Influence of memory in deterministic walks in random media: Analytical calculation within a mean-field approximation

César Augusto Sangaletti Terçariol[1,2,*] and Alexandre Souto Martinez[1,†]
[1]Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900 14040-901, Ribeirão Preto, São Paulo, Brazil
[2]Centro Universitário Barão de Mauá, Rua Ramos de Azevedo, 423 14090-180, Ribeirão Preto, São Paulo, Brazil
(Received 13 June 2008; published 9 September 2008)

Consider a random medium consisting of N points randomly distributed so that there is no correlation among the distances separating them. This is the random link model, which is the high dimensionality limit (mean-field approximation) for the Euclidean random point structure. In the random link model, at discrete time steps, a walker moves to the nearest point, which has not been visited in the last μ steps (memory), producing a deterministic partially self-avoiding walk (the tourist walk). We have analytically obtained the distribution of the number n of points explored by the walker with memory $\mu=2$, as well as the transient and period joint distribution. This result enables us to explain the abrupt change in the exploratory behavior between the cases $\mu=1$ (memoryless walker, driven by extreme value statistics) and $\mu=2$ (walker with memory, driven by combinatorial statistics). In the $\mu=1$ case, the mean newly visited points in the thermodynamic limit ($N \gg 1$) is just $\langle n \rangle = e = 2.72\ldots$ while in the $\mu=2$ case, the mean number $\langle n \rangle$ of visited points grows proportionally to $N^{1/2}$. Also, this result allows us to establish an equivalence between the random link model with $\mu=2$ and random map (uncorrelated back and forth distances) with $\mu=0$ and the abrupt change between the probabilities for null transient time and subsequent ones.

DOI: 10.1103/PhysRevE.78.031111 PACS number(s): 05.40.Fb, 05.60.−k, 05.90.+m, 02.50.−r

I. INTRODUCTION

Although not as thoroughly studied as random walks in disordered media [1] and complex media [2], deterministic walks in regular [3,4] and disordered media [5–7] present very interesting results, having for instance an application to foraging [8–10]. The memory in random walks has the effect of changing the behavior of the Gaussian asymptotic displacement distribution [11]. Here, we are interested in understanding the effect of memory in a partially self-avoiding deterministic walk algorithm, known as the tourist walk (TW) [12–14]. These walks, that are described below, have been applied to characterize thesaurus [14], as a pattern recognition algorithm [15] and image analysis [16,17].

Consider N points (sites, cities) randomly distributed inside a d-dimensional hypercube with unitary edges. The distance $D_{i,j}$ between any pair of points s_i and s_j is calculated via Euclidean metrics. A walker leaves a given point and moves, at each discrete time step, obeying the deterministic rule of going to the nearest point (shortest Euclidean distance), which has not been visited in the μ preceding steps. This rule produces trajectories with an initial transient part of t steps and a cycle of p steps as a final periodic part. Once trapped in a cycle, the walker does not visit new points any longer. Short transient times and short period cycles limit exploration of the medium by the walker. The analytical results [18] have been obtained for (i) memoryless walkers in the deterministic [19] and stochastic [20,21] versions of the TW and for (ii) deterministic walk with arbitrary memory in one-dimensional systems [22]. Here we consider the memory effect in deterministic walks in a mean-field approximation.

The deterministic TW, with memory $\mu=0$, is trivial since the walker does not move at each time step, so that the transient-time and period joint distribution is simply $S_{0,d}^{(N)}(t,p) = \delta_{t,0}\delta_{p,1}$, where $\delta_{i,j}$ is the Kronecker delta. With memory $\mu=1$, the walker must leave the current site at each time step. The joint distribution $S_{1,d}^{(N)}(t,p)$ is obtained considering the trajectories of a tourist leaving from all sites of a given map and statistics is performed for different realizations (maps). For $N \gg 1$, the transient-time and period joint distribution is obtained analytically for arbitrary dimensionality [19]: $S_{1,d}^{(\infty)}(t,p) = [(t+I_d^{-1})\Gamma(1+I_d^{-1})/\Gamma(t+p+I_d^{-1})]\delta_{p,2}$, where $\Gamma(z)$ is the gamma function and $I_d = I_{1/4}[1/2,(d+1)/2]$ is the normalized incomplete beta function. This case does not lead to exploration of the random medium since after a very short transient time, the tourist gets trapped in pairs of cities that are mutually nearest neighbors.

Interesting phenomena occur when the memory values are greater or equal to two ($\mu \geq 2$). In this case, the cycle distribution is no longer concentrated at $p_{min} = \mu+1$, but presents a whole spectrum of cycles with period $p \geq p_{min}$, with possible power-law decay [12,14], favoring exploration of the medium by the walker. The elucidation of this intriguing broadening of the cycle period distribution is our main objective in this paper.

As the medium dimensionality d increases, the correlations between the distances $D_{i,j}$ become weaker and weaker so that in the high dimensionality limit ($d \to \infty$), the distances can be considered as independent random variables, uniformly distributed in the interval [0,1] [23–28]. This is the mean-field model named random link (RL), where two Euclidean constraints still remain: (i) the distance from a point to itself is null, $D_{i,i}=0$, and (ii) the forward and backward distances are equal, $D_{i,j}=D_{j,i}$. Breaking these constraints

*cesartercariol@gmail.com
†asmartinez@ffclrp.usp.br

leads to the random map model (RM) [29], which is a mean-field approximation for the Kauffman's model [30]. The neighborhood statistics for these mean-field models have been analytically studied in Ref. [31].

In this paper, we obtain analytical results for the TW, with memory $\mu=2$ in the $d\to\infty$ medium, i.e., the RL approximation. These results enable us to explain the main mechanism which makes the $\mu=1$ and $\mu\geqslant 2$ situations so distinct. Also, they permit us to establish a relationship between the mean-field RL and RM models. The walks with memory $\mu=2$ in the symmetric independent random distance case (RL model) is equivalent to memoryless ($\mu=0$) walks in the asymmetric independent random distance case (RM model), which has been already solved in Ref. [19]. Throughout this relationship between RL and RM models, we show that the decay for the cycle period distribution in the RL model is also a power law $\propto p^{-1}$. Also we are able to explain the reason for the already observed numerically abrupt change in the transient and period joint distribution for null transient $t=0$.

The presentation of these results is briefly sketched in the following. In Sec. II, we calculate the probability $\widetilde{S}_{2,rl}^{(N)}(\widetilde{n})$ for the walker, with memory $\mu=2$, to visit $\widetilde{n}$ distinct sites before the first passage to any already visited site, walking on the RL model with N sites. We start calculating the complementary cumulative distribution $\widetilde{F}_{2,rl}^{(N)}(\widetilde{n})$ (upper-tail distribution). Next, through an analogy to the geometric distribution, we obtain the revisitation $\widetilde{p}_{2,rl}^{(N)}(j)$ (first passage) and exploration $\widetilde{q}_{2,rl}^{(N)}(j)$ probabilities. Using an alternative derivation, we obtain simpler expressions for these probabilities, which leads to a closed analytical expression for $\widetilde{F}_{2,rl}^{(N)}(\widetilde{n})$. In Sec. III, we show that the probability for the walker to get trapped into a cycle when revisiting a site along the trajectory is $2/3$, which is counterintuitive. This result (combined with previous ones) allows us to obtain the complementary cumulative distribution $F_{2,rl}^{(N)}(n)$ for the total number n of visited sites (until the walker enters an attractor) and explain the equivalent between RL and RM models. In Sec. IV, we obtain the joint distribution $S_{2,rl}^{(N)}(t,p)$ of transient time t and cycle period p and show the drastic difference between the $t=0$ and $t\neq 0$ cases. Universal probability distributions are obtained rescaling variables. A final discussion is presented in Sec. V and future studies are proposed.

II. DISTRIBUTION FOR THE NUMBER OF EXPLORED SITES BEFORE THE FIRST PASSAGE

Consider that the walker mentioned in the Introduction of this paper, who performs a walk with memory $\mu=2$ in the RL model with N points, has visited $\widetilde{n}\geqslant 3=\mu+1=\widetilde{n}_{min}$ distinct sites and then revisits one of these sites. Aiming to obtain the distribution $\widetilde{S}_{2,rl}^{(N)}(\widetilde{n})$ of the number $\widetilde{n}$ of sites visited before the first passage, we start calculating the complementary cumulative (upper-tail) distribution

$$\widetilde{F}_{2,rl}^{(N)}(\widetilde{n}) = \sum_{k=\widetilde{n}}^{N} \widetilde{S}_{2,rl}^{(N)}(k),$$

i.e., the probability for the tourist to explore at least $\widetilde{n}$ distinct sites, before the first revisitation.

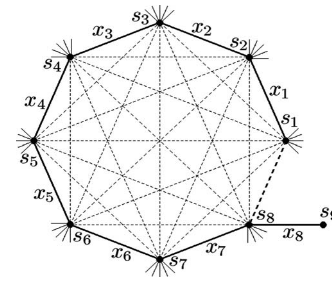

FIG. 1. Schematic representation of a walk with at least $\widetilde{n}=9$ sites visited before the first passage. The walker leaves from the site s_1 and follows the trajectory $s_1,s_2,s_3,\ldots,s_9$.

In the schema of Fig. 1, the tourist leaves from a given site s_1 (first step, $j=1$) and follows the trajectory $s_1,s_2,\ldots,s_{\widetilde{n}}$, exploring $\widetilde{n}=9$ distinct sites, with no revisitation. For $1\leqslant i\leqslant \widetilde{n}-1$, let us denote

(1) x_i the distance between the consecutive sites s_i and s_{i+1} in the trajectory (thick continuous lines of Fig. 1).

(2) $y_{i,k}$ the distances between the site s_i in the trajectory and other sites outside the trajectory (thin continuous lines partially depicted in Fig. 1).

(3) $z_{i,k}$ the distance between the nonconsecutive sites s_i and s_k in the trajectory (dashed lines of Fig. 1).

Using the definition of the RL model, all these distances x_i, $y_{i,k}$, and $z_{i,k}$ have a uniform deviate in the interval [0, 1].

The conditions for the tourist to follow the trajectory $s_1,s_2,\ldots,s_{\widetilde{n}}$ in the first $\widetilde{n}$ steps are as follows:

(1) In the case $\mu=1$ (already solved in Ref. [19]), the distances x_i must obey the relation $x_{\widetilde{n}-1}<x_{\widetilde{n}-2}<\cdots<x_1$, once the tourist stops exploring new sites when $x_{i+1}>x_i$, giving rise to a cycle of period $p=2$. But for the case $\mu=2$ addressed here, each distance x_i may vary unrestrictedly in the interval [0, 1], because the memory $\mu=2$ forbids the tourist to move backward from s_{i+1} to s_i (even if $x_{i+1}>x_i$).

(2) When the tourist is about to walk the distance x_i (and move from s_i to s_{i+1}) there are $N-i$ nonexplored sites at his or her disposal.

(3) For each site s_i, all $N-(\widetilde{n}-1)$ distances $y_{i,k}$ must be greater than x_i. The probability for this to occur is $[\int_{x_i}^{1} dy_{i,k}]^{N-\widetilde{n}+1}=(1-x_i)^{N-\widetilde{n}+1}$. The only exception is the site $s_{\widetilde{n}-1}$, which has $N-\widetilde{n}$ distances $y_{\widetilde{n}-1,k}$ connected to it (see Fig. 1, where $s_{\widetilde{n}-1}$ corresponds to s_8).

(4) To avoid shortcuts and revisits, each distance $z_{i,k}$ must be greater than both x_i and x_k.

These conditions lead to

$$\widetilde{F}_{2,rl}^{(N)}(\widetilde{n}) = \prod_{i=1}^{\widetilde{n}-2}\int_0^1 dx_i (N-i)(1-x_i)^{N-\widetilde{n}+1} \times \int_0^1 dx_{\widetilde{n}-1}(N-\widetilde{n}+1)(1-x_{\widetilde{n}-1})^{N-\widetilde{n}} \times \prod_{i=1}^{\widetilde{n}-3}\prod_{k=i+2}^{\widetilde{n}-1}\int_{\max(x_i,x_k)}^{1} dz_{i,k}. \quad (1)$$

It is worthwhile to mention that we have made no approxi-

TABLE I. Numerical validation of Eq. (1). The columns $\tilde{F}^{(6)}_{2,rl}(\tilde{n})$ and $\tilde{S}^{(6)}_{2,rl}(\tilde{n})$ refer to analytical values and the columns mean and standard error come from numeric simulation. Walks were performed on 300 000 000 maps with $N=6$ points each.

$\tilde{n}$	$\tilde{F}^{(6)}_{2,rl}(\tilde{n})$	$\tilde{S}^{(6)}_{2,rl}(\tilde{n})$	Mean	Standard error	Difference	Difference (in standard error)
3	1	0.15625	0.15624	1×10^{-5}	7×10^{-6}	0.62
4	$\frac{27}{32}$	0.29534	0.29535	1×10^{-5}	2×10^{-5}	1.13
5	$\frac{9\,459}{17\,248}$	0.33785	0.33784	1×10^{-5}	1×10^{-5}	0.82
6	$\frac{107\,301}{509\,600}$	0.21056	0.21056	1×10^{-5}	3×10^{-6}	0.22

mation yet, hence Eq. (1) yields exact results even for small values of N, as Table I shows.

Although Eq. (1) is exact, the function $\max(x_i,x_k)$ in the lower limits of the integrals in $z_{i,k}$ makes it difficult to solve, once one has to consider all possible $(\tilde{n}-1)!$ orderings of distances x_i. In the following, we will consider the thermodynamic limit $(N\gg1)$ and make some considerations to solve Eq. (1).

For a better visualization, notice that the integrals in $z_{i,k}$ refer to the dashed lines in Fig. 1. Observe that from each site exactly $\tilde{n}-4$ dashed lines leave, except for the sites s_1 and $s_{\tilde{n}-1}$, where $\tilde{n}-3$ dashed lines leave, due to the additional distance $z_{1,\tilde{n}-1}$ (thick dashed line in Fig. 1). To obtain a more regular expression, we can eliminate the integral in $z_{1,\tilde{n}-1}$ in Eq. (1) (without any harm) and then each variable x_i appears exactly $\tilde{n}-4$ times as argument of function max ( ). To justify this elimination, notice that, due to the deterministic rule of TW, each distance x_i is the minimum of $N-2$ random variables uniformly distributed in the interval [0,1]. Therefore, its probability density function (PDF) is given by [19]: $g(x_i)=(N-2)(1-x_i)^{N-3}$ and its mean and standard deviation are $\langle x_i\rangle=1/(N-1)\approx1/N$ and $\sigma_{x_i}=\sqrt{(N-2)/[N(N-1)^2]}\approx1/N$ so that in the limit $N\gg1$, x_i assumes values close to 0 and the value of the integral $\int^1_{\max(x_1,x_{\tilde{n}-1})}dz_{1,\tilde{n}-1}$ is close to 1.

Changing the exponent of $x_{\tilde{n}-1}$ from $N-\tilde{n}$ to $N-\tilde{n}+1$, all the variables x_i are raised to the same power. The resulting expression is algebraically symmetric with respect to the variables $x_1,x_2,\ldots,x_{\tilde{n}-1}$, which means that all possible $(\tilde{n}-1)!$ orderings occur with the same probability. Thus, one can consider the specific ordering $x_1<x_2<\cdots<x_{\tilde{n}-1}$ and rewrite Eq. (1) without using the inconvenient function max( ) as follows:

$$\frac{\tilde{F}^{(N)}_{2,rl}(\tilde{n})}{(\tilde{n}-1)!}=\prod_{i=1}^{\tilde{n}-1}(N-i)\int_0^1 dx_1(1-x_1)^{N-\tilde{n}+1}\times\prod_{i=2}^{\tilde{n}-2}\int_{x_{i-1}}^1 dx_i(1-x_i)^{N-\tilde{n}+i-1}\times\int_{x_{\tilde{n}-2}}^1 dx_{\tilde{n}-1}(1-x_{\tilde{n}-1})^{N-3},\qquad(2)$$

where we emphasize that the extra factor $(\tilde{n}-1)!$ takes into account all possible orderings of the variables x_i. The exponent of x_1 may be changed from $N-\tilde{n}+1$ to $N-\tilde{n}$ aiming the exponents of $x_1,x_2,\ldots,x_{\tilde{n}-2}$ to be in an arithmetic series. One then calculates the integrals of Eq. (2) to have

$$\tilde{F}^{(N)}_{2,rl}(\tilde{n})=\frac{(\tilde{n}-1)!(N-1)(N-2)(N-3)\cdots}{(N-2)(2N-4)(3N-7)(4N-11)\cdots}\times\frac{\cdots(N-\tilde{n}+1)}{\cdots\{(\tilde{n}-1)N-[(\tilde{n}-1)\tilde{n}/2+1]\}}=\prod_{k=1}^{\tilde{n}-1}\frac{k(N-k)}{kN-k(k+1)/2-1}=\prod_{j=4}^{\tilde{n}}\frac{N-j+1}{N-j/2-1/(j-1)},\qquad(3)$$

where we have called $j=k+1$ and the lower limit of the product has been changed from $j=2$ to $j=4$ because the factors for $j=2$ and $j=3$ are physically meaningless, as we shall argue in Sec. II A. The distribution of $\tilde{n}$ is calculated from the one step difference of the upper-tail distribution as follows:

$$\tilde{S}^{(N)}_{2,rl}(\tilde{n})=\tilde{F}^{(N)}_{2,rl}(\tilde{n})-\tilde{F}^{(N)}_{2,rl}(\tilde{n}+1)=\left[1-\frac{N-\tilde{n}}{N-(\tilde{n}+1)/2-1/\tilde{n}}\right]\prod_{j=4}^{\tilde{n}}\frac{N-j+1}{N-j/2-1/(j-1)}.\qquad(4)$$

The expression of Eq. (3) is similar to the one obtained for $\mu=1$ [using Eqs. (9) and (10) of Ref. [19] and calling $\tilde{n}=t+2$]: $\tilde{F}^{(N)}_{1,rl}(\tilde{n})=[\prod_{j=3}^{\tilde{n}}\frac{N-j+1}{N-j/2}]/(\tilde{n}-1)!$. The main difference is the presence of the factor $1/(\tilde{n}-1)!$, because, for $\mu=1$, one must consider only the specific ordering $x_{\tilde{n}-1}<x_{\tilde{n}-2}<\cdots<x_1$.

At this point we are able to understand the major role played by the memory in this partially self-avoiding walk. For $\mu=1$, the walker must go to the nearest neighbor. The extreme value statistics is behind this dynamics. But, for instance, forbidding the walker to return to the last visited site opens up the possibility to go to the first or second nearest neighbor, which transforms the extreme value statistics to the combinatorial statistics. Mathematically, this is expressed by the absence of $(\tilde{n}-1)!$ in Eq. (3).

A. Analogy to the geometric distribution

Making an analogy to the geometric distribution, we can write Eq. (4) as $\tilde{S}_{2,rl}^{(N)}(\tilde{n})=\tilde{p}_{2,rl}^{(N)}(\tilde{n}+1)\Pi_{j=4}^{\tilde{n}}\tilde{q}_{2,rl}^{(N)}(j)$, where

$$\tilde{q}_{2,rl}^{(N)}(j)=\frac{N-j+1}{N-j/2-1/(j-1)} \tag{5}$$

is the exploration probability in the jth step and $\tilde{p}_{2,rl}^{(N)}(j)=1-\tilde{q}_{2,rl}^{(N)}(j)$ is the revisitation probability in the jth step. We remark that the expression of Eq. (5) is similar to the one obtained for $\mu=1$ [adapting Eqs. (9) and (10) of Ref. [19] from their original concept of subsistence probability to the concept of exploration probability handled here]: $(j-1)\tilde{q}_{1,rl}^{(N)}(j)=(N-j+1)/(N-j/2)$. The main difference is the extra factor $j-1$, which is a consequence of the restriction $x_{\tilde{n}-1}<x_{\tilde{n}-2}<\cdots<x_1$. This extra factor explains the abrupt change in the exploratory behavior between $\mu=1$ and $\mu=2$ cases: on one hand, for $\mu=1$ the exploration probability (in the thermodynamic limit) decreases as $1/(j-1)$ along the trajectory; on the other hand, for $\mu=2$ this probability tends to 1, when $N\to\infty$.

Once the memory $\mu=2$ assures the tourist to explore at least $\tilde{n}_{\min}=\mu+1=3$ sites, it only makes sense to define exploration probability from the fourth step. In fact, for the first step $(j=1)$ Eq. (5) does not have a defined value, for the second step it yields $\tilde{q}_{2,rl}^{(N)}(2)=(N-1)/(N-2)>1$, which is absurd, and for the third step $\tilde{q}_{2,rl}^{(N)}(3)=1$. To take into account the proper physical content, we previously changed the lower limit of the products of Eq. (3) from $j=2$ to $j=4$. It is interesting to mention that for the step $j=N+1$ (after the tourist explores all the N sites), Eq. (5) correctly yields $\tilde{q}_{2,rl}^{(N)}(N+1)=0$.

Since in the jth step there are $j-3$ sites equally probable to be revisited and $\tilde{p}_{2,rl}^{(N)}(j)$ is the probability for the tourist to revisit any one of these sites, in the limit $N\gg j\gg 1$ the probability $\tilde{p}_{rl}$ for the tourist to revisit a specific site s_k is

$$\tilde{p}_{rl}=\frac{1}{j-3}\tilde{p}_{2,rl}^{(N)}(j)=\frac{1}{j-3}\frac{j/2-1-1/(j-1)}{N-j/2-1/(j-1)}\approx\frac{1}{2N}, \tag{6}$$

which is half the probability for him or her to explore a specific new site [namely, $\tilde{q}_{rl}=1/(N-j)\approx 1/N$].

B. Alternative derivation

In what follows, we obtain simpler expressions for the first passage and exploration probabilities for $N\gg 1$, via an alternative reasoning. From these probabilities, we obtain closed analytical expressions for $\tilde{F}_{2,rl}^{(N)}(\tilde{n})$.

1. First passage and exploration probabilities

Suppose that the tourist has traveled along the trajectory sites $s_1,s_2,\ldots,s_{\tilde{n}}$ $(\tilde{n}\geq 3)$ without any site revisitation. Let us first retrieve the probability $\tilde{p}_{rl}$ for the tourist to revisit a specific site s_k (outside the exclusion window, i.e., $k\leq\tilde{n}-2$) in the following step. To do this, consider the following constraints (see Fig. 1):

(1) The distance $z_{\tilde{n},k}$ must be smaller than $x_{\tilde{n}}$.

(2) Once in the $(k+1)$th step, the tourist came from site s_k to s_{k+1}; the distance $z_{\tilde{n},k}$ is greater than the distance x_k.

In brief, $z_{\tilde{n},k}$ must vary between x_k and $x_{\tilde{n}}$, so that, $0<x_k<z_{\tilde{n},k}<x_{\tilde{n}}<1$.

Once the PDF of each distance x_i is $g(x_i)=(N-2)(1-x_i)^{N-3}$ and $z_{\tilde{n},k}$ has uniform deviate (by definition of the RL model), for $N\gg 1$, the probability $\tilde{p}_{rl}$ is given by $\tilde{p}_{rl}=P(x_k<z_{\tilde{n},k}<x_{\tilde{n}})=\int_0^1 dx_k(N-2)(1-x_k)^{N-3}\int_{x_k}^1 dx_{\tilde{n}}\,(N-2)\,(1-x_{\tilde{n}})^{N-3}\int_{x_k}^{x_{\tilde{n}}}dz_{\tilde{n},k}=(N-2)/[(N-1)(2N-3)]\approx 1/(2N)$, which agrees with Eq. (6).

For a generic step j there are $j-3$ sites susceptible to be revisited so that the first passage and exploration probabilities for this step are $\tilde{p}_{2,rl}^{(N)}(j)=(j-3)/(2N)=1-\tilde{q}_{2,rl}^{(N)}(j)$, which is an approximation for Eq. (5), leading to

$$\tilde{F}_{2,rl}^{(N)}(\tilde{n})=\prod_{j=4}^{\tilde{n}}\tilde{q}_{2,rl}^{(N)}(j)=\prod_{j=4}^{\tilde{n}}\left[1-\frac{j-3}{2N}\right]$$
$$=\frac{\Gamma(2N)}{\Gamma(2N-\tilde{n}+3)(2N)^{\tilde{n}-3}}, \tag{7}$$

which is a closed analytical form for Eq. (3).

2. Exponential form (cumulative half Gaussian)

In the limit $N\gg 1$, the exploration probability may be written as $\tilde{q}_{2,rl}^{(N)}(j)=[1-1/(2N)]^{j-3}$, so that Eq. (7) assumes its exponential form

$$\tilde{F}_{2,rl}^{(N)}(\tilde{n})=\prod_{j=4}^{\tilde{n}}\tilde{q}_{2,rl}^{(N)}(j)$$
$$=\left(1-\frac{1}{2N}\right)^{\tilde{\omega}}\approx e^{-\tilde{\omega}/(2N)}$$
$$=e^{-[(\tilde{n}-3)^2/(4N)][1+1/(\tilde{n}-3)]}, \tag{8}$$

where

$$\tilde{\omega}=\sum_{j=4}^{\tilde{n}}(j-3)=\frac{(\tilde{n}-2)(\tilde{n}-3)}{2} \tag{9}$$

has a simple physical interpretation: it is just the number of distances $z_{i,k}$ between nonconsecutive sites of trajectory. Notice that the trajectory depicted in Fig. 1 is topologically equivalent to a $(\tilde{n}-1)$-sided polygon, which has $(\tilde{n}-1)(\tilde{n}-4)/2$ diagonals. All these diagonals plus the side $s_1s_{\tilde{n}-1}$ totalize $\tilde{\omega}=(\tilde{n}-2)(\tilde{n}-3)/2$ paths (dashed lines in Fig. 1), which allow revisitation.

For $\tilde{n}-3\gg 1$, one can disregard $1/(\tilde{n}-3)$ in Eq. (8), leading to a half Gaussian: $y=\tilde{F}_{2,rl}^{(N)}(\tilde{n})=e^{-[(\tilde{n}-3)/\sqrt{2N}]^2/2}$, indicating that the scaled variable is $x=(\tilde{n}-\tilde{n}_{\min})/\sqrt{2N}$ with $\tilde{n}_{\min}=\mu+1=3$, leading to the universal curve $y=e^{-x^2/2}$, with $x\geq 0$. We only have kept $\tilde{n}_{\min}$ to compare to a possible generalization of these calculations for the case of short memory $\mu\ll N$.

III. DISTRIBUTION OF THE TOTAL NUMBER OF EXPLORED SITES

Up to this point we have been focused on the number $\tilde{n}$ of sites explored before the first revisitation to a site. In the TW with $\mu=1$, the revisitation to a site implies the tourist has entered an attractor of period $p=2$ [19], but with $\mu=2$, this revisitation does not imply capture. In what follows, we calculate the probability p_t for the tourist to get trapped during a revisitation to a site and then obtain the capture $p_{2,rl}^{(N)}(j)$ and subsistence $q_{2,rl}^{(N)}(j)$ probabilities and also the upper-tail distribution $F_{2,rl}^{(N)}(n)$ for the number n of sites visited in the whole trajectory.

A. Trapping probability

Consider Fig. 1 and that the tourist has traveled along the trajectory $s_1,s_2,\ldots,s_{\tilde{n}}$ without any site revisitation. Assume that in the following step he or she revisits site s_k (outside the memory window, $k\leqslant\tilde{n}-2$). Due to the deterministic rule, two situations may occur: (i) if $x_k<x_{k-1}$, the tourist moves forward to site s_{k+1} and gets trapped by an attractor of period $p=\tilde{n}-k+1$; (ii) if $x_{k-1}<x_k$, the tourist moves backward to site s_{k-1} and escapes from the attractor. Therefore, the walker trapping or escaping depends on which distance x_{k-1} or x_k is shorter. The only exception is a revisitation to s_1, when the tourist is unconditionally trapped, leading to a trajectory with a null transient time ($t=0$) and a cycle of period $p=\tilde{n}$.

Taking into consideration that all $(\tilde{n}-1)!$ possible orderings of the distances $x_1,x_2,\ldots,x_{\tilde{n}-1}$ are equally probable, one could naively conclude that the trapping probability would be $p_t=P(x_k<x_{k-1})=1/2$. Nonetheless, numerical simulations of this system have refuted this expectation, pointing out that this probability is in fact $p_t=2/3$.

To understand this result, we first show that the probability $P_v(r)$ for the tourist to revisit a specific site s_k is proportional to the rank r occupied by the associated distance x_k (between sites s_k and s_{k+1}) when one reorders the distances $x_1,x_2,\ldots,x_{\tilde{n}-2}$ decreasingly (so that x_k is the rth greatest one). Secondly, we show that the probability $P_t(r)$ for the tourist to get trapped when revisiting the site s_k is proportional to $r-1$. Finally, from $P_v(r)$ and $P_t(r)$ we prove that $p_t=2/3$.

1. Order statistics

Let us recall some tools from the order statistics field. Given a sample of M random variables $X_1,X_2,\ldots,X_M$, reorder them so that $X_{(1)}>X_{(2)}>\cdots>X_{(M)}$, where the index r between parentheses is the rank of $X_{(r)}$. If X has PDF $g(x)$ and cumulative distribution $G(x)=\int_{-\infty}^{x}dx'g(x')$, then the PDF $h_r(x)$ of $X_{(r)}$ is $h_r(x)=M![G(x)]^{M-r}[1-G(x)]^{r-1}g(x)/[(r-1)!(M-r)!]$, for $r=1,2,\ldots,M$.

Resuming the TW with $\mu=2$ on the RL model, each distance x_i has PDF given by $g(x)=(N-2)(1-x)^{N-3}$, then its cumulative distribution is $G(x)=\int_0^x dx'g(x')=1-(1-x)^{N-2}$ and the PDF of $x_{(r)}$ is $h_r(x)=\tilde{n}![1-(1-x)^{N-2}]^{\tilde{n}-r}[(1-x)^{N-2}]^{r-1}(N-2)(1-x)^{N-3}/[(r-1)!(\tilde{n}-r)!]$.

2. Rank-revisitation and rank-trapping probabilities

Again, consider that the tourist has traveled along the trajectory sites $s_1,s_2,\ldots,s_{\tilde{n}}$ (without any site revisitation). Let us calculate the probability $P_v(r)$ for him or her to revisit the site $s_{(r)}=s_k$ (with associated distance $x_{(r)}=x_k$) in the next step. Once $s_{(r)}$ is the nearest site, the distance $z_{\tilde{n},(r)}$ has PDF given by $g(x)=(N-2)(1-x)^{N-3}$ and once the tourist came from site $s_{(r)}=s_k$ to s_{k+1} in the $(k+1)$th step, the distance $z_{\tilde{n},(r)}$ is certainly greater than $x_{(r)}$. Thus, $P_v(r)\propto P(z_{\tilde{n},(r)}>x_{(r)})=\tilde{n}!/[(r-1)!(\tilde{n}-r)!]\int_0^1 dx[1-(1-x)^{N-2}]^{\tilde{n}-r}[(1-x)^{N-2}]^{r-1}(N-2)(1-x)^{N-3}\int_x^1 dz(N-2)(1-z)^{N-3}$. Evaluating the integral in z and calling $y=(1-x)^{N-2}$ the above equation is rewritten as $P_v(r)\propto\tilde{n}!/[(r-1)!(\tilde{n}-r)!]B(\tilde{n}-r+1,r+1)=\tilde{n}!/[(r-1)!(\tilde{n}-r)!](\tilde{n}-r)!r!/(\tilde{n}+1)!=r/(\tilde{n}+1)$. This expression *is not* the probability P_v itself. Instead, it only gives the dependence of P_v on r.

Normalizing P_v over $1\leqslant r\leqslant\tilde{n}-2$, one has

$$P_v(r)=\frac{r}{\sum_{k=1}^{\tilde{n}-2}k}=\frac{2r}{(\tilde{n}-1)(\tilde{n}-2)}, \tag{10}$$

where $\tilde{n}-2$ is the number of sites available to revisitation (the sites $s_{\tilde{n}}$ and $s_{\tilde{n}-1}$ are forbidden by memory).

The result of Eq. (10) does not contradict Eq. (6), since Eq. (6) gives an approximated probability for the tourist to revisit a specific site s_k, regardless of its associated distance $x_k=x_{(r)}$, while Eq. (10) gives the conditional probability for the tourist to "choose" the r-ranked site $s_{(r)}$ during a revisitation after exploring $\tilde{n}$ distinct sites.

Once the tourist had revisited site s_k (or equivalently $s_{(r)}$), the probability $P_t(r)$ for him or her to get trapped also depends on the rank r. The trapping condition is that x_{k-1} must be greater than x_k. Since $x_k=x_{(r)}$ is the rth greater distance, there are only $r-1$ remaining distances (among $\tilde{n}-3$ ones) greater than x_k. Thus,

$$P_t(r)=\frac{r-1}{\tilde{n}-3}. \tag{11}$$

Combining Eqs. (10) and (11), the probability for the tourist to get trapped when revisiting a specific site $s_{(r)}$ is $P_v(r)P_t(r)=2r(r-1)/[(\tilde{n}-1)(\tilde{n}-2)(\tilde{n}-3)]$. Thus, the probability for the tourist to get trapped when revisiting any site is $p_t=\sum_{r=1}^{\tilde{n}-2}P_v(r)P_t(r)$. Calling $m=\tilde{n}-2$ and evaluating $\sum_{r=1}^{m}r(r-1)=m(m^2-1)/3$ one finds the trapping probability

$$p_t=2/3. \tag{12}$$

We remark that this result has been obtained without any approximation, and numerical simulations agree to it even for small values of N.

B. Capture and subsistence probabilities

Combining the probability $\tilde{p}_{rl}$ for the tourist to revisit a specific site s_k [Eq. (6)] and the trapping probability p_t [Eq. (12)], one obtains the probability p_{rl} for the tourist to revisit s_k and get trapped as follows:

$$p_{rl}=\tilde{p}_{rl}p_t=\frac{1}{2N}\frac{2}{3}=\frac{1}{3N}. \tag{13}$$

Since in the jth step there are $j-3$ sites available to revisitation, the capture (i.e., revisiting any site and getting trapped) and subsistence (i.e., exploring any new site or revisiting any site and not getting trapped) probabilities in the jth step are $p_{2,rl}^{(N)}(j)=(j-3)/(3N)=1-q_{2,rl}^{(N)}(j)$ and the upper-tail distribution for the number n of sites explored by the tourist in the whole trajectory is

$$F_{2,rl}^{(N)}(n) = \prod_{j=4}^{n} q_{2,rl}^{(N)}(j) = \frac{\Gamma(3N)}{\Gamma(3N-n+3)(3N)^{n-3}}, \qquad (14)$$

which is analogous to Eq. (7).

1. Comparison to the RM model with $\mu=0$

The expression of Eq. (14) is similar to the one obtained for the RM model with memory $\mu=0$ [19]: $F_{0,rm}^{(N)}(n) =\Gamma(N)/[\Gamma(N-n)N^n]$. This result explains the nontrivial equivalence observed between the RL model with N points and memory $\mu=2$ (memory effect) and the RM model with $3N$ points and memory $\mu=0$ (effect of distance symmetry break), when one compares the distributions for the total number n of sites explored by the tourist.

Notice that, taking both models with N points each, in RL with $\mu=2$, at each step, the probability for the tourist to revisit a specific site and get trapped is $p_{rl}\approx 1/(3N)$, and in RM with $\mu=0$, this probability is $p_{rm}=1/N$. Therefore, taking RL with N points and RM with $3N$ points equals these probabilities and justifies the equivalence.

2. Exponential form

In the limit $N\gg 1$, the subsistence probability is rewritten as $q_{2,rl}^{(N)}(j)=[1-1/(3N)]^{j-3}$ and one obtains the exponential form of Eq. (14), namely, $F_{2,rl}^{(N)}(n)=\prod_{j=4}^{n}q_{2,rl}^{(N)}(j)=[1-1/(3N)]^{\omega}\approx e^{-\omega/(3N)}$, with $\omega=(n-2)(n-3)/2$.

Rather than differentiating $F_{2,rl}^{(N)}(n)$, the distribution $S_{2,rl}^{(N)}(n)$ for the number n of sites explored in the whole trajectory is more precisely obtained by imposing the tourist to explore n distinct sites and then be captured in the next step (i.e., revisit any site and get trapped) as follows:

$$S_{2,rl}^{(N)}(n) = F_{2,rl}^{(N)}(n)p_{2,rl}^{(N)}(n+1) = \frac{n-2}{3N}e^{-[(n-2)(n-3)/2/(3N)]}. \qquad (15)$$

For $n\gg 1$, calling $y=\sqrt{3N}S_{2,rl}^{(N)}(n)$ and $x=(n-n_{\min})/\sqrt{3N}$ (with $n_{\min}=\mu+1=3$) one obtains the universal plot for this system as follows:

$$y = xe^{-x^2/2}, \qquad (16)$$

with $x\geq 0$ and mth moment $\langle x^m\rangle=2^{m/2}\Gamma(m/2+1)$, where we see that normalization is assured by $\langle x^0\rangle=1$. The mean value is $\langle x\rangle=\sqrt{\pi/2}$ and the variance $\langle x^2\rangle-\langle x\rangle^2=2-\pi/2$. Figure 2 exhibits a plot of Eq. (16) and experimental data. From this figure, or calculating analytically, one obtains that the mode is at $x=1$.

IV. TRANSIENT AND PERIOD JOINT DISTRIBUTION

The transient or period joint distribution $S_{2,rl}^{(N)}(t,p)$ can be obtained similarly to Eq. (15), by imposing the tourist to

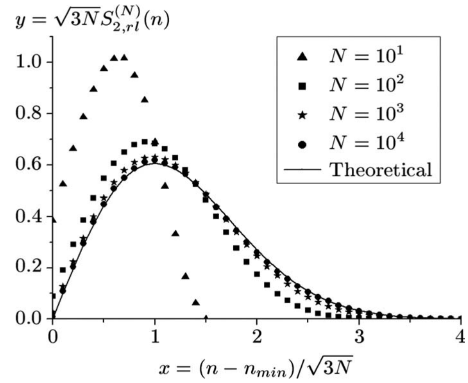

FIG. 2. Finite size effect for the distributions and convergence to the universal curve: $y=xe^{-x^2/2}$.

explore n distinct sites and then revisit the specific site s_k (instead of any site) and get trapped, giving rise to a trajectory with transient $t=k-1$ and period $p=n-k+1$. We notice that the relevant variable is $t+p=n$. Hence, $S_{2,rl}^{(N)}(t,p)$ is obtained multiplying $F_{2,rl}^{(N)}(t+p)$ by p_{rl} [Eq. (13)] [or by $\tilde{p}_{rl}$ [Eq. (6)] in the case $t=0$, since the tourist is unconditionally captured when revisiting the site s_1].

$$S_{2,rl}^{(N)}(t,p) = \frac{1}{(3-\delta_{t,0})N}e^{-[(t+p-2)(t+p-3)/2/(3N)]}, \qquad (17)$$

where $\delta_{i,j}$ is the Kronecker delta. Figure 3 exhibits a plot of Eq. (17) for $N=1000$ points.

A. Transient time marginal distribution

The transient time distribution is calculated summing Eq. (17) over all possible periods, i.e., $S_{2,rl}^{(N)}(t)=\Sigma_{p=3}^{N}S_{2,rl}^{(N)}(t,p)$. In the limit $N\gg 1$, this summation can be approximated by the integral

$$S_{2,rl}^{(N)}(t) = \int_{5/2}^{\infty} dp S_{2,rl}^{(N)}(t,p) = \left(1+\frac{\delta_{t,0}}{2}\right)\sqrt{\frac{\pi}{6N}}\,\mathrm{erfc}\left(\frac{t}{\sqrt{6N}}\right),$$

where the lower limit 5/2 is due to a Yates continuity correction (which other than improve the integral approximation, make the analytical form quite simpler) and the upper limit has been extended to infinity to make calculation easier

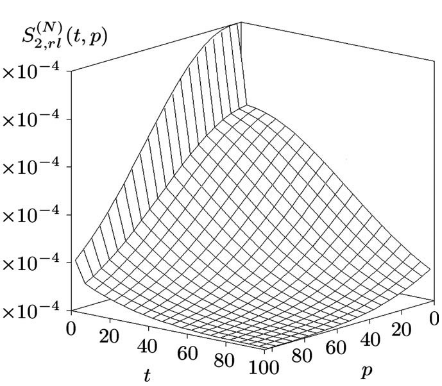

FIG. 3. Transient and period joint distribution for a map with $N=1000$ points in the RL model with $\mu=2$.

[with no harm, because $p>N$ in Eq. (17) yields values close to 0] and $\mathrm{erfc}(x)=(2/\sqrt{\pi})\int_z^\infty dx e^{-x^2}$ is the complementary error function.

B. Cycle period marginal distribution

Similarly, the period distribution is

$$\begin{aligned} S_{2,rl}^{(N)}(p) &= \sum_{t=0}^{N-3} S_{2,rl}^{(N)}(t,p) \\ &= \int_{-1}^{\infty} dt \frac{1}{3N} e^{\{-[(t+p-2)(t+p-3)/2/(3N)]\}} \\ &= \sqrt{\frac{\pi}{6N}}\,\mathrm{erfc}\left(\frac{p-7/2}{\sqrt{6N}}\right) \approx \frac{e^{-p^2/(6N)}}{p}, \end{aligned}$$

where the lower limit -1 is due to both Yates continuity correction and a compensation for the half extra degree in $t=0$. The mean period value is $\langle p\rangle=\sqrt{3\pi N/8}$ and standard deviation is $\sigma_p=\sqrt{(2-3\pi/8)N}$. For $p\ll\sqrt{6N}$, the decay follows a power law $S_{2,rl}^{(N)}(p)\propto p^{-1}$.

V. CONCLUSION

In this paper, we have analytically obtained the statistical distributions for the deterministic tourist walk with memory $\mu=2$ on the random link model. The distribution for the number of sites explored before the first passage has been compared to the one previously obtained for the case $\mu=1$, elucidating the mechanism that strongly increases the tourist's exploratory behavior. This mechanism is explained as follows. On one hand, for $\mu=1$ the distances traveled at each step must obey the ordering $x_1>x_2>\cdots$, leading to a localized exploration. In the thermodynamic limit, the mean number of explored sites is $\langle n\rangle=e=2.718\,28\ldots$ and the exploration probability in the jth step ($j\geqslant 4$) is $1/(j-1)$. This dynamics is due to the underlying extreme value statistics. On the other hand, for $\mu=2$ the distances $x_1,x_2,\ldots$ are unconstrained, leading to an extended exploration: $\langle n\rangle$ is proportional to $N^{1/2}$ and the exploration probability in the jth step (with $j\ll N$) tends to 1, as $N\to\infty$. This dynamics is due to the underlying combinatorial statistics. The factor $(\tilde{n}-1)!$ in Eq. (2) represents the change from the extreme value statistics to the combinatorial one, which makes the $\delta_{p,2}$ distribution of $\mu=1$ to broaden to a wide ($\propto 1/p$) distribution for $\mu\geqslant 2$.

Through the trapping probability $p_t=2/3$ (which value is counterintuitive), we have obtained the capture and subsistence probabilities and a closed form to the complementary cumulative distribution for the number of sites explored in the whole trajectory. This distribution is analogous to the one obtained for the random map model with $\mu=0$. This result explains the equivalence between these mean field models (RL with N points and memory $\mu=2$; and RM with $3N$ points and memory $\mu=0$). For a large number of sites ($N\gg 1$) in the random medium, the distribution $S_{2,rl}^{(N)}(n)$ of having n distinct sites visited by the tourist with memory $\mu=2$ in the random link model is universal $y=xe^{-x^2/2}$ with $y=\sqrt{3N}S_{2,rl}^{(N)}(n)$ and $x=(n-3)/\sqrt{3N}$.

The transient time t and cycle period p joint distribution $S_{2,rl}^{(N)}(t,p)=e^{[(t+p-3)^2/(3N)]/2}/[N(3-\delta_{t,0})]$ has been obtained noticing that the relevant variable is approximately given by $t+p=n$. The marginal distributions are also universal. For the transient time one has $y=[1+\delta(x)/2]\mathrm{erfc}(x)$ with $y=\sqrt{6N/\pi}S_{2,rl}^{(N)}(t)$ and $x=t/\sqrt{6N}$ and for the period distribution $y=\mathrm{erfc}(x)$, with $y=\sqrt{6N/\pi}S_{2,rl}^{(N)}(p)$ and $x=(p-7/2)/\sqrt{6N}$. We have shown that the discrepancy in the null transient time distribution ($t=0$), when compared to the subsequent ones ($t>0$), is due to the higher capture probability the starting site s_1 has [namely, $\tilde{p}_{rl}=1/(2N)$] when compared to the other ones [$p_{rl}=1/(3N)$]. We also have shown that the period distribution decays according to a power law $S_{2,rl}^{(N)}(p)\propto p^{-1}$. In short, it is the effect of shorter walker memory in the first steps along the trajectory.

Future studies concern the consideration of higher memory values in the random link model and the understanding of the connection with the random map model. As the memory increases, we expect a transition from the closed periods to nonclosed ones (chaotic phase). We are interested in understanding the role of finite dimensionality of the system.

ACKNOWLEDGMENTS

The authors thank R. S. Gonzalez for fruitful discussions. A.S.M. acknowledges the Brazilian agencies CNPq (Grant No. 303990/2007-4 and No. 476862/2007-8) for support.

[1] D. S. Fisher, Phys. Rev. A **30**, 960 (1984).
[2] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[3] H. Freund and P. Grassberger, Physica A **190**, 218 (1992).
[4] D. Gale, J. Propp, S. Sutherland, and S. Troubetzkoy, Math. Intell. **17**, 48 (1995).
[5] L. A. Bunimovich, Physica D **187**, 20 (2004).
[6] L. Lam, Int. J. Bifurcation Chaos Appl. Sci. Eng. **16**, 239 (2006).
[7] M. C. Santos, D. Boyer, O. Miramontes, G. M. Viswanathan, E. P. Raposo, J. L. Mateos, and M. G. E. da Luz, Phys. Rev. E **75**, 061114 (2007).
[8] D. Boyer, O. Miramontes, G. Ramos-Fernandez, J. L. Mateos, and G. Cocho, Physica A **342**, 329 (2004).
[9] D. Boyer and H. Larralde, Complexity **10**, 52 (2005).
[10] D. Boyer, G. Ramos-Fernandez, O. M. J. L. Mateos, G. Cocho, H. Larralde, H. Ramos, and F. Rojas, Proc. R. Soc. London, Ser. B **273**, 1743 (2006).
[11] J. C. Cressoni, Marco Antonio Alves da Silva, and G. M. Viswanathan, Phys. Rev. Lett. **98**, 070603 (2007).
[12] G. F. Lima, A. S. Martinez, and O. Kinouchi, Phys. Rev. Lett. **87**, 010603 (2001).
[13] H. E. Stanley and S. V. Buldyrev, Nature (London) **413**, 373

(2001).
[14] O. Kinouchi, A. S. Martinez, G. F. Lima, G. M. Lourenço, and S. Risau-Gusman, Physica A **315**, 665 (2002).
[15] M. G. Campiteli, P. D. Batista, O. Kinouchi, and A. S. Martinez, Phys. Rev. E **74**, 026703 (2006).
[16] M. G. Campiteli, A. S. Martinez, and O. M. Bruno, Lect. Notes Comput. Sci. **4140**, 159 (2006).
[17] A. R. Backes, O. M. Bruno, M. G. Campiteli, and A. S. Martinez, Lect. Notes Comput. Sci. **4225**, 784 (2006).
[18] C. A. S. Terçariol, R. S. Gonzalez, W. T. R. Oliveira, and A. S. Martinez, Physica A **386**, 678 (2007).
[19] C. A. S. Terçariol and A. S. Martinez, Phys. Rev. E **72**, 021103 (2005).
[20] S. Risau-Gusman, A. S. Martinez, and O. Kinouchi, Phys. Rev. E **68**, 016104 (2003).
[21] A. S. Martinez, O. Kinouchi, and S. Risau-Gusman, Phys. Rev. E **69**, 017101 (2004).
[22] C. A. S. Terçariol, R. S. Gonzalez, and A. S. Martinez, Phys. Rev. E **75**, 061117 (2007).
[23] M. Mézard and G. Parisi, Europhys. Lett. **2**, 913 (1986).
[24] A. G. Percus and O. C. Martin, Phys. Rev. Lett. **76**, 1188 (1996).
[25] N. J. Cerf, J. H. B. de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus, J. Phys. I **7**, 117 (1997).
[26] A. G. Percus and O. C. Martin, Adv. Appl. Math. **21**, 424 (1998).
[27] A. G. Percus and O. C. Martin, J. Stat. Phys. **94**, 739 (1999).
[28] C. A. S. Terçariol and A. S. Martinez, Braz. J. Phys. **36**, 232 (2006).
[29] B. Derrida and H. Flyvbjerg, J. Phys. (Paris) **48**, 971 (1987).
[30] S. A. Kauffman, J. Theor. Biol. **22**, 437 (1969).
[31] C. A. S. Terçariol, F. de Moura Kiipper, and A. S. Martinez, J. Phys. A **40**, 1981 (2007).