Inference of hopping rates of anisotropic random walk on a 2D lattice via covariance-based estimators of diffusion parameters

Masanori Mishima

Centre for Mechanochemical Cell Biology, Division of Biomedical Sciences, Warwick Medical School, University of Warwick

Abstract

Traditionally, time-development of the mean square displacement has been employed to determine the diffusion coefficient from the trajectories of single particles. However, this approach is sensitive to the noise and the motion blur upon image acquisition. Recently, Vestergaard et al. has proposed a novel method based on the covariance between the shifted displacement series. This approach gives a more robust estimator of the diffusion coefficient of one-dimensional diffusion without bias, i.e., when mean velocity is zero. Here, we extend this approach to a potentially biased random walk on a two-dimensional lattice. First, we describe the relationship between the hopping rates to the eight adjacent sites and the time development of the higher-order moments of the stochastic two-dimensional displacements. Then, we derive the covariance-based estimators for these higher-order moments. Numerical simulations confirmed that the procedure presented here allows inference of the stochastic hopping rates from two-dimensional trajectory data with location error and motion blur.
I. INTRODUCTION

The forward and backward hopping rates of a random walk on a one-dimensional lattice are linked to the average velocity ($v$) and diffusion coefficient ($D$). The latter macroscopic parameters can be estimated from the trajectory of the random walk, i.e., locations on the lattice measured with a regular time interval. Fitting of the mean square displacements (MSD) vs time lag data with $MSD = \sigma^2 + 2D\tau + (v\tau)^2$, where $\sigma$ is the localization noise, has been widely used to estimate the macroscopic parameters. However, it is known that there are practical issues such as the optimization of the range of the time lags and the influence of motion blurs when this approach is applied to the single particle dynamics of a biological molecule [1–3]. Recently, a novel approach based on the covariance between the displacements between adjacent time points (covariance-based estimator, CVE) has been
proposed to be a more simple, accurate and robust alternative to estimate the diffusion coefficient of a unbiased random walk [4–6].

Here, we extend the CVE-based approach to the random walk on a two-dimensional (2D) lattice, in which a particle takes stochastic hops to the eight surrounding sites with distinct rates. We start with the well-known relation between velocity and the diffusion coefficient and the 1D hopping rates and the CVE of the diffusion coefficient in this case. We then show that in a 2D random walk, x- and y-velocities and diffusion coefficients plus four (higher-order) co-moments of the observed two-dimensional displacement series are linked to the eight hopping rates. The procedure for calculating the CVEs for the macroscopic parameters, i.e., the co-moments of the 2D displacements, is provided. This allows us to infer the eight hopping rates from the trajectories of a 2D random walk even with temporal and spatial resolutions at which individual hopping events can’t be captured.

II. RESULTS

A. CVE-based approach for 1D diffusion with drift

Here we consider a potentially biased random walk $X(t)$ on a 1D lattice with the grid size, $a$ (Fig. 1a). Let $k_{+1}$ and $k_{-1}$ be the stochastic forward and backward hopping rates, respectively (we can assume $k_{+1} > k_{-1} > 0$ without loss of generality). $X(t)$ is the sum of $x_i$, the displacements in a small fraction of time $\delta t = t/n$, i.e., $X(t) = x_1 + x_2 + \ldots + x_n$.

Since $x_i$ are indistinguishable but independent of each other, the ensemble averages of $x_i$

$$\langle x_1 \rangle = \langle x_2 \rangle = \ldots = \langle x_n \rangle \equiv \langle x \rangle$$
$$\langle x_1^2 \rangle = \langle x_2^2 \rangle = \ldots = \langle x_n^2 \rangle \equiv \langle x^2 \rangle$$

where $\langle x \rangle$ and $\langle x^2 \rangle$ are the mean displacement and the mean square displacement of $X(t)$ in $\delta t$, respectively. In general, $\langle x_i^2 \rangle \neq (\langle x_i \rangle)^2$, while for $i \neq j$, $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$. The ensemble average of $X(t)$ is

$$\langle X(t) \rangle = \langle x_1 + x_2 + \ldots + x_n \rangle = n \langle x \rangle$$
$$= n \left[ (+a) \times k_{+1} \delta t + (-a) \times k_{-1} \delta t \right] = a(k_{+1} - k_{-1}) \cdot n \delta t$$
$$= a(k_{+1} - k_{-1}) t$$
FIG. 1. Hopping on 1D (a) and 2D (b) lattices. The gray dot represents the current position on the lattice. It takes a stochastic hop to the 2 and 8 nearby sites at the indicated rates ($k_{+1}$ and $k_{-1}$ on the 1D lattice (a) and $k_1 \sim k_8$ on the 2D lattice (b)). $k_F = k_1 + k_2 + k_3$, $k_B = k_6 + k_7 + k_8$, $k_L = k_1 + k_4 + k_6$, and $k_R = k_3 + k_5 + k_8$ are the rates of hopping towards forward, backward, left and right, respectively.

With $\langle X(t)^2 \rangle = \langle (x_1 + x_2 + \ldots + x_n)^2 \rangle = \langle \sum_i x_i^2 + \sum_j \sum_{i \neq j} x_i x_j \rangle = n \langle x^2 \rangle + n(n-1)\langle x \rangle^2$, the variance of $X(t)$ is

$$\text{var}(X(t)) = \langle (X(t) - \langle X(t) \rangle)^2 \rangle = \langle X(t)^2 \rangle - \langle X(t) \rangle^2$$

$$= n\langle x^2 \rangle + n(n-1)\langle x \rangle^2 - (n\langle x \rangle)^2 = n\langle x^2 \rangle - n\langle x \rangle^2$$

$$= n [(+a)^2 \times k_{+1} \delta t + (-a)^2 \times k_{-1} \delta t] - n [(+a) \times k_{+1} \delta t + (-a) \times k_{-1} \delta t]^2$$

$$= a^2(k_{+1} + k_{-1}) \cdot n \delta t - a^2(k_{+1} - k_{-1})^2 (n \delta t) \delta t$$

$$\rightarrow a^2(k_{+1} + k_{-1}) t \quad (\delta t \rightarrow 0)$$

Thus, the velocity of the constant drift and the diffusion coefficient, $v$ and $D$, respectively, are linked to the hopping rates

$$v = a(k_{+1} - k_{-1}) \quad (1)$$

$$2D = a^2(k_{+1} + k_{-1}) \quad (2)$$

As is well known, this indicates that we can infer the stochastic hopping rates by determining the macroscopic parameters, $v$ and $D$. Hereafter, for simplicity, we omit $a$ by assuming that $X(t)$ is a dimensionless value measured with $a$ as the unit, i.e., the physical location on the
lattice is $X(t) \cdot a$. Thus,

\begin{align*}
v &= k_{t+1} - k_{t-1} \\
2D &= k_{t+1} + k_{t-1}
\end{align*}

(3) (4)

In typical single molecule/particle observations, we can’t determine the true coordinate $X(t)$ as a continuous function. We can only measure the positions at a discrete time points $t = 0, \Delta t, 2\Delta t, \ldots, n\Delta t$ and the measured values $X_0, X_1, X_2, \ldots X_n$ suffer from the motion blur due to the movement during image capturing and the error in determining the position of the molecule/particle by image analysis. Under this circumstance, $X_k$ is related to the true $X(t)$ as

\[ X_k = \int_0^{\Delta t} s(t)X(t + (k-1)\Delta t)dt + \varepsilon_k \]

(5)

where $s(t)$ defines the state of the shutter ($s(t) = 0$ means closed shutter and $s(t) > 0$ means open shutter, for normalization $\int_0^{\Delta t} s(t)dt = 1$) and $\varepsilon_k$ the Gaussian error in localization by image analysis ($\langle \varepsilon_i \rangle = 0$ and $\langle \varepsilon_i^2 \rangle = \sigma^2$ represents the precision of the measurement when the target is immobile) \[2\]. The traditional approaches to determine $v$ and $D$, for example, by fitting a quadratic curve to the mean square displacement data, suffer from complications arising from non-zero $s(t)$ and $\varepsilon$. However, for unbiased diffusion (i.e. $v = 0$), it has recently been shown \[4\] that a combination of the adjacent displacements $\Delta X_k = X_{k+1} - X_k$ and $\Delta X_{k+1} = X_{k+2} - X_{k+1}$ cancels out the terms containing $s(t)$ and $\varepsilon$ and results in a simple relation

\[ \langle \Delta X_k^2 \rangle + 2\langle \Delta X_{k+1} \cdot \Delta X_k \rangle = 2D\Delta t. \]

(6)

$D$ calculated with this relation provides a more reliable estimator of the diffusion coefficients.

The above relation doesn’t hold in the presence of a bias in the hopping rates. However, if we consider $Z(t) = X(t) - vt$, the deviation from the constant drift, and its observed counterpart $Z_k = X_k - v \cdot k\Delta t$, we obtain $\langle Z_k \rangle = 0$, and thus the series of its displacements $\Delta Z_k = Z_{k+1} - Z_k = \Delta X_k - v\Delta t$ satisfy

\[ \langle \Delta Z_k^2 \rangle + 2\langle \Delta Z_{k+1} \cdot \Delta Z_k \rangle = 2D\Delta t. \]

(7)

Since $\langle (X(t) - \langle X(t) \rangle)^2 \rangle = \langle (Z(t) + vt - \langle Z(t) + vt \rangle)^2 \rangle = \langle (Z(t) - \langle Z(t) \rangle)^2 \rangle$, the value of $D$
determined with this formula gives the diffusion coefficient of the biased random walk $X(t)$.

B. Biased random walk on a 2D lattice: hopping rates and co-moments of the displacements

Here we discuss a biased random walk $(X(t), Y(t))$ on a 2D lattice measured with the x- and y-grid size $a$ and $b$, respectively. Hopping can occur to the eight surrounding sites with the distinct rates $k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8$. Let the velocities and diffusion coefficients of the movement along the grid axes be $v_x$ and $v_y$, and $D_x$ and $D_y$, respectively. Then

$$\langle X(t) \rangle = v_xt = (k_F - k_B)t = (k_1 + k_2 + k_3 - k_6 - k_7 - k_8)t$$
$$\langle (X(t) - \langle X(t) \rangle)^2 \rangle = 2D_xt = (k_F + k_B)t = (k_1 + k_2 + k_3 + k_6 + k_7 + k_8)t$$
$$\langle Y(t) \rangle = v_yt = (k_L - k_R)t = (k_1 + k_4 + k_6 - k_3 - k_5 - k_8)t$$
$$\langle (Y(t) - \langle Y(t) \rangle)^2 \rangle = 2D_yt = (k_L + k_R)t = (k_1 + k_4 + k_6 + k_3 + k_5 + k_8)t$$

as we discussed in the above section.

Here, we consider the covariance between $X(t)$ and $Y(t)$.

$$\text{cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$$
$$= \langle (x_1 + x_2 + \ldots + x_n)(y_1 + y_2 + \ldots + y_n) \rangle - \langle x_1 + x_2 + \ldots + x_n \rangle \langle y_1 + y_2 + \ldots + y_n \rangle$$
$$= \sum_{i=1}^{n} x_i y_i + \sum_{i \neq j} x_i y_j - n \langle x \rangle \cdot n \langle y \rangle$$
$$= n \langle xy \rangle + n(n - 1)\langle x \rangle \langle y \rangle - n^2 \langle x \rangle \langle y \rangle$$
$$= n \langle xy \rangle - n \langle x \rangle \langle y \rangle$$
$$= n \cdot \left[ (+1)(+1) \cdot k_1 \delta t + (-1)(+1) \cdot k_3 \delta t + (+1)(-1) \cdot k_6 \delta t + (-1)(-1) \cdot k_8 \delta t \right]$$
$$- n \cdot (k_F - k_B) \delta t \cdot (k_L - k_R) \delta t$$
$$= (k_1 - k_3 - k_6 + k_8)n \delta t - (k_F - k_B)(k_L - k_R)n \delta t \cdot \delta t$$
$$\to (k_1 - k_3 - k_6 + k_8)t \cdot (\delta t \to 0)$$
We move on to higher-order moments. Since

\[
\langle X^2Y \rangle = \left< \sum_i \sum_j \sum_k x_i x_j y_k \right>
\]

\[
= n \langle x^2y \rangle + n(n-1)\langle x^2 \rangle \langle y \rangle + 2n(n-1)\langle x \rangle \langle xy \rangle + n(n-1)(n-2)\langle x^2 \rangle \langle y \rangle
\]

we get

\[
\langle (X - \langle X \rangle)^2 (Y - \langle Y \rangle) \rangle = \langle X^2Y \rangle - \langle X^2 \rangle \langle Y \rangle - 2\langle X \rangle \langle XY \rangle - \langle X \rangle \langle Y \rangle
\]

\[
= n \langle x^2y \rangle - n\langle x^2 \rangle \langle y \rangle - 2n\langle x \rangle \langle xy \rangle + 2n\langle x^2 \rangle \langle y \rangle
\]

\[
= (k_1 - k_3 + k_6 - k_8)n\delta t - \left[ (k_F + k_B)(k_L - k_R) + 2(k_F - k_B)(k_1 - k_3 - k_6 + k_8) \right] n\delta t \cdot \delta t
\]

\[
+ 2(k_F - k_B)^2(k_L - k_R)n\delta t \cdot \delta t^2
\]

\[
\rightarrow (k_1 - k_3 + k_6 - k_8)t \quad (\delta t \rightarrow 0).
\]

Similarly,

\[
\langle (X - \langle X \rangle)(Y - \langle Y \rangle)^2 \rangle \rightarrow (k_1 + k_3 - k_6 - k_8)t \quad (\delta t \rightarrow 0)
\]

Finally, using

\[
\langle X^2Y^2 \rangle = \left< \sum_i \sum_k \sum_j \sum_l x_i x_j y_k y_l \right>
\]

\[
= n \langle x^2y^2 \rangle + 2n(n-1)\langle x \rangle \langle xy^2 \rangle + 2n(n-1)\langle y \rangle \langle x^2y \rangle + 2n(n-1)\langle xy \rangle \langle y \rangle
\]

\[
+ n(n-1)\langle x^2 \rangle \langle y^2 \rangle + n(n-1)\langle x \rangle \langle y \rangle \langle x^2 \rangle \langle y \rangle + 4n(n-1)(n-2)\langle x \rangle \langle y \rangle \langle x \rangle \langle y \rangle
\]

\[
+ n(n-1)(n-2)\langle x \rangle \langle y \rangle \langle x \rangle \langle y \rangle + n(n-1)(n-2)(n-3)\langle x \rangle \langle y \rangle \langle x \rangle \langle y \rangle
\]

we get

\[
\langle (X - \langle X \rangle)^2 (Y - \langle Y \rangle)^2 \rangle
\]

\[
= \langle X^2Y^2 \rangle - 2\langle X^2Y \rangle \langle Y \rangle - 2\langle XY^2 \rangle \langle X \rangle + \langle X^2 \rangle \langle Y^2 \rangle + 4\langle XY \rangle \langle X \rangle \langle Y \rangle + \langle Y^2 \rangle \langle X \rangle \langle Y \rangle - 3\langle X \rangle^2 \langle Y \rangle^2
\]

\[
= n \langle x^2y^2 \rangle + 2n(n-1)\langle x \rangle \langle xy^2 \rangle + n(n-1)\langle x^2 \rangle \langle y^2 \rangle
\]

\[
- 2n \left[ \langle x^2 \rangle \langle y \rangle + \langle x^2y \rangle \langle x \rangle \right] - n(n-2) \left[ \langle x^2 \rangle \langle y \rangle^2 + 4\langle xy \rangle \langle x \rangle \langle y \rangle + \langle y^2 \rangle \langle x \rangle \langle y \rangle - 3\langle x \rangle^2 \langle y \rangle^2 \right]
\]

\[
\rightarrow (k_1 + k_3 + k_6 + k_8)t + 2(At)^2 + 2D_x t \cdot 2D_y t \quad (\delta t \rightarrow 0)
\]

In summary, we have obtained a set of formulae that relate macroscopic observations to
the microscopic hopping rates:

\[ \langle X \rangle = v_x t \quad (8) \]
\[ \langle Y \rangle = v_y t \quad (9) \]
\[ \langle (X - \langle X \rangle)^2 \rangle = 2D_x t \quad (10) \]
\[ \langle (Y - \langle Y \rangle)^2 \rangle = 2D_y t \quad (11) \]
\[ \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = At \quad (12) \]
\[ \langle (X - \langle X \rangle)^2 (Y - \langle Y \rangle)^2 \rangle = Bt \quad (13) \]
\[ \langle (X - \langle X \rangle)(Y - \langle Y \rangle)^2 \rangle = Ct \quad (14) \]
\[ \langle (X - \langle X \rangle)^2 (Y - \langle Y \rangle)^2 \rangle - 2\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle^2 \]
\[ -\langle (X - \langle X \rangle)^2 \rangle \langle (Y - \langle Y \rangle)^2 \rangle = Et \quad (15) \]

where

\[
\begin{pmatrix}
  v_x \\
v_y \\
2D_x \\
2D_y \\
A \\
B \\
C \\
E
\end{pmatrix}
= \begin{pmatrix}
  1 & 1 & 1 & 0 & 0 & -1 & -1 & -1 \\
  1 & 0 & -1 & 1 & -1 & 1 & 1 & -1 \\
  1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\
  1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\
  1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\
  1 & 0 & -1 & 0 & 0 & 1 & 0 & -1 \\
  1 & 0 & 1 & 0 & 0 & -1 & 0 & -1 \\
  1 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
k_1 \\
k_2 \\
k_3 \\
k_4 \\
k_5 \\
k_6 \\
k_7 \\
k_8
\end{pmatrix}
\]

(16)

C. Covariance-based estimator for \( A \)

Based on formula 8 and 9, the velocities of constant drift, \( v_x \) and \( v_y \), can be estimated with

\[ \langle \Delta X \rangle = v_x \Delta t \quad (17) \]
\[ \langle \Delta Y \rangle = v_y \Delta t \quad (18) \]

As discussed in section A, drift-adjusted displacements \( \Delta Z_k = \Delta X_k - v_x \Delta t \) and \( \Delta W_k = \Delta Y_k - v_y \Delta t \) based on \( Z(t) = X(t) - v_x t \) and \( W(t) = Y(t) - v_y t \) give the diffusion coefficients
FIG. 2. Visual presentation of the relationships between the macroscopic parameters \(v_x, v_y, D_x, D_y, A, B, C\) and \(E\), which are the proportional coefficients of the time development of \(\langle X \rangle\) and \(\langle Y \rangle\) (\(v_x\) and \(v_y\)) or the moments of the drift-adjusted displacements, \(Z = X - v_x t, W = Y - v_y t\) (\(D_x, D_y, A, B, C\) and \(E\))\(^{[8-15]}\), and the microscopic hopping rates \((k_1, k_2, \ldots, k_8)\) on a 2D lattice. The red and blue colors on the grid indicate the signs (positive and negative, respectively) of the rates of hopping from the current position (gray) to the corresponding sites in assembling the macroscopic parameters \(^{[16]}\).

Here we derive analogous covariance-based estimators of \(A, B, C,\) and \(E\). Considering the symmetry between \(Z(t)\) and \(W(t)\), we guess the formula for \(A\) as

\[
\sum_{\alpha, \beta = 0 \text{ or } 1 \not\alpha = \beta = 1} \langle \Delta Z_{k+\alpha} \Delta W_{k+\beta} \rangle = \langle \Delta Z_k \Delta W_k \rangle + \langle \Delta Z_{k+1} \Delta W_k \rangle + \langle \Delta Z_k \Delta W_{k+1} \rangle = A \Delta t
\]

We shall now prove this.
With

$$\Delta Z_{k+\alpha} = \int_0^{\Delta t} s(t) \left[ Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right] dt + (\varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha})$$

$$\Delta W_{k+\beta} = \int_0^{\Delta t} s(t') \left[ W(t' + (k + \beta)\Delta t) - W(t' + (k + \beta - 1)\Delta t) \right] dt' + (\varepsilon'_{k+\beta+1} - \varepsilon'_{k+\beta})$$

$$\langle \Delta Z_{k+\alpha} \Delta W_{k+\beta} \rangle = \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t') \left\langle \left( Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right) \times \left( W(t' + (k + \beta)\Delta t) - W(t' + (k + \beta - 1)\Delta t) \right) \right\rangle dt dt' + \langle (\varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha})(\varepsilon'_{k+\beta+1} - \varepsilon'_{k+\beta}) \rangle$$

Since \((Z(t), W(t))\) is a random walk that fulfills \(\langle Z(t)W(t) \rangle = At, \langle Z(t)W(t') \rangle = A \cdot \min(t, t')\).

Using this, the factor to be integrated in the first term of \(\langle \Delta Z_{k+\alpha} \Delta W_{k+\beta} \rangle\) is evaluated to be

$$\langle \left( Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right) \times \left( W(t' + (k + \beta)\Delta t) - W(t' + (k + \beta - 1)\Delta t) \right) \rangle$$

$$= \langle Z(t + (k + \alpha)\Delta t)W(t' + (k + \beta)\Delta t) \rangle - \langle Z(t + (k + \alpha)\Delta t)W(t' + (k + \beta - 1)\Delta t) \rangle$$

$$- \langle Z(t + (k + \alpha - 1)\Delta t)W(t' + (k + \beta)\Delta t) \rangle + \langle Z(t + (k + \alpha - 1)\Delta t)W(t' + (k + \beta - 1)\Delta t) \rangle$$

$$= A \left[ \{k \Delta t + \min(t + \alpha \Delta t, t' + \beta \Delta t)\} - \{k \Delta t + \min(t + \alpha \Delta t, t' + (\beta - 1)\Delta t)\} \right]$$

$$\quad - \{k \Delta t + \min(t + (\alpha - 1) \Delta t, t' + \beta \Delta t)\} + \{(k - 1) \Delta t + \min(t + \alpha \Delta t, t' + \beta \Delta t)\} \right]$$

$$= A \left[ \min(t + \alpha \Delta t, t' + \beta \Delta t) - \min(t + \alpha \Delta t, t' + (\beta - 1)\Delta t) \right.$$

$$\left. - \min(t + (\alpha - 1) \Delta t, t' + \beta \Delta t) + \min(t + \alpha \Delta t, t' + \beta \Delta t) - \Delta t \right]$$

Here, considering \(0 \leq t \leq \Delta t, 0 \leq t' \leq \Delta t\), the factor within the above brackets

$$g(\alpha, \beta, t, t') = \min(t + \alpha \Delta t, t' + \beta \Delta t) - \min(t + (\alpha - 1) \Delta t, t' + \beta \Delta t)$$

$$- \min(t + (\alpha - 1) \Delta t, t' + \beta \Delta t) + \min(t + \alpha \Delta t, t' + \beta \Delta t) - \Delta t$$
is calculated to be

\[ g(0, 0, t, t') = \Delta t - (t + t' - 2 \cdot \min(t, t')) \]
\[ g(1, 0, t, t') = t' - \min(t, t') \]
\[ g(0, 1, t, t') = t - \min(t, t') \]

Thus, \( g(0, 0, t, t') + g(1, 0, t, t') + g(0, 1, t, t') = \Delta t \).

The x- and y-localization errors at the same time point are not independent. With Kronecker’s delta \( \delta(\alpha, \beta) = 1 \) (\( \alpha = \beta \)) or 0 (\( \alpha \neq \beta \)),

\[
\langle (\varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha}) (\varepsilon_{k+\beta+1}' - \varepsilon_{k+\beta}') \rangle \\
= \langle \varepsilon_{k+\alpha+1} \varepsilon_{k+\beta+1}' \rangle - \langle \varepsilon_{k+\alpha} \varepsilon_{k+\beta}' \rangle - \langle \varepsilon_{k+\alpha} \varepsilon_{k+\beta+1}' \rangle + \langle \varepsilon_{k+\alpha} \varepsilon_{k+\beta}' \rangle \\
= \sigma_{xy}^2 (\delta(\alpha, \beta) - \delta(\alpha + 1, \beta) - \delta(\alpha, \beta + 1) + \delta(\alpha, \beta)) \\
= \sigma_{xy}^2 (2\delta(\alpha, \beta) - \delta(\alpha + 1, \beta) - \delta(\alpha, \beta + 1)) = \sigma_{xy}^2 \Omega(\alpha, \beta),
\]

where we define \( \Omega(\alpha, \beta) = 2\delta(\alpha, \beta) - \delta(\alpha + 1, \beta) - \delta(\alpha, \beta + 1) \). Thus, \( \Omega(0, 0) + \Omega(0, 1) + \Omega(1, 0) = (2 \cdot 1 - 0 - 0) + (2 \cdot 0 - 1 - 0) + (2 \cdot 0 - 0 - 1) = 0 \).

Finally, we get

\[
\sum_{\alpha, \beta=0 \text{ or } 1} \langle \Delta Z_{k+\alpha} \Delta W_{k+\beta} \rangle = \int_0^{\Delta t} \int_0^{\Delta t} s(t) s(t') A \cdot \left[ g(0, 0, t, t') + g(1, 0, t, t') + g(0, 1, t, t') \right] dt dt' \\
+ \sigma_{xy}^2 \left[ \Omega(0, 0) + \Omega(1, 0) + \Omega(1, 1) \right] \\
= A \Delta t \
\]

\[(21)\]

### D. Covariance-based estimators for \( B \) and \( C \)

The above results imply that for random walks \( U(s), V(t), \ldots \), if a function \( f(x, y, \ldots) \) exists such that \( f(U(s), V(t), \ldots) = K \cdot \min(s, t, \ldots) \) for some constant \( K \) (for example, for \( f(x, y) = \langle xy \rangle, f(X(s), X(t)) = 2D_x \cdot \min(s, t) \), and \( f(X(s), Y(t)) = A \cdot \min(s, t) \)), then the
equation below holds

\[
\sum_{\alpha, \beta, \ldots = 0 \text{ or } 1 \text{ not } \alpha = \beta = \ldots = 1} f(\Delta U_{k+\alpha}, \Delta V_{k+\beta}, \ldots) = K \Delta t \tag{22}
\]

where \(\Delta U_k = U_{k+1} - U_k\), \(\Delta V_k = V_{k+1} - V_k\) ... are the observed displacements (adjusted for the constant drift), providing the general recipe for the CVE for \(K\). Below is a proof for the cases of \(B\) and \(C\).

Since \(\langle Z(t)^2W(t) \rangle = Bt\), we choose \(f(x, y, z) = \langle xyz \rangle\).

\[
f(\Delta Z_{k+\alpha}, \Delta Z_{k+\beta}, \Delta W_{k+\gamma}) = \langle \Delta Z_{k+\alpha} \cdot \Delta Z_{k+\beta} \cdot \Delta W_{k+\gamma} \rangle
\]

\[
= \left( \int_0^{\Delta t} s(t) \left[ Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right] dt + (\varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha}) \right)
\times \left( \int_0^{\Delta t} s(t') \left[ Z(t' + (k + \beta)\Delta t) - Z(t' + (k + \beta - 1)\Delta t) \right] dt' + (\varepsilon_{k+\beta+1} - \varepsilon_{k+\beta}) \right)
\times \left( \int_0^{\Delta t} s(t'') \left[ W(t'' + (k + \gamma)\Delta t) - W(t'' + (k + \gamma - 1)\Delta t) \right] dt'' + (\delta_{k+\gamma+1} - \delta_{k+\gamma}) \right)
\]

\[
= \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t')s(t'')
\left[ Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right] \times \left[ Z(t' + (k + \beta)\Delta t) - Z(t' + (k + \beta - 1)\Delta t) \right]
\times \left[ W(t'' + (k + \gamma)\Delta t) - W(t'' + (k + \gamma - 1)\Delta t) \right]dt \, dt' \, dt''
\]

in which we used \(\langle Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \rangle = 0\), etc. and \(\langle \varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha} \rangle = 0\), etc.

For unbiased random walks \(A(s), B(t), C(u)\) (\(\langle A(s) \rangle = \langle B(t) \rangle = \langle C(u) \rangle = 0\), and their displacements are independent for non-overlapping time sections), if we assume \(s < t < u\),

\[
\langle A(s)B(t)C(u) \rangle = \langle A(s) \rangle \langle (B(t) - B(s))(C(t) - C(s)) \rangle + \langle A(s)C(s) \rangle \langle (B(t) - B(s)) \rangle + \langle A(s)B(s) \rangle \langle (C(t) - C(s)) \rangle + \langle A(s)B(s)C(s) \rangle = \langle A(s)B(s)C(s) \rangle \tag{23}
\]

In general, with \(\tau = \min(s, t, u)\), \(\langle A(s)B(t)C(u) \rangle = \langle A(\tau)B(\tau)C(\tau) \rangle\). Applying this to \(Z(t)\) and \(W(t)\) that satisfy \(\langle Z(t)^2W(t) \rangle = Bt\), we get \(\langle Z(t)Z(t')W(t'') \rangle = B \cdot \min(t, t', t'')\). Using
this, we evaluate

\[
\left\langle \left[ Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t) \right] \times \left[ Z(t' + (k + \beta)\Delta t) - Z(t' + (k + \beta - 1)\Delta t) \right] \right. \\
\left. \times \left[ W(t'' + (k + \gamma)\Delta t) - W(t'' + (k + \gamma - 1)\Delta t) \right] \right\rangle
\]

\[= \langle Z(t + (k + \alpha \Delta t)) \cdot Z(t' + (k + \beta \Delta t)) \cdot W(t'' + (k + \gamma \Delta t)) \rangle \\
- \langle Z(t + (k + \alpha \Delta t)) \cdot Z(t' + (k + \beta \Delta t)) \cdot W(t'' + (k + \gamma - 1)\Delta t) \rangle \\
- \langle Z(t + (k + \alpha \Delta t)) \cdot Z(t' + (k + \beta - 1)\Delta t) \cdot W(t'' + (k + \gamma \Delta t)) \rangle \\
+ \langle Z(t + (k + \alpha \Delta t)) \cdot Z(t' + (k + \beta - 1)\Delta t) \cdot W(t'' + (k + \gamma - 1)\Delta t) \rangle \\
- \langle Z(t + (k + \alpha - 1)\Delta t) \cdot Z(t' + (k + \beta \Delta t)) \cdot W(t'' + (k + \gamma \Delta t)) \rangle \\
+ \langle Z(t + (k + \alpha - 1)\Delta t) \cdot Z(t' + (k + \beta \Delta t)) \cdot W(t'' + (k + \gamma - 1)\Delta t) \rangle \\
+ \langle Z(t + (k + \alpha - 1)\Delta t) \cdot Z(t' + (k + \beta - 1)\Delta t) \cdot W(t'' + (k + \gamma \Delta t)) \rangle \\
- \langle Z(t + (k + \alpha - 1)\Delta t) \cdot Z(t' + (k + \beta - 1)\Delta t) \cdot W(t'' + (k + \gamma - 1)\Delta t) \rangle
\]

\[= B \cdot g(\alpha, \beta, \gamma, t, t', t''),\]

where

\[g(\alpha, \beta, \gamma, t, t', t'')\]

\[= \min(t + \alpha \Delta t, t' + \beta \Delta t, t'' + \gamma \Delta t) - \min(t + \alpha \Delta t, t' + \beta \Delta t, t'' + (\gamma - 1)\Delta t) \\
- \min(t + (\alpha - 1)\Delta t, t' + (\beta - 1)\Delta t, t'' + \gamma \Delta t) + \min(t + (\alpha - 1)\Delta t, t' + (\beta - 1)\Delta t, t'' + (\gamma - 1)\Delta t) \\
- \min(t + (\alpha - 1)\Delta t, t' + \beta \Delta t, t'' + \gamma \Delta t) + \min(t + (\alpha - 1)\Delta t, t' + \beta \Delta t, t'' + (\gamma - 1)\Delta t) \\
+ \min(t + (\alpha - 1)\Delta t, t' + (\beta - 1)\Delta t, t'' + \gamma \Delta t) - \min(t + (\alpha - 1)\Delta t, t' + (\beta - 1)\Delta t, t'' + (\gamma - 1)\Delta t).
\]

g(\alpha, \beta, \gamma, t, t', t'') can be calculated to be

\[g(0, 0, 0, t, t', t'') = \Delta t - (t + t' + t'') + \min(t', t'') + \min(t, t') + \min(t, t'')\]
\[g(1, 0, 0, t, t', t'') = \min(t', t'') - \min(t, t', t'')\]
\[g(0, 1, 0, t, t', t'') = \min(t, t'') - \min(t, t', t'')\]
\[g(0, 0, 1, t, t', t'') = \min(t, t') - \min(t, t', t'')\]
\[g(1, 1, 0, t, t', t'') = t'' - \min(t, t'') - \min(t', t'') + \min(t, t', t'')\]
\[g(0, 1, 1, t, t', t'') = t' - \min(t, t') - \min(t', t'') + \min(t, t', t'')\]
\[g(0, 1, 1, t, t', t'') = t - \min(t, t') - \min(t, t'') + \min(t, t', t'').\]
Thus,

$$
\sum_{\alpha, \beta, \gamma = 0 \text{ or } 1, \alpha \neq \beta \neq \gamma = 1} g(\alpha, \beta, \gamma, t, t', t'') = \Delta t
$$

The error term

$$
\Omega(\alpha, \beta, \gamma) = \left( (\varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha})(\varepsilon_{k+\beta+1} - \varepsilon_{k+\beta})(\delta_{k+\gamma+1} - \delta_{k+\gamma}) \right)
$$

is calculated to be

$$
\Omega(0, 0, 0) = 0
$$

$$
\Omega(1, 0, 0) = \Omega(0, 1, 0) = \Omega(0, 0, 1) = -\langle \varepsilon_k^2 \delta_k \rangle
$$

$$
\Omega(0, 1, 1) = \Omega(1, 0, 1) = \Omega(1, 1, 0) = \langle \varepsilon_k^2 \delta_k \rangle.
$$

Thus,

$$
\sum_{\alpha, \beta, \gamma = 0 \text{ or } 1, \alpha \neq \beta \neq \gamma = 1} \Omega(\alpha, \beta, \gamma) = 0
$$

Finally, we obtain

$$
\sum_{\alpha, \beta, \gamma = 0 \text{ or } 1, \alpha \neq \beta = \gamma = 1} \langle \Delta Z_{k+\alpha} \cdot \Delta Z_{k+\beta} \cdot \Delta W_{k+\gamma} \rangle
$$

$$
= \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t')s(t'')B \Delta t dt dt' dt'' = B\Delta t. \quad (24)
$$

Similarly,

$$
\sum_{\alpha, \beta, \gamma = 0 \text{ or } 1, \alpha \neq \beta = \gamma = 1} \langle \Delta Z_{k+\alpha} \cdot \Delta W_{k+\beta} \cdot \Delta W_{k+\gamma} \rangle = C\Delta t \quad (25)
$$

E. Covariance-based estimators for $E$

$Z$ and $W$ are linked to $E$ by

$$
\langle Z(t)^2 W(t)^2 \rangle - 2\langle Z(t) W(t) \rangle^2 - \langle Z(t)^2 \rangle \langle W(t)^2 \rangle = Et
$$
With \( f(x, y, z, w) = \langle xyzw \rangle - \langle xy \rangle \langle zw \rangle - \langle xz \rangle \langle yw \rangle - \langle xw \rangle \langle yz \rangle \), this relation is written

\[
f(Z(t), Z(t), W(t), W(t)) = Et
\]

Thus, the relation that gives the covariance-based estimator of \( E \) is predicted to be

\[
\sum_{\alpha, \beta, \gamma, \lambda = 0 \text{ or } 1 \text{ not } \alpha = \beta = \gamma = \lambda = 1} f(\Delta Z_{k+\alpha}, \Delta Z_{k+\beta}, \Delta W_{k+\gamma}, \Delta W_{k+\lambda}) = E\Delta t \quad (26)
\]

This can be proven by some algebra similar to the above proofs for the CVEs of \( A, B \) and \( C \) (see appendix A for details), confirming that the recipe for constructing CVEs (22) holds for the necessary cases to infer the anisotropic hopping rates based on the relation (16) (Fig. 2).

**F. Inference of the hopping rates from the 2D trajectories simulated with motion blur and localization error**

We now examine the versatility of the covariance-based estimators of \( D_x, D_y, A, B, C, \) and \( E \) (formulas 19, 20, 21, 24, 25, and 26), the coefficients of the time development of the co-moments of the displacements, and whether we can infer the anisotropic hopping rates from the trajectory data of 2D random walks, using the relations that link the macroscopic coefficients to the microscopic hopping rates (formula 16). Let us first compare two examples of random walks, RW1 and RW2, respectively, that are generated by Monte Carlo simulations to have the same \( x \)- and \( y \)-velocities and diffusion coefficients \((v_x, v_y, D_x, D_y)\), but have different hopping rates (Fig. 3 a and b). The values of the velocities, \( v_x = 16 \text{ s}^{-1} \) and \( v_y = 1 \text{ s}^{-1} \) (hops per second) correspond, for example, to the helical motion of a kinesin-like motor protein with a helical pitch of \( \sim 1.7 \mu \text{m} \) around a microtubule that has a 2D lattice of the discrete motor-binding sites on its surface consisting of parallel 13 protofilaments, i.e., linear arrays of tubulin subunits aligned at 8 \( \text{nm} \) periodicity [7, 8]. To mimic a typical condition of image acquisition using an EM-CCD camera used for the single molecule/particle observation, a series of true positions are generated by the Gillespie algorithm [9] (blue lines in Fig. 3 c and d) and the observed positions at regular time points (every 0.1 s for 200 time points, corresponding to observation for 20 sec) were calculated by averaging the positions.
during the open shutter (90% of the cycle) and by adding Gaussian noises of the standard deviation of half the size of the lattice spacing (red lines in Fig. 3 c and d).

(a, b) Two random walks, RW1 and RW2, with the same x- and y- velocities and diffusion coefficients were generated based on the different sets of hopping rates. 

(c, d) Examples of simulated ‘true’ positions on a 2D lattice (blue) and the ‘observed’ trajectories (red) were shown. 

(e) Summary of the macroscopic parameters for the simulated RW1 and RW2 (magenta and green, respectively). For each random walk, 50 trajectories each consisting of 200 observations of the 0.1 s interval were simulated, and the CVEs of the parameters were calculated for each trajectory (gray dots). Means and standard errors are indicated in magenta and green. Black segments indicate the values expected by the theory (formula (16)).

Fifty trajectories of RW1 and RW2 were generated, respectively, and the macroscopic parameters $v_x, v_y, D_x, D_y, A, B, C,$ and $E$ for each trajectory were calculated based on the formulas 17, 18, 19, 20, 21, 24, 25, and 26. As shown in Fig. 3 e, the covariance-based estimators calculated from the simulated trajectories (gray dots) distributed around the theoretical values predicted by the formula (16) (solid lines). Although, due to the intrinsic
stochasticity of the random walk, the calculated values from individual trajectories spread around the theoretical ‘true’ values, the means of 50 trajectories were found close to the corresponding theoretical values within twice the standard error. Importantly, as expected, while there was no significant difference between RW1 and RW2 in $v_x$, $v_y$, $D_x$, and $D_y$, a clear difference between them was found in the distribution of $A$. This suggests that our recipe gives reasonable estimates of the coefficients of the temporal development of the co-moments of the x- and y-displacements in a realistic scenario. This allows us to detect a difference between the two random walks that have distinct preferences in the hopping direction but look exactly the same if we only consider the velocities and the diffusion coefficients along the two axes, separately.

We move on to examine the power of our approach to reveal the hopping behaviors of anisotropic random walks. With known macroscopic parameters ($v_x$, $v_y$, $D_x$, $D_y$, $A$, $B$, $C$, $E$), we can calculate $(k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8)$, the hopping rates to the 8 surrounding sites, by solving by the formula (16). However, due to the intrinsically stochastic behavior of the random walks, a simple solution might result in negative values of the hopping rates. To avoid this, we performed a Bayesian inference with a model that the observed macroscopic parameters are probabilistic variables that distribute around the true values that are defined by the microscopic hopping rates, which don’t take negative values (Fig. 4 a). This was implemented with Stan [10] on R [11] with the overall hopping rate, $k$ as a positive real number and the hopping preferences $p = (p_1, p_3, p_4, p_5, p_6, p_7, p_8)$ as simplex variables ($0 \leq p_i \leq 1$, $\sum p_i = 1$), where the individual hopping rates $k_i = k \cdot p_i$ with $k = \sum_i k_i$. As priors, uniform probabilities were used.

Fig. 4 b shows the distributions of the posterior probabilities of the hopping rates, obtained with the fifty sets of macroscopic parameters ($v_x$, $v_y$, $D_x$, $D_y$, $A$, $B$, $C$, $E$) for each Brownian motion from Fig. 3 e as the data for the Bayesian inference. As expected, RW1 (magenta) and RW2 (green) showed clearly distinct patterns in the hopping preferences, which are consistent with the theoretical values used for generation of the random walks by simulation. For example, the (posterior) probability of the hopping rate to the forward-left, $k_1$, of RW1 showed a distribution near the theoretical value $0 \text{ s}^{-1}$ while that of RW2 had a peak near $2 \text{ s}^{-1}$. This indicates that our approach can properly infer the microscopic parameters to the precision levels that are sufficient to distinguish the two example cases of random walk, which would look the same if we only analyze the movements in x- and
y-directions separately.

![Diagram](image-url)

**FIG. 4.** Inference of the hopping rates and hopping preferences. 

(a) Procedure of Bayesian inference of the hopping rates and hopping preferences from the simulated trajectories. Covariance-based estimators of the macroscopic coefficients (D) calculated for individual trajectories were used as data to compute the posterior probabilities of the hopping rates (k) and hopping preferences (p), which are linked to each other via the overall hopping rate (k), based on a model that the observed macroscopic coefficients are probabilistic variables that distribute around the theoretical values, Mk, where M is the matrix that appears in formula 16.

(b) Posterior probability distributions of the hopping rates (RW1:magenta, RW2:green). Curves represent the results of the four independent chains of Bayesian inference. Dashed lines indicate the theoretical values. (c to e) Influences of the data size (c), localization errors (d) and the correlation between the x- and y-localization errors (e). Dots linked with solid lines and shaded regions represent the averages of the means and 95% credible intervals of 100 independent trials, respectively. Horizontal dashed lines are true values. Vertical dashed lines indicate the parameter values used in the simulation in (b). (f) Loss of the estimation accuracy by increased localization errors and its recovery by expansion of the data size (the number of the observations per trajectory or the number of the trajectories). Means and 95% credible intervals.

If we look closer, we realize that the peak positions of the posterior probabilities of the
hopping rates don’t exactly match with the theoretical values. This is likely due to the probabilistic uncertainties in the observed random walks, which are influenced by the length and the number of the trajectories and by the precision of the measurements (the motion blurs and localization errors), in combination with the natural condition that the hopping rates can’t be less than zero. To assess the effects of the measurement conditions, the cycle of the generation of the observed trajectories by simulation, estimation of the coefficients for each trajectory, and inference of the rates and preferences of hopping was repeated a hundred times per condition. The average of the means and 95% credible intervals of the posterior distribution of the 100 independent trials were averaged and shown in Fig. 4c to e and Supplemental Figures.

As expected, increasing the size of data by increasing the number of trajectories sharpened the distribution of the posterior probabilities of the rates and preferences of hopping (Fig. 4c and Supplementary Figure 1). Importantly, their means, which showed deviations from the theoretical values when the number of input trajectories was small as mentioned above, asymptotically approached their respective theoretical values as the available trajectories increase. Similar sharpening of the distribution and approaching to the theoretical value were observed also when the length of each trajectory was increased instead (Supplementary Figure 2).

The localization errors don’t explicitly appear in the formulas of the CVEs (17, 18, 19, 21, 24, 25, and 26) because we took averages of infinitely many combinations of the displacements. With the finite displacement data, however, the noise terms would not be completely canceled out and affect the estimations. Indeed, increasing the amplitude of the localization errors significantly broadens the posterior probability distributions of the hopping preferences (and the hopping rates) (Fig. 4d and Supplementary Figure 3). For example, with the data of the 50 trajectories with 200 observations each, although the $p_4$ values of RW1 and RW2 were distinguishable in the presence of the localization errors up to half the size of the lattice unit, they became indistinguishable when the amplitude of the errors was increased to be the same as the lattice unit (Fig. 4d). This is in contrast with little impact of the extent of correlation between the x- and y-measurement errors for each observation (Fig. 4e, Supplementary Figure 4). Interestingly, the perturbation by the localization errors could be overcome by increasing the size of data, either by increasing the number of observations per trajectory or by observing more trajectories (Fig. 4f).
The difference between RW1 and RW2 that was once obscured by the increased localization errors (1 grid size) became distinguishable again with a 10-fold increase in the amount of data either by increasing the number of observations made per trajectory or by increasing the number of trajectories observed.

Although the hopping rates of RW1 and RW2 were set to have identical theoretical values of $v_x, v_y, D_x,$ and $D_y$, their theoretical values of $A$ were different (0 vs $3 \text{s}^{-1}$). Thus, we can’t exclude the possibility that the above success in discriminating between the hopping patterns of RW1 and RW2 might solely rely on the distinct values of $A$, irrespective of the coefficients for the other co-moments. To test whether our approach can distinguish two random walks that have identical $v_x, v_y, D_x, D_y,$ and $A$, we finally consider another random walk, RW3 (Fig. 5 a), whose theoretical $A$, in addition to the theoretical values of $v_x, v_y, D_x, D_y,$ is identical to that of RW1. As expected, the coefficients $v_x, v_y, D_x, D_y,$ and $A$ calculated from the 50 RW3 trajectories each with 200 observations showed distributions indistinguishable from the corresponding ones from the RW1 trajectories (Fig. 5 b). In contrast, the distributions of the calculated $C$ and $E$ values of RW3 were distinct from those of RW1. As expected, the posterior probabilities of the hopping rates of RW3 closely reproduced the values set for the simulation of the trajectories, exhibiting a difference from the corresponding one of RW1 ($k_1, k_4, k_6,$ and $k_7$) (Fig. 5 c). This suggests that our approach provides reasonable inference of the hopping behaviors under a realistic setting for the single particle observation even in a case where $B, C$ or $E$, the coefficients for the higher-order co-moments of the drift-adjusted x- and y-displacements, are the sole observable clues.

III. DISCUSSION

Here we studied the anisotropic random walk on a 2D lattice and derived the relationship between the macroscopic coefficients of the time-development of the drift-adjusted displacements and the microscopic hopping rates. We then extended the covariance-based estimators of the 1D diffusion coefficient to the higher-order co-moments of the drift-adjusted displacements and used them to infer the hopping rates from the trajectory data affected by motion blur and localization errors. The versatility of this novel approach was evaluated with the trajectory data generated by simulation. We could demonstrate that our approach can distinguish 2D random walks that have exactly the same x- and y-velocities and diffusion
FIG. 5. Further test of our approach with RW3, a random walk with theoretical values of $v_x$, $v_y$, $D_x$, $D_y$, and $A$ identical to those of RW1. (a) Hopping rates of random walk, RW3, designed to result in the same theoretical values of $v_x$, $v_y$, $D_x$, $D_y$, and $A$ as those of RW1. (b) Macroscopic coefficients of RW1 and RW3 calculated from the 50 trajectories, respectively, each with 200 observations generated by Monte Carlo simulation. The means and standard errors (magenta and blue, respectively) are shown with the values for individual trajectories (gray dots). Black segments indicate the theoretical values by the formula (16). (c) Distributions of the posterior probabilities of the hopping rates inferred with the coefficients in (b). Curves represent the results of four independent chains performed for each random walk. Dashed vertical lines indicate the theoretical values.

An important advantage of the covariance-based estimators [4–6] is that the equations to calculate them don’t explicitly contain terms for localization errors nor motion blur after taking the average of infinite terms of observed x- and y-displacements combined. This assumption, of course, is not true with the real-world data of finite size. With a finite size of data, the failure in cancellation of the interfering factors as well as the intrinsic uncertainly of the random process results in an error in the estimation of the coefficients that characterize the random walk. Although it had been demonstrated that the CVE-based approach is superior to the traditional ones in the case of 1D unbiased diffusion [4–6], it was unclear how robust a 2D version of the CVE-based approach would be since it is involved with the
higher-order terms of the displacements. However, through the analysis of the simulated model cases of 2D random walk, it was confirmed that the CVE-based approach combined with Bayesian inference can reasonably estimate the anisotropic hopping patterns, with a realistic size of data when the precision of the measurement is smaller than the size of the grid. Even with a lower precision of the measurements, increasing the size of data could restore the accuracy of the estimation. For real-world data, careful optimization will be necessary, considering rather complicated and intertwined influences of the frequency of image acquisition, exposure time, the intensity of illumination [4, 12–15].

It has recently been reported that some kinesin-like motor proteins show a helical motion around a microtubule [16–19]. The tubular structure of a microtubule consists of 13 or 14 protofilaments, i.e., the linear arrays of α- and β-tubulin heterodimers with the 8 nm periodicity, presenting the motor-binding sites as the 2D lattice on the surface [7, 8]. Interestingly, the reported pitches of the helical are intermediate between the shortest helical pitch of the lattice due to the staggered alignment of the protofilaments and the longest helical pitch observed in the 14-protofilament tubule. The observed helical motion corresponds to the protofilament switch that occurs once per \( \sim 10 \) forward steps on average, implying the stochastic stepping patterns to the neighboring sites. Our approach might be applicable to estimate such patterns based on the observed helical trajectories.

On a 3D lattice, there are 26\((= 9 + 8 + 9)\) choices for stochastic hopping to a nearby site. The relation between the macroscopic coefficients and the hopping rates for a 3D lattice analogous to the 2D version (16) will contain 26 hopping rates and 26 coefficients for the time-development of the combinations of (higher-order co-)moments of the drift-adjusted displacements up to the 6th order. Whether the recipe to derive the covariance-based estimators of the coefficients for the higher-order co-moments developed here for a 2D lattice (22) is applicable to a 3D lattice will be a future question.

ACKNOWLEDGMENTS

The author would like to thank Huong T. Vu (University of Warwick, UK), Matthew Turner (University of Warwick, UK), Junichiro Yajima (University of Tokyo, Japan) and Izumu Mishima (Durham University, UK) for insightful discussions and critical reading of
the manuscript.

[1] X. Michalet, Phys Rev E Stat Nonlin Soft Matter Phys 82, 041914 (2010).
[2] A. J. Berglund, Phys Rev E Stat Nonlin Soft Matter Phys 82, 011917 (2010).
[3] X. Michalet and A. J. Berglund, Phys Rev E Stat Nonlin Soft Matter Phys 85, 061916 (2012).
[4] C. L. Vestergaard, Phys Rev E 94, 022401 (2016).
[5] C. L. Vestergaard, J. N. Pedersen, K. I. Mortensen, and H. Flyvbjerg, Eur Phys J Special Topics 224, 1151 (2015).
[6] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg, Phys Rev E Stat Nonlin Soft Matter Phys 89, 022726 (2014).
[7] D. Chrétien and R. H. Wade, Biol. Cell 71, 161 (1991).
[8] R. H. Wade and D. Chrétien, J. Struct. Biol. 110, 1 (1993).
[9] D. T. Gillespie, J Phys Chem 81, 2340 (1977).
[10] Stan Development Team, “RStan: the R interface to Stan,” (2020), r package version 2.21.1.
[11] R Core Team, R: A Language and Environment for Statistical Computing, Vienna, Austria (2019).
[12] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes, Chem. Rev. 117, 7331 (2017).
[13] N. Hozé and D. Holcman, Ann Rev Stat Appl 4, 189 (2017).
[14] C. Manzo and M. F. Garcia-Parajo, Rep Prog Phys 78, 124601 (2015).
[15] N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. Ortiz-de Solorzano, J.-C. Olivo-Marin, and E. Meijering, Nat Methods 11, 281 (2014).
[16] V. Bormuth, B. Nitzsche, F. Ruhnow, A. Mitra, M. Storch, B. Rammner, J. Howard, and S. Diez, Biophys J 103, L4 (2012).
[17] A. Mitra, M. Suñé, S. Diez, J. M. Sancho, D. Oriola, and J. Casademunt, Biophys J 116, 2266 (2019).
[18] M. Bugiel and E. Schäffer, Biophys J 115, 1993 (2018).

[19] A. Mitra, F. Ruhnow, S. Girardo, and S. Diez, Proceedings of the National Academy of Sciences 115, E7950 (2018).
Appendix A: Proof of the equation for the covariance-based estimator of $E$

Here we will prove Equation 26. Let us define notations, $\Delta \hat{Z}(t, k, \alpha) = Z(t + (k + \alpha)\Delta t) - Z(t + (k + \alpha - 1)\Delta t)$ and $\Delta \varepsilon_{k,\alpha} = \varepsilon_{k+\alpha+1} - \varepsilon_{k+\alpha}$. With these,

$$f(\Delta Z_{k+\alpha}, \Delta Z_{k+\beta}, \Delta W_{k+\gamma}, \Delta W_{k+\lambda})$$

$$= \left\langle \left[ \int_0^{\Delta t} s(t) \Delta \hat{Z}(t, k, \alpha) dt + \Delta \varepsilon_{k,\alpha} \right] \times \left[ \int_0^{\Delta t} s(t') \Delta \hat{Z}(t', k, \beta) dt' + \Delta \varepsilon_{k,\beta} \right] \times \left[ \int_0^{\Delta t} s(t'') \Delta \hat{W}(t'', k, \gamma) dt'' + \Delta \varepsilon_{k,\gamma} \right] \right\rangle$$

$$= \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t')s(t'') \times$$

$$f(\Delta \hat{Z}(t, k, \alpha), \Delta \hat{Z}(t', k, \beta), \Delta \hat{W}(t'', k, \gamma), \Delta \hat{W}(t'', k, \gamma)) dt dt' dt''$$

$$+ f(\Delta \varepsilon_{k,\alpha}, \Delta \varepsilon_{k,\beta}, \Delta \varepsilon_{k,\gamma}, \Delta \varepsilon_{k,\lambda})$$

(A1)

In the above transformation, we used that the true displacements and the measurement errors are independent of each other and that the ensemble means of the true displacements and those of the measurement errors are both zero.

Now we evaluate $f(\Delta \hat{Z}(t, k, \alpha), \Delta \hat{Z}(t', k, \beta), \Delta \hat{W}(t'', k, \gamma), \Delta \hat{W}(t'', k, \gamma))$ in the first term. For unbiased random walks $A(s), B(t), C(u), D(v)$ ($\langle A(s) \rangle = \langle B(t) \rangle = \langle C(u) \rangle = \langle D(v) \rangle = 0$, and their displacements are independent for non-overlapping time sections), we can show, by a similar calculation to [23], that

$$f(A(s), B(t), C(u), D(v)) = f(A(\tau), B(\tau), C(\tau), D(\tau))$$

(A2)
where $\tau = \min(s, t, u, v)$. Thus, with $\tau = \min(t + (k + \alpha - i)\Delta t, t' + (k + \beta - i')\Delta t, t'' + (k + \gamma - i'')\Delta t),$

$$f(\Delta Z(t, k, \alpha), \Delta Z(t', k, \beta), \Delta \hat{W}(t'', k, \gamma), \Delta \hat{W}(t'''', k, \lambda))$$

$$= \sum_{i=0}^{1} \sum_{i' = 0}^{1} \sum_{i'' = 0}^{1} \sum_{i'''}^{1} (-1)^{i + i' + i'' + i'''} \times$$

$$f(Z(t + (k + \alpha - i)\Delta t), Z(t' + (k + \beta - i')\Delta t), W(t'' + (k + \gamma - i'')\Delta t), W(t''' + (k + \lambda - i''')\Delta t))$$

$$= \sum_{i=0}^{1} \sum_{i' = 0}^{1} \sum_{i'' = 0}^{1} \sum_{i'''}^{1} (-1)^{i + i' + i'' + i'''} f(Z(\tau), Z(\tau), W(\tau), W(\tau))$$

$$= E \cdot \sum_{i=0}^{1} \sum_{i' = 0}^{1} \sum_{i'' = 0}^{1} \sum_{i'''}^{1} (-1)^{i + i' + i'' + i'''} \times$$

$$\min\left[+(k + \alpha - i)\Delta t, t' + (k + \beta - i')\Delta t, t'' + (k + \gamma - i'')\Delta t, t''' + (k + \lambda - i''')\Delta t\right]$$

Here, let us consider

$$g(\alpha, \beta, \gamma, \lambda, t, t', t'', t''') = \sum_{i=0}^{1} \sum_{i' = 0}^{1} \sum_{i'' = 0}^{1} \sum_{i'''}^{1} (-1)^{i + i' + i'' + i'''} \times$$

$$\min\left[+(k + \alpha - i)\Delta t, t' + (k + \beta - i')\Delta t, t'' + (k + \gamma - i'')\Delta t, t''' + (k + \lambda - i''')\Delta t\right].$$

The values of 15 combinations of $\alpha, \beta, \gamma,$ and $\lambda = 0$ or 1 (except for $\alpha = \beta = \gamma = \lambda = 1)$ are evaluated as

$$g(0, 0, 0, 0, t, t', t'', t''') = \Delta t - (t + t' + t'' + t''')$$

$$+ \min(t, t') + \min(t, t'') + \min(t, t''') + \min(t', t'') + \min(t', t''') + \min(t'', t''')$$

$$- \left[\min(t', t'', t''') + \min(t, t'', t''') + \min(t', t', t''') + \min(t', t', t''')\right] + 2 \cdot \min(t, t', t'', t''')$$

$$g(1, 0, 0, 0, t, t', t'', t''') = \min(t', t'', t''') - \min(t, t', t''', t''')$$

$$g(1, 1, 0, 0, t, t', t'', t''') = \min(t'', t''') - \min(t', t'', t''') - \min(t, t', t''') + \min(t, t', t''')$$

$$g(1, 1, 1, 0, t, t', t'', t''') = t''' - \left[\min(t, t''', t''') + \min(t, t', t''') + \min(t', t', t''')\right] + \min(t, t', t''') + \min(t, t'', t''') + \min(t', t', t''') - \min(t, t', t''')$$

$$\min(t, t', t''', t''') - \min(t, t', t''', t'''),$$

etc.

All the terms except for $\Delta t$ cancel out (Table. I).
TABLE I. Calculation of $g(\alpha, \beta, \gamma, \lambda, t, t', t'', t''')$. Counts of the consisting terms for various combinations of $\alpha, \beta, \gamma$, and $\lambda$.

| Term | $t, t'$, etc. | $\min(t, t')$, etc. | $\min(t, t', t'')$, etc. | $\min(t, t', t'', t''')$, etc. |
|------|---------------|----------------------|-----------------------------|---------------------------------|
| $g(0, 0, 0, 0, t, t', t''', t''')$ | $-1$ each 4 | $+1$ each 6 | $-1$ each 4 | $+2$ |
| $g(1, 0, 0, 0, t, t', t''', t''')$ etc. 4 | $+1$ each 4 | $+1$ each 4 | $-1$ each 4 | $-1$ each 4 |
| $g(1, 1, 0, 0, t, t', t''', t''')$ etc. 6 | $+1$ each 6 | $-2$ each 6 | $+1$ each 6 | $+1$ each 6 |
| $g(1, 1, 1, 0, t, t', t''', t''')$ etc. 4 | $+1$ each 4 | $-3$ each 4 | $+3$ each 4 | $-1$ each 4 |
| sum | $-4 + 4 = 0$ | $6 + 6 - 12 = 0$ | $-4 + 4 - 12 + 12 = 0$ | $+2 - 4 + 6 - 4 = 0$ |

Thus,

$$\sum_{\alpha, \beta, \gamma, \lambda \neq 0 \text{ or } 1 \atop \text{not } \alpha = \beta = \gamma = \lambda = 1} g(\alpha, \beta, \gamma, \lambda, t, t', t'', t''') = \Delta t$$

i.e.,

$$\sum_{\alpha, \beta, \gamma, \lambda \neq 0 \text{ or } 1 \atop \text{not } \alpha = \beta = \gamma = \lambda = 1} f(\Delta \tilde{Z}(t, k, \alpha), \Delta \tilde{Z}(t', k, \beta), \Delta \tilde{W}(t'', k, \gamma), \Delta \tilde{W}(t''', k, \lambda)) = E\Delta t \quad (A3)$$

The second term of $f(\Delta Z_{k+\alpha}, \Delta Z_{k+\beta}, \Delta W_{k+\gamma}, \Delta W_{k+\lambda})$ (the formula [A1]) can be decomposed into a sum of the terms with various combinations of $i, i', i'', i'''$. Most of the $15 \times 16$ terms are zero. As in Table. II, the other non-zero terms are cancelled out each other.

Thus,

$$\sum_{\alpha, \beta, \gamma, \lambda \neq 0 \text{ or } 1 \atop \text{not } \alpha = \beta = \gamma = \lambda = 1} f(\Delta \varepsilon_{k, \alpha}, \Delta \varepsilon_{k, \beta}, \Delta \delta_{k, \gamma}, \Delta \delta_{k, \lambda}) = 0 \quad (A4)$$
TABLE II. Calculation of \( f(\Delta \varepsilon_{k,\alpha}, \Delta \varepsilon_{k,\beta}, \Delta \delta_{k,\gamma}, \Delta \delta_{k,\lambda}) \). All the possible combinations of \( \alpha, \beta, \gamma, \) and \( \lambda \) (excluding \( \alpha = \beta = \gamma = \lambda = 1 \)) that make \( \langle \varepsilon_{k+\alpha-i\varepsilon_{k+\beta-i'}\delta_{k+\gamma+i''}\delta_{k+\lambda-i''''} \rangle \) or \( \langle \varepsilon_{k+\alpha-i\varepsilon_{k+\beta-i'}\delta_{k+\gamma+i''}\delta_{k+\lambda-i''''} \rangle \) non-zero for different combinations of \( i, i', i'', \) and \( i'''' \).

| \( i \) | \( i' \) | \( i'' \) | \( i'''' \) | \( \varepsilon \) | \( \gamma \) | \( \delta \) | \( \lambda \) | number of terms for \( \langle \varepsilon \rangle \neq 0 \) | \( \langle \gamma \rangle \) for \( \langle \gamma \rangle \neq 0 \) | \( \langle \delta \rangle \) for \( \langle \delta \rangle \neq 0 \) | \( \langle \lambda \rangle \) for \( \langle \lambda \rangle \neq 0 \) | number of terms for \( \langle \varepsilon \rangle \langle \gamma \rangle \langle \delta \rangle \langle \lambda \rangle \neq 0 \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | +1a | 1β+1 | 1γ+1 | 1λ+1 | \((0, 0, 0, 0)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 0)\) | \((0, 0, 0, 0)\) |
| 0 | 0 | 0 | 1 | -1a | 1β+1 | 1γ+1 | 0λ | \((0, 0, 0, 1)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 1)\) | \((0, 0, 0, 1)\) |
| 0 | 0 | 1 | 0 | -1a | 1β+1 | 0γ+1 | 1λ | \((0, 0, 1, 0)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 1, 0)\) | \((0, 0, 1, 0)\) |
| 0 | 0 | 1 | 1 | +1a | 1β+1 | 0γ+1 | 0λ | \((0, 0, 1, 1)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 0)\) | \((0, 0, 0, 0)\) |
| 0 | 0 | 1 | 1 | -1a | 1β+1 | 0γ+1 | 0λ | \((0, 1, 0, 0)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 0)\) | \((0, 0, 0, 0)\) |
| 0 | 0 | 1 | 1 | +1a | 1β+1 | 0γ+1 | 0λ | \((0, 1, 1, 0)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 0)\) | \((0, 0, 0, 0)\) |
| 0 | 0 | 1 | 1 | -1a | 1β+1 | 0γ+1 | 0λ | \((0, 1, 1, 1)\) | \((0, 0)\) | \((0, 0)\) | \((0, 0, 0, 0)\) | \((0, 0, 0, 0)\) |
| 0 | 1 | 0 | 0 | -1a | 0β+1 | 1γ+1 | 1λ | \((1, 0, 0, 0)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 0 | 1 | +1a | 0β+1 | 1γ+1 | 0λ | \((1, 0, 1, 0)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 0 | +1a | 0β+1 | 0γ+1 | 1λ | \((1, 0, 1, 0)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 0, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 0, 0)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | +1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 0)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | +1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | -1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |
| 0 | 1 | 1 | 1 | +1a | 0β+1 | 0γ+1 | 0λ | \((1, 1, 1, 1)\) | \((1, 0)\) | \((1, 0)\) | \((1, 0, 0, 0)\) | \((1, 0, 0, 0)\) |

**sum** | 0 | **sum** | 0
Finally, by plugging (A3) and (A4) into (A1),

\[
\sum_{\alpha,\beta,\gamma,\lambda = 0 \text{ or } 1 \atop \alpha\neq\beta\neq\gamma\neq\lambda = 1} f(\Delta Z_{k+\alpha}, \Delta Z_{k+\beta}, \Delta W_{k+\gamma}, \Delta W_{k+\lambda}) \\
= \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t')s(t'')s(t''') \times \\
\sum_{\alpha,\beta,\gamma,\lambda = 0 \text{ or } 1 \atop \alpha\neq\beta\neq\gamma\neq\lambda = 1} f(\Delta \hat{Z}(t, k, \alpha), \Delta \hat{Z}(t', k, \beta), \Delta \hat{W}(t'', k, \gamma), \Delta \hat{W}(t''', k, \lambda))dt dt' dt'' dt'''
\]

\[
+ \sum_{\alpha,\beta,\gamma,\lambda = 0 \text{ or } 1 \atop \alpha\neq\beta\neq\gamma\neq\lambda = 1} f(\Delta \varepsilon_{k,\alpha}, \Delta \varepsilon_{k,\beta}, \Delta \delta_{k,\gamma}, \Delta \delta_{k,\lambda})
\]

\[
= \int_0^{\Delta t} \int_0^{\Delta t} \int_0^{\Delta t} s(t)s(t')s(t'')s(t''')E \cdot \Delta t dt dt' dt'' dt''' = E \Delta t
\]

we obtain the formula (26).