

# Tensor approximation of the self-diffusion matrix of tagged particle processes

Jad Dabaghi<sup>1,2</sup>, Virginie Ehrlacher<sup>1,2</sup>, and Christoph Strössner<sup>3</sup>

<sup>1</sup>Ecole des Ponts ParisTech, Marne-la-Vallée, France

<sup>2</sup>INRIA Paris, France, MATERIALS team-project

<sup>3</sup>Ecole Polytechnique Fédérale de Lausanne (EPFL), Institute of Mathematics, CH-1015 Lausanne, Switzerland

`jad.dabaghi@enpc.fr, virginie.ehrlacher@enpc.fr, christoph.stroessner@epfl.ch`

April 8, 2022

## Abstract

The objective of this paper is to investigate a new numerical method for the approximation of the self-diffusion matrix of a tagged particle process defined on a grid. While standard numerical methods make use of long-time averages of empirical means of deviations of some stochastic processes, and are thus subject to statistical noise, we propose here a tensor method in order to compute an approximation of the solution of a high-dimensional quadratic optimization problem, which enables to obtain a numerical approximation of the self-diffusion matrix. The tensor method we use here relies on an iterative scheme which builds low-rank approximations of the quantity of interest and on a carefully tuned variance reduction method so as to evaluate the various terms arising in the functional to minimize. In particular, we numerically observe here that it is much less subject to statistical noise than classical approaches.

## 1 Introduction

The aim of this paper is to propose a new numerical method for the approximation of the self-diffusion matrix of a tagged particle process on a grid [20] which is less subject to statistical noise than standard Monte-Carlo strategies.

The coefficients of the self-diffusion matrix of this process are defined as long-time averages of the mean-square displacement of the tagged particle [3, 7, 17, 29]. The knowledge of the latter enables one to identify the hydrodynamic limits of multi-species symmetric exclusion processes [29] and is thus of tremendous importance for applications. However, computing a numerical approximation of these coefficients is not an easy task since standard Monte-Carlo approaches suffer from the large amount of statistical noise in the corresponding approximations [8, 9, 13, 32, 34, 35].

In this paper, we propose an alternative numerical method based on the fact that the self-diffusion coefficients can be equivalently reformulated using the solution of a deterministic high-dimensional optimization problem [3, 21, 29]. The main mathematical ingredient of the numerical method we investigate here is the use of low-rank tensor decompositions to obtain a numerical approximation of this solution, which enables to bypass the curse of dimensionality. This approach leads to an iterative scheme, each elementary step of which boils down to a minimization problem over the set of low-rank tensors, which is solved using a classical

alternating linear scheme [2, 11, 30]. Our numerical experiments demonstrate that this low-rank approach leads to very accurate approximations of the self-diffusion coefficients on finite-size grids.

As an illustration of the method, we use the obtained numerical approximations in order to compute the self-diffusion matrix of a two-dimensional tagged particle process defined on a Cartesian grid.

This article is organized as follows. In Section 2, we introduce the lattice-based stochastic hopping model, its hydrodynamic limit and the two equivalent definitions of the self-diffusion matrix of the tagged particle process our work is based upon. The numerical approach we propose here in order to compute a numerical approximation is presented in Section 3. Finally, the efficiency of our approach is illustrated through several numerical experiments presented in Section 4.

## 2 Self-diffusion matrix

### 2.1 Infinite-dimensional definition

Let  $d = 1, 2, 3$  denote the dimension of the problem. We consider a symmetric tagged particle process defined on the infinite grid  $\mathbb{Z}^d$ . Let  $K \in \mathbb{N}^*$  denote the number of possible jump directions for the particles and  $(\mathbf{v}_k)_{1 \leq k \leq K} \subset \mathbb{Z}^d \setminus \{0\}$  the set of possible jump directions. For all  $1 \leq k \leq K$ , the probability of jumping in the direction  $\mathbf{v}_k$  is denoted by  $p_k \in ]0, 1]$ . This jumping scheme is assumed to be symmetric in the sense that if the jump in the direction  $\mathbf{v}_k$  occurs with probability  $p_k$ , then the jump in the direction  $-\mathbf{v}_k$  occurs with the same probability.

The self-diffusion matrix application

$$\mathbb{D}_s : \begin{cases} [0, 1] & \rightarrow \\ \rho & \mapsto \end{cases} \mathbb{D}_s(\rho) := (\mathbb{D}_{s,ij}(\rho))_{1 \leq i,j \leq d} \in \mathbb{R}^{d \times d}$$

can be defined in the following two ways.

**Definition as optimization problem.** Let us first introduce some notation. Let  $S := \mathbb{Z}^d \setminus \{0\}$ . For all  $\boldsymbol{\eta} := (\eta_{\mathbf{s}})_{\mathbf{s} \in S} \in \{0, 1\}^S$  and all  $\mathbf{y} \neq \mathbf{z} \in S$ , we define by  $\boldsymbol{\eta}^{\mathbf{y}, \mathbf{z}} := (\eta_{\mathbf{s}}^{\mathbf{y}, \mathbf{z}})_{\mathbf{s} \in S}$  the element of  $\{0, 1\}^S$  such that

$$\eta_{\mathbf{s}}^{\mathbf{y}, \mathbf{z}} := \begin{cases} \eta_{\mathbf{s}} & \text{if } \mathbf{s} \neq \mathbf{y}, \mathbf{z}, \\ \eta_{\mathbf{y}} & \text{if } \mathbf{s} = \mathbf{z}, \\ \eta_{\mathbf{z}} & \text{if } \mathbf{s} = \mathbf{y}. \end{cases}$$

Furthermore, for all  $\mathbf{w} \in S$ , we define by  $\boldsymbol{\eta}^{\mathbf{0}, \mathbf{w}} := (\eta_{\mathbf{s}}^{\mathbf{0}, \mathbf{w}})_{\mathbf{s} \in S}$  the element of  $\{0, 1\}^S$  such that

$$\eta_{\mathbf{s}}^{\mathbf{0}, \mathbf{w}} := \begin{cases} \eta_{\mathbf{s}+\mathbf{w}} & \text{if } \mathbf{s} \neq -\mathbf{w}, \\ 0 & \text{if } \mathbf{s} = -\mathbf{w}. \end{cases}$$

For all  $\rho \in [0, 1]$ , we denote by  $\rho^{\otimes}$  the Bernoulli product measure on  $\{0, 1\}^S$  with marginals given by

$$\rho^{\otimes}(\{\boldsymbol{\eta} := (\eta_{\mathbf{s}})_{\mathbf{s} \in S} : \eta_{\hat{\mathbf{s}}} = 1\}) = \rho,$$

for all  $\hat{\mathbf{s}} \in S$ . Let us also define the set of functions  $H_{\rho} := L_{\rho^{\otimes}}^2(\{0, 1\}^S)$ , where  $L_{\rho^{\otimes}}^2$  denotes the  $L^2$  Lebesgue space with measure  $\rho^{\otimes}$ .

For all  $\mathbf{u} \in \mathbb{R}^d$  and  $\rho \in [0, 1]$ , the self-diffusion coefficient  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$  is given by:

$$\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u} := 2 \inf_{\Psi \in H_{\rho}} \mathbb{E}_{\rho^{\otimes}} \left[ \sum_{k=1}^K p_k \left( (1 - \eta_{\mathbf{v}_k}) (\mathbf{u} \cdot \mathbf{v}_k + \Psi(\boldsymbol{\eta}^{\mathbf{0}, \mathbf{v}_k}) - \Psi(\boldsymbol{\eta}))^2 + \frac{1}{2} \sum_{\substack{\mathbf{y} \in S \setminus \{0\} \\ \mathbf{y} + \mathbf{v}_k \neq 0}} (\Psi(\boldsymbol{\eta}^{\mathbf{y} + \mathbf{v}_k, \mathbf{y}}) - \Psi(\boldsymbol{\eta}))^2 \right) \right], \quad (1)$$

where the notation  $\mathbb{E}_{\rho^{\otimes}}$  refers to the fact that the expectation is taken over the product measure  $\rho^{\otimes}$  [20, 29]. Problem (1) thus reads as an infinite-dimensional optimization problem over the set  $H_{\rho}$ .

**Remark 2.1.** Naturally, for all  $\rho \in [0, 1]$ , since  $\mathbb{D}_s(\rho)$  is a symmetric matrix, one can easily deduce the full matrix  $\mathbb{D}_s(\rho)$  from the knowledge of  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$  for a few vectors  $\mathbf{u} \in \mathbb{R}^d$ .

**Definition as long time mean square deviation.** The quantity  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$  can be equivalently expressed as the long time limit of the following expectation [29, Theorem 2.3]. Let us assume that, at time  $t = 0$ , the tagged particle is located at position  $\mathbf{0}$  and all other sites are occupied with probability  $\rho$  following a Bernoulli distribution. The duration between two consecutive jumping events follow an exponential law with parameter 1. The jumping directions (respectively rates) are given by  $\{\mathbf{v}_1, \dots, \mathbf{v}_K\}$  (respectively  $p_1, \dots, p_K$ ). Jumps are not allowed if the final site of the jumping particle is already occupied. Let  $\mathbf{w}(t)$  denote the position of the tagged particle at time  $t \geq 0$ . It then holds that, for all  $\mathbf{u} \in \mathbb{R}^d$  and  $\rho \in [0, 1]$ , the quantity  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$  can be equivalently formulated as

$$\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u} = 2 \lim_{t \rightarrow \infty} \frac{\mathbb{E}_{\rho \otimes} [\langle \mathbf{u}, \mathbf{w}(t) \rangle^2]}{t}, \quad (2)$$

where  $\langle \cdot, \cdot \rangle$  denotes the euclidean scalar product of  $\mathbb{R}^d$ .

**Remark 2.2.** Starting from a Bernoulli-product measure and symmetric transition rates, it is a classical problem in probability theory to study the motion of a tagged particle on  $\mathbb{Z}^d$  [18, 22, 31], see also [5, 10] for recent work on tagged particles on trees. For  $d \geq 2$  and symmetric nearest neighbor transition rates, the tagged particle is known to satisfy a central limit theorem with non-degenerate limiting variance. However, to our best knowledge, a general closed formula for the limiting variance is not available.

Notice that when starting from a Bernoulli-product measure with a tagged particle in the origin, the resulting environment process is stationary with respect to the Bernoulli-product measure conditioned to contain a particle in the origin, usually called the Palm measure. This suggests that the limiting variance for the tagged particle can be described in terms of the sum of separable functions, i.e. in terms of a low-rank function. In turn, this indicates that the equivalent characterization of the limiting variance as in Equation (1) is also related to low-rank functions.

Let us point out that the quantity  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$  cannot be computed exactly in practice, neither using expression (1) nor expression (2). On the one hand, (1) reads as an infinite-dimensional optimization problem and has to be approximated by a finite-dimensional optimization problem in practice [15, 21]. On the other hand, (2) requires the computation of the long-time average of the stochastic process defined on the infinite lattice grid  $\mathbb{Z}^d$ . Equivalently, it has to be approximated in a finite-dimensional setting, as long-time average of the mean-square deviation of the tagged particle associated to a stochastic process defined on a finite-size grid with periodic boundary conditions. Both finite-dimensional approximations are presented in detail in the next section.

## 2.2 Finite-dimensional approximation

**Discretized minimization problems.** Let  $M \in \mathbb{N}^*$  denote a discretization parameter and introduce the finite grid  $S_M := \{-M, \dots, M\}^d \setminus \{\mathbf{0}\}$ . For the sake of simplicity, we assume that  $M \geq \max_{k \in \{1, \dots, K\}} \|v_k\|_1$ , where  $\|\cdot\|_1$  denotes the taxicab norm. For any  $\boldsymbol{\eta} \in \{0, 1\}^{S_M}$ , we can construct by periodicity an extension  $\tilde{\boldsymbol{\eta}} := (\tilde{\eta}_{\mathbf{s}})_{\mathbf{s} \in S} \in \{0, 1\}^S$  by assuming with a slight abuse of notation that the site  $\mathbf{0}$  is occupied, i.e.  $\tilde{\eta}_{\mathbf{s}} = 1$  for  $\mathbf{s} \in (2M+1) \cdot \mathbb{Z}^d \setminus \{\mathbf{0}\}$ . Using this notation, for all  $\boldsymbol{\eta} \in \{0, 1\}^{S_M}$  and all  $\mathbf{y}, \mathbf{z}, \mathbf{w} \in S$ , we define  $\boldsymbol{\eta}^{\mathbf{y}, \mathbf{z}} \in \{0, 1\}^{S_M}$  and  $\boldsymbol{\eta}^{\mathbf{0}, \mathbf{w}} \in \{0, 1\}^{S_M}$  as

$$\boldsymbol{\eta}^{\mathbf{y}, \mathbf{z}} := (\tilde{\eta}_{\mathbf{s}}^{\mathbf{y}, \mathbf{z}})_{\mathbf{s} \in S_M} \quad \text{and} \quad \boldsymbol{\eta}^{\mathbf{0}, \mathbf{w}} := (\tilde{\eta}_{\mathbf{s}}^{\mathbf{0}, \mathbf{w}})_{\mathbf{s} \in S_M}.$$

Figure 1 shows an illustration of  $\boldsymbol{\eta}^{\mathbf{y}, \mathbf{z}}$  and  $\boldsymbol{\eta}^{\mathbf{0}, \mathbf{w}}$ .

Let  $N := (2M+1)^d - 1 = \text{Card}(S_M)$ . For all  $\ell \in \{0, \dots, N\}$ , let  $C_{M, \ell} := \{\boldsymbol{\eta} \in \{0, 1\}^{S_M} \mid \sum_{\mathbf{s} \in S_M} \eta_{\mathbf{s}} = \ell\}$  denote the set of all possible configurations of the particles on  $S_M$  so that the total number of occupied sites is equal to  $\ell$ .

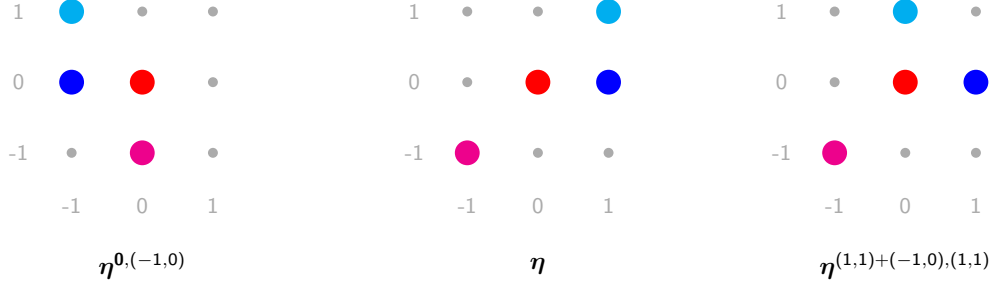


Figure 1: Middle: Visualization of one particular  $\eta \in S_1$  for  $d = 2$  with three occupied sites marked in blue, cyan and magenta. Additionally, we mark the tagged particle  $\mathbf{0}$  in red. Left: Visualization of the occupied sites of  $\eta^{0,(-1,0)}$ . This can be seen as a jump of the imaginary red particle one step to the left, followed by immediately relabeling of the sites such that red particle remains at  $\mathbf{0}$ . By exploiting the periodicity, we obtain  $\eta^{0,(-1,0)} \in S_1$ . Right: Visualization of the occupied sites of  $\eta^{(1,1)+(-1,0),(1,1)}$ . This can be seen as jump of the cyan particle one step to the left.

Let us define  $H_M := \{\Psi : \{0, 1\}^{S_M} \rightarrow \mathbb{R}\}$ . For every  $\mathbf{u} \in \mathbb{R}^d$  and  $l \in \{0, \dots, N\}$ , we introduce the quadratic functional  $A_{M,\ell}^{\mathbf{u}} : H_M \rightarrow \mathbb{R}$  defined by

$$A_{M,\ell}^{\mathbf{u}}(\Psi) := \frac{1}{|C_{M,\ell}|} \sum_{\eta \in C_{M,\ell}} \sum_{k=1}^K p_k \left( (1 - \eta_{\mathbf{v}_k}) (\mathbf{u} \cdot \mathbf{v}_k + \Psi(\eta^{0,\mathbf{v}_k}) - \Psi(\eta))^2 + \frac{1}{2} \sum_{\substack{\mathbf{y} \in S_M \\ \mathbf{y} + \mathbf{v}_k \neq \mathbf{0}}} (\Psi(\eta^{\mathbf{y} + \mathbf{v}_k, \mathbf{y}}) - \Psi(\eta))^2 \right).$$

Then, assuming that  $\bar{\rho} = \frac{\ell}{N}$  for some  $0 \leq \ell \leq N$ , one can define [3] for all  $\mathbf{u} \in \mathbb{R}^d$ ,

$$\mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u} := 2 \min_{\Psi \in H_M} A_{M,\ell}^{\mathbf{u}}(\Psi). \quad (3)$$

It is proved in [21] that  $\lim_{\substack{M \rightarrow +\infty \\ \frac{\ell}{N} \rightarrow \bar{\rho}}} \mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u} = \mathbf{u}^T \mathbb{D}_s(\bar{\rho}) \mathbf{u}$ .

**Combined minimization problem.** The collection of sets  $C_{M,0}, \dots, C_{M,N}$  form a partition of the set  $\{0, 1\}^{S_M}$ . Observe for a given  $\Psi \in H_M$ ,  $A_{M,\ell}^{\mathbf{u}}(\Psi)$  only depends on the values of  $\Psi(\eta)$  for  $\eta \in C_{M,\ell}$ , since  $\eta \in C_{M,\ell}$  implies that also  $\eta^{0,\mathbf{v}_k} \in C_{M,\ell}$  and  $\eta^{\mathbf{y} + \mathbf{v}_k, \mathbf{y}} \in C_{M,\ell}$  for all  $1 \leq k \leq K$ . As a consequence, if  $\Psi_{\text{opt}}^{M,\mathbf{u}} \in H_M$  is a minimizer of

$$\min_{\Psi \in H_M} A_M^{\mathbf{u}}(\Psi), \quad (4)$$

where

$$A_M^{\mathbf{u}}(\Psi) := \sum_{\eta \in \{0,1\}^{S_M}} \sum_{k=1}^K p_k \left( (1 - \eta_{\mathbf{v}_k}) (\mathbf{u} \cdot \mathbf{v}_k + \Psi(\eta^{0,\mathbf{v}_k}) - \Psi(\eta))^2 + \frac{1}{2} \sum_{\substack{\mathbf{y} \in S_M \\ \mathbf{y} + \mathbf{v}_k \neq \mathbf{0}}} (\Psi(\eta^{\mathbf{y} + \mathbf{v}_k, \mathbf{y}}) - \Psi(\eta))^2 \right), \quad (5)$$

then it holds that  $A_{M,\ell}^{\mathbf{u}}(\Psi_{\text{opt}}^{M,\mathbf{u}}) = \min_{\Psi \in H_M} A_{M,\ell}^{\mathbf{u}}(\Psi)$  for all  $\ell \in \{0, \dots, N\}$ . The knowledge of  $\Psi_{\text{opt}}^{M,\mathbf{u}}$  then allows us to compute  $\mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u}$  for all  $0 \leq \ell \leq N$  as  $2A_{M,\ell}^{\mathbf{u}}(\Psi_{\text{opt}}^{M,\mathbf{u}})$ . Note that the minimization problem (4) is then independent of  $\ell$ , in contrast to (3). However, the minimization problem (4) has  $2^N$  degrees of freedom and is thus intractable for large values of  $N$ .

The main contribution of our paper is to propose a low-rank tensor approximation algorithm to compute an approximation of  $\Psi_{\text{opt}}^{M,u}$  in order to mitigate this issue. Our proposed method is presented in Section 3.

**Estimation of long-time mean square deviation.** The quantity  $\mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u}$  can be equivalently expressed as the long-time mean square deviation of a tagged particle evolving in a periodic environment [21]. The Monte-Carlo algorithm is the standard method of choice in practice to compute an approximation of  $\mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u}$  by means of empirical averages. The quality of the obtained approximations then depends on the choice of two numerical parameters, namely the number of Monte-Carlo samples (i.e. stochastic realizations of the periodized tagged particle process) and the value of the chosen final time.

We aim to compare the low-rank tensor approximation algorithm we propose in this work with such Monte-Carlo methods. To this aim, we detail below the algorithm used to compute such empirical averages, and which we will use as a comparison with the method we develop here and which is presented in Section 3.

Let  $1 \leq \ell \leq N$ . An initial environment  $\boldsymbol{\eta}$  is obtained by sampling uniformly from  $C_{M,\ell}$ . We sample from exponential distributions to determine the next time a particle in the environment or the tagged particle performs a jump. Whenever a jump is performed the environment  $\boldsymbol{\eta}$  is updated accordingly. Let  $\mathbf{w} = \mathbf{0} \in \mathbb{Z}^d$  denote the initial position of the tagged particle. Throughout the simulation, we track the position of the tagged particle in  $\mathbb{Z}^d$ , i.e. whenever the tagged particle successfully jumps in direction  $\mathbf{v}_k$ , we set  $\mathbf{w} = \mathbf{w} + \mathbf{v}_k$ . To approximate the expectation of  $\mathbf{w}$ , we repeat this simulation  $N_s$  times with different samples. Each of these simulations is stopped at the same stopping time  $T$  at which we compute the approximation of  $\mathbf{u}^T \mathbb{D}_s(\rho) \mathbf{u}$ . This approach is formalized in Algorithm 1.

---

**Algorithm 1** Long-term Monte Carlo estimation

---

- 1: **Input:**  $M \in \mathbb{N}^*$ ,  $\mathbf{u} \in \mathbb{R}^d$ , number of occupied sites  $1 \leq \ell \leq N$  with  $N = (2M+1)^d - 1$ , final time  $T > 0$ ,  $\hat{N}_s \in \mathbb{N}^*$
  - 2: **Output:** approximation  $\alpha \approx \frac{1}{2} \mathbf{u}^T \mathbb{D}_s^M \left( \frac{\ell}{N} \right) \mathbf{u} = A_{M,\ell}^{\mathbf{u}}(\Psi_{\text{opt}}^{M,\mathbf{u}})$
  - 3:  $\mathbf{w} = \mathbf{0}$ ,  $\alpha = 0$ ,  $N_s = \lceil \hat{N}_s / (\ell + 1) \rceil$
  - 4: **for**  $i = 1, 2, 3, \dots, N_s$
  - 5:     randomly initialize  $\boldsymbol{\eta} \in C_{M,\ell}$  and set  $t_{\text{TOTAL}} = 0$ ,
  - 6:     **while** true
  - 7:         sample  $t_{\text{NEW}}$  from an exponential distribution with mean  $\frac{1}{\ell+1}$  and set  $t_{\text{TOTAL}} = t_{\text{TOTAL}} + t_{\text{NEW}}$
  - 8:         **if**  $t_{\text{TOTAL}} > T$
  - 9:             break
  - 10:         uniformly select either one occupied site  $\mathbf{y} \in S_M$  with  $\boldsymbol{\eta}_{\mathbf{y}} = 1$  or the tagged particle  $\mathbf{y} = \mathbf{0}$
  - 11:         select a jump direction  $\mathbf{v}_k$  with probability  $p_k$
  - 12:         **if**  $\tilde{\boldsymbol{\eta}}_{\mathbf{y}+\mathbf{v}_k} = \mathbf{0}$ , i.e. the target location is not occupied
  - 13:             when  $\mathbf{y} \neq \mathbf{0} \Rightarrow$  set  $\boldsymbol{\eta} = \boldsymbol{\eta}^{\mathbf{y}+\mathbf{v}_k, \mathbf{y}}$
  - 14:             when  $\mathbf{y} = \mathbf{0} \Rightarrow$  set  $\boldsymbol{\eta} = \boldsymbol{\eta}^{0, \mathbf{v}_k}$  and update  $\mathbf{w} = \mathbf{w} + \mathbf{v}_k$
  - 15:         update  $\alpha = \alpha + \langle \mathbf{u}, \mathbf{w} \rangle^2$
  - 16:  $\alpha = \frac{\alpha}{TN_s}$
- 

**Remark 2.3.** In Algorithm 1, we set  $N_s = \lceil \hat{N}_s / (\ell + 1) \rceil$ , where  $\hat{N}_s$  is a given input parameter. This ensures that the expected runtime is approximately equal for all choices of  $\ell$ . Additionally Figure 5 shows that the variance of the output is comparable independently of  $\ell$ . If we were to use the same  $N_s$  for all  $\ell$ , we would observe a much larger variance for smaller  $\ell$  and the runtime would increase for larger  $\ell$ .

### 3 Low-rank solutions for the optimization problem

The aim of this section is to present the numerical method based on low-rank tensor approximations we propose for the resolution of the high-dimensional optimization problem (4). We first define low-rank functions

in  $H_M$  and introduce a fast and stable algorithm for the evaluation of  $A_M^{\mathbf{u}}$ . We then develop a successive minimization scheme to compute low-rank solutions of (4). Each minimization step is performed using an alternating linear scheme, which relies on the fast and stable evaluations of  $A_M^{\mathbf{u}}$ . Lastly, we discuss the evaluation of  $A_{M,\ell}^{\mathbf{u}}$  for the computation of the self-diffusion coefficient (3).

### 3.1 Low-rank functions

A function  $R \in H_M$  is called a rank-1 or pure tensor product function when it can be written as

$$R(\boldsymbol{\eta}) = \prod_{\mathbf{s} \in S_M} R_{\mathbf{s}}(\eta_{\mathbf{s}}), \quad \forall \boldsymbol{\eta} = (\eta_{\mathbf{s}})_{\mathbf{s} \in S_M} \in \{0, 1\}^{S_M},$$

for some  $R_{\mathbf{s}} : \{0, 1\} \rightarrow \mathbb{R}$  for  $\mathbf{s} \in S_M$ . Let  $H_M^1 \subset H_M$  denote the set of pure tensor product functions of  $H_M$ . Let  $r \in \mathbb{N}$ . A function  $\Phi \in H_M$  is called a rank- $r$  function when it can be written as  $\Phi = R^{(1)} + \dots + R^{(r)}$ , where  $R^{(k)} \in H_M^1$  for  $1 \leq k \leq r$ . We denote by  $H_M^r \subset H_M$  the set of all rank- $r$  functions. For all  $r \in \mathbb{N}^*$ , it holds

$$\min_{\Psi \in H_M} A_M^{\mathbf{u}}(\Psi) = \min_{\Phi \in H_M^{2^N}} A_M^{\mathbf{u}}(\Phi) \leq \min_{\Phi \in H_M^{r+1}} A_M^{\mathbf{u}}(\Phi) \leq \min_{\Phi \in H_M^r} A_M^{\mathbf{u}}(\Phi) \leq \min_{R \in H_M^1} A_M^{\mathbf{u}}(R).$$

In the next sections, we derive an algorithm to compute an approximation of  $\min_{\Phi \in H_M^r} A_M^{\mathbf{u}}(\Phi)$ .

### 3.2 Fast and stable evaluation for low-rank functions

In this section, we introduce a fast and stable method to evaluate  $A_M^{\mathbf{u}}(\Psi)$  for  $\Psi \in H_M^r$ . A naive evaluation would require to sum over  $2^N$  terms, which is intractable for large values of  $N$ . In principle, the order of summation can be exchanged leading to terms of the form  $\sum_{\boldsymbol{\eta} \in \{0,1\}^{S_M}} \Psi(\boldsymbol{\eta})^2$ , which can be evaluated individually. However, subtracting such terms leads to a lot of numerical cancellation for  $M > 1$ . We circumvent these issues by treating the evaluation of  $A_M^{\mathbf{u}}(\Psi)$  as the computation of the Frobenius norms of certain tensors. These Frobenius norms can then be evaluated in a fast and stable way.

#### 3.2.1 Reformulation as tensor norm

For  $\Psi \in H$ , we consider the tensor  $\mathcal{T}^{(\Psi)} \in \mathbb{R}^{2^N}$  with entries  $\mathcal{T}_{i_1, \dots, i_N}^{(\Psi)} = \Psi(\boldsymbol{\eta})$ , where  $\boldsymbol{\eta} = (\eta_{\mathbf{s}_1}, \dots, \eta_{\mathbf{s}_N})$  for some enumeration  $\mathbf{s}_1, \dots, \mathbf{s}_N$  of the sites in  $S_M$  and  $\eta_{\mathbf{s}_j} = i_j - 1$  for  $1 \leq j \leq N$ . Note that  $\mathcal{T}^{(\Psi)}$  is a tensor of at most rank- $r$  when  $\Psi \in H_M^r$ . In particular, when  $\Psi$  is given in term of the functions  $R_{\mathbf{s}}^{(k)}$  for  $1 \leq j \leq N$ ,  $1 \leq k \leq r$ , we can write  $\mathcal{T}^{(\Psi)}$  in the so called CP-format [16, 19] as

$$\mathcal{T}_{i_1, \dots, i_N}^{(\Psi)} = \sum_{k=1}^r \prod_{j=1}^N \mathbf{a}_{i_j}^{(j,k)}, \quad (6)$$

where  $\mathbf{a}^{(j,k)} \in \mathbb{R}^2$  is defined as  $\mathbf{a}^{(j,k)} = (R_{\mathbf{s}_j}^{(k)}(0), R_{\mathbf{s}_j}^{(k)}(1))$ .

For  $\mathbf{v}, \mathbf{y} \in S_m$ , we analogously define the tensor  $\mathcal{T}^{(\Psi^{(0,\mathbf{v})})} \in \mathbb{R}^{2^N}$  with entries  $\mathcal{T}_{i_1, \dots, i_N}^{(\Psi^{(0,\mathbf{v})})} = \Psi(\boldsymbol{\eta}^{0,\mathbf{v}})$ . When  $\mathbf{y} + \mathbf{v} \neq \mathbf{0}$  we additionally define the tensor  $\mathcal{T}^{(\Psi^{(\mathbf{y}+\mathbf{v},\mathbf{y})})} \in \mathbb{R}^{2^N}$  with entries  $\mathcal{T}_{i_1, \dots, i_N}^{(\Psi^{(\mathbf{y}+\mathbf{v},\mathbf{y})})} = \Psi(\boldsymbol{\eta}^{\mathbf{y}+\mathbf{v},\mathbf{y}})$ . We observe that these are again rank- $r$  tensors when  $\Psi \in H_M^r$ . Lastly, we define the rank-1 tensor  $\mathcal{T}^{(\mathbf{u} \cdot \mathbf{v})} \in \mathbb{R}^{2^N}$  with entries  $\mathcal{T}_{i_1, \dots, i_N}^{(\mathbf{u} \cdot \mathbf{v})} = \mathbf{u} \cdot \mathbf{v}$ .

For a tensor  $\mathcal{T} \in \mathbb{R}^{2^N}$  and a site  $\mathbf{s}$  we define the projection operators  $\mathcal{P}_{\mathbf{s}} : \mathbb{R}^{2^N} \rightarrow \mathbb{R}^{2^N}$  as

$$(\mathcal{P}_{\mathbf{s}}(\mathcal{T}))_{i_1, \dots, i_N} := \begin{cases} \mathcal{T}_{i_1, \dots, i_N} & i_{\mathbf{s}} = 1 \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } 1 \leq i_1, \dots, i_N \leq 2,$$

where  $i_{\mathbf{s}}$  denotes to the index assigned to site  $\mathbf{s}$ . We can now rewrite (5) as

$$A_M^{\mathbf{u}}(\Psi) = \sum_{k=1}^K p_k \left( \|P_{\mathbf{v}_k}(\mathcal{T}^{(\mathbf{u} \cdot \mathbf{v}_k)} + \mathcal{T}^{(\Psi^{(0,\mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y}+\mathbf{v}_k,\mathbf{y})})})\|_F^2 + \frac{1}{2} \sum_{\substack{\mathbf{y} \in S_M \\ \mathbf{y} + \mathbf{v}_k \neq \mathbf{0}}} \|\mathcal{T}^{(\Psi^{(0,\mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y}+\mathbf{v}_k,\mathbf{y})})}\|_F^2 \right), \quad (7)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm.

### 3.2.2 Efficient evaluation of Frobenius norms for sums of low-rank tensors

Let  $k \in \{1, \dots, K\}$  and  $\mathbf{y} \in S_M$  such that  $\mathbf{y} + \mathbf{v}_k \neq \mathbf{0}$  be fixed. We derive an algorithm inspired by tensor train (TT) orthogonalization [26] to evaluate  $\|\mathcal{T}^{(\Psi^{(0, \mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y} + \mathbf{v}_k, \mathbf{y})})}\|_F^2$ .

We start by rewriting Equation (6) as

$$T_{i_1, \dots, i_N}^{(\Psi)} = \sum_{k_1=1}^r \cdots \sum_{k_{N-1}=1}^r \mathcal{C}_{1, i_1, k_1}^1 \mathcal{C}_{k_1, i_2, k_2}^2 \mathcal{C}_{k_2, i_3, k_3}^3 \cdots \mathcal{C}_{k_{N-1}, i_N, 1}^N, \quad (8)$$

with tensors  $\mathcal{C}^1 \in \mathbb{R}^{1 \times 2 \times r}$ ,  $\mathcal{C}^N \in \mathbb{R}^{r \times 2 \times 1}$  and  $\mathcal{C}^j \in \mathbb{R}^{r \times 2 \times r}$  for  $1 < j < N$ , whose entries are given by

$$\begin{aligned} \mathcal{C}_{k_1, i, k_2}^j &= \begin{cases} \mathbf{a}_i^{j, k_1} & k_1 = k_2 \\ 0 & \text{otherwise} \end{cases}, & \text{for } 1 < j < N, \ 1 \leq i \leq 2, \ 1 \leq k_1, k_2 \leq N, \\ \mathcal{C}_{1, i, k}^1 &= \mathbf{a}_i^{1, k} & \text{for } 1 \leq i \leq 2, \ 1 \leq k \leq N, \\ \mathcal{C}_{k, i, 1}^N &= \mathbf{a}_i^{N, k} & \text{for } 1 \leq i \leq 2, \ 1 \leq k \leq N. \end{aligned}$$

The representation format (8) can be generalized to other low-rank tensors. Let  $\mathcal{T}^{(\Psi^{(\mathbf{y} + \mathbf{v}_k, \mathbf{y})})}$  be analogously represented in the format (8) by tensors  $\mathcal{D}^j$ . The tensor  $\mathcal{T}^{(\Psi^{(0, \mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y} + \mathbf{v}_k, \mathbf{y})})}$  is at most rank- $2r$ . Its representation in the form of (8) with tensors  $\mathcal{E}^j$  can be constructed from the tensors  $\mathcal{C}^j, \mathcal{D}^j$ , i.e. the tensors  $\mathcal{E}^1 \in \mathbb{R}^{1 \times 2 \times 2r}$ ,  $\mathcal{E}^N \in \mathbb{R}^{2r \times 2 \times 1}$  and  $\mathcal{E}^j \in \mathbb{R}^{2r \times 2 \times 2r}$  for  $1 < j < N$  are defined as  $\mathcal{E}_{1, :, 1:r}^1 = \mathcal{C}^1$ ,  $\mathcal{E}_{1, :, r+1:2r}^1 = -\mathcal{D}^1$ ,  $\mathcal{E}_{1:r, :, 1}^N = \mathcal{C}^N$ ,  $\mathcal{E}_{r+1:2r, :, 1}^N = \mathcal{D}^N$  and  $\mathcal{E}_{1:r, :, 1:r}^j = \mathcal{C}^j$ ,  $\mathcal{E}_{r+1:2r, :, r+1:2r}^j = \mathcal{D}^j$ . Note that for  $1 < j < N$  these tensors are sparse by construction.

We can compute the norm  $\|\mathcal{T}^{(\Psi^{(0, \mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y} + \mathbf{v}_k, \mathbf{y})})}\|_F^2$  directly from the tensors  $\mathcal{E}^j$  in a fast and stable way using TT orthogonalization [26] as defined in Algorithm 2. We want to emphasize that the sparsity structure of  $\mathcal{E}^j$  is preserved in Algorithm 2. The evaluation of  $\|\mathcal{E}^N\|_F^2$  in line 13 requires summing up  $4r$  terms, whereas a naive evaluation of  $\|\mathcal{T}^{(\Psi^{(0, \mathbf{v}_k)})} - \mathcal{T}^{(\Psi^{(\mathbf{y} + \mathbf{v}_k, \mathbf{y})})}\|_F^2$  involves  $2^N$  terms.

---

#### Algorithm 2 Frobenius norm evaluation

---

- 1: **Input:** tensors  $\mathcal{E}^1 \in \mathbb{R}^{1 \times 2 \times 2r}$ ,  $\mathcal{E}^N \in \mathbb{R}^{2r \times 2 \times 1}$  and  $\mathcal{E}^j \in \mathbb{R}^{2r \times 2 \times 2r}$  for  $1 < j < N$
  - 2: **Output:**  $\|\mathcal{T}\|_F^2$ , where  $\mathcal{T} \in \mathbb{R}^{2^N}$  has entries  $T_{i_1, \dots, i_N} = \sum_{k_1=1}^r \cdots \sum_{k_{N-1}=1}^r \mathcal{E}_{1, i_1, k_1}^1 \mathcal{E}_{k_1, i_2, k_2}^2 \cdots \mathcal{E}_{k_{N-1}, i_N, 1}^N$
  - 3: Compute  $QR$  decomposition of  $\mathcal{E}^1$  reshaped as element in  $\mathbb{R}^{2 \times 2r}$ .
  - 4: Set  $\mathcal{E}^1$  to  $Q$  reshaped as element in  $\mathbb{R}^{1 \times 2 \times 2r}$ .
  - 5: Update  $\mathcal{E}^2$ : reshape to  $\mathbb{R}^{2r \times 2 \times 2r}$ , multiply with  $R$  from the left, reshape back to  $\mathbb{R}^{2r \times 2 \times 2r}$ .
  - 6: **for**  $j = 2, \dots, N-1$
  - 7:     Compute  $QR$  decomposition of  $\mathcal{E}^j$  reshaped as element in  $\mathbb{R}^{2r \times 2 \times 2r}$ .
  - 8:     Set  $\mathcal{E}^j$  to  $Q$  reshaped as element in  $\mathbb{R}^{2r \times 2 \times 2r}$ .
  - 9:     **if**  $j < N-1$
  - 10:         Update  $\mathcal{E}^{j+1}$ : reshape to  $\mathbb{R}^{2r \times 2 \times 2r}$ , multiply with  $R$  from the left, reshape back to  $\mathbb{R}^{2r \times 2 \times 2r}$ .
  - 11:     **else**
  - 12:         Update  $\mathcal{E}^N$ : reshape to  $\mathbb{R}^{2r \times 2}$ , multiply with  $R$  from the left, reshape back to  $\mathbb{R}^{2r \times 2 \times 1}$ .
  - 13:  $\|\mathcal{T}\|_F^2 = \|\mathcal{E}^N\|_F^2$
- 

**Remark 3.1.** The sum of several low-rank tensors can be represented in the format (8). The representation of the sum can again be constructed from the representations of the individual low-rank tensors. This allows us to apply TT orthogonalization to evaluate all Frobenius-norms in Equation (7) in a fast and stable way.

### 3.3 Successive minimization

Let  $r \in \mathbb{N}$ . In order to approximate the solution of the minimization problem  $\min_{\Phi \in \mathcal{T}_M^r} A_M^u(\Phi)$ , we decompose the problem into a sequence of successive minimization problems [1, 11, 25]. Let  $\Phi \in H_M^r$  be represented as  $\Phi = R^{(1)} + \dots + R^{(r)}$  by rank-1 functions  $R^{(k)} \in H_M^1$ . The main idea of successive minimization is to first determine  $R^{(1)}$  as solution of  $\min_{R^{(1)} \in \mathcal{T}_M^1} A_M^u(R^{(1)})$ . In a successive step  $R^{(2)}$  is determined as solution of  $\min_{R^{(2)} \in \mathcal{T}_M^1} A_M^u(R^{(1)} + R^{(2)})$ . This is continued successively until  $\Phi$  is determined. This idea is formalized in Algorithm 3.

---

#### Algorithm 3 Successive minimization

---

- 1: **Input:** rank  $r$
  - 2: **Output:** approximation  $\Phi \approx \operatorname{argmin}_{\Phi \in \mathcal{T}_M^r} A_M^u(\Phi)$
  - 3:  $\Phi = 0$
  - 4: **for**  $k = 1, \dots, r$
  - 5:      $R^{(k)} = \operatorname{argmin}_{R \in \mathcal{T}_M^1} A_M^u(\Phi + R)$
  - 6:      $\Phi = \sum_{i=1}^k R^{(i)}$
- 

### 3.4 Alternating least squares

In the successive minimization algorithm 3, we need to solve minimization problems of the form

$$\min_{R \in \mathcal{T}_M^1} A_M^u(\Phi + R) \quad (9)$$

for given  $\Phi \in H_M^r$ . In the following, we introduce an alternating least squares algorithm [4, 6, 19, 27] to solve such minimization problems. The main idea is to approximate the solution of (9) by an iterative scheme which amounts to solving a sequence of small-dimensional linear problems. We start from an initial  $R(\boldsymbol{\eta}) := \Pi_{\mathbf{s} \in S_M} R_{\mathbf{s}}(\eta_{\mathbf{s}})$ . We first minimize  $A^u(\Phi + R)$  only with respect to a selected  $R_{\mathbf{s}_0} : \{0, 1\} \rightarrow \mathbb{R}$  for some  $\mathbf{s}_0 \in S_M$  leaving the other  $R_{\mathbf{s}}, \mathbf{s} \neq \mathbf{s}_0$  fixed. By partially evaluating  $A^u(\Phi + R)$  for all terms not depending on  $R_{\mathbf{s}_0}$ , we obtain that  $\min_{R_{\mathbf{s}_0} \in \{\{0, 1\} \rightarrow \mathbb{R}\}} A^u(\Phi + R)$  with  $R(\boldsymbol{\eta}) := R_{\mathbf{s}_0}(\eta_{\mathbf{s}_0}) \Pi_{\mathbf{s} \in S_M \setminus \{\mathbf{s}_0\}} R_{\mathbf{s}}(\eta_{\mathbf{s}})$  is equivalent to a quadratic optimization problem

$$\min_{R_{\mathbf{s}_0} \in \{\{0, 1\} \rightarrow \mathbb{R}\}} \alpha_1 R_{\mathbf{s}_0}(1)^2 + \alpha_2 R_{\mathbf{s}_0}(0)^2 + \alpha_3 R_{\mathbf{s}_0}(1) R_{\mathbf{s}_0}(0) + \alpha_4 R_{\mathbf{s}_0}(1) + \alpha_5 R_{\mathbf{s}_0}(0) + \alpha_6,$$

with constants  $\alpha_1, \dots, \alpha_6 \in \mathbb{R}$  depending on the fixed  $R_{\mathbf{s}}, \mathbf{s} \neq \mathbf{s}_0$  and  $\Phi$ . This quadratic optimization problem always admits a unique optimal  $R_{\mathbf{s}_0}$ , which is given by  $R_{\mathbf{s}_0}(1) = a$  and  $R_{\mathbf{s}_0}(0) = b$ , where  $a, b \in \mathbb{R}$  are the solution of the linear system

$$\begin{pmatrix} 2\alpha_1 & \alpha_3 \\ \alpha_3 & 2\alpha_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -\alpha_4 \\ -\alpha_5 \end{pmatrix}.$$

This allows us to optimize  $A_M^u(\Phi + R)$  with respect to individual  $R_{\mathbf{s}_0}$ . By alternating the the selected  $\mathbf{s}_0 \in S_M$ , we obtain the alternating least squares algorithm, which is formalized in Algorithm 4.

**Remark 3.2.** To compute the constants  $\alpha_i$ , we can either explicitly implement the partial evaluations of  $A_M^u(\Phi + R)$ . Alternatively, we can treat  $A_M^u(\Phi + R)$  as a function in  $\mathbb{R}^2 \rightarrow \mathbb{R}$  depending on the values  $R_{\mathbf{s}_0}(0)$  and  $R_{\mathbf{s}_0}(1)$ . We know that this function is a multivariate-polynomial of the form  $\alpha_1 R_{\mathbf{s}_0}(1)^2 + \alpha_2 R_{\mathbf{s}_0}(0)^2 + \alpha_3 R_{\mathbf{s}_0}(1) R_{\mathbf{s}_0}(0) + \alpha_4 R_{\mathbf{s}_0}(1) + \alpha_5 R_{\mathbf{s}_0}(0) + \alpha_6$ . The constants can be computed using multivariate-polynomial interpolation in six points. This interpolation has the advantages that it is non-intrusive and that the evaluations of  $A_M^u(\Phi + R)$  can be performed efficiently using the ideas of Section 3.2.

**Remark 3.3.** Throughout this work, we use approximations in the CP-format. The matrix product state representation [28] also known as tensor train format [26] would present an alternative low-rank format, for which minimization problems can also be solved using alternating algorithms [11, 14, 33].



---

**Algorithm 4** Alternating least squares

---

- 1: **Input:** initial functions  $R_s^0 : \{0, 1\} \rightarrow \mathbb{R}$  for  $s \in S_M$ , function  $\Phi \in H_M^r$ , vector  $\mathbf{u}$ , tolerance  $\varepsilon$
  - 2: **Output:** approximation  $R_{\text{opt}}(\boldsymbol{\eta}) = \Pi_{s \in S_M} R_s(\eta_s)$  of  $\text{argmin}_{R \in \mathcal{T}_M^1} A_M^u(\Phi + R)$
  - 3:  $v_{\text{old}} = \infty$ ,  $v_{\text{new}} = A_M^u(\Phi + R^0)$  with  $R^0(\boldsymbol{\eta}) := \Pi_{s \in S_M} R_s^0(\eta_s)$
  - 4:  $\forall s \in S_M$ ,  $R_s := R_s^0$ .
  - 5: **while**  $|v_{\text{old}} - v_{\text{new}}| > \varepsilon |v_{\text{new}}|$ .
  - 6:      $v_{\text{old}} = v_{\text{new}}$
  - 7:     **for**  $s_0 \in S_M$
  - 8:          $R_{s_0} = \text{argmin}_{\tilde{R}_{s_0} : \{0,1\} \rightarrow \mathbb{R}} A^u(\Phi + \tilde{R})$  where  $\tilde{R}(\boldsymbol{\eta}) = \tilde{R}_{s_0}(\eta_{s_0}) \Pi_{s \in S_M \setminus \{s_0\}} R_s(\eta_s)$  for all  $\boldsymbol{\eta} = (\eta_s)_{s \in S_M}$
  - 9:      $v_{\text{new}} = A_M^u(\Phi + R)$
- 

### 3.5 Monte Carlo methods

Let  $\Phi \in H_M^r$  denote an approximation of the solution of (4). In the following, we discuss how to evaluate  $A_{M,\ell}^u(\Phi)$ . In the definition of  $A_{M,\ell}^u$  the function

$$f(\boldsymbol{\eta}) := \sum_{k=1}^K p_k \left( (1 - \eta_{\mathbf{v}_k}) (\mathbf{u} \cdot \mathbf{v}_k + \Psi(\boldsymbol{\eta}^{0, \mathbf{v}_k}) - \Psi(\boldsymbol{\eta}))^2 + \frac{1}{2} \sum_{\substack{\mathbf{y} \in S \setminus \{\mathbf{0}\} \\ \mathbf{y} + \mathbf{v}_k \neq \mathbf{0}}} (\Psi(\boldsymbol{\eta}^{\mathbf{y} + \mathbf{v}_k, \mathbf{y}}) - \Psi(\boldsymbol{\eta}))^2 \right) \quad (10)$$

is evaluated for all  $\boldsymbol{\eta} \in C_{M,\ell}$ . For larger  $N$  and most choices of  $\ell$  evaluating  $|C_{M,\ell}| = \binom{N}{\ell}$  terms is intractable. We thus propose to use a Monte Carlo method [23] to approximate  $A_{M,\ell}^u(\Phi)$ .

In a naive Monte Carlo method we compute  $N_s$  samples  $\boldsymbol{\eta}^{(i)} \in C_{M,\ell}$  for  $1 \leq i \leq N_s$  and replace  $\frac{1}{|C_{M,\ell}|} \sum_{\boldsymbol{\eta} \in C_{M,\ell}}$  by  $\frac{1}{N_s} \sum_{i=1}^{N_s} \boldsymbol{\eta}^{(i)}$ . In Algorithm 5, we define a Monte Carlo method with additional variance reduction [12], which as demonstrated in Section 4 reduces the number of samples needed to obtain a given approximation accuracy. The main idea is to observe that the sites  $\mathbf{v}_1, \dots, \mathbf{v}_K$  play a special role since they are the most relevant for jumps of the tagged particle. Instead of sampling the  $\boldsymbol{\eta}$  uniformly in  $C_{M,\ell}$ , we now ensure that all possible states of the sites  $\mathbf{v}_1, \dots, \mathbf{v}_K$  are occurring equally often in the set of sample points.

---

**Algorithm 5** Monte Carlo method with variance reduction

---

- 1: **Input:** function  $f : C_{M,\ell} \rightarrow \mathbb{R}$ , parameter  $\tilde{N}_s$
  - 2: **Output:** approximation  $S$  of  $\frac{1}{|C_{M,\ell}|} \sum_{\boldsymbol{\eta} \in C_{M,\ell}} f(\boldsymbol{\eta})$
  - 3:  $S = 0$
  - 4: **for**  $\tilde{\mathbf{w}}^{(1)} \in \{0, 1\}^K$
  - 5:      $n_1 := \|\tilde{\mathbf{w}}^{(1)}\|_1$ ,  $n_2 := \ell - n_1$
  - 6:     **if**  $0 \leq n_2 \leq N - \ell$
  - 7:          $\tilde{S} = 0$
  - 8:         **for**  $i = 1, \dots, \tilde{N}_s$
  - 9:             Sample  $\tilde{\mathbf{w}}^{(2)} \in \{0, 1\}^{S_M \setminus \{\mathbf{v}_1, \dots, \mathbf{v}_K\}}$  such that  $\|\tilde{\mathbf{w}}^{(2)}\|_1 = n_2$ .
  - 10:             Construct  $\boldsymbol{\eta} \in C_{M,\ell}$  such that  $\boldsymbol{\eta}(\mathbf{s}) = \begin{cases} \tilde{\mathbf{w}}_k^{(1)} & \mathbf{s} = \mathbf{v}_k \\ \tilde{\mathbf{w}}_{\mathbf{s}}^{(2)} & \text{otherwise} \end{cases}$ .
  - 11:              $\tilde{S} = \tilde{S} + f(\boldsymbol{\eta})$
  - 12:              $S = S + \binom{N-K}{n_2} \cdot \tilde{S} / \tilde{N}_s$
  - 13:  $S = \frac{S}{\binom{N}{\ell}}$
-

**Remark 3.4.** In line 9 of Algorithm 5, we sample from the set  $\{\vec{\mathbf{w}}^{(2)} \in \{0, 1\}^{S_M \setminus \{\mathbf{v}_1, \dots, \mathbf{v}_k\}} \mid \|\vec{\mathbf{w}}^{(2)}\|_1 = n_2\}$  which contains  $\binom{N-K}{n_2}$  elements. When the number of elements in this set is smaller than  $\tilde{N}_s$ , we can compute  $\tilde{S}$  based on all possible  $\vec{\mathbf{w}}^{(2)}$  instead of sampling  $\tilde{N}_s$  times. This decreases the variance of the approximation further and reduces the number of required evaluations of  $f$ .

**Remark 3.5.** Lines 12 and 13 of Algorithm 5 potentially lead to stability issues in floating point arithmetic when  $\binom{N-K}{n_2}$  or  $\binom{N}{\ell}$  are large. This problem can be mitigated using the log-sum-exp trick.

## 4 Numerical Experiments

In the following numerical experiments <sup>1</sup>, we consider the a jumping model with  $K = 4$  displacement vectors  $\mathbf{v}_1 = (1, 0)$ ,  $\mathbf{v}_2 = (-1, 0)$ ,  $\mathbf{v}_3 = (0, 1)$ , and  $\mathbf{v}_4 = (0, -1)$  and with associated probability  $p_k = 1/4$  for  $k = 1, 2, 3, 4$ . All computing times are measured without parallelization in MATLAB R2018b on a Lenovo Thinkpad T480s with Intel Core i7-8650U CPU and 15.4 GiB RAM. Note that both the sampling Algorithm 1 and Algorithm 5 as well as the evaluation of the different Frobenius-norms in Equation (7) can be parallelized to speed up computations.

### 4.1 Computation of the self-diffusion coefficient

#### 4.1.1 Solving the optimization problem

In the following, we analyze the numerical approximation of the self-diffusion coefficient by solving the optimization problem (3).

**Low-rank approximation error** First, we study how the rank  $r$  affects  $\min_{\Phi \in H_M^r} A_M^{\mathbf{u}}(\Phi)$  for  $\mathbf{u} = (1, 0)$ . Let  $\Phi_{\text{ALS}}^{r, \mathbf{u}} \in H_M^r$  denote the function obtained by using Algorithm 3 with rank  $r$ , where the minimization problem in each iteration is solved using Algorithm 4. The initial functions  $R_s^0$  in Algorithm 4 are selected randomly by assigning random values drawn from the uniform distribution on  $[0, 1]$  to  $R_s^0(0), R_s^0(1)$ . In Algorithm 4, we set  $\varepsilon = 10^{-12}$  and additionally stop after line 8 has been executed 420 times.

We depict the error of  $\Phi_{\text{ALS}}^{r, \mathbf{u}}$  for various  $r$  in Figure 2. The computation with  $r = 10$  takes 4 minutes for  $M = 1$  and 82 minutes for  $M = 2$ . We observe that the error decays quickly with increasing rank. In particular, for  $M = 1$  an approximation with  $r = 1$  is less than 0.1% away from a direct least squares solution of the minimization problem (4). Such a direct solution is intractable for  $M = 2$ .

**Sampling-based evaluation** Let  $M = 2$  and  $\mathbf{u} = (1, 0)$ . For a given approximate solution  $\Phi_{\text{ALS}}^{10, \mathbf{u}}$ , we want to evaluate  $A_{2, \ell}^{\mathbf{u}}(\Phi_{\text{ALS}}^{10, \mathbf{u}})$ . A single evaluation of (10) with  $\Psi$  replaced by  $\Phi_{\text{ALS}}^{10, \mathbf{u}}$  requires on average around  $6.6 \cdot 10^{-4}$  seconds. Computing  $A_{2, \ell}^{\mathbf{u}}(\Phi_{\text{ALS}}^{10, \mathbf{u}})$  for  $0 \leq \ell \leq N$  directly would require  $2^{24} \approx 1.6 \cdot 10^7$  evaluations of (10), i.e. around 7.5 hours. In Section 3.5, we proposed two Monte Carlo algorithms to obtain approximations of  $A_{2, \ell}^{\mathbf{u}}(\Phi_{\text{ALS}}^{10, \mathbf{u}})$ . We visualize the variance in the approximation obtained by these algorithms in Figure 3. The studied quantity  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  approximates  $\int_0^1 \text{Tr}(\mathbb{D}_s(\bar{\rho})) d\bar{\rho}$ , where  $\text{Tr}$  denotes the trace operator. Algorithm 5 clearly leads to a variance reduction. With only  $10^5$  samples, which can be evaluated in about one minute, we can already reach a variance of  $10^{-6}$ .

#### 4.1.2 Estimation of long-time mean square deviation

In the following, we study estimating the long-term limit using Algorithm 1.

Figure 4 motivates our choice of  $T = 300$  and  $\hat{N}_s = 30\,000$  for the following numerical experiments. The error caused by stopping with  $T = 300$  is negligible compared to the stochastic variance when stopping with  $\hat{N}_s = 30\,000$ .

<sup>1</sup>The code to reproduce all experiments is available on <https://github.com/cstroessner/SelfDiffusion>

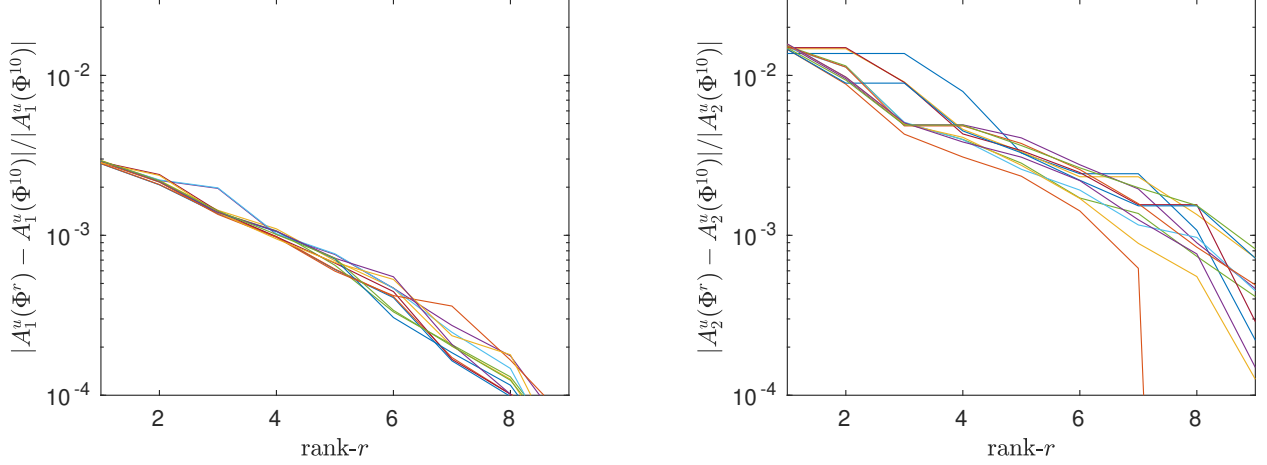


Figure 2: Algorithm 3 yields successive approximations  $\Phi_{\text{ALS}}^{r,\mathbf{u}} \in H_M^r$ . For  $r \in \{1, \dots, 9\}$ , we plot the relative error of  $\Phi_{\text{ALS}}^{k,\mathbf{u}}$  compared to  $\Phi_{\text{ALS}}^{10,\mathbf{u}}$ . We repeat this experiment 12 times with different random initial functions in Algorithm 4. The resulting relative errors are displayed in different colors. Left:  $M = 1$ . Right:  $M = 2$ .

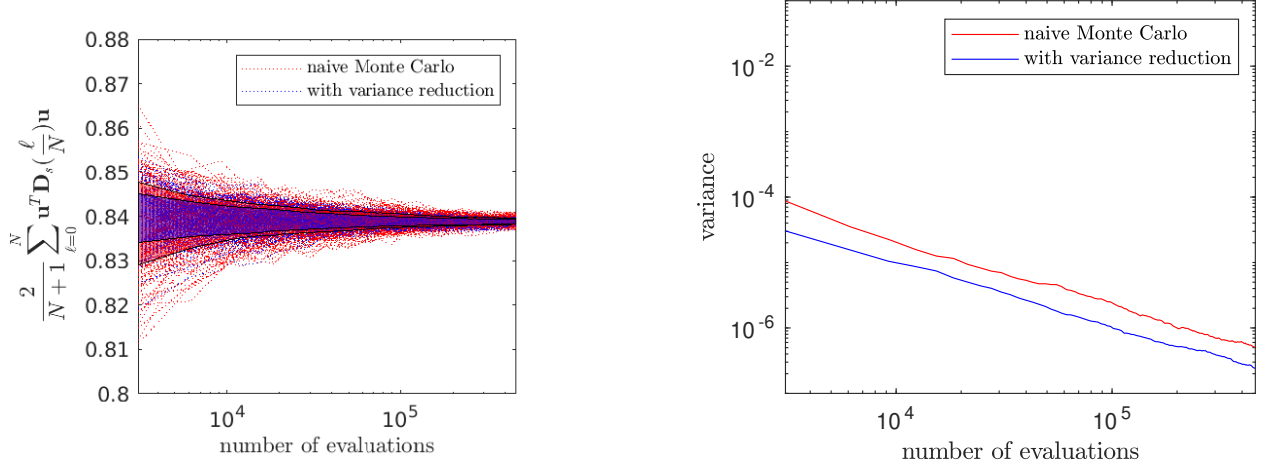


Figure 3: For  $M = 2$  and  $\mathbf{u} = (1, 0)$  we compute  $\Phi_{\text{ALS}}^{10,\mathbf{u}}$  using Algorithm 3. We then approximate  $\mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u} = 2A_{2,\ell}^{\mathbf{u}}(\Phi_{\text{ALS}}^{10,\mathbf{u}})$  for  $0 \leq \ell \leq N$  using sampling as in Section 3.5. We sample  $\boldsymbol{\eta} \in C_{M,\ell}$  using two different approaches; from the uniform distribution (naive Monte Carlo) and using Algorithm 5 (with variance reduction). We use a total of 460 800 evaluations of (10) for both sampling methods. After every 3 072 evaluations, we evaluate  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  based on the current approximation. The whole experiment is repeated 250 times. Left: Evolution of  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  with increasing number of samples. The shaded areas mark one standard deviation from the mean for the respective algorithm. Right: Evolution of the variance of  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  with increasing number of samples.

In Figure 5 we analyze how the parameter  $\hat{N}_s$  affects an approximation of  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$ . We observe that the quantity of interest converges to the same value around 0.84 as in Figure 3. Note that a single run with  $\hat{N}_s = 30\,000$  requires 33 minutes and reaches a variance of  $10^{-4}$ . Compared to the results in Figure 3 the computational time needed to achieve the same variance with Algorithm 1 compared to Algorithm 5 is much higher.

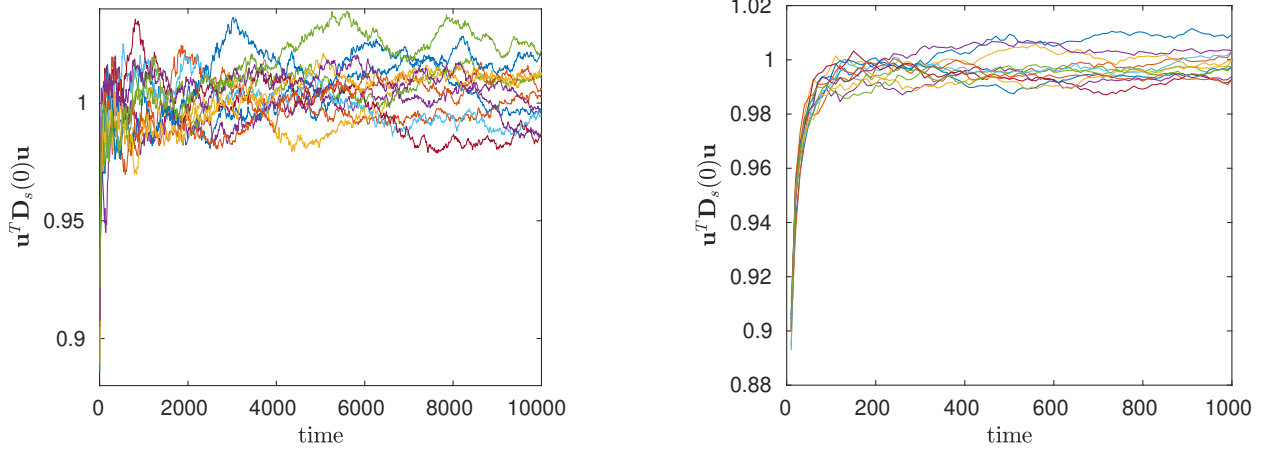


Figure 4: We run Algorithm 1 with  $M = 2$ ,  $\mathbf{u} = (1, 0)$ ,  $\ell = 0$  and different values for  $T$  and  $\hat{N}_s$ . For each setting, we plot how the approximation of the value  $\mathbf{u}^T \mathbb{D}_s(0) \mathbf{u}$ , which is known to be equal to 1, evolves over time for 12 random initializations visualized in different colors. Left:  $\hat{N}_s = 10\,000$ ,  $T = 10\,000$ . Right:  $\hat{N}_s = 100\,000$ ,  $T = 1\,000$ .

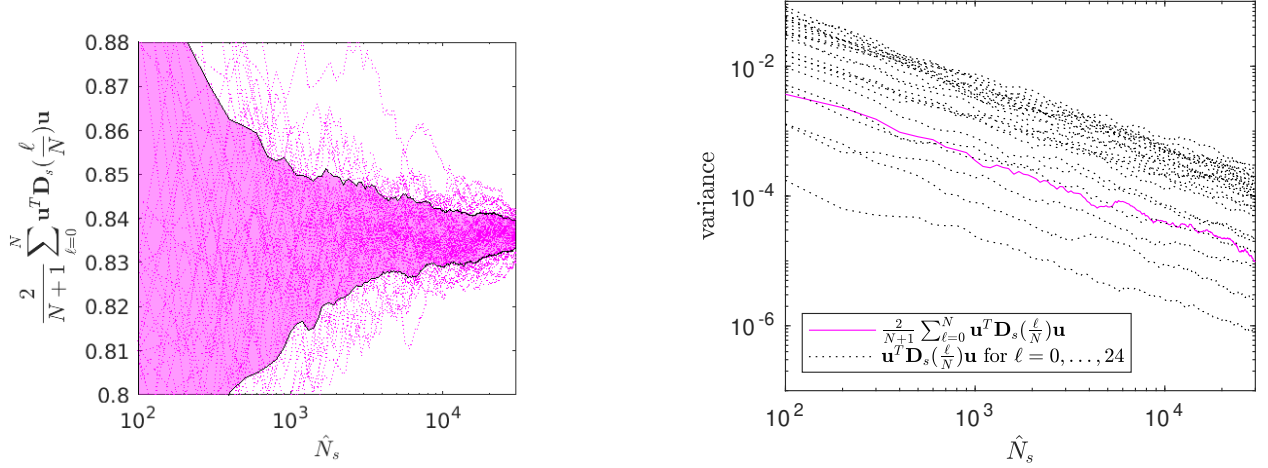


Figure 5: We run Algorithm 1 with  $T = 300$ ,  $\mathbf{u} = (1, 0)$  for  $\ell = 0, \dots, 24$  to approximate  $\mathbb{D}_s(\frac{\ell}{N})$ . We use  $\hat{N}_s = 30\,000$  and store intermediate values for  $\hat{N}_s = 100, 200, 300, \dots$ . Based on these approximations we compute  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$ . This procedure is repeated 50 times with different random initializations. Left: Evolution of  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  with increasing values of  $\hat{N}_s$ . The shaded area marks one standard deviation from the mean. Right: Evolution of the variance of  $\frac{2}{N+1} \sum_{\ell=0}^N \mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  with increasing values of  $\hat{N}_s$ . We additionally display the evolution of the variance of  $\mathbf{u}^T \mathbb{D}_s(\frac{\ell}{N}) \mathbf{u}$  for individual values of  $\ell$ .

#### 4.1.3 Comparison of algorithms

We now compare the two approaches of approximating the self-diffusion coefficient. We evaluate evaluating  $\mathbf{u}^T \mathbb{D}_s(\ell/N) \mathbf{u}$  for  $\mathbf{u} \in \{(1, 0), (0, 1), (1, 1)\}$  to recover all entries of the symmetric matrix  $\mathbb{D}_s(\ell/N)$ . Interpolation of the entries yields an approximation of  $\mathbb{D}_s(\rho)$  for  $\rho \in [0, 1]$ . Alternatively, interpolation in the space of symmetric positive definite matrices could be used [24].

For the following experiments, we use Algorithm 1 with a slight modification to approximate  $\mathbf{u}^T \mathbb{D}_s(\bar{\rho}) \mathbf{u}$  for different values of  $\mathbf{u}$  simultaneously. This is achieved by keeping track of different  $\alpha$  for different  $\mathbf{u}$  in

lines 15 and 16. The optimization problem (4) needs to be solved for every  $\mathbf{u}$  separately, but the solution can be evaluated for different  $\ell$ . In contrast, the sampling in Algorithm 1 needs to be redone completely for every  $\ell$ .

In Figure 6, we display the trace of  $\mathbb{D}_s(\rho)$  computed using different approaches. Computing the whole matrix  $\mathbb{D}_s(\ell/N)$  for all  $0 \leq \ell \leq N$  took 3 minutes (37 minutes) for the approach based on solving the minimization problem, whereas using Algorithm 1 took 15 minutes (45 minutes) for  $M = 1$  ( $M = 2$ ). The much higher variance of Algorithm 1 for approximating the long-term limit leads to clearly visible changes in the graph of trace at the values  $\rho = \frac{\ell}{N}$ . The sampled solution of the minimization problem changes less due to a smaller sampling variance in Algorithm 5.

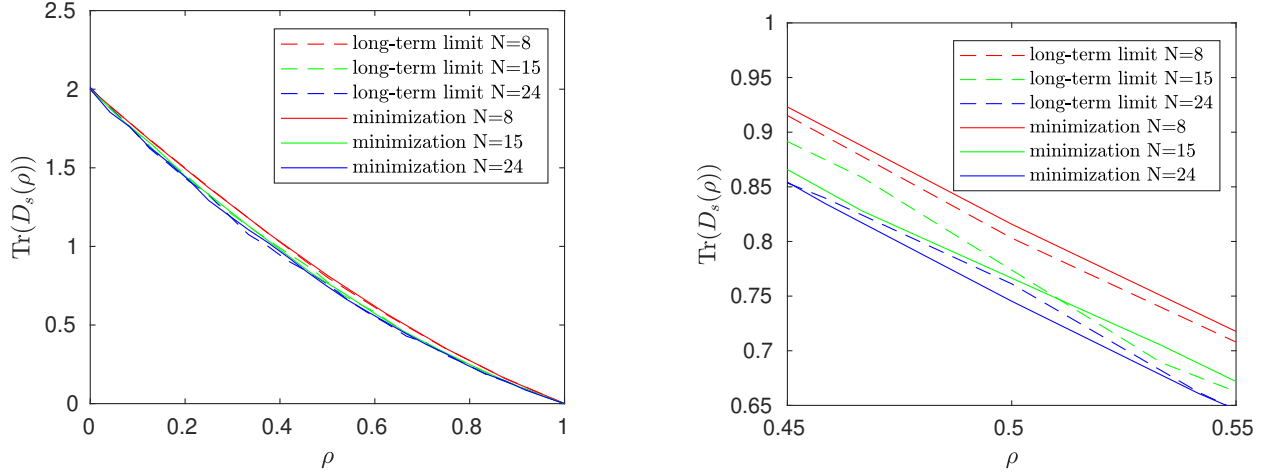


Figure 6: For  $M = 1$  ( $N = 8$ ) and  $M = 2$  ( $N = 24$ ) we solve the minimization problem (4) using Algorithm 3 with  $r = 3$ . We evaluate  $A_{M,\ell}^{\mathbf{u}}$  directly for  $M = 1$ , whereas for  $M = 2$  we use Algorithm 5 with  $\tilde{N}_s = 100$ . We compare to Algorithm 1 to estimate the long-term limit (2) with  $\hat{N}_s = 30\,000$  and  $T = 300$  for  $M = 1$  and  $M = 2$  to compute  $\mathbb{D}_s(\ell/N)$ . Repeating this for  $\mathbf{u} \in \{(1,0), (1,1), (0,1)\}$  yields  $\mathbb{D}_s(\ell/N)$ . Element-wise linear interpolation allows us to evaluate  $\mathbb{D}_s(\rho)$  for  $\rho \in [0,1]$ . We plot the trace of  $\mathbb{D}_s(\rho)$ . In addition, we run both algorithms on a  $4 \times 4$  periodic grid ( $N = 15$ ) using the same parameters as for  $N = 24$ . Note that this grid lies in between the  $3 \times 3$  grid for  $M = 1$  and the  $5 \times 5$  grid for  $M = 2$ . Right: Zoom on part of the graphs.

## 5 Conclusion

In this work, we compare two different approaches to compute the self-diffusion matrix: one is based on the solution of a minimization problem (1) and the other on the interpretation of this coefficient as the long-time limit of the mean-square deviation of the tagged particle process (2).

The second formulation is traditionally used in standard Monte Carlo methods. In this work, we propose a novel method which consists of solving a high-dimensional minimization problem using low-rank solutions combined with an alternating optimization scheme. Our numerical experiments demonstrate that our approach is very advantageous compared to the classical Monte Carlo method. Indeed, with the same amount of computational resources, the variance in the obtained self-diffusion matrix is much smaller when using our new approach. Some numerical issues related to the new method we present here are still to be solved when applying this method to three-dimensional problems or larger values of  $M$  and we intend to investigate them in a future work.

## Acknowledgments

The work presented here was initiated during the CEMRACS 2021 summer school at CIRM, Luminy, Marseille. The authors acknowledge funding by the ANR project COMODO (ANR-19-CE46-0002), the Center on Energy and Climate Change (E4C) and the I-Site FUTURE.

The authors would like to thank Mi-Song Dupuy, Guillaume Enchéry, Daniel Kressner and Dominik Schmid for stimulating discussions on this work.

## References

- [1] A. AMMAR, F. CHINESTA, E. CUETO, AND M. DOBLARÉ, *Proper generalized decomposition of time-multiscale models*, Internat. J. Numer. Methods Engrg., 90 (2012), pp. 569–596.
- [2] G. BEYLKIN AND M. J. MOHLENKAMP, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.
- [3] O. BLONDEL AND C. TONINELLI, *Kinetically constrained lattice gases: tagged particle diffusion*, Ann. Inst. Henri Poincaré Probab. Stat., 54 (2018), pp. 2335–2348.
- [4] J. D. CARROLL AND J.-J. CHANG, *Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of “Eckart-Young” decomposition*, Psychometrika, 35 (1970), pp. 283–319.
- [5] D. CHEN, P. CHEN, N. GANTERT, AND D. SCHMID, *Limit theorems for the tagged particle in exclusion processes on regular trees*, Electron. Commun. Probab., 24 (2019), pp. 1–10.
- [6] L. DE LATHAUWER AND D. NION, *Decompositions of a higher-order tensor in block terms. III. Alternating least squares algorithms*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1067–1083.
- [7] A. DE MASI AND E. PRESUTTI, *Mathematical methods for hydrodynamic limits*, vol. 1501 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1991.
- [8] H. P. DEUTSCH AND K. BINDER, *Interdiffusion and self-diffusion in polymer mixtures: A Monte Carlo study*, J. Chem. Phys., 94 (1991), pp. 2294–2304.
- [9] R. FERRANDO AND E. SCALAS, *Self-diffusion in a 2D lattice gas with lateral interactions*, Surf. Sci., 281 (1993), pp. 178–190.
- [10] N. GANTERT AND D. SCHMID, *The speed of the tagged particle in the exclusion process on Galton-Watson trees*, Electron. J. Probab., 25 (2020), pp. Paper No. 71, 27.
- [11] L. GRASEDYCK, D. KRESSNER, AND C. TOBLER, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt., 36 (2013), pp. 53–78.
- [12] J. M. HAMMERSLEY AND D. C. HANDSCOMB, *Monte Carlo methods*, Methuen & Co., Ltd., London; Barnes & Noble, Inc., New York, 1965.
- [13] S. HAO AND D. S. SHOLL, *Self-diffusion and macroscopic diffusion of hydrogen in amorphous metals from first-principles calculations*, J. Chem. Phys., 130 (2009), p. 244705.
- [14] S. HOLTZ, T. ROHWEDDER, AND R. SCHNEIDER, *The alternating linear scheme for tensor optimization in the tensor train format*, SIAM J. Sci. Comput., 34 (2012), pp. A683–A713.
- [15] M. JARA, *Finite-dimensional approximation for the diffusion coefficient in the simple exclusion process*, Ann. Probab., 34 (2006), pp. 2365–2381.
- [16] H. A. L. KIERS, *Towards a standardized notation and terminology in multiway analysis*, J. Chemom., 14 (2000), pp. 105–122.
- [17] C. KIPNIS, C. LANDIM, AND S. OLLA, *Hydrodynamical limit for a nongradient system: the generalized symmetric exclusion process*, Commun. Pure Appl. Math., 47 (1994), pp. 1475–1545.
- [18] C. KIPNIS AND S. R. S. VARADHAN, *Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions*, Comm. Math. Phys., 104 (1986), pp. 1–19.
- [19] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.

- [20] T. KOMOROWSKI, C. LANDIM, AND S. OLLA, *Fluctuations in Markov processes*, vol. 345 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer, 2012.
- [21] C. LANDIM, S. OLLA, AND S. R. S. VARADHAN, *Finite-dimensional approximation of the self-diffusion coefficient for the exclusion process*, Ann. Probab., 30 (2002), pp. 483–508.
- [22] T. M. LIGGETT, *Stochastic interacting systems: contact, voter and exclusion processes*, vol. 324 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1999.
- [23] N. METROPOLIS AND S. ULAM, *The Monte Carlo method*, J. Amer. Statist. Assoc., 44 (1949), pp. 335–341.
- [24] M. MOAKHER AND P. G. BATCHELOR, *Symmetric positive-definite matrices: from geometry to applications and visualization*, in Visualization and processing of tensor fields, Math. Vis., Springer, Berlin, 2006, pp. 285–298, 452.
- [25] A. NOUY, *Low-rank methods for high-dimensional approximation and model order reduction*, in Model reduction and approximation, vol. 15 of Comput. Sci. Eng., SIAM, Philadelphia, PA, 2017, pp. 171–226.
- [26] I. V. OSELEDETS, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.
- [27] I. V. OSELEDETS, M. V. RAKHUBA, AND A. USCHMAJEV, *Alternating least squares as moving subspace correction*, SIAM J. Numer. Anal., 56 (2018), pp. 3459–3479.
- [28] D. PEREZ-GARCIA, F. VERSTRAETE, M. M. WOLF, AND J. I. CIRAC, *Matrix product state representations*, Quantum Inf. Comput., 7 (2007), pp. 401–430.
- [29] J. QUASTEL, *Diffusion of color in the simple exclusion process*, Comm. Pure Appl. Math., 45 (1992), pp. 623–679.
- [30] T. ROHWEDDER AND A. USCHMAJEV, *On local convergence of alternating schemes for optimization of convex problems in the tensor train format*, SIAM J. Numer. Anal., 51 (2013), pp. 1134–1162.
- [31] E. SAADA, *A limit theorem for the position of a tagged particle in a simple exclusion process*, Ann. Probab., 15 (1987), pp. 375–381.
- [32] E. SANZ AND D. MARENDUZZO, *Dynamic Monte Carlo versus Brownian dynamics: A comparison for self-diffusion and crystallization in colloidal fluids*, J. Chem. Phys., 132 (2010), p. 194102.
- [33] S. SZALAY, M. PFEFFER, V. MURG, G. BARCZA, F. VERSTRAETE, R. SCHNEIDER, AND O. LEGEZA, *Tensor product methods and entanglement optimization for ab initio quantum chemistry*, Int. J. Quantum Chem., 115 (2015), pp. 1342–1391.
- [34] A. L. THORNEYWORK, R. E. ROZAS, R. P. A. DULLENS, AND J. HORBACH, *Effect of hydrodynamic interactions on self-diffusion of quasi-two-dimensional colloidal hard spheres*, Phys. Rev. Lett., 115 (2015), p. 268301.
- [35] C. TONINELLI, G. BIROLI, AND D. S. FISHER, *Spatial structures and dynamics of kinetically constrained models of glasses*, Phys. Rev. Lett., 92 (2004), p. 185504.