=1}^{p}\sum_{j<i}\left(\frac{p+1}{2}-i\right)\left(\frac{p+1}{2}-j\right)g(i,j,N,x)\right]$$

(B11)

where functions *f*(*i, N, x*) and *g*(*i, j, N, x*) have been introduced in Ref. [21] and are defined by

$$f(i,N,x)=\frac{d}{2}(4D\Delta t)^2/\sigma_i^2,$$

$$g(i,j,N,x)=\sigma_{ij}^2/\frac{d}{2}(4D\Delta t)^2, \quad \text{(B12)}$$

where the quantities σ_{ij}^2 in the above equation are the covariance (variance when $i=j$) of the MSD derived in Refs. [20·21] and reported in Appendix C.

The variance of $2d\sigma^2$ can be either computed directly or using the properties of multivariate normal distributions:

$$\mathrm{Var}(2d\sigma^2)=\mathrm{Var}\ (a)+(2R\Delta t)^2\ \mathrm{Var}\ (b)+4R\Delta t\ \mathrm{Cov}\ (a,b), \quad \text{(B13)}$$

where Cov(*a, b*) is the covariance of *a* and *b* defined by

$$\mathrm{Cov}\ (a,b)=\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}^2\left(\frac{\partial a}{\partial \bar{p}_i}\right)\left(\frac{\partial b}{\partial \bar{p}_j}\right). \quad \text{(B14)}$$

We obtain

$$\mathrm{Cov}\ (a,b)=\frac{(2dD\Delta t)^2}{2\Delta t}\frac{2}{d(p+1)}\left[\frac{6}{p(p-1)}\right]^2\left[\sum_{i=1}^{p}\left(\frac{2p+1}{3}-i\right)\left(i-\frac{p+1}{2}\right)f(i,N,x)^{-1}\right.$$
$$\left.+\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}\left(\frac{2p+1}{3}-i\right)\left(j-\frac{p+1}{2}\right)g(i,j,N,x)\right]=\frac{(2dD\Delta t)^2}{2\Delta t}U(p,N,x).$$

(B15)

The relative variance of σ^2 can thus be rewritten:

$$\frac{\mathrm{Var}(\sigma^2)}{\sigma^4}=(x+2R)^{-2}\left[x^2\frac{\mathrm{Var}(a)}{a^2}+4R^2\frac{\mathrm{Var}(b)}{b^2}+8RU(p,N,x)\right], \quad \text{(B16)}$$

where the expressions for $\mathrm{Var}(a)/a^2$ and $\mathrm{Var}(b)/b^2$ are given by Eq. (B11) and $U(p,N,x)$ by Eq. (B15).

The relative variance of *D* is simply given by the relative variance of *b* [Eq. (B11)]. These results were used to plot the graphs of Figs. 2 and 9.

APPENDIX C

VARIANCE AND COVARIANCE OF THE MSD FOR ARBITRARY SPATIAL DIMENSION *d*

The variance $\sigma_i^2=\langle\rho_i^2\rangle-\langle\rho_i\rangle^2$ and covariance $\sigma_{ij}^2=\langle\rho_i\rho_j\rangle-\langle\rho_i\rangle\langle\rho_j\rangle$ of the MSD were computed in Refs. [20,21]. However, Ref. [20] only provided a compact form for $d=1$, whereas Ref. [21] provided an expanded form for $d=2$. Both expressions are equivalent and can be easily generalized for arbitrary dimension d (assuming an identical dynamic localization uncertainty σ for all spatial dimensions) as follows:

$$\begin{aligned}\sigma_i^2=&8d(D\Delta t)^2\left(\frac{n}{6K^2}(4n^2K+2K-n^3+n)+\frac{1}{K}\left\{2nx+\left[1+\frac{1}{2}(1-\frac{n}{K})\right]x^2\right\}\right),n\le K=N-n\\ =&8d(D\Delta t)^2\left[\frac{1}{6K}(6n^2K-4nK^2+4n+K^3-K)+\frac{1}{K}(2nx+x^2)\right],n>K\end{aligned}$$

$$\begin{aligned}\sigma_{ij}^2=&8d(D\Delta t)^2\left\{\frac{n}{6KP}[4n^2K+2K-n^3+n+(m-n)(6nP-4n^2-2)]\right.\\ &\left.+\frac{1}{K}\left[2nx+(1-\frac{n}{2P})x^2\right]\right\},m+n\le N,K=N-n,P=N-m\\ =&8d(D\Delta t)^2\left(\frac{1}{6K}\{6n^2K-4nK^2+K^3+4n-K\right.\\ &\left.+(m-n)[(n+m)(2K+P)+2nP-3K^2+1]\}+\frac{1}{K}\left[2nx+\frac{x^2}{2}\right]\right),m+n>N.\end{aligned}\quad\text{(C1)}$$

References

1. Crocker JC, Grier DG. J. Colloid Interface Sci. 1996; 179:298.
2. Saxton MJ, Jacobson K. Annu. Rev. Biophys. Biomol. Struct. 1997; 26:373. [PubMed: 9241424]
3. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Science. 2006; 313:1642. [PubMed: 16902090]
4. Hess ST, Girirajan TPK, Mason MD. Biophys. J. 2006; 91:4258. [PubMed: 16980368]
5. Huang B, Wang W, Bates M, Zhuang X. Science. 2008; 319:810. [PubMed: 18174397]
6. Antelman J, Wilking-Chang C, Weiss S, Michalet X. Nano Lett. 2009; 9:2199. [PubMed: 19374434]
7. Pertsinidis A, Zhang YX, Chu S. Nature (London). 2010; 466:647. [PubMed: 20613725]
8. Yildiz A, Forkey JN, McKinney SA, Ha T, Goldman YE, Selvin PR. Science. 2003; 300:2061. [PubMed: 12791999]
9. Dahan M, Levi S, Luccardini C, Rostaing P, Riveau B, Triller A. Science. 2003; 302:442. [PubMed: 14564008]
10. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, Spector DL, Singer RH. Science. 2004; 304:1797. [PubMed: 15205532]
11. Cao J. Phys. Rev. E. 2001; 63:041101.
12. Witkoskie JB, Yang S, Cao J. Phys. Rev. E. 2002; 66:051111.
13. Dietrich C, Yang B, Fujiwara T, Kusumi A, Jacobson K. Biophys. J. 2002; 82:274. [PubMed: 11751315]

14. Kusumi A, Nakada C, Ritchie K, Murase K, Suzuki K, Murakoshi H, Kasai RS, Kondo J, Fujiwara T. Annu. Rev. Biophys. Biomol. Struct. 2005; 34:351. [PubMed: 15869394]
15. Montiel D, Cang H, Yang H, Phys J. Chem. B. 2006; 110:19763.
16. Huet S, Karatekin E, Tran VS, Fanget I, Cribier S, Henry J-P. Biophys. J. 2006; 91:3542. [PubMed: 16891360]
17. Pinaud F, Michalet X, Iyer G, Margeat E, Moore H-P, Weiss S. Traffic. 2009; 10:691. [PubMed: 19416475]
18. Thompson RE, Larson DR, Webb WW. Biophys. J. 2002; 82:2775. [PubMed: 11964263]
19. Ram S, Ward ES, Ober RJ. Proc. Natl. Acad. Sci. USA. 2006; 103:4457. [PubMed: 16537357]
20. Berglund AJ. Phys. Rev. E. 2010; 82:011917.
21. Michalet X. Phys. Rev. E. 2010; 82:041914. 83, 059904(E) (2011).
22. Wong Y, Lin ZP, Ober RJ. IEEE Trans. Signal Process. 2011; 59:895. [PubMed: 24665193]
23. Michalet X, Lacoste TD, Weiss S. Methods. 2001; 25:87. [PubMed: 11559000]
24. Qian H, Sheetz MP, Elson EL. Biophys. J. 1991; 60:910. [PubMed: 1742458]
25. Schütz GJ, Schindler H, Schmidt T. Biophys. J. 1997; 73:1073. [PubMed: 9251823]
26. Matsuoka S, Shibata T, Ueda M. Biophys. J. 2009; 97:1115. [PubMed: 19686659]
27. Wong WP, Halvorsen K. Opt. Express. 2006; 14:12517. [PubMed: 19529687]
28. Savin T, Doyle PS. Phys. Rev. E. 2007; 76:021501.
29. Goulian M, Simon SM. Biophys. J. 2000; 79:2188. [PubMed: 11023923]
30. Savin T, Doyle PS. Biophys. J. 2005; 88:623. [PubMed: 15533928]
31. Zacks, S. The Theory of Statistical Inference. New York: Wiley; 1971.
32. Kusumi A, Sako Y, Yamamoto M. Biophys. J. 1993; 65:2021. [PubMed: 8298032]
33. Saxton MJ. Biophys J. 1997; 72:1744. [PubMed: 9083678]
34. See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.85.061916 for a PDF file showing a comparison of the MLE and OLSF algorithms performance and a Matlab routine implementing both algorithms.
35. Saxton MJ. Biophys. J. 1993; 64:1766. [PubMed: 8369407]
36. Umemura YM, Vrljic M, Nishimura SY, Fujiwara TK, Suzuki KGN, Kusumi A. Biophys. J. 2008; 95:435. [PubMed: 18339737]
37. Marsaglia G, Stat J. Software. 2006; 16:1. http://www.jstatsoft.org/v16/i04/paper.
38. Jain AK. IEEE Trans. Commun. 1976; 24:1023.
39. Wang Z. IEEE Trans. Acoust. 1984; 32:803.

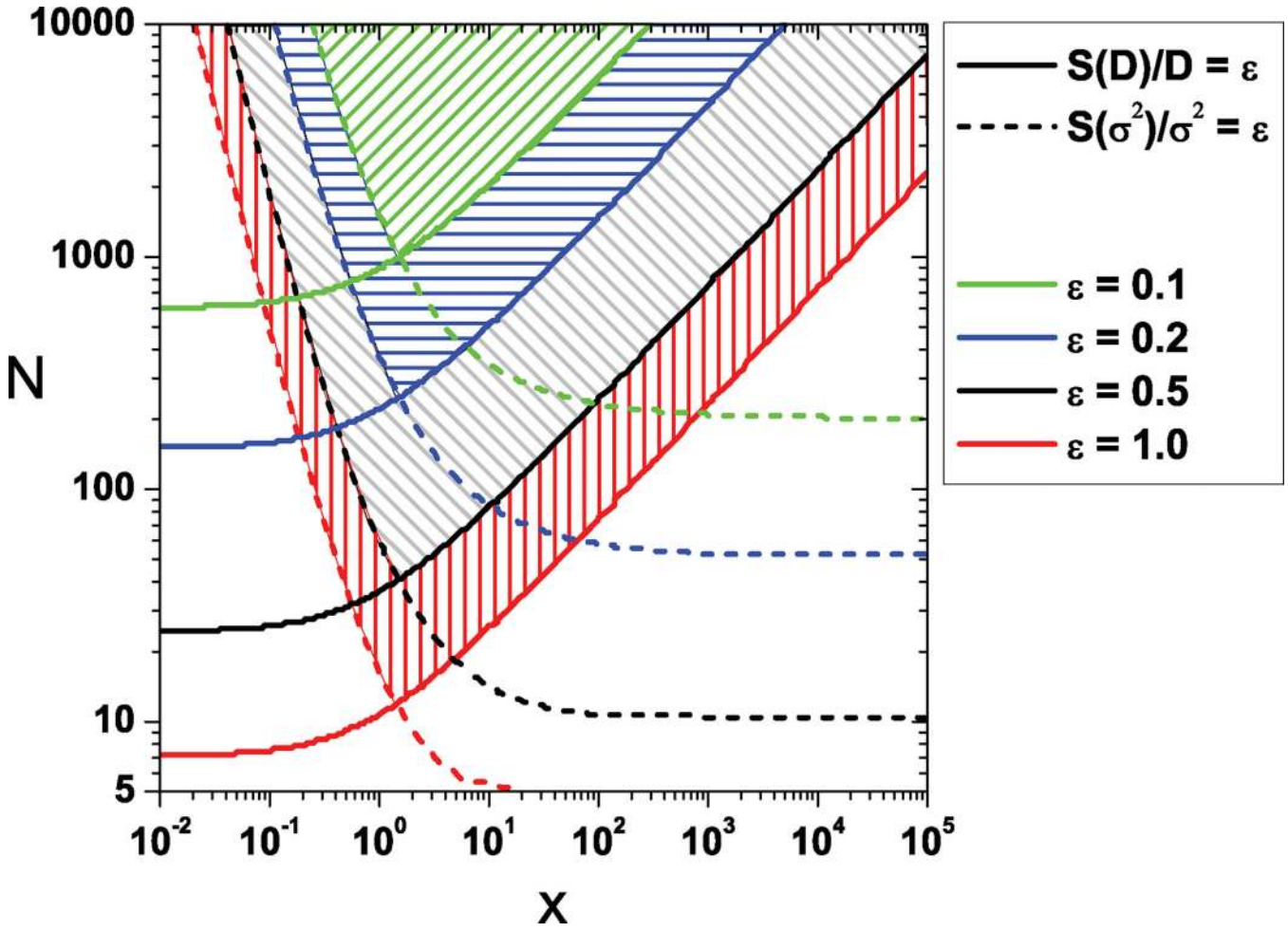

FIG. 1.

(Color online) Contours of constant Cramér-Rao lower bounds ε [Eqs. (7) and (8)] for the relative uncertainty on D (plain curves) or σ^2 (dashed curves), for $R = 0$ and $d = 1$. From top to bottom, ε = 0.1, 0.2, 0.5, and 1. Pairs of values (x, N) located above a given contour have a lower relative uncertainty than the value defined by this contour (reported in the legend). For instance, pairs in the top green hatched region have both relative uncertainties on D and σ^2 smaller than 10%. By looking at the intercept of these contours with the N axis when $x \to 0$ (respectively, $x \to \infty$), the minimum number of trajectory points needed to achieve a desired level of precision for D (respectively, σ^2) can be determined. For instance, a minimum of 600 (respectively, 150 or 24) trajectory points is needed to obtain a 10% (respectively, 20% or 50%) uncertainty on D in the absence of localization error ($x = 0$). Significantly more trajectory points are needed if $x > 1$. S is the standard deviation. Note that values of N reported to the left need to be replaced by $1 + (N - 1)/d$ in $d > 1$ dimensions.

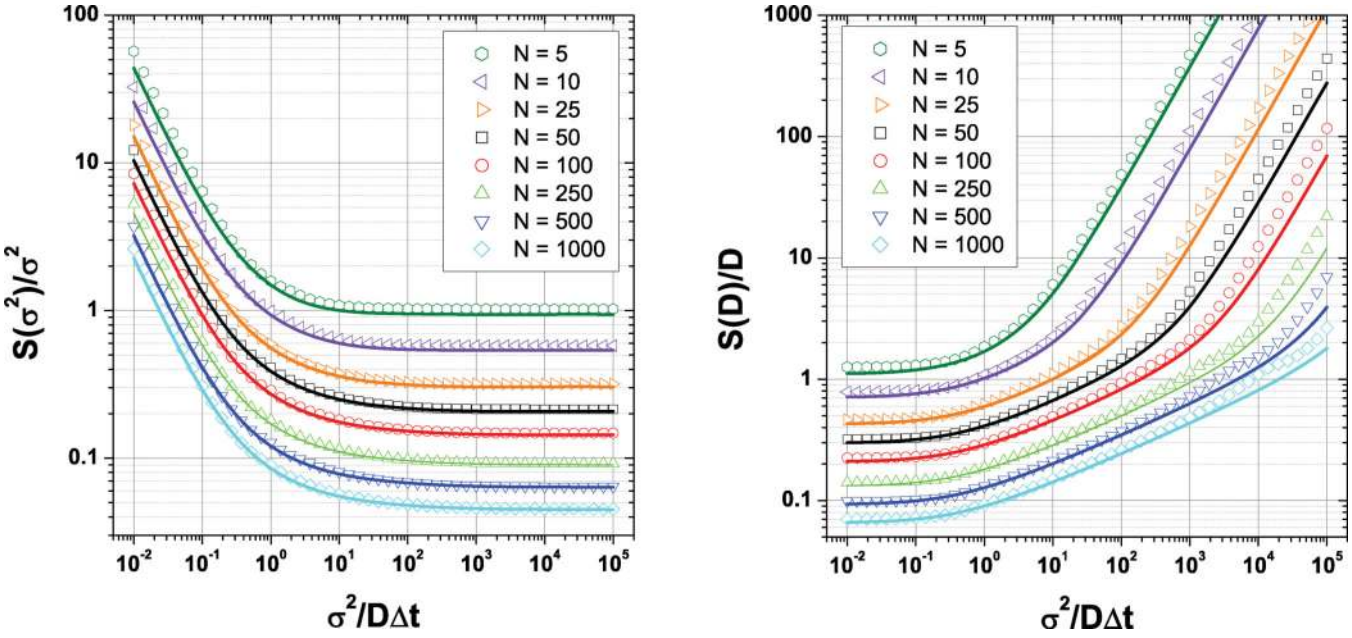

FIG. 2.

(Color online) Comparison between the relative error on OLSF fit parameters [open symbols, Eqs. (B11) and (B16)] and the corresponding Cramér-Rao lower bounds [plain curves, Eqs. (7) and (8)] for various number of trajectory points N, in the case $R = 1/6$ (camera integration during the whole frame duration) and $d = 1$. $N = 5$ to 1000 from top to bottom. There is a remarkable agreement between the two, demonstrating that OLSF provides close to optimal estimate of the parameters. S is the standard deviation.

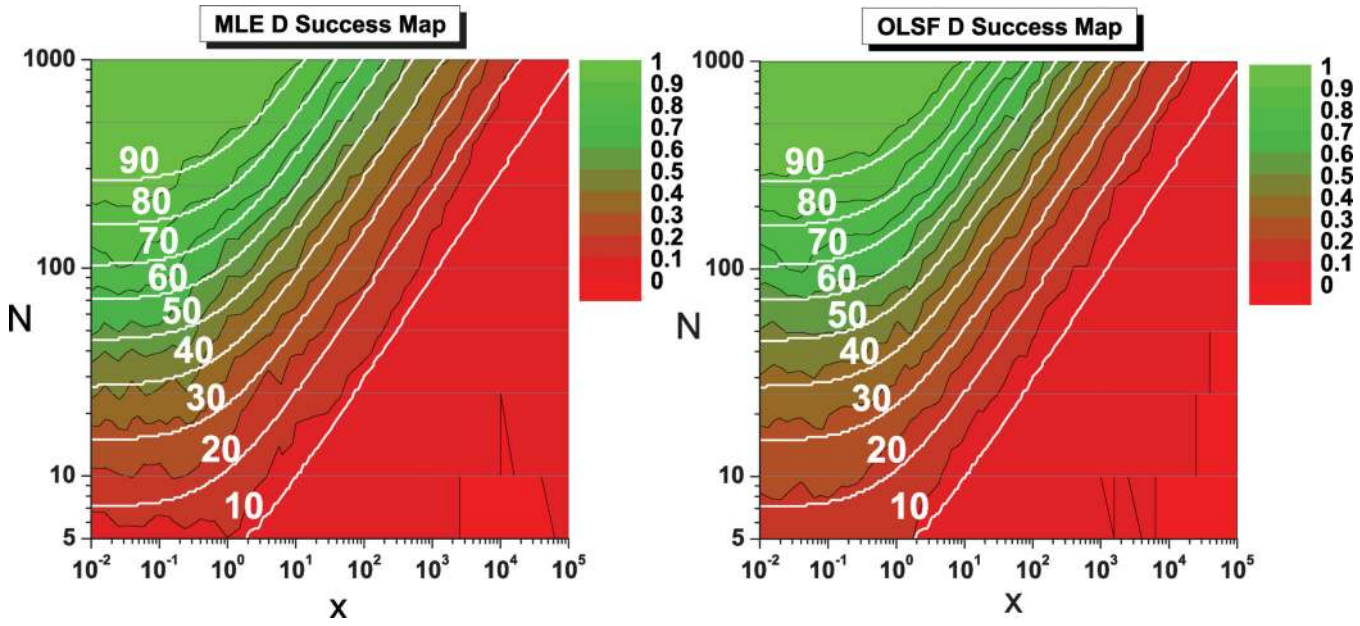

FIG. 3.

(Color online) Success maps for the MLE and OLSF algorithms for $d = 1$, showing the fraction of trajectories resulting in an estimated D value within 25% of the true value, as a function of the reduced localization error x and the number of points N. Green indicates 100% success, while red indicates failure. White curves are contours where an optimal unbiased estimator reaches success 10% to 90% of the time (from bottom to top). Note that these curves are different from (though related) to those represented in Fig. 1. As in Fig. 1, values of N reported to the left need to be replaced by $1 + (N - 1)/d$ in $d > 1$ dimensions.

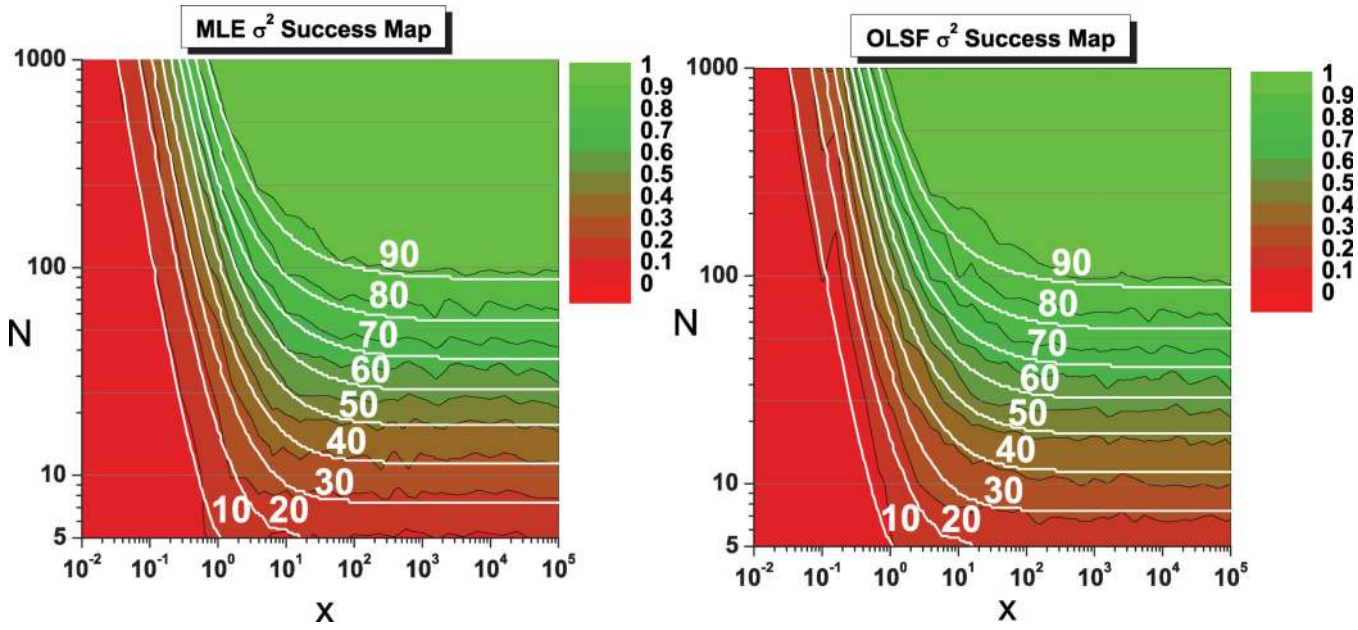

FIG. 4.

(Color online) Success map for the MLE and OLSF algorithms for $d = 1$, showing the fraction of trajectories resulting in an estimated σ^2 value within 25% of the true value, as a function of the reduced localization error x and the number of points N. White curves are contours where an optimal unbiased estimator reaches success 10% to 90% of the time (from bottom to top). The same comment as for Fig. 3 applies.

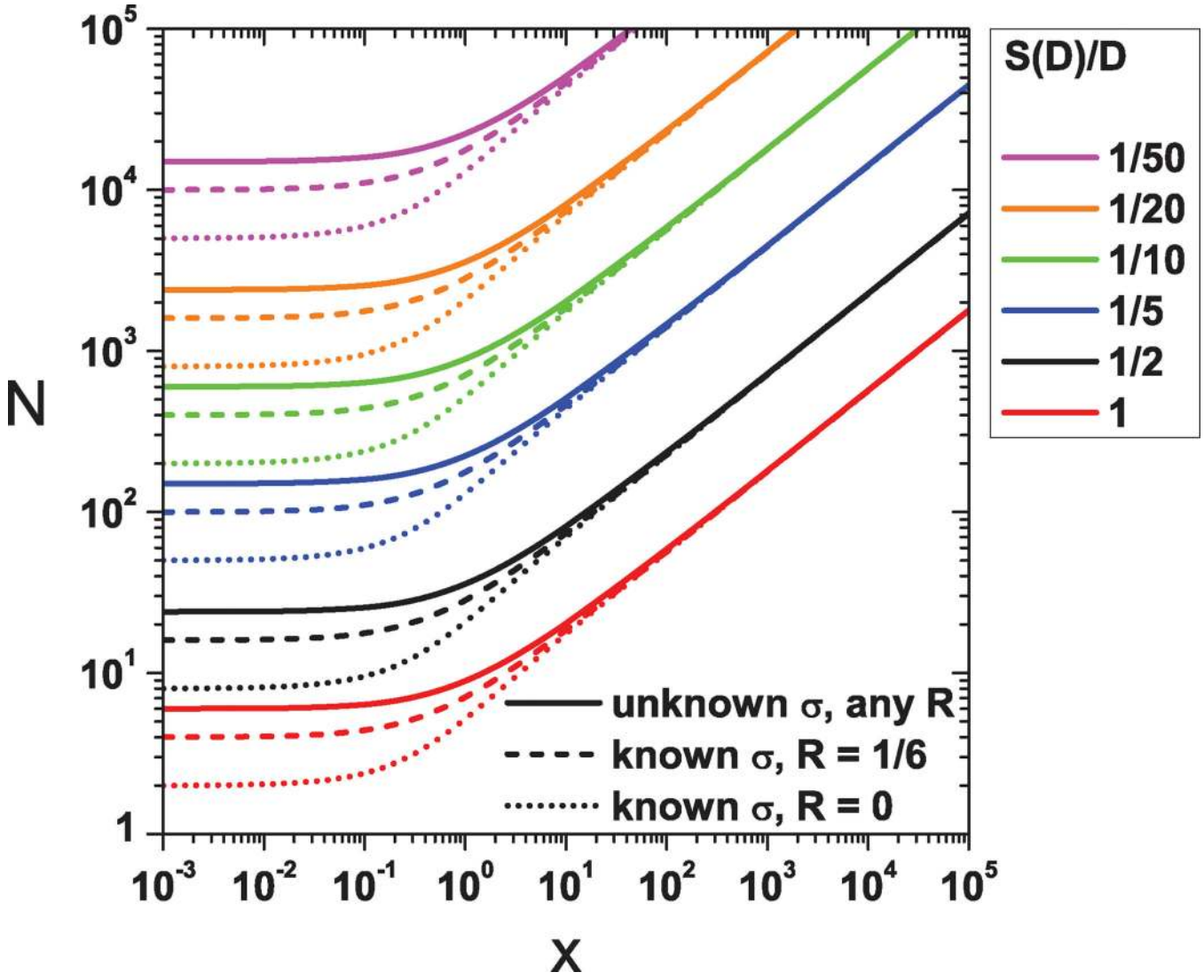

FIG. 5.

(Color online) Curves of constant Cramér-Rao lower bounds (CRLB) for the relative uncertainty on D for d=1 dimension in different experimental situations: unknown σ and arbitrary R (plain curves), known and uniform exposure (R = 1/6, dashed curves), and known σ and instantaneous exposure (R = 0, dotted curves). The first experimental situation requires more points to obtain a given uncertainty on D, but this disadvantage is negligible for large relative localization uncertainty. More precisely, noting N_k the number of trajectory points to achieve a given relative uncertainty on D when σ is known and N_{uk} that number when it is unknown, $N_{uk}/N_k(R = 0, x = 0) = 3$ and $N_{uk}/N_k(R = 1/6, x = 0) = 3/2$. Note that for $R > 0$, negative values of x can be obtained but are not represented here. S is the standard deviation.

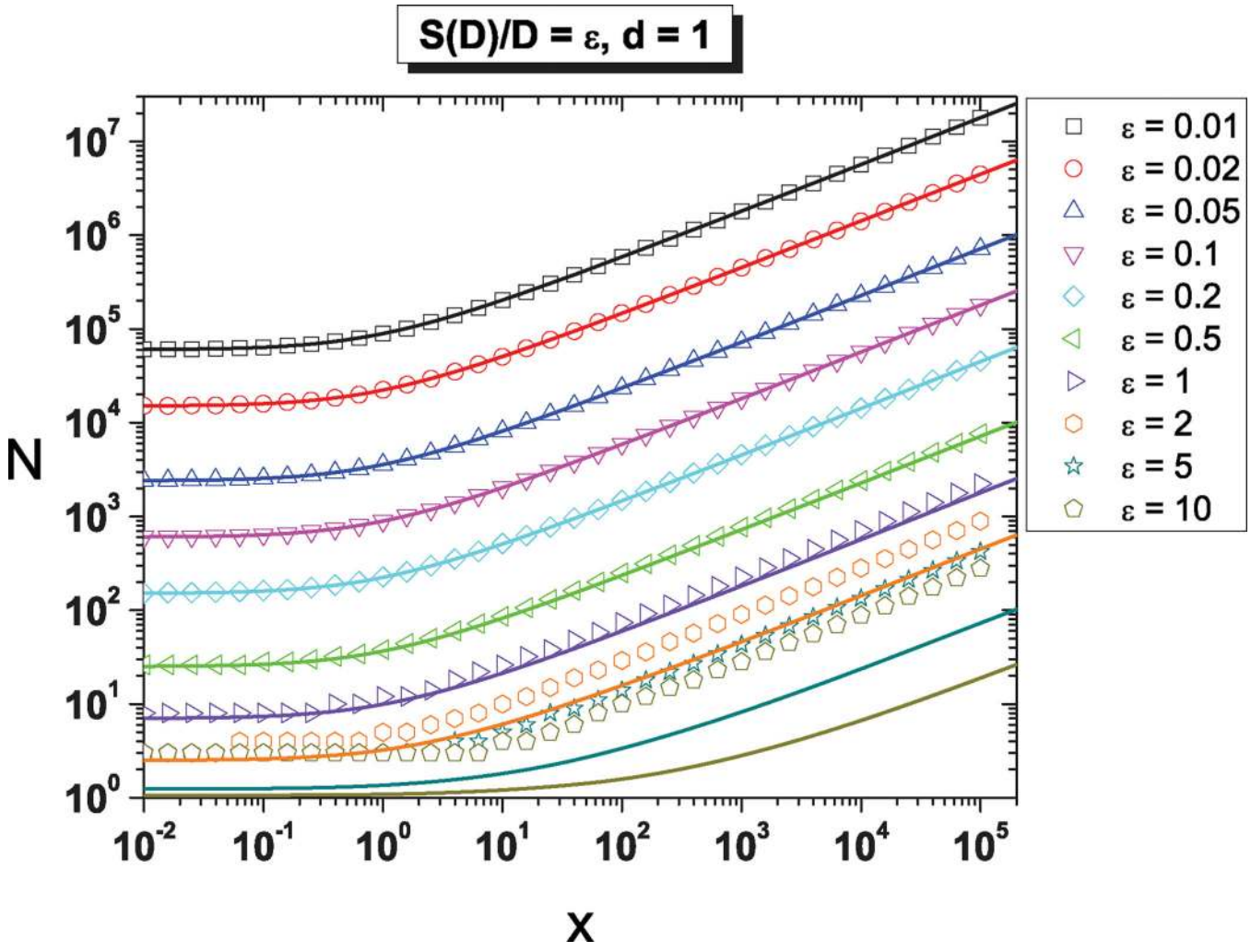

FIG. 6.

(Color online) Contours of constant Cramér-Rao lower bounds for the relative uncertainty on D, for $d = 1$. Open symbols correspond to the exact formula (7) [right-hand side of Eq. (7) = ε], whereas the plain curves correspond to the approximate formula [right-hand side of Eq. (12) = ε]. As in Fig. 1, values of N reported to the left need to be replaced by $1 + (N - 1)/d$ in $d > 1$ dimensions. S is the standard deviation.

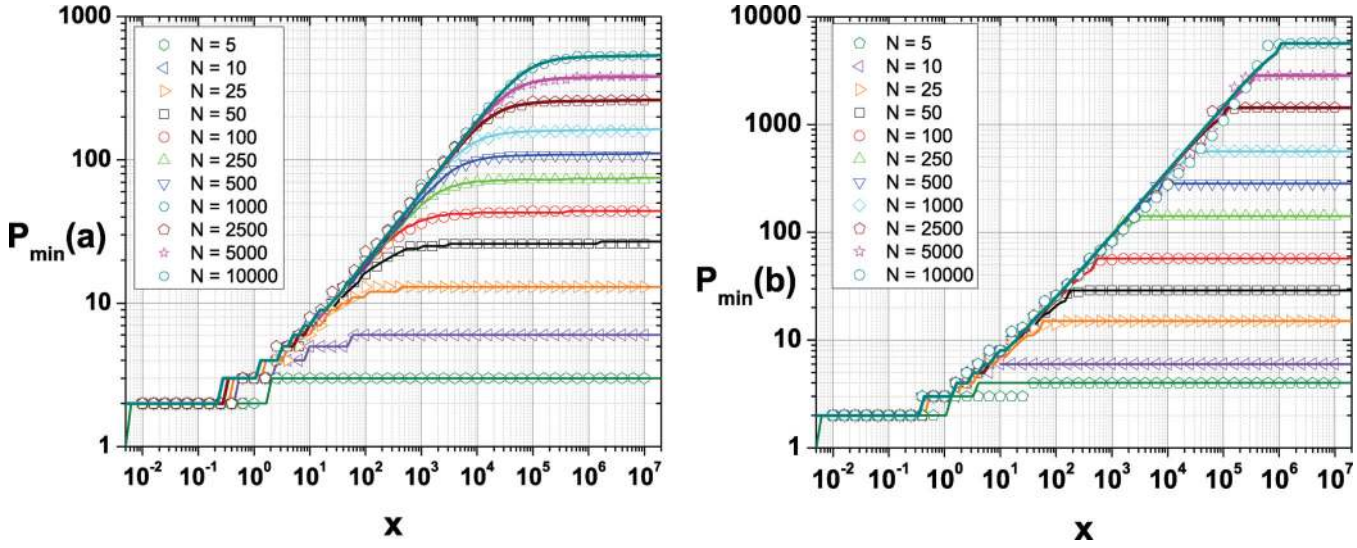

FIG. 7.

(Color online) Optimal number of points for least-square fitting of (left) the intercept a (i.e., σ^2) and (right) the slope b (i.e., D) of the MSD. Each curve corresponds to a different number of trajectory points N (N = 5 to 10000 from bottom to top). The plain curves correspond to Eq. (B3) (left) and Eq. (B4) (right). These results are valid for any value of the blur factor R. For $x < 0$, $p_{\min} = 2$.

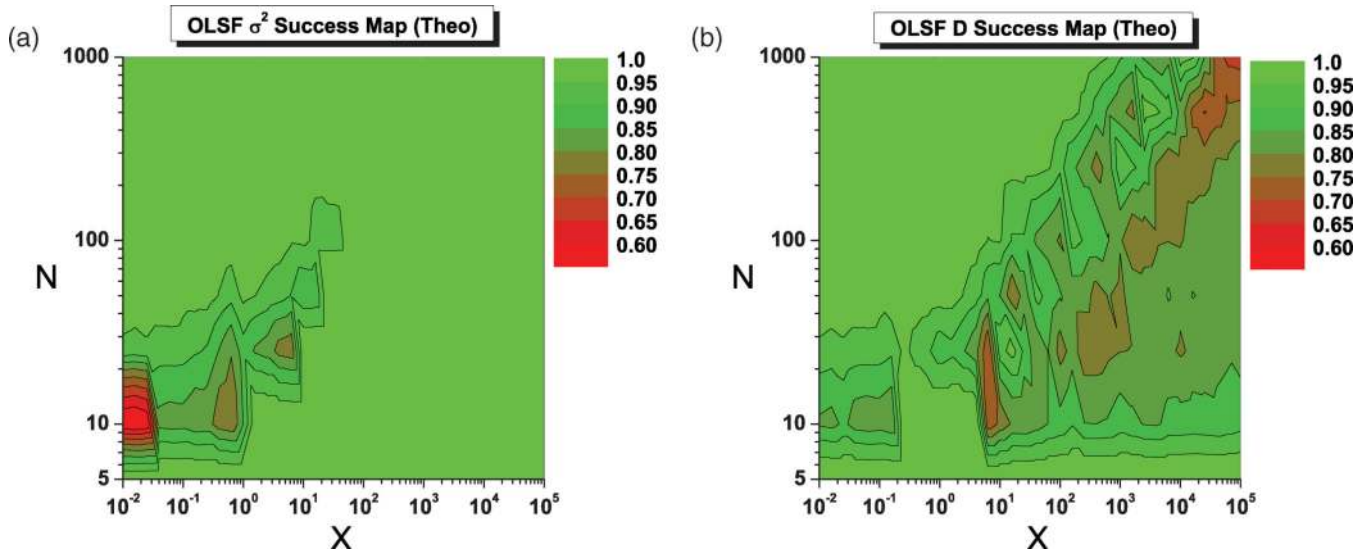

FIG. 8.

(Color online) Success rate of the OLSF algorithm in determining the optimal number of fitting point needed to optimally fit (a) σ^2 and (b) *D*.

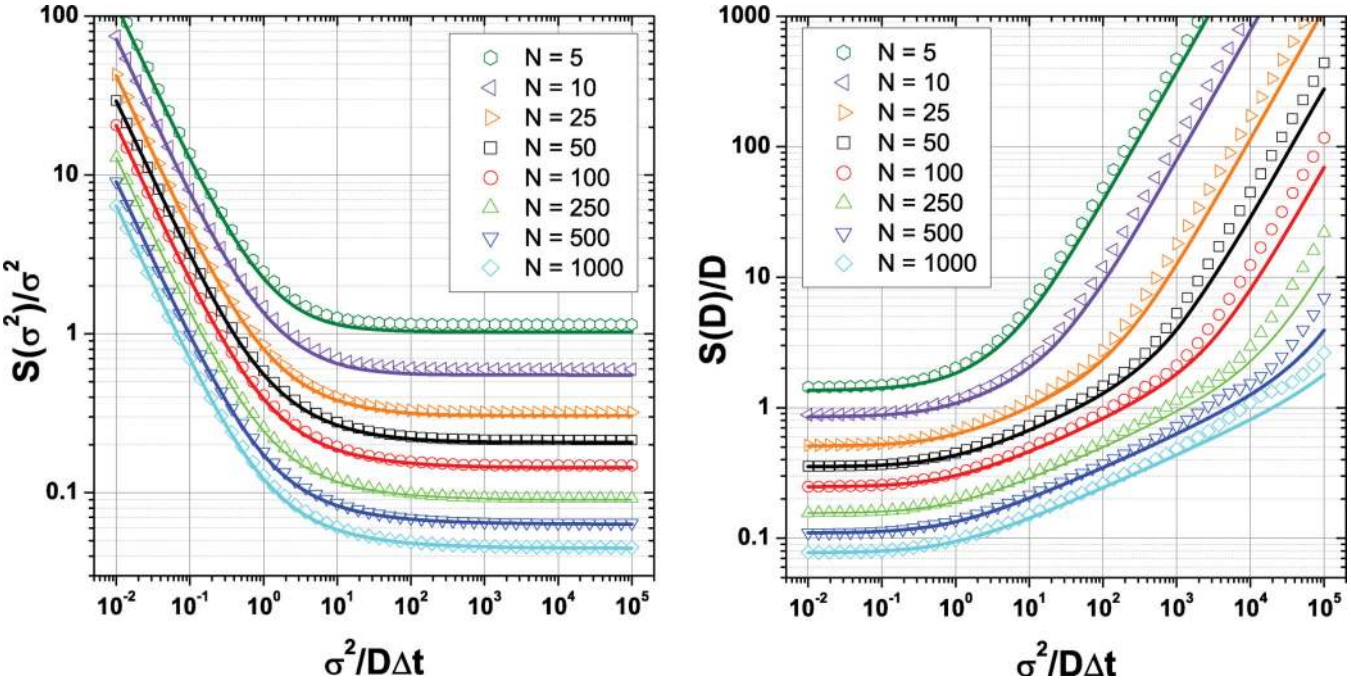

FIG. 9.

(Color online) Comparison between the relative error on OLSF fit parameters and the corresponding Cramér-Rao lower bounds [plain curves, Eqs. (8) and (7) of the main text] for various number of trajectory points N, in the case $R = 0$ (no camera integration) and $d = 1$. $N = 5$ to 1000 from top to bottom.