Polynomially Filtered Exact Diagonalization Approach to Many-Body Localization

Piotr Sierant,1,2,* Maciej Lewenstein,2,3,1 and Jakub Zakrzewski1,4,†

1Institute of Theoretical Physics, Jagiellonian University in Kraków, Łojasiewicza 11, 30-348 Kraków, Poland
2ICFO - Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain
3ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain
4Mark Kac Complex Systems Research Center, Jagiellonian University in Kraków, Łojasiewicza 11, 30-348 Kraków, Poland

(Received 27 May 2020; accepted 13 September 2020; published 9 October 2020)

Polynomially filtered exact diagonalization method (POLFED) for large sparse matrices is introduced. The algorithm finds an optimal basis of a subspace spanned by eigenvectors with eigenvalues close to a specified energy target by a spectral transformation using a high order polynomial of the matrix. The memory requirements scale better with system size than in the state-of-the-art shift-invert approach. The potential of POLFED is demonstrated examining many-body localization transition in 1D interacting quantum spin-1/2 chains. We investigate the disorder strength and system size scaling of Thouless time. System size dependence of bipartite entanglement entropy and of the gap ratio highlights the importance of finite-size effects. We discuss possible scenarios regarding the many-body localization transition obtaining estimates for the critical disorder strength.

DOI: 10.1103/PhysRevLett.125.156601

Introduction.—Quantum many-body systems are generically expected to approach equilibrium according to eigenstate thermalization hypothesis [1–3]. The phenomenon of many-body localization (MBL) [4–6] provides a robust class of many-body systems which fail to reach thermal equilibrium [7–16]. Further examples of nonergodic behavior include Stark localization [17,18], persistent oscillations [19–23], the presence of confinement [24,25], Hilbert space fragmentation [26–28], or lack of thermalization in lattice gauge theories [29–34].

Classification of many-body systems according to their ergodic properties is a fascinating new direction of research, however, it poses serious technical challenges as exact methods are restricted either to small system sizes [35] or allow to trace time evolution only within a short time interval [36–38]. Hence, a fully consistent theory of MBL transition is missing, with recent approaches pointing toward Kosterlitz-Thouless scaling [39–43]. The finite-size effects strongly influence exact diagonalization (ED) results, leading to a recent debate [44–47] about discriminating between finite-size effects and asymptotic features of disordered many-body systems.

The example of MBL transition shows that development of ED techniques allowing the study of thermalization properties of possibly large many-body systems is in demand. In this Letter, we introduce a polynomially filtered exact diagonalization (POLFED) as a tool to calculate eigenvectors of large sparse matrices with eigenvalues close to a specified energy target. The polynomial spectral transformation preserves the sparse structure of matrices avoiding the main bottleneck of shift-invert method of exact diagonalization (SIMED) [35]. We employ POLFED in the study of MBL transition in disordered quantum spin chains unveiling new aspects of system size scaling of Thouless time, entanglement entropy, and level statistics. Our results provide novel qualitative and quantitative arguments in favor of the existence of MBL transition in the thermodynamic limit.

Benchmark models.—We consider 1D disordered spin chains with Hamiltonian:

$$\hat{H} = \sum_{i=1}^{L} J_{i} (\tilde{S}_{i}^{x} \tilde{S}_{i+1}^{x} + \tilde{S}_{i}^{y} \tilde{S}_{i+1}^{y} + \Delta \tilde{S}_{i}^{z} \tilde{S}_{i+1}^{z}) + \sum_{i=1}^{L} h_{i} \tilde{S}_{i}^{z},$$ (1)

where $\tilde{S}_{i}$ are spin-1/2 matrices, $L$ is the system size, $J_{1} = 1$ is fixed as the energy unit, periodic boundary conditions are assumed, and $h_{i} \in [-W,W]$ are independent, uniformly distributed random variables. The XXZ model, widely studied in the MBL context [48–53], is obtained for $J_{2} = 0$ and $\Delta = 1$. The choice $J_{2} = 1$ and $\Delta = 0.55$ leads to the $J_{1}$-$J_{2}$ model studied in [44]. The Hamiltonian (1) becomes a real symmetric sparse matrix $H \in \mathbb{R}^{N \times N}$ in basis of eigenvectors of $\tilde{S}_{i}^{z}$ operator; the matrix size $N$ in the zero magnetization $\sum_{i} \tilde{S}_{i}^{z} = 0$ sector is given by $N \sim (\frac{L}{L_{T/2}}) \propto e^{L \ln 2}/\sqrt{L}$.

Calculation of eigenpairs.—Hamiltonians of many-body systems are typically characterized by exponential scaling of matrix size $N$, with the system size $L$ and sparsity in appropriately chosen basis. For a sparse matrix the number
of nonzero entries, \( N_{nz} \) is much smaller than \( N^2 \) implying that matrix vector multiplication requires much less operations than for a dense matrix. The Lanczos algorithm [54] utilizes this fact to find exterior eigenpairs (corresponding to highest and lowest eigenvalues). However, due to an increasing density of states and reorthogonalization costs, Lanczos algorithm becomes inefficient if many eigenpairs are requested. In contrast, a full ED procedure [55] allows one to determine all eigenpairs of \( H \) but, with present day computers, it is limited to \( N \approx 5 \times 10^3 \) corresponding to \( L = 18 \) in (1). Larger matrix sizes are tractable by SIMED [35]. The Hamiltonian is transformed via \( H \to R_\sigma(H) = (\sigma - H)^{-1} \) so that eigenvalues close to \( \sigma \) become exterior eigenvalues of the matrix \( R_\sigma(H) \), see Fig. 1. Consequently, the Lanczos algorithm for the matrix \( R_\sigma(H) \) converges to eigenpairs close to the target \( \sigma \). The Lanczos iteration with \( R_\sigma(H) \) is performed by calculating \( \sigma \)-tridiagonal matrices [56,57] of the matrix \( H \). That has a significant drawback: the sparsity pattern of \( H \) is lost resulting in a very severe for large \( N \) phenomenon of fill-in of the matrix. This was identified as the main bottleneck of SIMED when applied to quantum many-body systems [35].

**POLFED algorithm.**—To avoid the fill-in phenomenon and utilize the sparsity of the \( H \) matrix in an efficient way, we use the polynomial spectral transformation

\[
H \to P_\sigma^K(H) = \frac{1}{D} \sum_{n=0}^K c_n T_n(H),
\]

where \( T_n(x) \) denotes the \( n \)th Chebyshev polynomial, the coefficients \( c_n = \sqrt{4 - 3\delta_{0,n}} \cos(n \arccos(\sigma)) \) are obtained from expanding a Dirac delta function centered at \( \sigma \) in Chebyshev polynomials and normalization \( D \) assures that \( P_\sigma^K(\sigma) = 1 \). The eigenvalues close to the target energy \( \sigma \) are the largest eigenvalues of the transformed matrix \( P_\sigma^K(H) \) as shown in Fig. 1(b). Hence, a block Lanczos method [58,59] applied to matrix \( P_\sigma^K(H) \) converges to eigenpairs close to the target \( \sigma \). We note that eigensolvers employing polynomial spectral transformations were considered also in [60–63].

The POLFED consists of the following steps. Lanczos algorithm is used to find the lowest (highest) eigenvalue \( E_0 (E_1) \) of matrix \( H \) which is then rescaled to \( \tilde{H} = [2H - (E_0 + E_1)]/(E_1 - E_0) \). The order \( K \) of transformation (2) is specified by requiring that the number of eigenvalues \( \theta_i \) of \( P_\sigma^K(\tilde{H}) \) accessible to Lanczos algorithm (belonging to the shaded area in Fig. 1) is equal to a number of requested eigenvalues \( N_{ev} \).—as the condition we take \( \theta_i \geq p = 0.17 \). To find the value of \( K \), an estimate of density of states \( \tilde{\rho}(\sigma) \) at energy \( \sigma \) of the matrix \( \tilde{H} \) is needed. The \( \tilde{\rho}(\sigma) \) can be found efficiently for arbitrary sparse matrices using iterative methods [64,65]. For the benchmark models (1), the density of states is Gaussian and is well approximated by an analytic expression \( \tilde{\rho}(0) = (E_1 - E_0)N/\Gamma \) at the center of spectrum \( \sigma = 0 \) where \( \Gamma \propto \sqrt{LW} \). Having found \( K \), the POLFED algorithm, starting with a matrix of orthonormalized random vectors \( \mathcal{Q}_1 \in \mathbb{R}^{N \times s} \), performs the block Lanczos iteration

\[
U_j = P_\sigma^K(\tilde{H}) \mathcal{Q}_j - Q_{j-1}B_j^T, \quad A_j = Q_j^T U_j, \quad j = 1, \ldots, m.
\]

where \( Q_0 = 0, B_0 = 0 \) and the second operation in (4) is QR decomposition. The iteration is repeated for \( j = 1, \ldots, m \) resulting in \( Q_j, U_j, R_j \in \mathbb{R}^{N \times s} \) and \( A_j, B_j \in \mathbb{R}^{s \times s} \) matrices. In exact arithmetic, columns of \( Q_j \) matrices form an orthonormal set of vectors. This property is gradually lost with increasing \( m \) during calculations with a finite precision. Hence, between (3) and (4), we perform a reorthogonalization of columns of matrix \( U_j \) against the columns of matrices \( \{Q_i\}_{i=1}^j \). The product of \( P_\sigma^K(\tilde{H}) \) with each column of \( Q_j \) in (3) is computed with the Clessenaw algorithm [66]. The orthogonal matrix \( \mathcal{Q}_m = [\mathcal{Q}_1, \ldots, \mathcal{Q}_m] \in \mathbb{R}^{N \times ms} \) defines a block tridiagonal matrix \( T_m = \mathcal{Q}_m^T P_\sigma^K(\tilde{H}) \mathcal{Q}_m \) with \( A_j \) matrices on the diagonal and \( B_j \) below (above) the diagonal. The eigenvectors \( t_i \in \mathbb{R}^{ms} \) of \( T_m \) are used to calculate \( u_i = \mathcal{Q}_m t_i \) which converge, with increasing \( m \), to exterior eigenvectors of \( P_\sigma^K(\tilde{H}) \) [67], that is to eigenvectors of \( \tilde{H} \) with eigenvalues close to the target \( \sigma \). The convergence is reached after \( m \) steps when the residual norm \( \|B_{j+1} t_i\| \) [59] (where \( t_i \) are the last \( s \) components of the vector \( t_i \) and \( \|u_i\| = \sqrt{u_i^T u_i} \) vanishes within the numerical precision for each eigenvector \( t_i \) corresponding to eigenvalue \( \theta_i \geq p \). The eigenvalues of the matrix \( \tilde{H} \) are found as \( \theta_i = u_i^T \tilde{H} u_i \) and the convergence is verified by a direct calculation of the residual norms \( \|\tilde{H} u_i - \theta_i u_i\| \). Each of our tests shows that eigenvalues \( \theta_i \) are, within numerical precision, equal to \( N_{ev} \) eigenvalues of \( \tilde{H} \) closest to the target \( \sigma \). For technical details of the algorithm see [68].
The POLFED is tailored for maximal efficiency in calculations for quantum many-body systems. The order $K$ of the polynomial transformation (2) scales linearly with the density of states $\tilde{\rho}(\sigma)$ that increases exponentially with system size $L$. Thus, the product $P^{k}_{\sigma}(\tilde{H})Q_{j}$ in (3) is the most time consuming step of the calculation. POLFED offers high scalability as the product can be parallelized in two manners: (i) it splits into independent multiplications of subsequent columns of $Q_{j}$ by $P^{k}_{\sigma}(\tilde{H})$; (ii) each of the matrix vector multiplications can be parallelized. The reorthogonalization step between (3) and (4) can be parallelized in a similar manner. The number $m$ of iterations after which the algorithm converges is proportional to $N_{ev}$. Hence, the memory consumption, dominated by $Q_{m}$ scales as $N_{ev}N$. The memory requirements of SIMED are larger and scale as $c(L)N^{2}$ where the factor $c(L)$ is due to the fill-in of the matrix. For XXZ model $c(L) \propto 3^{L/2}$ [35]. Moreover, $c(L)$ grows rapidly with number $N_{nc}$ of nonzero elements of the matrix significantly increasing the resources needed in calculations for $J_{1}-J_{2}$ model. In contrast, computation time of POLFED increases linearly with $N_{nc}$.—resources for XXZ and $J_{1}-J_{2}$ models are comparable; for detailed benchmarks see [68]. POLFED allows to find larger number of eigenpairs in a single run than the recently proposed eigen solver [69]. This reduces fluctuations of averages over eigenstates and is essential in calculation of the Thouless time.

**Thouless time.**—The spectral form factor is defined as $K(\tau) = \langle |\sum_{j=1}^{N} g(E_{j}) e^{-iE_{j}\tau}|^{2} \rangle / Z$, where $E_{j}$ are eigenvalues of $H$ after an unfolding procedure [70], $g(\epsilon)$ is a Gaussian function, the average is taken over disorder realizations and $Z$ is a normalization constant ensuring $K(\tau) \rightarrow 1$. The spectral form factor of the many-body system (with time reversal invariance) follows Gaussian orthogonal ensemble (GOE) prediction $K(\tau) = K_{\text{GOE}}(\tau)$ only for $\tau > t_{Th}$ defining the Thouless time $t_{Th} = t_{nf}t_{H}$, where $t_{nf} = 2\pi/\langle 0 \rangle$ is the Heisenberg time.

The Thouless time $t_{Th}$ calculated for $J_{1}-J_{2}$ spin chains of length $L \leq 18$ [44] scales as $t_{Th} \propto L^{2}W^{1/3}$ where $W$ is the disorder strength and $\Omega$ is constant. If this scaling prevailed in the $L \rightarrow \infty$ limit, it would imply $t_{Th}/t_{H} \rightarrow 0$ so that the system would be well described by GOE and the MBL phase would be absent for arbitrary disorder strength in the thermodynamic limit. To verify this surprising conclusion we supplement results of full ED of the $J_{1}-J_{2}$ model with Thouless times obtained with POLFED for $L = 20, 22, 24$, respectively, for 800, 200, 50 disorder realizations. Since we calculate $N_{ev} = 2500$ eigenvalues in the middle of spectrum (\sigma = 0), the sum in the definition of spectral form factor $K(\tau)$ is truncated. However, this does not influence the value of $t_{Th}$ as long as it is larger than a certain threshold value determined by $N_{ev}$ [68]. The obtained Thouless times are shown in Fig. 2. Data for $L \leq 20$ follow the scaling $t_{Th} \propto L^{2}W^{1/3}$ deviating from it at disorder strength $W(L)$ which increases with the system size, for instance $\tilde{W}(18) \approx 3.7$ or $\tilde{W}(20) \approx 4.6$. This behavior changes qualitatively for $L = 22, 24$ data breaking the scaling $t_{Th} \propto L^{2}W^{1/3}$. Similar behavior heralds Anderson localization transition in single particle disordered systems [45], hence, our data suggest the presence of the transition to MBL phase in the $J_{1}-J_{2}$ model. Therefore, one has to reach a sufficiently large $L$ to see the correct scaling of Thouless time, which raises the question about the finite-size effects at MBL transition.

**Entanglement entropy and level statistics.**—The entanglement entropy allows for insights in nature of MBL transition [71–73]. The entanglement entropy of an eigenstate is defined as $S_{E} = -\sum_{i} \alpha_{i}^{2} \log(\alpha_{i}^{2})$, where $\alpha_{i}$ are Schmidt basis coefficients (see, e.g., [74]) associated with the bipartition of the lattice into subsystems containing sites $\{x, x + L/2\}$ and $\{x + L/2, x + L\}$ (the sites are numbered modulo $L$). We average $S_{E}$ over the position of the cut $x$, over $N_{ev} \leq \min\{N/100, 2000\}$ eigenstates in the middle of the spectrum (\sigma = 0) of the $J_{1}-J_{2}$ model for system sizes $12 \leq L \leq 24$ (for $L = 8, 10$ we take $N_{ev} = 5$) as well as over more than 5000, 200, 50 disorder realizations, respectively, for $L \leq 20$, $L = 22, 24$. Finally, we obtain the scaled entanglement entropy $s_{E} = S_{E}/S_{\text{RMT}}(L)$ where $S_{\text{RMT}}(L) = (L/2)\ln(2) + [1/2 + \ln(1/2)]/2 - 1/2$ corresponds to a chaotic spin chain in the total $\sum_{i} S_{i}^{z} = 0$ sector [75]. The resulting $s_{E}$ is shown in Fig. 3(a). For available system sizes, the scaled entanglement entropy $s_{E}$: (i) monotonically increases with $L$ for $W \lesssim 3.4$; (ii) monotonically decreases for $W \gtrsim 11$; (iii) decreases for smaller $L$ and starts increasing for larger system sizes (a similar reentrant behavior was observed, e.g., in [47,76]). The behavior (i) clearly leads to an ergodic system at large $L$. In contrast, for large disorder strengths, e.g., $W = 15$, an area law of entanglement entropy [77,78] $s_{E} \propto 1/L$ arises due to the emergent integrability of MBL phase [79–85]. Averaging $r_{i} = \min\{g_{i}, g_{i+1}\} / \max\{g_{i}, g_{i+1}\}$ (where $g_{i} = E_{i+1} - E_{i}$) over eigenvalues corresponding to eigenstates from which $s_{E}$ was calculated, we obtain a mean gap ratio $\bar{r}$ shown in Fig. 3(b). The mean gap ratio $\bar{r}$ probes level statistics of the system, admitting values characteristic for
GOE and Poisson statistics for ergodic and localized systems [7,86]. Similarly as for $s_E$, the mean gap ratio $\bar{r}$ follows the three types of behavior with system size depending on disorder strength $W$.

To understand whether and at which disorder strength the MBL transition takes place one has to study the interplay between the (ii) and (iii) trends. To this end, we find the disorder strength $W^*_E(L)$ such that $s_E(L - 1) = s_E(L + 1)$ for odd $L$ and $s_E(L - 2) = s_E(L + 2)$ for even $L$, for details see [68]. Smooth changes of $s_E$ with $L$ and $W$ assure that $W^*_E(L)$ is the largest disorder strength, for a given system size $L$, at which the volume-law $S_E \propto L$, expected for an ergodic system, is still obeyed. Consequently, the disorder strength $W^*_E(L)$ is a lower bound for the critical disorder strength $W_C$ of the transition to MBL phase. Figure 3(c) shows the relation between $W^*_E$ and 1/L along with disorder strength $W^*_J$ obtained in analogous manner for the average gap ratio $\bar{r}$. Another aspect of finite-size effects at MBL transition is revealed when, for given $L$, one finds a disorder strength $W_T(L)$ for which the scaled entanglement entropy $s_E$ is close to the ergodic limit, e.g., $s_E(W_T) = 0.8$. Such a criterion yields $W_T \propto L$. Equivalently, $W_T$ can be found as a disorder strength for which the average gap ratio $\bar{r}$ departs from the GOE limit [44]. This allows us to identify the following regimes: (A) thermal, for $W < W_T$, with entanglement entropy fulfilling the volume-law and close to the value for chaotic spin chain $S_E \approx S_RMT(L)$ and level statistics well described by GOE; (B) critical, for $W < W^*_E < W < W_T$, with $S_E < S_{RMT}(L)$ but scaling super linearly with $L$ and value of $\bar{r}$ increasing with $L$ toward the GOE limit; (C) MBL, for $W < W^*_E$, with both scaled entanglement entropy $s_E(L)$ and average gap ratio $\bar{r}$ decreasing with system size $L$. Figure 3(d) shows that behavior of the XXZ model is similar (data for $s_E$ and $\bar{r}$ can be found in [68]). The three regimes resemble the qualitative picture of MBL transition proposed in [73].

The asymptotic features of disordered spin chains depend on how $W^*_E$ and $W_T$ behave in the thermodynamic limit. For available system sizes, $8 \leq L \leq 24$, the linear scaling of $W^*_E$ with $L$ as well as the linear scaling of $W^*_E$ with inverse of system size $1/L$, denoted by solid lines in Figs. 3(c) and 3(d), are accurately obeyed. Extrapolating to $L \rightarrow \infty$ limit, we get estimates of critical disorder strength $W^*_E \approx 3.7$ and $W^*_C \approx 5.4$. (5) respectively, for $J_1-J_2$ and XXZ models. Our estimate for $W^*_C^{XXZ}$ is larger than the value $W_C \approx 3.7$ for the XXZ model [87] (which yields the critical exponent violating the Harris criterion [88–90]) or the estimates obtained after an asymmetric scalings on both sides of the transition: $W_C \approx 3.8$ [91], $W_C \approx 4.2$ [42]. Since our approach relies on analysis of the drift of crossing points of $s_E(W)$ and $\bar{r}(W)$ curves, it does not rely on any finite-size scaling procedure. Our estimate for $W^*_C^{XXZ}$ is consistent with the lower bound $W_C > 4.5$ of [92] as well as with $W_C > 5$ obtained in analysis of quench dynamics of large XXZ spin chain [37].

Conclusions.—The POLFED algorithm, thanks to the employed polynomial spectral transformation, has a better scaling of computation time with matrix size $N$ than the state-of-the-art SIMED algorithm. Avoiding the fill-in phenomenon, POLFED has a significantly lower memory consumption than SIMED, moreover, its performance...
decreases only linearly with increasing the number of nonzero off-diagonal matrix entries. For those reasons POLFED opens new pathways in studies of highly excited states of many-body systems with potential applications to systems with long-range interactions realized in experiments with polar molecules [93], Rydberg atoms [94], trapped ions [95–97] and problems of MBL or information spreading in the presence of power-law interactions [98–117]. Understanding the relation of POLFED to alternative eigensolvers [69,118,119] is an interesting task for a further research.

POLFED allowed us to study MBL transition in \( J_1-J_2 \) model of size \( L \leq 24 \). Such a system size is sufficient to demonstrate the breakdown of the scaling \( t_{\text{Th}} \propto L^2 e^{W/\Omega} \) of Thouless time [44]. Studying the system size scaling of entanglement entropy \( S_E \) of eigenstates we estimated the critical disorder strength of transition to MBL phase.

We thank Fabien Alet and Dominique Delande for insightful discussions. The computations have been performed within PL-Grid Infrastructure, its support is acknowledged. M. L. acknowledges support from ERC AdG NOQIA, Spanish Ministry of Economy and Competitiveness (“Severo Ochoa” program for Centres of Excellence in R&D (CEX2019-000910-S), Plan National FISICATEAMO and FIDEUA PID2019-106901GB-I00/10.13039/501100011033, FPI), Fundació Privada Cellnex, Fundació Mir-Puig, and from Generalitat de Catalunya (AGAUR Grant No. 2017 SGR 1341, CERCA program, QuantumCAT_U16-011424, co-funded by ERDF Operational Program of Catalonia), MINECO-EU QUANTERA MAQS (funded by State Research Agency (AEI) PCI2019-111828-2/10.13039/ 501100011033), EU Horizon 2020 FET-OPEN OPTOLogic (Grant No. 899794), and the National Science Centre, Poland-Symphonia Grant No. 2016/20/W/ST4/00314. The support of National Science Centre, Poland under Unisono Grant No. 2017/25/Z/ST2/03029 (Quantera: QTFLAG) is also acknowledged (J. Z.). P. S. acknowledges National Science Centre, Poland: ETIUDA Grant No. 2018/28/T/ST2/00401. P. S. acknowledges the support by Foundation for Polish Science.

*piotr.sierant@uj.edu.pl
†maciej.lewenstein@icfo.eu
‡jakub.zakrzewski@uj.edu.pl