Metropolis Monte Carlo sampling: convergence, localization transition and optimality

Alexei D. Chepelianskii

Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405 Orsay, France.

Satya N. Majumdar

Université Paris-Saclay, CNRS, LPTMS, 91405 Orsay, France.

Hendrik Schawe

LPTM, UMR 8089, CY Cergy Paris Université, CNRS, 95000 Cergy, France.

Emmanuel Trizac

Université Paris-Saclay, CNRS, LPTMS, 91405 Orsay, France.

Among random sampling methods, Markov Chain Monte Carlo algorithms are foremost. Using a combination of analytical and numerical approaches, we study their convergence properties towards the steady state, within a random walk Metropolis scheme. We show that the deviations from the target steady-state distribution feature a localization transition as a function of the characteristic length of the attempted jumps defining the random walk. This transition changes drastically the error which is introduced by incomplete convergence, and discriminates two regimes where the relaxation mechanism is limited respectively by diffusion and by rejection.

I. INTRODUCTION

Although Buffon's needle problem [1] may be considered as the earliest documented use of Monte Carlo sampling (18th century), the method was developed at the end of the second world war and dates from the early days of computer use [2, 3]. With the increase in computational power, it has become a pervasive and versatile technique in basic sciences and engineering. It uses random sampling for solving both deterministic and stochastic problems, as found in physics, biology, chemistry, or artificial intelligence [4–10]. Monte Carlo techniques also allow to assess risk in quantitative analysis and decision making [11, 12], and their methodological developments provide tools for economy, epidemiology or archaeology [13]. It is then crucial to understand the type of errors which can be introduced as a consequence of the incomplete convergence of such algorithms.

Our interest goes to Markov Chain Monte Carlo techniques [14, 15], that create correlated random samples from a target distribution; a special emphasis goes to the relaxation rate of these methods. From the target probability distribution, a sequence of samples is obtained by a random walk, with appropriate transition probabilities. The walker's density evolves at long times towards the target distribution, and quantities of interest follow from the law of large numbers and other methods of statistical inference [11–19]. The Monte-Carlo method and its modern developments [20–30] have now been adopted in many topics in and outside physics. A key issue deals with the speed of convergence of the algorithm: the larger the convergence time, the larger the error bars for the computed quantities. For most practical applications, if the amplitude a of the random jumps is small, phase space is not sufficiently explored, even though most attempted jumps are accepted. Conversely, large jumps will lead to a large rejection probability, and to an equally ineffective method, see Appendix A. In between, one expects an optimal jump size a_{opt} at which the convergence rate is maximal. There have been theoretical attempts in deriving a_{opt} for specific models [31]. In practice, without a precise knowledge of a_{opt}, a widely accepted rule of thumb is to choose a such that the acceptance probability is close to 50% for the attempted moves [16, 17, 32]. In this paper, we show the existence of a new critical value a^*, where an unexpected localization transition occurs such that the relaxation mechanism is drastically different for $a < a^*$ and $a > a^*$. This deeply modifies the nature and amplitude of the error. The existence of the critical value a^* is our main finding, and the novelty of this work. Besides, although there is no reason to expect any relation between a_{opt} and a^*, we report, rather interestingly, a number of examples, where they coincide precisely. We emphasize that while a number of results have been proven for relaxation rates [33–35], we are interested in this work also in relaxation eigenmodes, which in turn open the way to accurate approximations for relaxation rates.

II. MASTER EQUATION FOR METROPOLIS MONTE-CARLO SAMPLING AND RELAXATION TO EQUILIBRIUM

We start with by reminding general results on the master equation describing relaxation of the Metropolis Monte-Carlo algorithm. The spectral properties of Markov-Chains have been extensively studied in the Mathematical

literature [36, 37]. Here we instead focus on the eigenvectors (or their absence in the case of a localization transition) which have received much less attention. This introduction will allow us to introduce notations and to contrast the relaxation of a discrete Markov-chain with the relaxation properties that we find for the continuous case.

The Markov Chain Monte Carlo method amounts to considering a random walker with position x (here on the line), in the presence of an external confining potential $U(x)$. We adopt the framework of the Metropolis algorithm [15–17, 38, 39]. The position of the particle evolves in discrete time steps n following the rule

$$x_n = \begin{cases} x_{n-1} + \eta_n & \text{with prob. } p = \min\left(1, e^{-\beta\,\Delta U}\right) \\ \\ x_{n-1} & \text{with prob. } 1 - p\,, \end{cases} \tag{1}$$

where $\Delta U = U(x_{n-1} + \eta_n) - U(x_{n-1})$ and $\beta = 1/(k_B T)$ denotes inverse temperature. The random jumps η_n at different times are independent, drawn from a continuous and symmetric probability distribution $w(\eta)$. In other words, the particle attempts at time n a displacement η_n from its current location x_{n-1}, which is definitely accepted (with probability 1) if it leads to an energy decrease, but is accepted with a lesser probability $e^{-\beta\Delta U}$ if the move leads to an energy increase $\Delta U > 0$. A key quantity in what follows is the amplitude a of the attempted jumps, that we introduce as the characteristic length associated with $w(\eta)$, taken to obey the scaling form

$$w(\eta) = \frac{1}{a} f\left(\eta/a\right). \tag{2}$$

Normalization demands that $\int f = \int w = 1$.

The dynamics encoded in Eq. (1) can be written in terms of a Master equation for $P_n(x)$, the probability density of the walker at time n,

$$P_n(x) = \int_{-\infty}^{\infty} F_\beta(x, y) P_{n-1}(y)\, dy. \tag{3}$$

The explicit form of the temperature-dependent kernel F_β is given below in Eq. (6). Generically, $P_n(x)$ converges towards the target distribution [40, 41], given by the (equilibrium) Gibbs-Boltzmann expression $P_\infty(x) \propto \exp(-\beta\, U(x))$, see Appendix B. We assume that U is confining enough so that $\exp(-\beta U(x))$ is integrable, and for simplicity that $U(x) = U(-x)$. Our main interest is to find how quickly the dynamics converges towards the target density, and with which error $\delta P_n(x) = P_n(x) - P_\infty(x)$. The convergence rate can be defined from the large time limit of the deviation from equilibrium of some observable $\mathcal{O}(x)$:

$$\log \Lambda = \sup_{\{\mathcal{O}(x), P_0(x)\}} \lim_{n \to \infty} \frac{1}{n} \log \left| \int \mathcal{O}(x) \delta P_n(x) dx \right| \tag{4}$$

where the maximum is taken over all possible functions $\mathcal{O}(x)$ and initial distributions $P_0(x)$. If $\Lambda < 1$ (given that $\Lambda \leq 1$) the probability distribution $P_n(x)$ converges exponentially fast to the equilibrium distribution for large n, i.e., $|P_n(x) - P_\infty(x)| \propto \Lambda^n \propto e^{-n/\tau}$ where $\tau = -(\log \Lambda)^{-1}$ denotes the convergence time (in number of Monte-Carlo algorithm steps unit). The convergence rate $-\log(\Lambda) > 0$ is the figure of merit of the algorithm; the smaller the Λ, the larger the rate, the smaller the convergence time and the more efficient the sampling is.

A quantity of central importance in the approach is the rejection probability $R(x)$ (see Appendix B), or more precisely the fraction of rejected moves per attempted jump:

$$R(x) = \int_{-\infty}^{\infty} dy\, w(y - x) \left(1 - e^{-\beta\,(U(y) - U(x))}\right) \theta\left(U(y) - U(x)\right), \tag{5}$$

where $\theta(z)$ is the Heaviside function: $\theta(z) = 1$ for $z > 0$ and $\theta(z) = 0$ for $z < 0$. Thus, the rejection probability $R(x)$ from the current location x is zero if the new position y occurs downhill. In the opposite case of an uphill move, it is $1 - \langle e^{-\beta\Delta U} \rangle$ where the average is over all attempted jump lengths. The integral kernel of the Master equation is then given by:

$$F_\beta(x, y) = \delta(x - y) R(x) + w(x - y) \left[\theta\left(U(y) - U(x)\right) + e^{-\beta\,(U(x) - U(y))} \theta\left(U(x) - U(y)\right)\right]. \tag{6}$$

Averaging over the position of the particle yields the mean rejection probability $R_n = \int R(x) P_n(x)\, dx$ which is monitored by default in all rejection-based algorithms. This is the quantity that the practitioner aims at keeping close to 50%, following a time honored rule of thumb stating that this provides efficient sampling [17, 32]. In the limit $n \to \infty$, this mean rejection probability approaches the stationary value R_∞. Rigorous studies, in a one-dimensional

harmonically confined setting with a Gaussian jump distribution, have found that the optimal acceptance probability $1 - R_\infty$ is close to 44%, while this quantity may decay when increasing space dimension [31]. On intuitive grounds, one may expect a relation between $R(x)$ and the convergence rate of the algorithm. Indeed, starting from an arbitrary point x_0 at time $n = 0$, the density at time n, given x_0, can be written

$$P_n(x|x_0) = R(x_0)^n \delta(x - x_0) + p_n(x|x_0) \tag{7}$$

where $p_n(x|x_0)$ is a smooth function. Thus, an observable $\mathcal{O}$ that would only measure the walker's presence in the immediate vicinity of x_0, for instance $\mathcal{O}_n(x_0) = \lim_{\epsilon \to 0} \int_{x_0 - \epsilon}^{x_0 + \epsilon} P_n(x) dx$, would decay as $R(x_0)^n$. The system as a whole cannot relax faster, and we obtain from Eq. (4) a lower bound for the convergence rate, corresponding to $\Lambda > R(x_0)$, which holds for all choices of x_0:

$$\Lambda \geq \max_{x_0} R(x_0). \tag{8}$$

Our objective is to study Λ as a function of a, for a fixed choice of $U(x)$ and $f(z)$. We expect Λ to be minimum at a well defined value $a = a_{\text{opt}}$.

More precisely, for a given confining potential $U(x)$ and type of jumps $f(z)$, the convergence rate and the resulting error are encoded in the spectral properties of the kernel $F_\beta(x, y)$ in Eq. (3). We have attacked this question by four complementary techniques: the derivation of exact results, numerical diagonalization, numerical iteration of the Master equation, and direct Monte Carlo simulation of the random walk dynamics, with proper averaging over multiple realizations to gather statistics, see Appendix C. We begin with a *discretized* approximation to the Master equation (3), for which Perron-Frobenius theorem shows that the equilibrium state, reached at large n (formally $n \to \infty$), is unique [41]: it is given by $P_\infty(x)$. At any time, the probability density can furthermore be decomposed as

$$P_n(x) = \sum_\lambda \mathcal{A}_\lambda \mathcal{P}_\lambda(x) \lambda^n \tag{9}$$

where the eigenvectors of F_β are denoted by $\mathcal{P}_\lambda(x)$, and the eigenvalues λ can be proven to be real [36], see also Appendix B. Indeed, detailed balance [15–17] allows to transform the Master equation into a self-adjoint problem, similarly to the mapping between the Fokker-Planck and Schrödinger equations [42]. The precise form of the projection coefficients $\mathcal{A}_\lambda$ is not essential. Ordering eigenvalues in decreasing order ($\lambda_0 > \lambda_1 \geq \lambda_2 \ldots$), the eigenvalue $\lambda_0 = 1$ is associated with equilibrium, with eigenvector $P_\infty(x)$. We also see that the asymptotic error δP_n behaves like $\mathcal{P}_{\lambda_1}(x)$, and decays to 0 like λ_1^n (also meaning that $\Lambda = \lambda_1$). Finding the optimal a is a minmax problem, where one should minimize $\Lambda = \lambda_1$, i.e. the maximum eigenvalue, leaving aside the top (equilibrium) eigenvalue $\lambda_0 = 1$.

III. RELAXATION TO EQUILIBRIUM AND LOCALIZATION

While the above results hold for the discretized version of Eq. (3), explicit analytical calculations of the spectrum for a number of potentials $U(x)$ reveal that the eigenvector decomposition (9) fails in the continuum limit. In addition to the discrete spectrum with well defined eigenfunctions, a continuum of eigenvalues appears, with singular localized eigenfunctions which in the continuum limit collapse to a point x_0 where they take a finite value. The corresponding eigenvalue is $R(x_0)$. We call this continuum of singular eigenvalues the singular spectrum, which is therefore bounded from below and above by $\min_x R(x)$ and $\max_x R(x)$. Equation (9) now takes the form

$$P_n(x) = \sum_{\lambda \in \{\lambda_0 \ldots \lambda_{\mathcal{N}}\}} \mathcal{A}_\lambda \mathcal{P}_\lambda(x) \lambda^n + \mathcal{L}_n(x), \tag{10}$$

where $\mathcal{L}_n(x)$ stems from the singular continuum. Here, the discrete summation runs over a finite (and possibly small) number of $1 + \mathcal{N}$ terms: $\mathcal{N} \geq 0$ since the term $\lambda_0 = 1$ is necessarily present in the expansion, to ensure the proper steady state. The remaining term $\mathcal{L}_n(x)$ localizes at large times $n \to \infty$ around a finite number of points x_l where the rejection rate $R(x)$ in (5) is maximal: $\lim_{n \to \infty} \mathcal{L}_n(x)/\mathcal{L}_n(x_l) = 0$ for any $x \neq x_l$. This property of the localizing term $\mathcal{L}_n$ is valid only for the non-discretized master equation and is thus most directly established by analytical means. From our analytical computations, two possible scenarios emerge: (i) $\mathcal{N} > 0$ for all a and (ii) $\mathcal{N} = 0$ for $a > a^*$ where a^* gives the position for the localization transition; a^* marks the transition from a diffusion governed evolution to a phase where relaxation is limited by rejected moves. In case (i), the eigenvalue λ_1 lies above the singular continuum and $\Lambda = \lambda_1$. The error is ruled by a "regular" eigenmode akin to what would be found in the discretized approximation. In case (ii) on the contrary, λ_1 merges with the singular continuum at $a = a^*$ and the
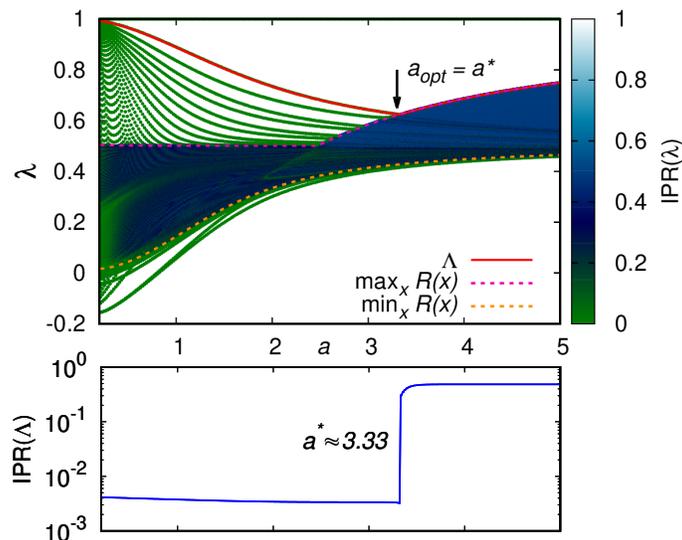
FIG. 1. The top panel shows the spectrum of F_β for harmonic confinement $U(x) = x^2/2$ as a function of jump amplitude a, for a uniform jump distribution of range $(-a, a)$. The color code, provided on the right-hand-side, is for the Inverse Participation Ratio of the eigenvector associated to the eigenvalue displayed (see Appendix C). The upper envelope of the relaxation spectrum defines Λ, see Eq. (4); shown by the red line, it reaches its minimum for $a = a_{\rm opt} \simeq 3.33$. This value coincides with the threshold a^* for localization. Here, $\mathcal{N}$, denoting the number of discrete relaxation modes (excluding the stationary state), is 0 for large jumps ($a > a^*$), while $\mathcal{N}$ quickly grows as a diminishes. Dashed lines show the bounds for the singular continuum, that appears in dark blue, see Eqs. (5) and (8). The bottom panel is for the IPR associated to Λ, as a function of a (same abscissa as the upper panel). The localization transition is signaled by the sharp jump at $a = a^*$. This threshold does not depend on the number of sites N_d, as long as N_d is large enough. Here $N_d = 1000$. The length unit is the thermal length, meaning the standard deviation of $P_\infty(x)$.

error is dominated by the localizing term $\mathcal{L}_n(x)$. Numerical simulations suggest that this localized scenario (ii) is the generic case, see also Appendix D.

In Fig. 1, we illustrate the merging between regular and singular spectrum for the harmonic potential with a flat jump distribution. To distinguish numerically the regular spectrum as in Eq. (9) from the singular one, we have discretized $F_\beta(x, x')$ into a matrix of size $N_d \times N_d$, and computed the spectrum. Two methods have then been employed, both relying on a large N_d analysis. For the regular part, the spacing between successive eigenvalues stay non-zero as $N_d \to \infty$ while they do vanish in the singular part. Another signature can be found with the eigenvectors by computing the inverse participation ratio (IPR) (see Appendix C for the definition), usually used to quantify localization of quantum states [43]. For a regular eigenvalue with a well defined continuum eigenvector, the IPR $\to 0$ as $1/N_d$ for large N_d, while the IPR is much larger within the singular continuum as evidenced by the color code in Fig. 1. At $a = a^*$, this singular part crosses the regular λ_1 branch, leading to a gap closure. For $a > a^*$, the singular continuum is dominant and governs relaxation. In Fig. 1, a^* is shown by an arrow. Furthermore here, the structure of the spectrum ensures that $a^* = a_{\rm opt}$, see Fig. 1; at this point, $\Lambda(a)$ features a cusp. Quite remarkably, the acceptance probability $1 - R_n$ at $a = a^* = a_{\rm opt}$ tends at long times towards 0.455, close to the 50% rule of thumb alluded to above.

The critical nature of the parameter $a = a^*$ can be appreciated by the behavior of the IPR of the slowest decay mode, as displayed in Fig. 1-bottom. The large value of the IPR for $a > a^*$ indicates that δP_n ceases to be spread over the whole system, but rather gets more and more "pinned" onto a discrete set of points; in the present case, this set reduces to a single point, $x_l = 0$. This results in the central dip in the error $\delta P_n(x)$ observed in the main graph in Fig. 2, that becomes more narrow as time n increases (see below). Fig. 2 also reveals that a complete change of symmetry goes with the crossing of a^*. For $a < a^*$, the longest lived perturbation in the system is antisymmetric, see the inset of Fig. 2: given the symmetry of the confining potentials considered ($U(x) = U(-x)$), such a mode takes indeed longer to relax than symmetric ones. This can be understood from the mapping of our problem to a Schrödinger equation, for small a, see Appendix B3: the first excited state, meaning the λ_1 branch, has only one zero and is anti-symmetric. On the other hand, for $a > a^*$, δP_n becomes symmetric after a transient (see the evolution from an early asymmetric situation towards symmetry in Fig. 2).

To gain more insight into the localization phenomenon and its dynamics, we studied analytically the Master equation for confinement in a box, i.e. when $U(x) = 0$ for $|x| < L$ and $U(x) = \infty$ for $|x| > L$. Such a case is rich enough
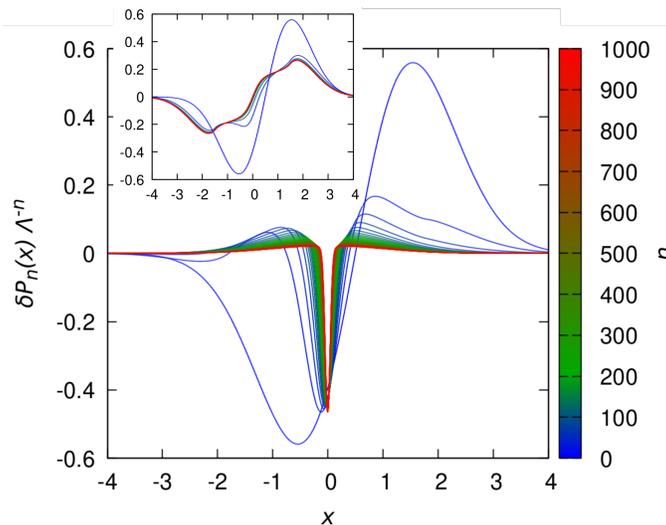
FIG. 2. Scaled evolution of the error $\delta P_n(x) = P_n(x) - P_\infty(x)$ vs x for different times n (indicated by the color code on the right), for the same system as in Fig. 1, with the same choice of length unit. Initial $P_0(x) = (2\pi)^{-1/2}\exp(-(x-1)^2/2)$. Comparison between $a = 3.6 > a^*$ (main graph) and $a = 3 < a^*$ (inset). Although the values of a and the convergence rates are similar in the two graphs, the asymptotic errors are significantly different.
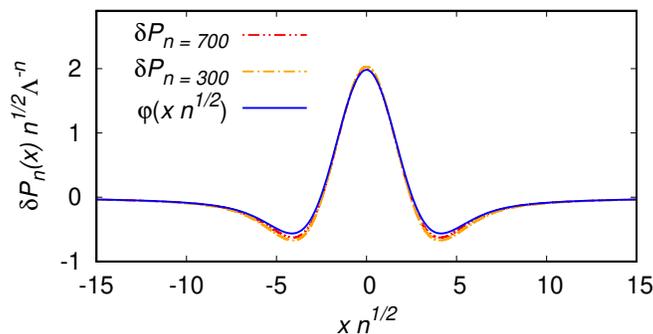

FIG. 3. Comparison between the exact calculation presented in Appendix D and the numerical data for the scaling behavior of localization. Box potential confinement with $w(\eta) = 3(1 + a^{-2}\eta^2)\theta(a - |\eta|)/(8a)$. Lengths are expressed in unit of the box size L, convergence to the scaling function is shown for $a = 2.1$ and $P_0(x) = 2\theta(1/2 - |x|)$. Our analytical expression for the scaling function $\varphi(z)$ and the proof that $\mathcal{N} = 0$, for this choice of $w(\eta)$, are obtained for $a > 2 > a^* \simeq 1.79$.

to display the generic phenomenology of localization, while remaining sufficiently simple to allow for the derivation of exact results for several jump distributions $w(\eta)$, see Appendix D. For cases where $w(\eta)$ is minimum at $\eta = 0$, we proved that $\mathcal{N} = 0$ for sufficiently large a as in the case of harmonic confinement. As in Fig. 2, the localization transition then manifests as a progressive collapse of the error $\delta P_n = P_n(x) - P_\infty(x)$ onto the point where rejection probability is maximal ($x_l = 0$), with a spread which decays as $1/\sqrt{n}$. More precisely, in the vicinity of this point, we obtained the asymptotic form

$$\delta P_n(x) = \Lambda^n n^{-\gamma} \varphi(x\sqrt{n}) \tag{11}$$

where $\varphi(z)$ is a regular scaling function, and the exponent γ depends on $w(\eta)$ and $U(x)$. We found that a scaling function ansatz with $\Lambda = 1$ also describes the relaxation of a zero temperature Metropolis Monte-Carlo algorithms towards a minimum [44]. In the zero temperature limit the steady state is a Dirac delta function localized at the minimum of the potential and such a scaling form evolving into a Dirac delta function can be expected. At finite temperature however, the steady state has a finite width which is given by the thermal length, and thus the scaling-function ansatz does not directly follow from the ground state. The difference between the two cases can also be seen from the vanishing integral $\int \varphi(x)dx = 0$ in (11) while this integral is normalized to unity at zero temperature. Figure 3 shows that such a form is well obeyed in the simulations, and that $\gamma = 1/2$ for the case displayed, in full agreement

with our exact treatment that also explicitly provides $\varphi(z)$ in Appendix D (Eq. (D35)), shown by the continuous line. Numerical evidence shows that for the harmonic potential, $\gamma = 0$. Analytical studies of the box potential where $w(\eta)$ is maximum at $\eta = 0$, provide examples where we can prove that $\mathcal{N} = 1$, in the large a limit. The localisation transition in the error can consequently not be seen for a generic observable, but special choices of the observable or initial conditions allow to reveal a hidden localization transition, even in this case.

IV.   ANALYTICAL APPROXIMATION TO RELAXATION RATES FROM FOKKER-PLANCK EIGENVECTORS

We already mentioned that the critical amplitude a^* separates two regimes, a regime $a < a^*$ where the dynamics is governed by the relaxation of diffusion eigenmodes and a regime $a > a^*$ where the relaxation is governed by the highest rejection probability. Surprisingly this knowledge provides a very precise approximation scheme to find quantitatively the full dependence of the relaxation rate $\Lambda(a)$ on the jump length a. In the limit of small a, the master equation reduces to a Fokker-Planck equation, and it is possible to use the lowest eigenmodes of this Fokker-Planck equation to project the full master equation on a small finite dimensional-basis; the details of this procedure are described in Appendix E and illustrated in Fig 4. We find that for $a < a^*$, a very small number of diffusion eigenmodes provide a very accurate estimation of $\Lambda(a)$ or good analytical approximations when the diagonalization of the reduced matrix is possible. On the contrary, for $a > a^*$, the convergence of this procedure is very slow, and $\Lambda(a)$ coincides with the maximum rejection probability. For a flat jump distribution $w(\eta)$ in a harmonic potential, for which $a_{opt} = a^*$, this procedure also provides an analytical estimate of the optimal mean acceptance probability, $1 - R_\infty \simeq 0.455$, close to values obtained by numerical diagonalization. A similar computation can be done for Gaussian jumps (see Appendix E), for which we get analytically $1 - R_\infty \simeq 0.467$, which improves the previously reported numerical estimate of 0.44 [31], alluded to above.
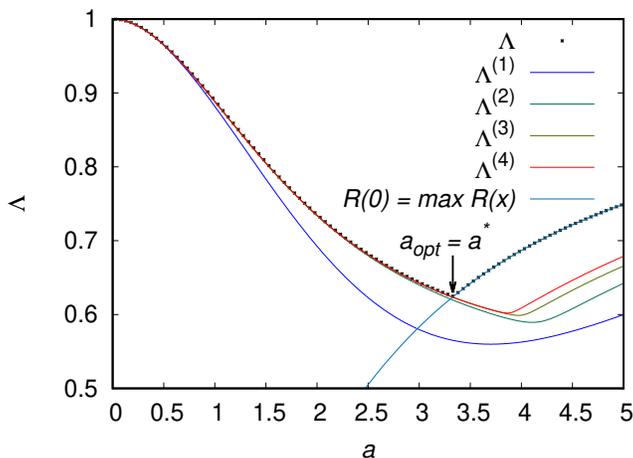

FIG. 4. Harmonic confinement with flat jump distribution. Plot of the largest non-stationary eigenvalue as function of jump amplitude a, obtained by 1) exact numerical diagonalization (dots) and 2) analytical approximation in the truncated Schrödinger equation basis, with increasing N_s, the number of anti-symmetric diffusion modes retained (see Appendix E). With $N_s = 2$, the analytical predictions are already very accurate for $a < a^*$, convergence is very slow on the other side of the localization transition $a > a^*$

An interesting issue is to assess how robust is the localization transition found: does it survive in higher dimensions or in the presence of interactions between particles? To investigate these, we have studied a) a non interacting model in dimensions 2 and 3, and b), an interacting system in dimension 1 and c) the situation where the confining potential features multiple local minima, see Appendix F. In all cases, we found a localization transition, demonstrating its wider applicability. Analyzing the fate of the present localization transition for more complex potential landscapes, as found in disordered systems, is an interesting open problem.

V. CONCLUSIONS

To summarize, we have uncovered that a localization transition does generically take place in Monte Carlo sampling, for a critical value a^* of the amplitude of the jump distribution. A central result of this paper is to show that at $a = a^*$, a singular continuum takes over the regular spectrum as the leading relaxation mode. We found that below $a < a^*$, the relaxation rate can be determined very accurately by the projection of the full master equation on the leading relaxation modes of the Fokker-Planck dynamics. This opens the way to analytic calculation for the relaxation rate. For $a > a^*$, the convergence of this expansion becomes much slower and the relaxation rate is instead given by maximal rejection probability. This results in a dynamical collapse, evidenced by a sharp increase of the Inverse Participation Ratio at $a = a^*$, reminiscent of Anderson localization [45]. However the underlying physical pictures differ. In the Anderson scenario, the localization length is given by the mean free path of a disordered potential. Here, the error progressively shrinks to a point when increasing time n, without any corresponding limiting eigenvector $\psi(x)$ with non zero norm $\int |\psi(x)|^2 dx > 0$. Thus our study shows an example of a well known Markov process whose relaxation is not determined by the contribution of discrete eigenmodes, but by a progressive localization (collapse) on discrete points. Furthermore, we found that a^*, when it exists, coincides with the optimal jump length a_{opt}, although we are not able to prove it. We may surmise that the localization phenomenon has been overlooked so far for the reason that the upper part of the spectrum, $\Lambda(a)$, which rules relaxation, is continuous for all a including the transition point a^*; it is the derivative $d\Lambda/da$ that is discontinuous at a^*. Yet, the error incurred, due to unavoidable lack of convergence at finite time, does change nature when crossing a^*: its symmetry, amplitude, and scaling are deeply affected. The understanding of the localization transition in Monte Carlo relaxation modes may help to avoid excess events on the localization sites in the applications of Monte Carlo random walks.

Appendix A: Why an optimal jump amplitude?

The relaxation time of the Metropolis algorithm, for a given functional form w of the jumps (see Eq. (2) below), depends on the jump amplitude a. On general grounds, this time should exhibit a non-monotonous behavior with a well-defined minimum at some specific amplitude a_{opt} (corresponding to a minimum convergence time τ (minimum Λ, i.e. a maximum rate $-\ln \Lambda = 1/\tau$). This is the so-called Goldilock's principle [46]. The rationale behind this expectation goes as follows:

- In the diffusive limit where a is small, though most of the jumps are accepted, the particle moves over a limited region of space which results in a long time for exploring the full available space. Hence, we expect τ to diverge, i.e. $\Lambda \to 1$. We can be more specific, assuming a confinement potential of the form $U(x) = |x|^\alpha$ with $\alpha > 0$. At equilibrium, the walker's density P will be concentrated within the thermal length $\ell \propto \beta^{-1/\alpha}$ around the origin, and equilibrium will be reached after a characteristic time τ such that $D\tau = \ell^2$, where D is the diffusion coefficient. For our discrete time dynamics, we have $D \propto a^2$, so that we expect here $\tau \propto \beta^{-2/\alpha} a^{-2}$, meaning $\Lambda - 1 \propto a^2$

- In the opposite long jump limit with large a, most of the moves are rejected and the particle hardly moves. As long as $w(0)$ is non-vanishing, increasing the jump amplitude a simply reduces the displacement probability by a factor $1/a$, while leading to the same sampling of phase space on the scale of the confinement length $\ell \ll a$. Hence we expect the system to relax very slowly, i.e., the relaxation time τ to diverge as $\tau \propto a$, so that $\Lambda - 1 \propto 1/a$. This scaling law can only be altered for $w(0) = 0$.

We thus expect an optimal finite jump amplitude $a = a_{\text{opt}}$, for a given functional form w, where $\Lambda(a)$ is minimal and hence the convergence is the fastest.

Appendix B: The Master equation and its reformulations

1. The formalism

From the dynamics defined in the main text, we can write the Master equation obeyed by the walker's density as

$$P_n(x) = \int_{-\infty}^{\infty} dx' \, P_{n-1}(x') \, w(x-x') \min\left(1, e^{-\beta\left(U(x)-U(x')\right)}\right) + \left[1 - \int_{-\infty}^{\infty} dy \, w(y-x) \min\left(1, e^{-\beta\left(U(y)-U(x)\right)}\right)\right] P_{n-1}(x),$$

(B1)

where $w(\eta)$ is the jump distribution and $U(x)$ the confining potential. At a given time step n, the first term describes the probability flux to x from all other positions x'. The second term is for the probability that all attempted moves made by the particle at x (to another arbitrary position y) are rejected. It proves convenient to replace the 'min' function above by the identity

$$\min\left(1, e^{-\beta\left(U(x)-U(x')\right)}\right) = \theta\left(U(x')-U(x)\right) + e^{-\beta\left(U(x)-U(x')\right)}\theta\left(U(x)-U(x')\right) \tag{B2}$$

where $\theta(z)$ is the Heaviside theta function. The Master equation (B1) can then be written as

$$P_n(x) = \int_{-\infty}^{\infty} F_\beta(x, x')\, P_{n-1}(x')\, dx' \tag{B3}$$

where the temperature dependent kernel is given by

$$F_\beta(x, x') = w(x - x')\left[\theta\left(U(x')-U(x)\right) + e^{-\beta\left(U(x)-U(x')\right)}\theta\left(U(x)-U(x')\right)\right]$$

$$+\delta(x - x')\underbrace{\left[1 - \int_{-\infty}^{\infty} dy\, w(y - x')\left[\theta\left(U(x')-U(y)\right) + e^{-\beta\left(U(y)-U(x')\right)}\theta\left(U(y)-U(x')\right)\right]\right]}_{R(x')}. \tag{B4}$$

The kernel $F_\beta(x, x')$ can be interpreted as the probability of a jump from x' to x at inverse temperature β. The term in square brackets on the second line of Eq. (B4) is the rejection probability, that can be recast in

$$R(x') = \int_{-\infty}^{\infty} dy\, w(y - x')\left(1 - e^{-\beta\left(U(y)-U(x')\right)}\right)\theta\left(U(y)-U(x')\right). \tag{B5}$$

Written as such, it directly expresses the fact that among all attempted moves from x' to y, only a fraction $1 - e^{-\beta\left(U(y)-U(x')\right)}$ of those leading to an energy increase ($U(y) > U(x)$), is effectively rejected. All others attempts are accepted and thus do not contribute to $R(x')$.

A first check for the validity of the Master equation is that it should conserve the total probability $\int_{-\infty}^{\infty} P_n(x)\, dx = 1$. From (B3), this means that kernel $F_\beta(x, x')$ must satisfy the condition

$$\int_{-\infty}^{\infty} F_\beta(x, x')\, dx = 1 \quad \text{for all } x'. \tag{B6}$$

Indeed, substituting $F_\beta(x, x')$ from (B4) into the integral (B6), it is easy to check that it satisfies the probability conservation for all x'.

Next, we verify explicitly that the Master equation (B3), with $F_\beta(x, x')$ given in (B4), admits, as $n \to \infty$, a stationary solution that is of the Gibbs-Boltzmann equilibrium form

$$P_\infty(x) = \frac{1}{Z} e^{-\beta U(x)}, \tag{B7}$$

where the partition function Z is a normalization constant. Assuming a stationary solution exists as $n \to \infty$ in (B3), it must satisfy the integral equation

$$P_\infty(x) = \int_{-\infty}^{\infty} F_\beta(x, x')\, P_\infty(x')\, dx'. \tag{B8}$$

To verify this equality, we substitute $P_\infty(x') = (1/Z)\, e^{-\beta U(x')}$ on the right hand side (rhs) of (B8) and use the explicit form of $F_\beta(x, x')$ from (B4). By writing down each term on the rhs explicitly, it is straightforward to check that indeed for arbitrary symmetric jump distributions such that $w(x - x') = w(x' - x)$, the rhs gives (after a few cancellations) $(1/Z)e^{-\beta U(x)}$ for arbitrary confining potential $U(x)$. This is of course expected since the Metropolis rule indeed does satisfy detailed balance with respect to the Gibbs-Boltzmann stationary state.

2. Transformation to a self-adjoint problem

Solving the Master equation (B3) analytically for arbitrary potential is out of reach. A first difficulty one encounters is that the kernel $F_\beta(x, x')$ in (B4) is non-symmetric under the exchange of x and x': the integral operator $F_\beta(x, x')$

is not self-adjoint. This problem can be circumvented by applying the following 'symmetrizing' trick [42]. Let us first define a new quantity $Q_n(x)$ related simply to $P_n(x)$ via the relation

$$P_n(x) = e^{-\beta\,U(x)/2}\,Q_n(x)\,. \tag{B9}$$

Substituting this relation in (B3), we see that $Q_n(x)$ satisfies the following integral equation

$$Q_n(x) = \widehat{K}_\beta Q_{n-1}(x) = \int_{-\infty}^{\infty} K_\beta(x, x')\,Q_{n-1}(x')\,dx' \tag{B10}$$

where the action of the integral operator $\widehat{K}_\beta$ is described by its kernel $K_\beta(x, x')$:

$$K_\beta(x, x') = w(x - x')e^{-\beta|U(x)-U(x')|/2} + \delta(x - x')R(x) \tag{B11}$$

and the rejection probability $R(x)$ is defined in Eq. (B5). Thus, for symmetric jump distribution $w(x-x') = w(x'-x)$, $K_\beta(x, x')$ is symmetric and we can consider $\widehat{K}_\beta$ as a real self-adjoint integral operator (operating on the real line) whose matrix element $\langle y|\widehat{K}_\beta|y'\rangle = K_\beta(y, y')$ is given by Eq. (B11). Besides, Eq. (B10) admits a stationary solution

$$Q_\infty(x) = \frac{1}{Z}\,e^{-\beta\,U(x)/2}. \tag{B12}$$

The solution of the integral equation (B10) can be written as a linear combination of the eigenmodes of the operator $\widehat{K}_\beta$, i.e.,

$$Q_n(x) = \sum_\lambda \mathcal{A}_\lambda\,\psi_\lambda(x)\,\lambda^n \tag{B13}$$

where $\psi_\lambda(x)$ satisfies the eigenvalue equation

$$\int_{-\infty}^{\infty} K_\beta(x, x')\,\psi_\lambda(x')\,dx' = \lambda\,\psi_\lambda(x) \tag{B14}$$

and the $\mathcal{A}_\lambda$'s are arbitrary at this point. Consequently, from Eq. (B9),

$$P_n(x) = \sum_\lambda \mathcal{A}_\lambda\,\psi_\lambda(x)\,e^{-\beta U(x)/2}\lambda^n = \sum_\lambda \mathcal{A}_\lambda\,\mathcal{P}_\lambda(x)\lambda^n \quad \text{with} \quad \mathcal{P}_\lambda(x) = \psi_\lambda(x)\,e^{-\beta U(x)/2}, \tag{B15}$$

as written in Eq. (9) in the main text.

Since the operator $\widehat{K}_\beta$ is real self-adjoint, both its eigenvalues and eigenvectors are real valued [36]. This property extends to the operator defined from $F_\beta(x, x')$, since

$$e^{-\beta U(x)/2}\,F_\beta(x, x') = e^{-\beta U(x')/2}\,K_\beta(x, x'). \tag{B16}$$

Having a real spectrum is a non-trivial property, as the eigenvalues of Frobenius-Perron type of operators to which the original integral equation Eq (B8) belongs are in general complex numbers inside the unit circle $|\lambda| < 1$. The detailed balance rules which are used to derive the Metropolis algorithm actually constrain the eigenvalue of the associated integral equation to be real (at non zero temperatures) [36]. The eigenvalue $\lambda_0 = 1$ corresponds to the steady state solution $Q_\infty(x)$ in (B12); all other eigenvalues are real and strictly below 1. We have labeled the spectrum so that $1 > \lambda_1 \geq \lambda_2 \ldots$. A particular interest goes into the eigenvalue λ_1 that is closest to 1 from below, since it rules the long time dynamics.

3. The diffusive limit: Schrödinger reformulation and symmetry

The distribution w of attempted jumps is taken of the form (2) with a representing a characteristic length. The limit of small a is informative: the original Master equation reduces to a diffusive-like Fokker-Planck equation [42]. In line with our preceding treatment, it is more convenient to work with the self-adjoint dynamics, which is described by an equivalent Schrödinger equation, as we proceed to show.

For $a \to 0$, it is possible to Taylor-expand the eigenvalue equation (B14). Introducing the second moment of the jump distribution

$$\sigma^2 = \int dy\, y^2 w(y), \tag{B17}$$

and making use of the identity $\int dy\, w(x-y)(x-y) = 0$ together with the symmetry relations, we get

$$\int dy\, w(x-y)\frac{(x-y)^2}{2}\text{sign}(U(x)-U(y)) = \int dy\, w(x-y)\frac{(x-y)^2}{2}\text{sign}(U'(x)(x-y)) = 0 \tag{B18}$$

valid when $U'(x) \neq 0$. These cancellations stem from the symmetry $w(x) = w(-x)$. We thereby get:

$$(1-\lambda)\psi_\lambda(x) = \sigma^2\left(-\frac{1}{2}\psi_\lambda''(x) + \frac{\beta^2}{8}U'(x)^2\psi_\lambda(x) - \frac{\beta}{4}U''(x)\psi_\lambda(x)\right) \tag{B19}$$

The relaxation rates of this equation can thus be determined from the eigenvalues ϵ_n and eigenvectors of the effective Schrödinger equation

$$\widehat{H}\psi = -\frac{1}{2}\psi''(x) + \frac{\beta^2}{8}U'(x)^2\psi(x) - \frac{\beta}{4}U''(x)\psi(x) = \epsilon_n\psi(x). \tag{B20}$$

The connection reads

$$\lambda_n = 1 - \sigma^2\epsilon_n, \tag{B21}$$

providing an explicit expression for the spectrum $\{\lambda_n\}$.

We stress that a truncation of the Taylor expansion behind the derivation of the Schrödinger equation is justified if the length-scale on which the wavefunctions vary is large compared to a, the typical amplitude of the jumps generated by $w(\eta)$. Thus Eq. (B19) is not valid in the limit of the high energy modes ϵ_n of the Schrödinger equation. This limitation of the Schrödinger picture can be anticipated from the fact that the eigenvalues of the original Master equation are in the interval $\lambda \in [-1, 1]$ while the eigenvalues predicted by Eq. (B19) extend to all the range $(-\infty, 1]$. The ground state of the Hamiltonian Eq. (B20) has a vanishing ground state eigenvalue $\epsilon_0 = 0$ with an eigenvector given by $\psi_0 = e^{-\beta U(x)/2}$. This eigenvector describes the equilibrium probability distribution and is identical to the ground state of the original Eqs. (B10,B11), without the assumption of a small jump length.

Note that since the original confining potential is symmetric in x (even), so is the Schrödinger effective potential in Eq. (B20), $\beta^2 U'(x)^2/8 - \beta U''(x)/4$. The Schrödinger reformulation then allows to understand why the longest lived eigenmode, for small a, is antisymmetric: it corresponds to the first excited state, with an eigenfunction featuring a unique zero.

4. Analytical solutions in the truncated Schrödinger eigenbasis

Since Eq. (B20) is a Schrödinger equation, its (normalized) excited state eigenvectors $\psi_n(x)$ ($n \geq 1$) are all orthogonal to $\psi_0(x)$ and provide a natural basis for a variational estimation of the relaxation rate. Indeed, the definition of Λ in the main text as the leading relaxation mode (upper value of the relaxation spectrum, leaving aside the top eigenvalue $\lambda = 1$ corresponding to the equilibrium state) can be recast as

$$\Lambda = \max_{\Phi \perp \psi_0} \frac{\int dy \int dy'\, \Phi(y)K_\beta(y,y')\Phi(y')}{\int \Phi^2(y)\,dy}. \tag{B22}$$

As a consequence, by restricting to the first N excited states (which are perpendicular to the ground state $\psi_0(x)$), we get a lower bound in the form

$$\Lambda \geq \max_{c_1,...c_{N_s}} \frac{\int dy \int dy'\, \Phi(y)K_\beta(y,y')\Phi(y')}{\int \Phi^2(y)\,dy} \tag{B23}$$

$$\Phi = c_1\psi_1 + ... + c_{N_s}\psi_{N_s} \tag{B24}$$

When the Schrödinger equation limit is valid, Eqs. (B19)-(B21) allow to approximate the relaxation rates of the Metropolis algorithm from the eigenvalues of the Schrödinger equation:

$$\lambda_n = 1 - \sigma^2\epsilon_n. \tag{B25}$$

Upon increasing of the typical size of jump length, the operator $\widehat{K}_\beta$ will mix different Schrödinger eigenmodes and this estimate will no longer be valid.

Solving the present optimization problem is equivalent to finding the largest eigenvalue of the reduced $N_s \times N_s$ matrices $K^{(N_s)}$ with matrix elements

$$K_{nm} = \int dy \int dy' \, \psi_n(y) K_\beta(y, y') \psi_m(y'), \tag{B26}$$

with the truncation $1 \leq n, m \leq N_s$ where the positive integer N_s gives the number of retained eigenfunctions. We will show in section E that with a few modes only, very good quantitative estimates for Λ can be obtained by this approach, even where Eq. (B19) is no longer valid, far from the small jump amplitude limit.

In cases where the potential $U(x)$ is even (as assumed here), the eigenbasis $\psi_n(x)$ will split into symmetric and anti-symmetric eigenfunctions. The master equation kernel K_β inherits the symmetry properties of the potential $U(x)$ and the matrix elements Eq. (B26) will be non-zero only for wavefunctions from the same parity. The truncated matrix will thus split into a direct sum of even-even and odd-odd matrices. The mapping to the Schrödinger equation ensures that at least in the small jump limit, Λ will be in the odd sector, but we will show in section E an example where this is not necessarily true for large a.

Appendix C: Overview of the cases investigated and main tools of analysis

1. Potentials, sampling choice, and observables

The claims put forward in the main text rely on the study of a number of confining potentials of the form $U(x) \propto |x|^\alpha$, with $\alpha > 0$. Some emphasis has also been put in the study of confinement by hard walls, the box potential, where $U(x) = 0$ for $x \in [-L, L]$ and $U(x) = \infty$ for $|x| > L$.

In these potential landscapes, we have changed the sampling method, varying the distribution $f(\eta)$ of attempted jumps. Scaling out the jump's typical length a, we obtain the dimensionless distribution $f(z)$:

$$w(\eta) = \frac{1}{a} f\left(\frac{\eta}{a}\right). \tag{C1}$$

Different choices were made, symmetric for simplicity ($f(z) = f(-z)$):

- Gaussian distribution of jumps

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{C2}$$

- Exponential distribution

$$f(z) = \frac{1}{2} e^{-|z|} \tag{C3}$$

- Flat distribution

$$f(z) = \theta\left(\frac{1}{2} - |z|\right) \tag{C4}$$

- Other more specific choices, as introduced to analyze the box confinement, see section D.

In order to study convergence to equilibrium, it is important to pay attention to the symmetry of the observables used, for it affects relaxation rates. This can be understood from the Schrödinger reformulation, where excited states of increasing order are alternatively even and odd in x, while their energy is directly related to the relaxation rate, see Eq. (B21). Therefore, we can use even observables (with even initial conditions) to suppress a slower relaxation rate corresponding to an odd mode, allowing to estimate λ_1 and λ_2 from sampling. In particular, we measured

$$\mathcal{O}_1(x) = (x - 0.5)^2, \quad \text{and} \quad \mathcal{O}_2(x) = |x|. \tag{C5}$$

2. Probing localization with the Inverse Participation Ratio

Since the transition we identify amounts to a localization of the convergence error onto well defined positions, it is essential to discriminate delocalized states, from localized ones. To this end, we discretize the integral in the Master equation Eq. (B1) into a sum of N_d terms, with a running position index to denote lattice sites $1 \leq i \leq N_d$. We then introduce the inverse participation ratio for an eigenvector $\Psi_\lambda(x)$ as

$$\mathrm{IPR}(\lambda) = \sum_{i=1}^{N_d} |\Psi_\lambda(i)|^4 \left/ \left(\sum_i |\Psi_\lambda(i)|^2 \right)^2 \right. . \tag{C6}$$

This quantity can vary between two extremes. If the eigenfunction is completely delocalized over the whole system, so that $\Psi_\lambda(i)$ is a constant (normalization is irrelevant here), then $\mathrm{IPR}(\lambda) = 1/N_d$, with $N_d \gg 1$. If on the other hand, $\Psi_\lambda(i)$ vanishes on all sites but one, then $\mathrm{IPR}(\lambda) = 1$, irrespective of N_d. If $\Psi_\lambda(i)$, the discretization of an eigenvector $\psi_\lambda(x)$, is well defined in the continuum limit, then $\mathrm{IPR} \to 0$ as N_d^{-1}. On the other hand, if part of the eigenfunction localizes, a slower decay as a function of N_d will be observed and the discrete eigenfunction $\Psi_\lambda(i)$ will not converge to a well defined continuum limit.

3. Numerical diagonalization

Numerical diagonalization of the discretized form of the Master equation Eq. (B1) allows to find the spectrum of eigenvalues. The master equation was discretized by a uniform mesh with N_d sites. The integration was replaced by a sum over accessible neighbors, ensuring probability conservation. For particles in a box $x \in [-L, L]$, the first and last points of the mesh were set to $-L$ and L respectively. For the harmonic potential, the first and last points were set to $\pm X_{\max}$ where $X_{\max}$ is the largest $|x|$ allowed by the mesh. The results in Fig. 1 from the main text were obtained for $X_{\max} = 10$ (in units of thermal length in the harmonic potential). We increased $X_{\max}$ up to 30 to check that the results were independent on this choice of $X_{\max}$. To obtain eigenvalues and eigenvectors, we used the diagonalization routines from the eigen++ library. To avoid the appearance of spurious complex eigenvalues due to rounding errors in diagonalization algorithms, we took advantage of detailed balance to transform the kernel of the integral Master equation into its symmetric form Eq. (B11). Considering the fast increase of the numerical time required for full diagonalization with matrix size, we used this approach for $N_d \leq 10^4$.

4. Numerical iteration of the Master equation

To study the relaxation of the error $\delta P_n(x)$, it is also possible to follow the evolution of a fixed initial state by successive iterations of the Master equation. This approach is computationally less demanding than full diagonalization. With this method, we ran simulations up to $N_d = 2 \times 10^5$.

5. Monte Carlo simulations

To put our analytical calculations to the test and assess the accuracy of the predicted bounds, we have directly simulated the dynamics defined by the Master equation. The Metropolis rule, spelled out in the main text, defines a Markov chain which can be readily simulated by means of classical Monte Carlo. For large enough time n, equilibrium will be reached and the walker's position will sample the Gibbs-Boltzmann distribution (B7). The sampling scheme obeys detailed balance [16], which guarantees the existence of a steady state, that is furthermore unique for an ergodic irreducible chain [18]. We are interested in the long-time approach towards the equilibrium distribution. To gather statistics, we perform the simulation until $n = 30$ typically, and repeat this for $m = 10^{10}$ or 10^{11} independent samples. At every time step n, we compute a number of observables, see section C 1. An observable $\mathcal{O}$ is then averaged over all m samples at fixed time n, leading to $\overline{\mathcal{O}}$.

$$\overline{\mathcal{O}}(n) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{O}^{(i)}(n) \tag{C7}$$

where the observable measured at time n in the ith sample is $\mathcal{O}^{(i)}(n)$.

Our Monte Carlo estimates of the largest eigenvalue are obtained by fits to the deviation from the equilibrium value at time n of the form $|\overline{\mathcal{O}}(n) - \langle\mathcal{O}\rangle_{\mathrm{eq}}| = c_1\lambda_1^n + c_2\lambda_2^n$, where $\langle\mathcal{O}\rangle_{\mathrm{eq}}$ is the equilibrium value, reached after long times. We exclude the first values (typically $n < 5$) to minimize influence of transient behavior and c_1 and c_2 are free constants. Technically, we use an analytical value for $\langle\mathcal{O}\rangle_{\mathrm{eq}}$, if known, or Monte Carlo results at $n = 200$, where the statistical error dominates over the systematic deviation. Performing this procedure for multiple jump distributions parametrized by a allows us to gather measurements of relaxation rates, which we can compare to our analytical results and the other computational approaches. For a reliable fit, it is necessary to have good estimates of the standard errors of the measured mean values; *Welford's algorithm* has been used [47, p. 232]. The acceptance probability is computed during an independent simulation of a single particle over $1.1 \cdot 10^6$ Metropolis steps, where the first 10^5 steps are ignored for the average.

We verify the quality of the three numerical approaches by comparing them to each other, and to the analytical results for the case of the box potential. Figure 5 shows the estimates of Λ obtained from the Monte Carlo simulations and λ_1 obtained from the diagonalization (the upper envelope of the spectrum).
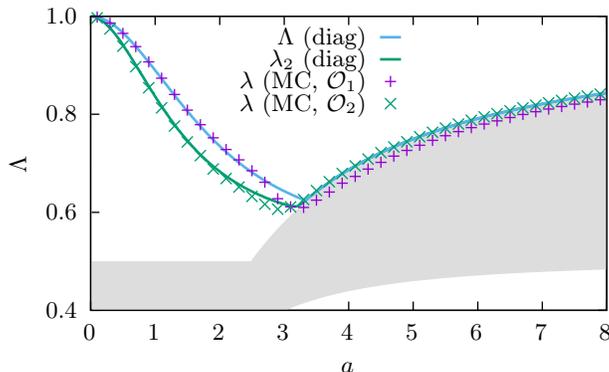

FIG. 5. Convergence rate Λ vs. jump amplitude a for a harmonic confinement ($U(x) = x^2/2$) and a flat $w(\eta)$ distribution as in Eq. (C4) (cf. Fig. 1 of the main text). Comparison between the direct Monte Carlo simulations measure (MC) and the numerical diagonalization technique. Both methods agree very well. The Monte Carlo approach needs to look at symmetric and asymmetric observables to measure the different branches of the spectrum. The observables used are provided in Eq. (C5). For $a < a^* \simeq 3.33$, using the asymmetric observable $\mathcal{O}_1$ provides a very good estimate of the largest eigenvalue $\lambda_1 = \Lambda$. For $a < a^*$, using the even observable $\mathcal{O}_2$ yields the second largest eigenvalue λ_2. For $a > a^*$, the largest of both (λ from $\mathcal{O}_2$) gives a very good estimation of Λ. The singular continuum is shown by the grey region. As in the main text, a is given in units of the thermal length at equilibrium.

Figure 6 shows the shape of the deviation $\delta P_n(x)$ from the equilibrium distribution at a finite time, for uniform jumps in the harmonic potential. The direct Monte Carlo results and the results from the iteration of the Master equation are compatible within statistical fluctuations. This lends a high confidence in the results of the iteration for longer times, which are shown in the main manuscript.

Appendix D: The box potential

Obtaining exact analytical results in the general case of a potential $U(x)$ in $|x|^\alpha$ seems out of reach. Yet, the box potential, where the random walker moves freely between hard walls at $\pm L$, is a useful model system that presents the whole range of phenomena observed generically. A key aspect lies in the choice of the jump distribution scaling function $f(z)$, that can lead to any of the two scenarios mentioned in the main text: ⓐ a gapped spectrum for which the discrete branch λ_1 is above the singular continuum, for all jump amplitudes a (regular case, where the singular continuum, although present, does not play a role in the long time error, and there is no localization); ⓑ a gapless spectrum where the singular continuum becomes the dominant relaxation mode for $a > a^*$. This situation ⓑ where localization appears is the generic case. This is why we focused on case ⓑ in the main text.

It is useful here to introduce the late-time rejection probabilities $R_a(0)$ at $x = 0$, and $R_a(\text{edge})$ at the system's edge, meaning $x = \pm L$ in the box case. Both depend on a. For $a \to 0$, we have $R_a(0) < R_a(\text{edge})$: all moves from $x = 0$ are accepted (vanishing rejection probability), while only half of them are, starting from the edge (both in the box case, and when exponent $\alpha > 1$, leading to a convex-up confining potential). A careful inspection of all the numerical data we gathered shows that case ⓐ corresponds to $R_a(0) < R_a(\text{edge})$ for all a; ⓑ is for the situation where $R_a(0)$ and $R_a(\text{edge})$ do cross for $a = a^*$, so that $R_a(0) > R_a(\text{edge})$ for $a > a^*$. It is then straightforward to realize that
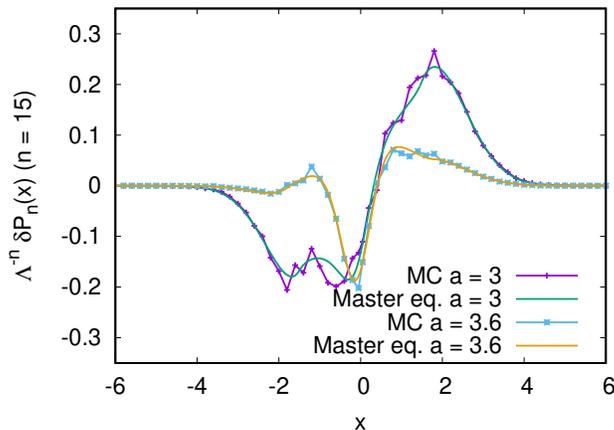
FIG. 6. Shape of the rescaled deviation $\delta P_n(x)$ vs x from the equilibrium distribution at finite times (here $n = 15$). The symbols show results of direct Monte Carlo and the lines show the results of the iteration of the Master equation; both methods are in good agreement. Same confinement and jump distribution as in Fig. 5.

the behavior of $w(\eta)$ at small η discriminates the two regimes: if $w(\eta)$ decreases when increasing $|\eta|$, we have case (a); if $w(\eta)$ increases when increasing $|\eta|$, we have case (b). We considered the family of polynomial w-functions, for instance piecewise linear or quadratic such as

$$w^{(1)}(\eta) = \frac{1}{a(2b+c)} \left(b + c\frac{a - |\eta|}{a} \right) \theta(a - |\eta|) \tag{D1}$$

$$w^{(2)}(\eta) = \frac{1}{a(2b+4c/3)} \left(b + c\frac{a^2 - \eta^2}{a^2} \right) \theta(a - |\eta|) \tag{D2}$$

parameterized by the constants b and c, in addition to the jump size a: positive values of c define convex-down functions, pictorially written $w^{\cap}(\eta)$ and associated to case (a); $c < 0$ defines convex-up functions, denoted $w^{\cup}(\eta)$, associated to case (b). We take hereafter $L = 1$, without loss of generality.

1. Numerical results

We show in Fig. 7 the spectrum of $F_\beta(x, x')$ obtained by numerical diagonalization with the parabolic jump distribution $w^{(2)}(\eta)$, either of the type $w^{\cap}$ or $w^{\cup}$. The distinction between the gapped ((a), with $w^{\cap}$) and gapless ((b), with $w^{\cup}$) cases appears. At variance with case (a), (b) shows a regime for $a > a^* \simeq 1.79$ where the singular continuum defines the dominant relaxation mode, so that localization ensues. The crossing of the curves $R_a(0)$ and $R_a(\text{edge})$ for a slightly below a^* is also visible. For the present box potential, $U(x)$ either vanishes inside the box, or diverges outside. Hence, the value of inverse temperature is irrelevant. We have checked that the qualitative results remain unchanged for all monotonous (for $\eta > 0$) jump distributions $w(\eta)$ even in η, in particular using the piecewise linear distribution $w^{(1)}(\eta)$. Thus for the box potential $U(x) = 0$ ($x \in [-1, 1]$) the presence or absence of localization is determined by whether $w(\eta)$ is either minimum or maximum at $\eta = 0$.

2. Exact results on eigenvalues of the Monte-Carlo Master equation for a box potential

We have seen above that the box potential subsumes the gapless/gapped spectra dichotomy, corresponding to the (a) absence/(b) presence of localization. Besides, the shape of the gapless spectrum shown in Fig. 7 is closely reminiscent of its counterpart presented in the main text. We thus take advantage of the fact that exact results can be obtained with the box confinement, to shed new light on the localization phenomenon and its scaling properties.

For a box potential $U(x) = 0$ ($x \in [-1, 1]$) the eigenvalue problem of the Monte-Carlo Master equation simplifies into:

$$\lambda \Psi_\lambda(x) = \int_{-1}^{1} \Psi_\lambda(y)\, w(x - y)\, dy + \left[1 - \int_{-1}^{1} w(y - x)\, dy \right] \Psi_\lambda(x) \tag{D3}$$
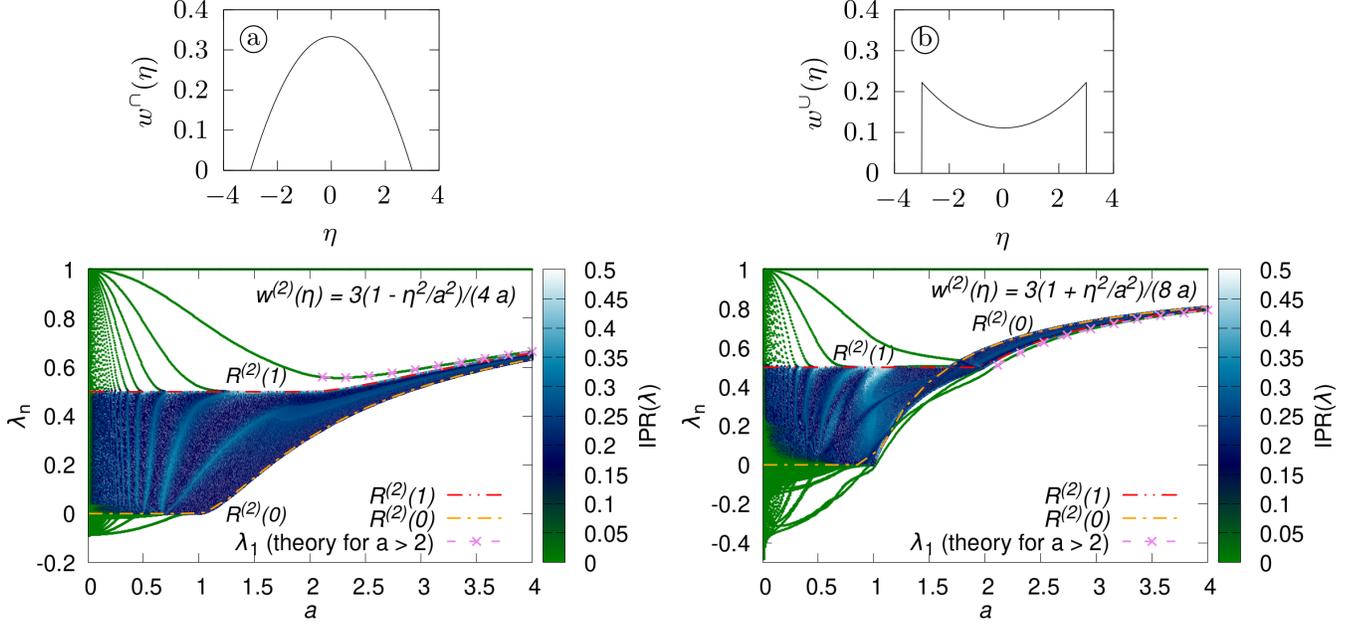
FIG. 7. Spectrum of the kernel $F_\beta(x, x')$ for the box potential, with either ⓐ $w^{(2)}(\eta) = w^\cap(\eta)$, left column ($b = 0, c = 1$), or ⓑ $w^{(2)}(\eta) = w^\cup(\eta)$, right column ($b = 2, c = -1$). In both cases, the singular continuum appears in dark blue. In the gapped case ⓐ where there is no transition, it remains below the discrete λ_1 branch; the optimal jump can be found by minimizing λ_1 in Eq. (D8) which gives $a_{\rm opt} = \sqrt{c(1 + 3k^{(2)})(b + c)^{-1}} \simeq 2.251$. In the gapless case ⓑ on the right, the singular continuum becomes the dominant relaxation mode for $a > a^*$. Localization ensues, for $a > a^* = a_{\rm opt} \simeq 1.79$. Lengths are in units of the box size ($L = 1$).

where λ is the eigenvalue and $\Psi_\lambda(x)$ is the eigenvector. For $a > 2$ and the two choices $w^{(p)}(\eta)$ ($p = 1, 2$) from Eq. (D2), this equation reduces to a second order differential equation which can be solved to yield a single eigenvalue $\lambda < 1$. This eigenvalue can be written in the form:

$$\lambda_1^{(p)} = R^{(p)}(1) + (k^{(p)} - 1)[R^{(p)}(1) - R^{(p)}(0)]. \tag{D4}$$

Here, the rejection probability $R^{(p)}(x)$, with index $p = 1$ and $p = 2$ is given by

$$R^{(p)}(x) = 1 - \int_{-1}^{1} w^{(p)}(x - y)dy. \tag{D5}$$

The constants $k^{(p)}$ read $k^{(1)} \simeq 1.439$ and $k^{(2)} \simeq 1.356$; they are the solutions of

$$\frac{1}{\sqrt{k^{(1)}}} \operatorname{arccoth} \sqrt{k^{(1)}} = 1 \ , \ \sqrt{k^{(2)}} \operatorname{arccoth} \sqrt{k^{(2)}} = \frac{3}{2}. \tag{D6}$$

In both cases, $k^{(p)} > 1$, which implies that if $R^{(p)}(1) > R^{(p)}(0)$, the eigenvalue $\lambda_1^{(p)} > R^{(p)}(1) = \max_x R^{(p)}(x)$ (case ⓐ). On the contrary, if $R^{(p)}(1) < R^{(p)}(0)$, $\lambda_1^{(p)} < R^{(p)}(1) = \min_x R^{(p)}(x)$. Thus, it is indeed the comparison between $R^{(p)}(1)$ and $R^{(p)}(0)$ which determines if $\lambda_1^{(p)}$ is above or below the singular continuum, thereby discriminating between ⓐ and ⓑ.

Evaluating the integrals in Eq. (D5) we find the explicit expressions (we remind that they are valid for $a > 2$):

$$\lambda_1^{(1)} = \frac{-2ab + 2a^2b + c - 2ac + a^2c + k^{(1)}c}{a^2(2b + c)} \tag{D7}$$

$$\lambda_1^{(2)} = \frac{-3a^2b + 3a^3b + c - 3a^2c + 2a^3c + 3k^{(2)}c}{a^3(3b + 2c)}. \tag{D8}$$

To summarize at this point, for both parametrizations of the jump distribution function and for $a > 2$, there is a single eigenvalue $\lambda < 1$ (besides the singular continuum). This eigenvalue lies above $\max_x R(x)$ or below $\min_x R(x)$ depending on whether $w(\eta)$ is (a) maximum or (b) minimum at $\eta = 0$. Considering that the interval $(\min_x R(x), \max_x R(x))$ is actually filled with singular eigenvalues $\lambda = R(x)$, we describe this situation as λ_1 lying above or below the singular continuum.

3. Exact results on the localization of the error $\delta P_n(x)$

We wish to describe analytically the relaxation of the error $\delta P_n(x)$ when λ_1 lies below $\max_x R(x)$. We remind that for the two parametrizations of $w(\eta)$ from the previous section, a stronger result holds and that in this case, $\lambda_1 < \min_x R(x)$. The Master equation for the error $\delta P_n = P_n - P_\infty$ reads:

$$\delta P_{n+1}(x) = \int_{-1}^{1} \delta P_n(y) w(x-y) dy + R(x) \delta P_n(x) \qquad \text{with} \qquad R(x) = 1 - \int_{-1}^{1} w(x-y) dy. \tag{D9}$$

Normalization implies $\int \delta P_0(x) dx = \int \delta P_n(x) dx = 0$.

For $a > 2$ and focusing on the case of a parabolic jump distribution in Eq. (D2), it is possible to simplify notations:

$$w^{(2)}(\eta) = w_0 + w_2 \eta^2, \qquad R(x) = r_0 - r_2 x^2, \qquad r_0 = 1 - 2w_0 - \frac{2w_2}{3}, \qquad r_2 = 2 w_2, \quad \text{with } w_2 > 0. \tag{D10}$$

We then look at symmetric initial conditions $\delta P_0(x) = \delta P_0(-x)$:

$$\delta P_{n+1}(x) = w_2 \int_{-1}^{1} \delta P_n(y) y^2 dy + R(x) \delta P_n(x). \tag{D11}$$

We introduce the generating function:

$$G(x,z) = \sum_{n=0}^{\infty} \delta P_n(x) z^n = \delta P_0(x) + w_2 z \int_{-1}^{1} G(y,z) y^2 dy + z R(x) G(x,z) \tag{D12}$$

from which we get

$$G(x,z) = \frac{1}{1 - zR(x)} \left(\delta P_0(x) + w_2 z \int_{-1}^{1} G(y,z) y^2 dy \right). \tag{D13}$$

We then solve for $G_2(z) = \int_{-1}^{1} G(x,z) x^2 dx$. Integrating the Master equation we find:

$$G_2(z) = \frac{\sqrt{r_2 z}}{\sqrt{1 - r_0 z} \arctan \frac{\sqrt{r_2 z}}{\sqrt{1-r_0 z}}} \int_{-1}^{1} \frac{\delta P_0(x) x^2}{1 - zR(x)} dx. \tag{D14}$$

To make further progress, we choose as an initial condition

$$\delta P_0(x) = r^{-1} \theta(r - |x|) - 1, \tag{D15}$$

which allows to compute the integral in Eq. (D14) explicitly:

$$G_2(z) = \int_{-1}^{1} G(x,z) x^2 dx = \frac{2}{r_2 z} \left(1 - \frac{1}{r} \frac{\arctan rZ}{\arctan Z} \right) \qquad \text{with} \qquad Z = \sqrt{\frac{r_2 z}{1 - r_0 z}}. \tag{D16}$$

We note that $G_2(z)$ is the generating function for the series

$$S_n = \int_{-1}^{1} \delta P_n(x) x^2 dx \tag{D17}$$

which can be viewed as the error in the variance x^2 at step n. The general method of singularity analysis [48] allows to find the asymptotic behavior of a series from the analysis of the singularities of its generating function in the complex
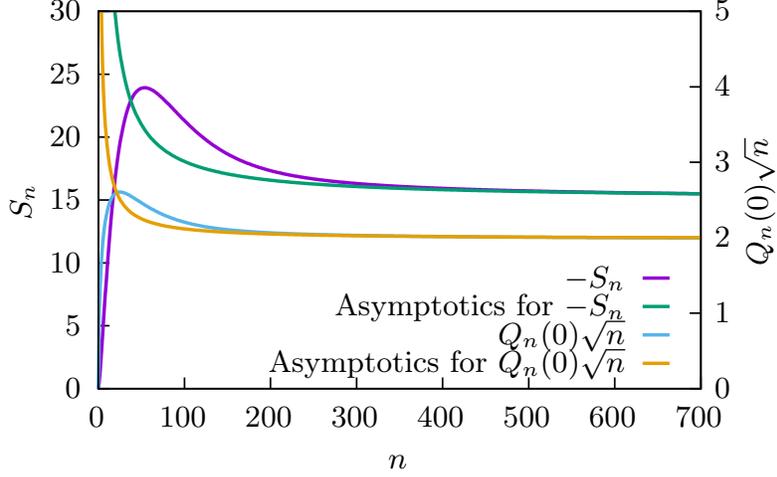
FIG. 8. Comparison of the numerical results for $S_n = \int_{-1}^{1} \delta P_n(x) x^2 dx$ and $Q_n(x) = R(x)^{-n} \delta P_n(x)$ with the asymptotic estimates in Eqs. (D18,D25). We chose $w^{(2)}(\eta) = 3(1 + \eta^2/a^2)/(8a)$ and $a = 2.1$; the number of sites for the discretization of the Master equation was $N_d = 2 \times 10^5$. The initial conditions were $\delta P_0(x) = 2\theta(1/2 - |x|) - 1$ $(r = 1/2)$. We note that even if simulations with a finite N_d cannot reproduce the asymptotic power in the $n \to \infty$ behavior (because of the discrete spectrum), the agreement at finite but large n is nevertheless very good.

plane which are nearest to the origin $z = 0$. For $G_2(z)$ the singularity closest to the origin is $z = r_0^{-1}$. The asymptotic expansion of the generating function near this singularity allows us to find:

$$S_n \simeq -\frac{2(1-r)}{r^2} \frac{r_0^{3/2}}{(\pi r_2)^{3/2}} \frac{r_0^n}{n^{3/2}} - \frac{(1-r)r_0^{3/2}[-48r^2 r_0 + 4\pi^2(1+r+r^2)r_0 - 15\pi^2 r^2 r_2]}{4\pi^{7/2} r^4 r_2^{5/2}} \frac{r_0^n}{n^{5/2}}. \tag{D18}$$

We then introduce the functions $Q_n(x)$ as

$$\delta P_n(x) = R(x)^n Q_n(x), \tag{D19}$$

and the recurrence equation Eq. (D11) becomes:

$$Q_{n+1}(x) = \frac{w_2}{R(x)} \int_{-1}^{1} \delta P_n(y) R(x)^{-n} y^2 dy + Q_n(x) \tag{D20}$$

$$= \frac{w_2}{R(x)} \sum_{m=0}^{n} \int_{-1}^{1} P_m(y) R(x)^{-m} y^2 dy + Q_0(x). \tag{D21}$$

Taking the limit $n \to \infty$ and under the proviso that the series converges, we find

$$Q_\infty(x) = \frac{w_2}{R(x)} G_2(R(x)^{-1}) + \delta P_0(x) \tag{D22}$$

where $G_2(z)$ is defined in Eq. (D14). The problem with this expression is that $R(x)^{-1} \geq r_0^{-1}$ lies outside the radius of convergence $|z| \leq r_0$ of $G_2(z)$, so this formula is valid only at $x = 0$ when $R(0) = r_0$. Using the obtained value of $G_2(r_0^{-1})$, we find

$$Q_\infty(0) = 0. \tag{D23}$$

We can then obtain an asymptotic estimate for $Q_n(0)$:

$$Q_n(0) = Q_\infty(0) - \frac{r_2}{2r_0} \sum_{m=n}^{\infty} \int_{-1}^{1} \delta P_m(y) r_0^{-m} y^2 dy \tag{D24}$$

$$\simeq \frac{1-r}{r^2} \frac{r_0^{1/2}}{\pi^{3/2} r_2^{1/2}} \frac{4n+1}{2n^{3/2}} + \frac{(1-r)r_0^{1/2}[-48r^2 r_0 + 4\pi^2(1+r+r^2)r_0 - 15\pi^2 r^2 r_2]}{12\pi^{7/2} r^4 r_2^{3/2}} \frac{r_0^n}{n^{3/2}}. \tag{D25}$$

From Eq. (D19), it follows that $\delta P_n(0) = r_0^n Q_n(0)$ where we used $R(0) = r_0$. Equation (D25) indicates that $Q_n(0)$ decays as a power law $n^{-1/2}$ for large n:

$$\delta P_n(0) \simeq \frac{2(1-r)}{\pi^{3/2} r^2} \sqrt{\frac{r_0}{r_2}} \frac{r_0^n}{n^{1/2}}. \tag{D26}$$

Comparison of Eqs. (D18,D25) with numerical simulations of discretized approximation of the Master equation are shown in Fig. 8.

For $x \neq 0$ the series becomes diverging and $Q_\infty(x)$ does not exist. The leading asymptotic behavior can be extracted from the singular behavior of $Q(x,z)$ near $z = r_0^{-1}$:

$$\delta P_n(x) \simeq -\frac{1-r}{r^2 x^2} \left(\frac{r_0}{\pi r_2} \right)^{3/2} \frac{r_0^n}{n^{3/2}}. \tag{D27}$$

Interestingly, we find that the ratio $\delta P_n(x)/\delta P_n(0)$ (for $x \neq 0$) does not decay exponentially but as a power law n^{-1}. Comparing Eqs. (D26) and (D27), we thus proved the main property of the localizing contribution to the error $\delta P_n(x)$:

$$\lim_{n \to \infty} \delta P_n(x)/\delta P_n(0) = 0 \quad \text{whenever} \quad x \neq 0. \tag{D28}$$
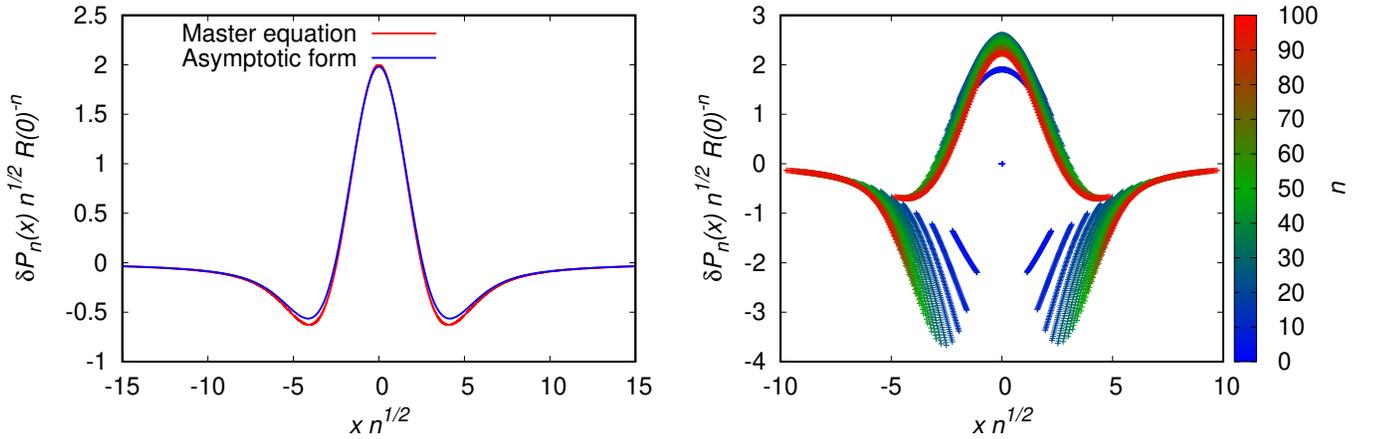

FIG. 9. Relaxation of $\delta P_0(x) = 2\theta(1/2 - |x|) - 1$ (i.e. $r = 1/2$) for $w^{(2)}(\eta) = 3(1 + \eta^2/a^2)/(8a)$ and $a = 2.1$; same parameters as Fig. 8. This figure compares the rescaled $\delta P_n(x)$ as obtained by direct iteration of the Master equation at step $n = 700$, with the scaling prediction of Eqs. (D29) and (D35). The right panel shows convergence for $n \leq 100$; the moving discontinuity is a trace of the initial distribution $P_0(x)$, that is discontinuous at $x = \pm 1/2$. Hence, as time proceeds, the central peak extends further, and ultimately reaches the scaling form shown on the left plot. We also illustrated the convergence to the scaling function Eq. (D35) on Fig. 3 from the main text using the same dataset.

To find a uniform approximation to $\delta P_n(x)$, we assume the following scaling form

$$\delta P_n(x) = \frac{R(0)^n}{\sqrt{n}} \varphi(x\sqrt{n}). \tag{D29}$$

Using Eq. (D18), we can approximate

$$r_0^{-n} (\delta P_{n+1}(x) - R(x)\delta P_n(x)) = \frac{r_2 r_0^{-n}}{2} \int_{-1}^{1} \delta P_n(y) y^2 dy \simeq -\frac{1-r}{r^2} \sqrt{\frac{r_0^3}{\pi^3 r_2 n^3}}. \tag{D30}$$

On the other hand using Eq. (D29), we find:

$$r_0^{-n} (\delta P_{n+1}(x) - R(x)\delta P_n(x)) = \frac{r_0}{\sqrt{n+\epsilon}} \varphi(x\sqrt{n+\epsilon}) - (r_0 - r_2 x^2)\varphi(x\sqrt{n}) \tag{D31}$$

$$\simeq \frac{r_2 x^2}{n^{1/2}} \varphi(x\sqrt{n}) - \frac{\epsilon r_0}{2n^{3/2}} \varphi(x\sqrt{n}) + \frac{\epsilon r_0 x}{2n} \varphi'(x\sqrt{n}), \tag{D32}$$

where we introduced a formal small expansion parameter $\epsilon = 1$ and expanded to first order in ϵ. Introducing $\widetilde{x} = x\sqrt{n}$ and combining Eqs. (D30,D32), we find a first order differential equation on the scaling function $\varphi(\widetilde{x})$:

$$\frac{r_0\widetilde{x}}{2}\varphi'(\widetilde{x}) + r_2\widetilde{x}^2\varphi(\widetilde{x}) - \frac{r_0}{2}\varphi(\widetilde{x}) = -\frac{1-r}{r^2}\sqrt{\frac{r_0^3}{\pi^3 r_2}}. \tag{D33}$$

Equation (D33) admits a single symmetric solution which can be expressed in a compact form introducing the Dawson function:

$$D_+(x) = e^{-x^2} \int_0^x e^{y^2} dy. \tag{D34}$$

We get

$$\varphi(\widetilde{x}) = \frac{2(1-r)}{\pi^{3/2}r^2}\sqrt{\frac{r_0}{r_2}}\left[1 - 2\widetilde{x}\sqrt{\frac{r_2}{r_0}}D_+\left(\widetilde{x}\sqrt{\frac{r_2}{r_0}}\right)\right]. \tag{D35}$$

From the results

$$D_+(0) = 0 \qquad \text{and} \qquad D_+(x) \sim \frac{1}{2x} + \frac{1}{4x^3} \quad \text{for} \quad x \to \infty, \tag{D36}$$

we recover Eqs. (D26) and (D27). Hence, the scaling assumption (D29) appears fully consistent. The comparison between $\delta P_n(x)$ obtained by iteration of the Master equation and the prediction of the scaling form is shown in Fig. 9.

Appendix E: Analytical calculation of the MC relaxation rate for an Harmonic potential

In this section, we show two examples of analytic calculations in the truncated Schrödinger eigenbasis, as introduced in sub-sections B 3, B 4.

For a harmonic potential $U(x) = x^2/2$, the (dimensionless) Schrödinger equation reduces to the celebrated eigenvalue equation of a quantum harmonic oscillator:

$$\epsilon_n \psi_n(x) = -\psi_n''(x) + \frac{x^2}{4}\psi_n(x). \tag{E1}$$

The corresponding eigenfunctions can be expressed through Hermite polynomials H_n:

$$\psi_n(x) = \frac{1}{N_n}e^{-x^2/4}H_n(2^{-1/2}x) \tag{E2}$$

$$N_n = \sqrt{\int dx H_n(2^{-1/2}x)^2 e^{-x^2/2}dx} \tag{E3}$$

where N_n is the normalization. To obtain an approximation (and lower bound) for Λ, we calculate the matrix elements.

$$K_{nm} = \int dy \int dy' \psi_n(y)K_{\beta=1}(y,y')\psi_m(y') \tag{E4}$$

where the integral kernel is given by Eq. (B11). Here, as in the main text, we have expressed positions in units of thermal length which amounts to setting $\beta = 1$. This gives the following expression for K_{nm}

$$K_{nm} = \int_{-\infty}^{\infty} dy\, w(y) \int_{-\infty}^{\infty} dx\, \psi_m(x - y/2)\, \psi_n(x + y/2)e^{-|xy|/2} + \int_{-\infty}^{\infty} dx\, \psi_n(x)\psi_m(x)R(x) \tag{E5}$$

where $R(x)$ is the rejection probability. For sufficiently simple expressions of $w(\eta)$ and low values of indices n and m, the integrals can be evaluated analytically.

For a symmetric potential $U(x)$, the truncated matrix splits into a direct sum of odd-even subspaces, as discussed in B 4. The sequence of $N_s \times N_s$ truncated matrices built from odd eigenfunctions will be noted $K_o^{(N_s)}$. For example, the matrix $K_o^{(1)}$ reduces to a single scalar K_{11} while $K_o^{(2)}$ is the 2×2 symmetric matrix with matrix elements corresponding to K_{nm} for $n, m \in \{1, 3\}$; higher order approximations are obtained similarly. Likewise with the even sector: the

sequence of $N_s \times N_s$ matrices $K_e^{(N_s)}$ is constructed from even wavefunctions. Since $\psi_0(x)$ is an exact eigenvector for any value of the jump amplitude a, the lowest order $K_e^{(1)}$ is given by the scalar K_{22}; the next order $K_e^{(2)}$ is given by K_{nm} for $n, m \in \{2, 4\}$ and so forth with increasing order N_s. The relaxation rate is then approximated by

$$\Lambda_o^{(N_s)} = \max \text{eigenvalues}(K_o^{(N_s)}) \ , \ \Lambda_e^{(N_s)} = \max \text{eigenvalues}(K_e^{(N_s)}) \tag{E6}$$

$$\Lambda^{(N_s)} = \max \left\{ \Lambda_o^{(N_s)}, \Lambda_e^{(N_s)} \right\}. \tag{E7}$$

Below, we considered the case of an harmonic potential $U(x)$ for several possible shapes of $w(\eta)$. In all cases, we found that the following approximation is very accurate:

$$\Lambda \simeq \max \left\{ \Lambda^{(N_s)}, \max_x R(x) \right\} \tag{E8}$$

where $R(x)$ is the rejection probability. This expression is operational even for small values $N_s = 2$, and indistinguishable from numerical diagonalization at $N_s = 6$.

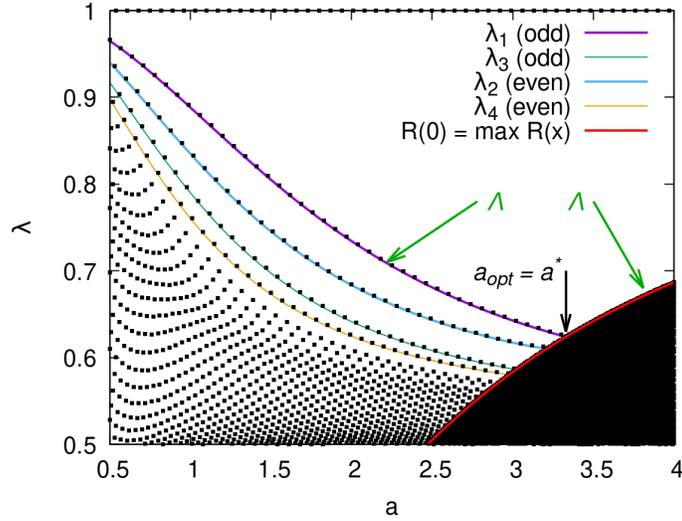### 1. Harmonic potential with a flat jump distribution $w(\eta)$

FIG. 10. Comparison between numerical eigenvalues and the analytical approximation for $U(x) = x^2/2$ with flat jump distribution $w(\eta) = \theta(a - |\eta|)/(2a)$. The symbols show numerical eigenvalues of the discretized master equation ($N_d = 2000$, $X_{max} = 10$) while the continuous curves display the analytical results for the four slowest relaxation eigenmodes: $\lambda_1, \lambda_2, \lambda_3$ and λ_4. The analytical approximation is obtained from the two highest eigenvalues of $K_o^{(6)}$ (yielding λ_1 and λ_3) and $K_e^{(6)}$ (yielding λ_2 and λ_4). The analytical top eigenvalue Λ is thus here $\Lambda^{(6)}$. We see that below Λ, the other eigenvalues are also indistinguishable from the numerical eigenvalues, until the crossing with the singular continuum, bounded by $R(0) = \max_x R(x)$. This shows that lower eigenvalues are also very accurately reproduced in this approximation, for values of a below the localisation transition of the corresponding mode. The maximum rejection probability is given by Eq. (E13).

We report explicit results for the lowest order terms for $w(\eta) = \theta(a - |\eta|)/(2a)$, indicating that the scaling function f reads $f(z) = \theta(1 - |z|)/2$. We find:

$$K_{11}(a) = 1 - \frac{a^3 \text{erfc}\left(\frac{a}{2\sqrt{2}}\right) - 2\sqrt{\frac{2}{\pi}} \left(a^2 + 8\right) e^{-\frac{a^2}{8}} + 16\sqrt{\frac{2}{\pi}}}{6a} \tag{E9}$$

$$K_{13}(a) = \frac{-\sqrt{2\pi} a^5 \text{erfc}\left(\frac{a}{2\sqrt{2}}\right) + 4\left(a^4 + a^2 + 8\right) e^{-\frac{a^2}{8}} - 32}{20\sqrt{3\pi}a} \tag{E10}$$

$$K_{33}(a) = 1 - \frac{1}{420} \left(5a^4 + 63a^2 + 210\right) a^2 \text{erfc}\left(\frac{a}{2\sqrt{2}}\right) + \frac{\left(20a^6 + 207a^4 + 372a^2 + 2976\right) e^{-\frac{a^2}{8}}}{420\sqrt{2\pi}a} - \frac{124\sqrt{\frac{2}{\pi}}}{35a} \tag{E11}$$

The steady state rejection probability R_∞ is given by:

$$R_\infty = \frac{2}{a}\sqrt{\frac{2}{\pi}}\left(e^{-\frac{a^2}{8}} - 1\right) + \text{erf}\left(\frac{a}{2\sqrt{2}}\right),$$

(E12)

and the maximum rejection probability reads:

$$R(0) = \max_x R(x) = 1 - \frac{\sqrt{\pi/2}}{a}\text{erf}\left(\frac{a}{\sqrt{2}}\right).$$

(E13)

This gives explicit expressions for the first two orders:

$$\Lambda^{(1)} = \Lambda_o^{(1)} = K_{11}(a)$$

(E14)

$$\Lambda^{(2)} = \Lambda_o^{(2)} = \frac{K_{11}(a) + K_{33}(a)}{2} + \sqrt{K_{13}(a)^2 + \left(\frac{K_{11}(a) - K_{33}(a)}{2}\right)^2}.$$

(E15)

We do not report explicit expressions for higher K_{nm} matrix elements, as expressions become more cumbersome. From Fig. 4 (in the main text), we see that this approximation quickly converges for $a < a^*$ and that $\Lambda^{(2)}$ is already very close to the value of Λ obtained by numerical diagonalization. For $a > a^*$, the convergence of this expansion is much slower and Λ is instead given by the maximum rejection probability, as explained in the main text.

The rapid convergence of the approximation Eq. (E8) with increasing N_s was already illustrated on Fig. 4 from the main text for the slowest relaxation mode Λ. Here we show that this approximation allows also to obtain accurate expressions for other sub-leading relaxation modes (see Fig. 10).

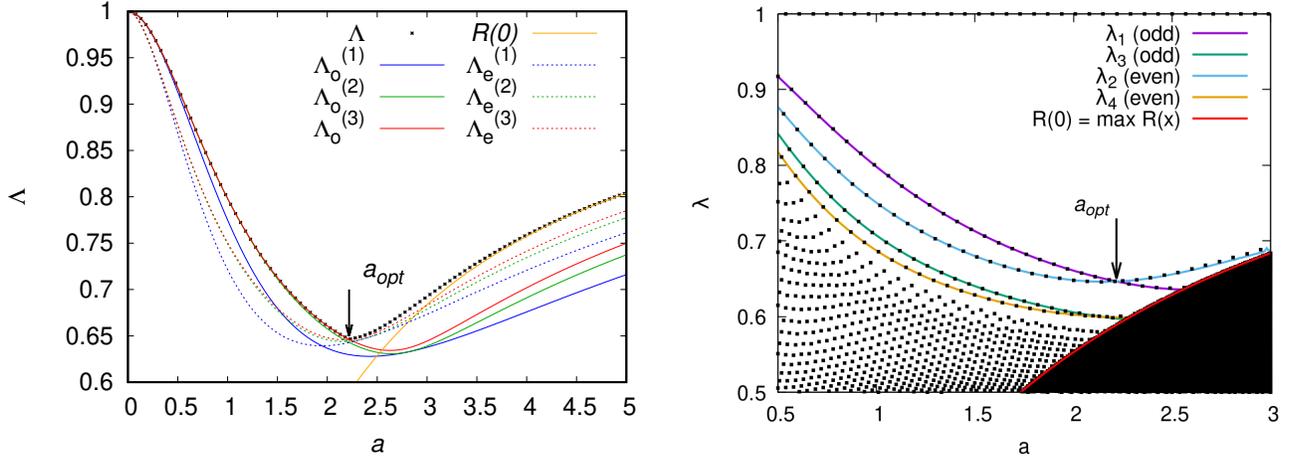## 2. Harmonic potential with a Gaussian jump distribution $w(\eta)$

FIG. 11. The left panel shows analytic approximations for Λ for $U(x) = x^2/2$ with a Gaussian jump distribution $f(z) = (\sqrt{2\pi})^{-1}\exp\left(-z^2/2\right)$. In this case, the matrices $K_o^{(N_s)}$ provide a good approximation for Λ for $a < a_{opt} \simeq 2.21845$. At $a = a_{opt}$, the symmetry of the leading relaxation mode changes from odd to even and for $a > a_{opt}$, Λ is instead well approximated by $\Lambda_e^{(N_s)}$, becoming increasingly close to the maximum of the rejection probability as a increases further. We note, however, that there does not seem to be a localization transition in this model: the IPR of the eigenvector for Λ follows a N_d^{-1} scaling suggesting that for this case $\mathcal{N} = 1$. The right panel is similar to Fig. (10), confirming that $K_o^{(6)}$ and $K_e^{(6)}$ provide very good approximations for the leading eigenvalues up to the crossing with the maximum of the rejection probability, given by Eq. (E23). Using this method, we find position of the localization transition for slowest antisymmetric relaxation mode λ_2 at $a \simeq 2.55657$.

For a Gaussian jump distribution

$$w(\eta) = \frac{1}{a\sqrt{2\pi}}\exp\left(-\frac{\eta^2}{2a^2}\right) \quad \Leftrightarrow \quad f(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right),$$

(E16)

we find the matrix elements for the odd subspace of the Schrödinger eigenbasis:

$$K_{11}(a) = 1 - \frac{a^2}{2} + \frac{a^2}{\pi} \arctan\left(\frac{a}{2}\right) + \frac{2a^3}{\pi\left(a^2 + 4\right)} \tag{E17}$$

$$K_{13}(a) = \sqrt{\frac{3}{2}}\frac{a^4}{\pi} \arctan\left(\frac{a}{2}\right) + \frac{\left(12a^4 + 80a^2 - 3\pi\left(a^2 + 4\right)^2 a + 96\right)a^3}{2\sqrt{6}\pi\left(a^2 + 4\right)^2} \tag{E18}$$

$$K_{33}(a) = 1 - \frac{\left(5a^4 + 9a^2 + 6\right)a^2}{2\pi} \arctan\left(\frac{2}{a}\right) + \frac{\left(15a^8 + 187a^6 + 834a^4 + 1560a^2 + 1152\right)a^3}{3\pi\left(a^2 + 4\right)^3} \tag{E19}$$

For the Gaussian jumps, Λ also depends on the matrix elements in the even subspace:

$$K_{22}(a) = \frac{2\left(3a^3 + \pi\right) - a^2\left(3a^2 + 4\right)\arctan\left(\frac{2}{a}\right)}{2\pi} \tag{E20}$$

$$K_{24}(a) = \frac{a^3}{4\sqrt{3}\pi\left(a^2 + 4\right)}\left[30a^4 + 128a^2 - 3\left(a^2 + 4\right)\left(5a^2 + 8\right)a\arctan\left(\frac{2}{a}\right) + 64\right] \tag{E21}$$

$$K_{44}(a) = 1 - \frac{\left(35a^6 + 80a^4 + 72a^2 + 32\right)a^2 \arctan\left(\frac{2}{a}\right)}{8\pi} + \frac{\left(105a^8 + 940a^6 + 2712a^4 + 3072a^2 + 1920\right)a^3}{12\pi\left(a^2 + 4\right)^2} \tag{E22}$$

We also get:

$$R(0) = \max_x R(x) = 1 - \frac{1}{\sqrt{1 - a^2}} \tag{E23}$$

$$R_\infty = 1 - \frac{2}{\pi} \arctan\frac{2}{a} \tag{E24}$$

Figure 11 compares the result of the Schrödinger eigenbasis approximation for a Gaussian $w(\eta)$ to numerical eigenvalues for the discretized master equation. We do not find evidence of a localization transition for Λ, but instead a change of parity at $a = a_{opt} \simeq 2.21845$. As for the case of a flat jump distribution shown on Fig. 10, the Schrödinger eigenbasis approximation works very accurately for all the slowest relaxation modes until they cross the singular continuum. It seems that even if $\mathcal{N} = 1$ for this case, the maximum rejection probability $\max_x R(x)$ is still a very good approximation for Λ at large a ($a \geq 4$).
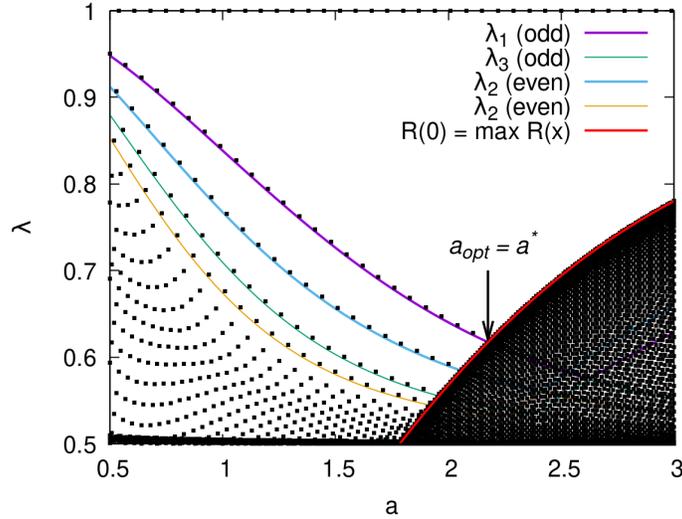

FIG. 12. Results for $U(x) = x^2/2$ with a jump distribution $w(\eta) = |\eta|\theta(a - |\eta|)/a^2$. The continuous curves show the results of the analytical calculation $K_o^{(6)}$ and $K_e^{(6)}$; they provide a very good approximation for the leading eigenvalues up to the crossing with the maximum of the rejection probability.

3. Harmonic potential with jump distribution $w(\eta) = a^{-2}|\eta|\theta(a - |\eta|)$

Again for a harmonic potential, analytical results for this shape of $w(\eta)$ can be obtained in the same way as above. We do not report them here, and only provide a comparison between numerical and analytical calculations on Fig. 12.

4. Comparing the different jump distributions

$w(\eta)$	a_{opt}	a^*	$\Lambda(a_{\mathrm{opt}})$	$1 - R_\infty(a_{\mathrm{opt}})$
$(2a)^{-1}\theta(a - |\eta|)$	3.32878	$a^* = a_{\mathrm{opt}}$	0.62382	0.45543
$(a\sqrt{2\pi})^{-1}\exp(-\eta^2/(2a^2))$	2.21845	none	0.64638	0.467
$a^{-2}|\eta|\theta(a - |\eta|)$	2.17613	$a^* = a_{\mathrm{opt}}$	0.61723	0.48193

TABLE I. Summary of the results on the optimal jump length a for several shapes of the jump distribution $w(\eta)$ in a harmonic potential. These three cases correspond to Figures 10, 11, 12

Among the three jump distributions worked out above, the last one provides the value $\Lambda(a_{\mathrm{opt}}) = 0.61723$, which is the lowest among the studied examples. In this respect, this jump distribution, at the optimal jump amplitude a_{opt}, yields the fastest method for sampling the equilibrium distribution. Results are summarized in Table I.
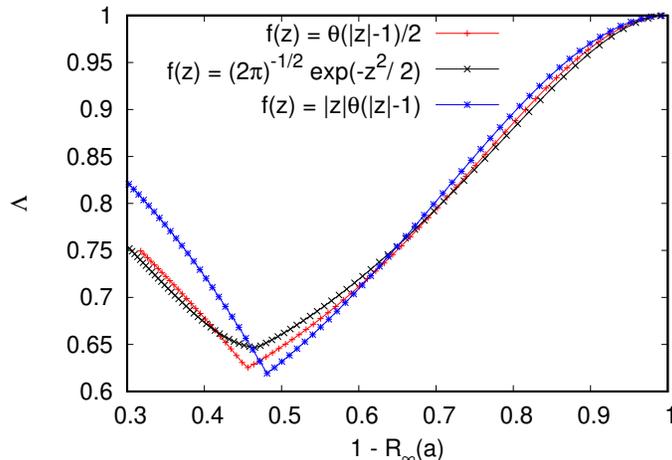

FIG. 13. Comparing different jump distributions. Plots of the leading relaxation eigenvalue Λ as a function of the acceptance probability, for the three cases summarized in Table I, corresponding to Figures 10, 11, 12.

Figure 13 compares the spectral results for the three jump distributions. We note that they correspond to a $w(\eta)$ that is either increasing, flat, or decreasing with $|\eta|$. In spite of these differences, the leading relaxation eigenvalue Λ displays the same behaviour as a function of the acceptance probability $1 - R_\infty$. In particular, the three cases feature optimality (smallest Λ, fastest convergence) for an acceptance probability close to 50%.

Appendix F: Generalization: beyond one dimension and inclusion of interactions

While the results presented so far focused on one-dimensional dynamics, we here put to the test the generality of the localization transition by considering more generic models, beyond 1D or with interacting degrees of freedom. The analysis is here mostly numerical.

1. Beyond 1D

Simulations in higher dimensions rapidly become demanding in terms of numerical resources. In two dimensions, it is still possible to use direct diagonalization to obtain the full eigenspectrum of the master equation and the IPR of the eigenvectors. An example of such a simulation is shown on Fig. 14: the results are very similar to the one dimensional simulation in Fig. 1 (main text) except that $a^* \simeq 2.6$ instead of $a^* \simeq 3.3$ due to the two dimensional nature of attempted jumps.
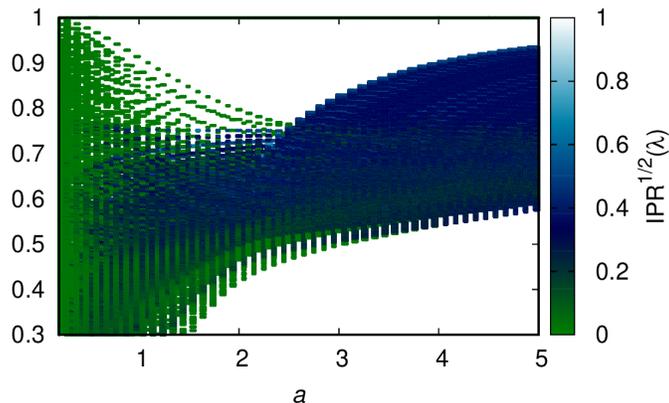

FIG. 14. Spectrum of the Monte-Carlo master equation kernel for a two dimensional particle confined to the cell $(-5, 5) \times (-5, 5)$ (discretized to 100^2 boxes) and in the potential $U(x, y) = (x^2 + y^2)/2$ ($\beta = 1$). The attempted jumps are two dimensional changing both x and y in an interval $(-a, a)$ centered around their initial values. Color shows $\text{IPR}^{1/2}$, where the square root is used to enhance contrast (the lower contrast in IPR values is related to the high symmetry of the potential $U(x, y)$, see for example the higher contrast in Fig. 16 where all symmetries are broken).

Simulations in 3D are numerically more accessible if jumps are attempted in only one of the directions x, y, z at a time. This makes the matrix representing the Master equation kernel sparse, allowing to find the time evolution of the error distribution $\delta P_n = P_n - P_\infty$. We show in Fig. 15 the evolution of the IPR of δP_n with the number of algorithm steps (time). A sharp transition from decreasing to increasing IPR as a function of time is seen around $a = 3.3$. Since the attempted jumps are 1D, the localization transition takes place at the same value as for the 1D harmonic potential.
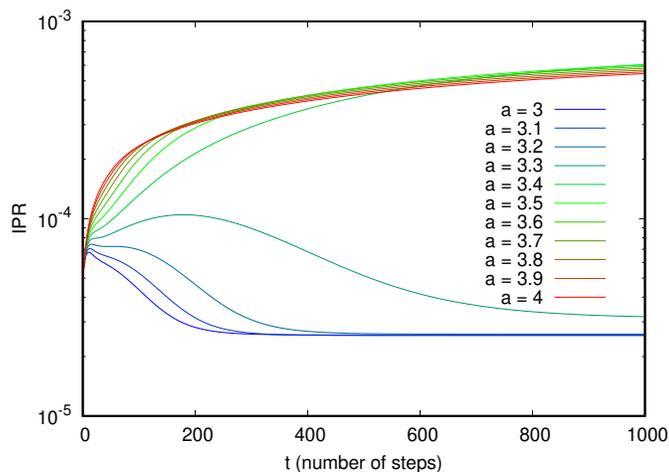

FIG. 15. A 3D example with the potential $U(x, y, z) = (x^2 + y^2 + z^2)/2$ and a confinement volume $(-5, 5)^3$ discretized in 100^3 boxes. The initial distribution $P_0(x, y, z)$ is an off-centred Gaussian.

2. Interactions

We provide a numerical example illustrating the localization transition in the Monte Carlo relaxation of interacting particles. We consider a case which is numerically tractable by full diagonalization, in analogy with Fig. 1 from the main text and with Fig. 7. We consider two particles at positions x_1 and x_2 in a one dimensional box, with $x_1, x_2 \in [-5, 5]$. The energy of a configuration (x_1, x_2) is given by the potential:

$$U_{\pm}(x_1, x_2) = \frac{x_1^2 + x_2^2}{2} \pm \frac{2}{0.1 + |x_1 - x_2|} + x_1 - x_2 \qquad (F1)$$

where, depending on the plus or minus signs, the interaction between x_1 and x_2 is repulsive (U_+) or attractive (U_-). We simulate the steady state of this system using a Monte-Carlo algorithm, with jumps where we attempt to simultaneously change x_1 and x_2 in an interval $(-a, a)$ around their initial position. The spectrum of the corresponding master equation is shown in Fig. 16, indicating that a localization transition occurs in this case even when interactions are present. Switching from repulsive to attractive interaction changes the value of the optimal jump length a^*, and the spread of the eigenspectrum. In both cases however, the IPR drastically increases for $a > a^*$, indicating a localization transition.
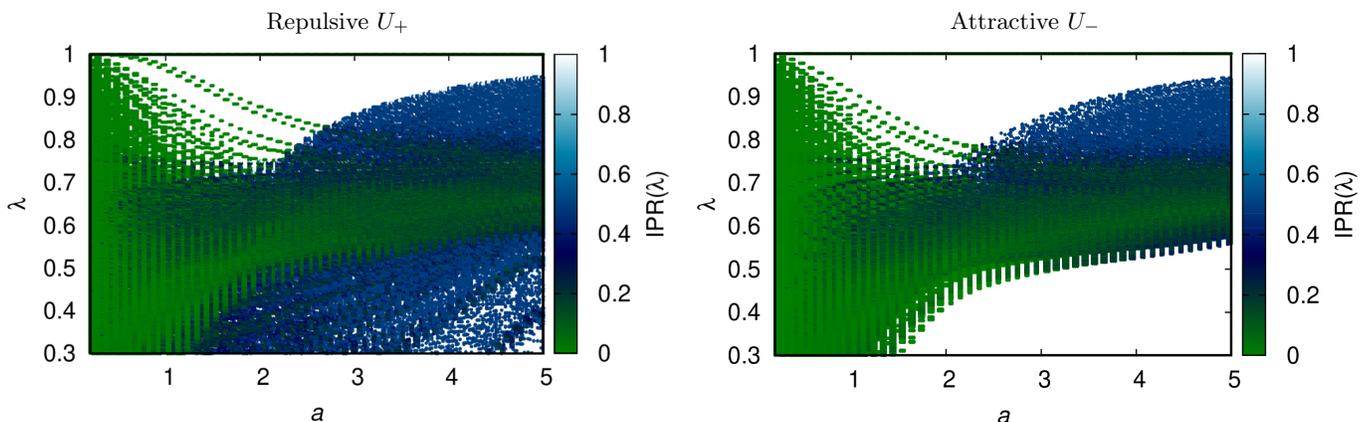

FIG. 16. Spectrum of the Monte-Carlo master equation kernel for two interacting particles, with interaction potential given by Eq. (F1). Two situations were investigated, with a repulsive (left panel) or an attractive potential (right panel). The configuration space, restricted to the interval $(-5, 5)$, was discretized in 100×100 cells.

3. Relaxation in presence of multiple local minima

Finally, we illustrate numerically the relaxation spectrum for a Monte Carlo simulation in a 1D potential with many local minima. We take the potential:

$$U(x) = x^2/2 + 3 \sin 9x \qquad (F2)$$

inside a box $x \in (-5, 5)$. This potential has many local minima as illustrated in the left panel of Fig. 17. The eigenspectrum (see Fig. 17 right panel) features a localization transition at $a^* \simeq 2.1$ as in the prototype cases with only a single minium. At variance with the spectrum for $U(x) = x^2/2$ (see Fig. 1 from the main text), many quasi-degenerate eigenvalues are present near $\lambda = 1$, for low values of the jump amplitude a. In this regime indeed, hopping over the barrier is thermally activated and the mimima become metastable.

---

[1] When a needle of length ℓ is randomly tossed over a floor with parallel equidistant strips of size 2ℓ, the probability that the needle crosses a line between two strips is $1/\pi$.

[2] N. Metropolis and S. Ulam, The Monte Carlo method, J. Am. Stat. Assoc. **44**, 335 (1949).

[3] R. Eckhardt, Stan Ulam, John von Neumann, and the Monte Carlo method, Los Alamos Science **15**, 30 (1987).
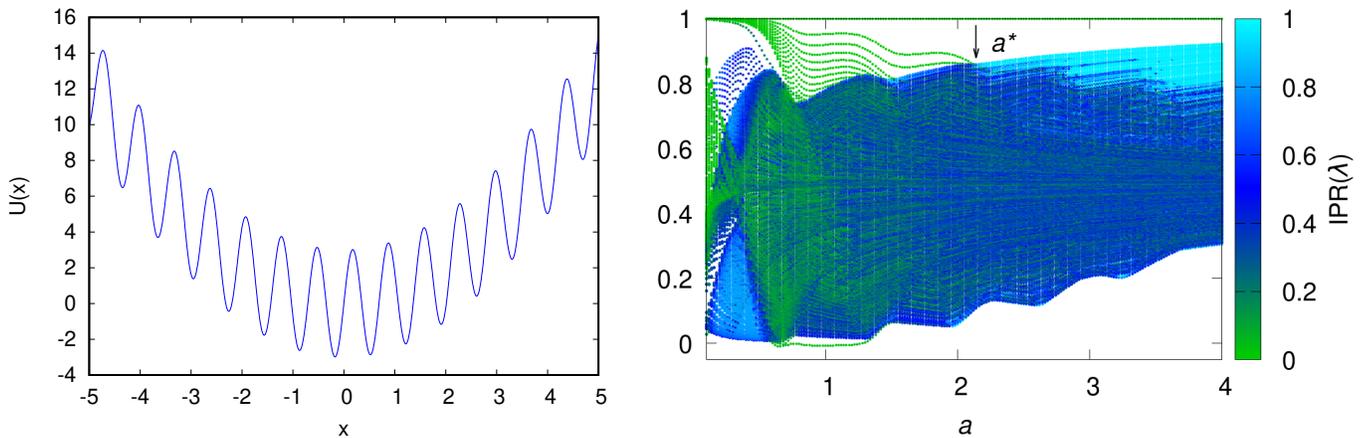
FIG. 17. (left panel) Example of a potential with many local minima given by Eq. (F2). (right panel) Monte Carlo relaxation eigenspectrum for this potential with a flat jump distribution as in Fig. 1 from the main text. The box $(-5, 5)$ was discretized in 10^3 sites.

[4] A. B. Owen, *Monte Carlo Theory, Methods and Examples* (2013).

[5] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, 2021).

[6] C. J. Mode, *Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science* (IntechOpen, 2011).

[7] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark, Annu. Rev. Phys. Chem. **53**, 291 (2002).

[8] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017).

[9] C. M. Bishop, *Pattern recognition and machine learning* (Springer New York, 2006).

[10] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger, Computational models for active matter, Nat. Rev. Phys. **2**, 181 (2020).

[11] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method* (Wiley, 2016, 1996).

[12] P. Glasserman, *Monte Carlo Methods in Financial Engineering* (Springer, New York, 2003).

[13] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall/CRC, 1996).

[14] G. O. Roberts, A. Gelman, and W. R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. **7**, 110 (1997).

[15] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, 1999).

[16] D. Frenkel and B. Smith, *Understanding Molecular Simulations*, 2nd ed. (Adademic Press, 2002).

[17] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford Master Series in Physics, 2006).

[18] L. Wasserman, *All of Statistics*, 2nd ed. (Springer-Verlag, 2004).

[19] M. Bédard, Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234, Stochastic Processes and their Applications **118**, 2198 (2008).

[20] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, Phys. Lett. B **195**, 216 (1987).

[21] M. Creutz and A. Gocksch, Higher-order hybrid Monte Carlo algorithms, Phys. Rev. Lett. **63**, 9 (1989).

[22] U. Wolff, Collective Monte Carlo updating for spin systems, Phys. Rev. Lett. **62**, 361 (1989).

[23] F. L. J. S. Liu and W. H. Wong, The multiple-try method and local optimization in Metropolis sampling, J. Am. Stat. Assoc. **95**, 121 (2000).

[24] G. O. Roberts and J. S. Rosenthal, Optimal scaling for various Metropolis-Hastings algorithms, Stat. Sci. **16**, 351 (2001).

[25] E. P. Bernard, W. Krauth, and D. B. Wilson, Event-chain Monte Carlo algorithms for hard-sphere systems, Phys. Rev. E **80**, 056704 (2009).

[26] H.-P. Hsu and P. Grassberger, A review of Monte Carlo simulations of polymers with perm, J. Stat. Phys. **144**, 597 (2011).

[27] M. Michel, S. C. Kapfer, and W. Krauth, Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps, J. Chem. Phys. **140**, 054116 (2014).

[28] S. C. Kapfer and W. Krauth, Irreversible local Markov chains with rapid convergence towards equilibrium, Phys. Rev. Lett. **119**, 240603 (2017).

[29] D. Oliinychenko and V. Koch, Microcanonical particlization with local conservation laws, Phys. Rev. Lett. **123**, 182302 (2019).

[30] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Equivariant flow-based sampling for lattice gauge theory, Phys. Rev. Lett. **125**, 121601 (2020).

[31] A. Gelman, G. O. Roberts, and W. R. Gilks, Efficient Metrolopis jumping rules, in *Bayesian Statistics*, edited by J. Bernardo, J. Berger, A. Dawid, and F. Smith (Oxford University Press, Oxford, 1996) pp. 599–607.

[32] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed. (Oxford University Press, 2017).

[33] K. K. Dey and S. Bhattacharya, A brief review of optimal scaling of the main MCMC approaches and optimal scaling of additive TMCMC under non-regular cases, Brazilian Journal of Probability and Statistics **33**, 222 (2019).

[34] B. Jourdain, T. Lelièvre, and B. Miasojedow, Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit, The Annals of Applied Probability **25**, 2263 (2015).

[35] P. Peskun, Optimum Monte-Carlo sampling using Markov chains, Biometrika **60**, 607 (1973).

[36] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed. (American Mathematical Society, 2017).

[37] P. P. P. S. Randal Douc, Eric Moulines, *Markov Chains* (Springer, 2018).

[38] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[39] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika **57**, 97 (1970).

[40] S. D. Hill and J. C. Spall, Stationarity and convergence of the Metropolis-Hastings algorithm: Insights into theoretical aspects, IEEE Control Syst. **39**, 56 (2019).

[41] R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications* (Cambridge University Press, 1997).

[42] H. Risken, *The Fokker-Planck equation*, 3rd ed. (Springer, 2007).

[43] F. Wegner, Inverse participation ratios in $2 + \epsilon$ dimensions, Z. Physik B **36**, 209 (1980).

[44] A. D. Chepelianskii, S. N. Majumdar, H. Schawe, and E. Trizac, One-dimensional monte carlo dynamics at zero temperature, Journal of Physics A: Mathematical and Theoretical **54**, 485001 (2021).

[45] E. Abrahams, *50 Years of Anderson Localization* (World Scientific, 2010).

[46] G. O. Roberts and J. S. Rosenthal, General state space Markov chains and MCMC algorithms., Probability Surveys [electronic only] **1**, 20 (2004).

[47] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, Vol. 2 (Addison-Wesley, 1998).

[48] P. Flajolet and R. Sedgewick, *Analytic Combinatorics* (Cambridge University Press, 2009).