Editors' Suggestion

Onsager's regression hypothesis adjusted to quantum systems

Peter Reimann and Christian Eidecker-Dunkel
Faculty of Physics, Bielefeld University, 33615 Bielefeld, Germany

(Received 17 April 2024; revised 28 June 2024; accepted 28 June 2024; published 15 July 2024)

Onsager's regression hypothesis connects the temporal relaxation of close-to-equilibrium systems with their dynamical correlation functions at thermal equilibrium. While the hypothesis is provably correct in classical systems, it is known to fail in the quantum regime. Here, we derive a suitably adjusted quantum version of Onsager's original hypothesis. Rigorous analytical results are complemented by a variety of numerical examples.

DOI: 10.1103/PhysRevB.110.014306

I. INTRODUCTION AND MAIN RESULTS

Onsager originally postulated and utilized the so-called regression hypothesis to establish his celebrated reciprocal relations between the kinetic coefficients of irreversible processes close to thermal equilibrium [1]. Qualitatively, the basic physical content of the hypothesis is that the temporal correlations of a system at thermal equilibrium are sufficient to describe how a slightly perturbed system returns to equilibrium. The quantitative formulation of the hypothesis will be provided later [see Eq. (15) below]. For classical systems, Onsager's hypothesis has been rigorously deduced from a microscopic description, for instance, in Ref. [2]. In the realm of quantum mechanics, the hypothesis is known to fail, as pointed out, for example, in Refs. [3–5] (see also Sec. V below). The main objective of our present paper is to deduce a properly modified quantum version of Onsager's regression hypothesis from basic microscopic considerations [see Eq. (18) below].

Concretely, let us consider a quantum system with Hamiltonian H at thermal equilibrium, described by a canonical ensemble

$$\rho := Z^{-1}\, e^{-\beta H}, \tag{1}$$

$$Z := \mathrm{Tr}\{e^{-\beta H}\}, \tag{2}$$

$$\beta := 1/k_{\mathrm{B}}T, \tag{3}$$

where T is the system's temperature and k_{B} Boltzmann's constant. The corresponding thermal equilibrium expectation value of an observable (Hermitian operator) A is denoted as

$$A_{\mathrm{th}} := \mathrm{Tr}\{\rho A\}. \tag{4}$$

Similarly, the temporal correlation (also called, among others, dynamic or two-point correlation function) of two Hermitian operators V and A at thermal equilibrium is given by

$$C_{VA}(t) := \mathrm{Tr}\{\rho V A(t)\} - V_{\mathrm{th}}A_{\mathrm{th}}, \tag{5}$$

where $A(t)$ is the observable A at time t in the Heisenberg picture:

$$A(t) := e^{iHt/\hbar}Ae^{-iHt/\hbar}. \tag{6}$$

In general, $C_{VA}(t)$ in (5) is a complex valued function, whose real and imaginary parts are denoted as

$$R_{VA}(t) := \mathrm{Re}(C_{VA}(t)), \tag{7}$$

$$I_{VA}(t) := \mathrm{Im}(C_{VA}(t)). \tag{8}$$

Next we turn to a slightly different situation, namely, the system is prepared at time $t = 0$ in an initial state of the form

$$\rho_0 := Z_g^{-1}\, e^{-\beta H_g}, \tag{9}$$

$$Z_g := \mathrm{Tr}\{e^{-\beta H_g}\}, \tag{10}$$

corresponding to the canonical ensemble of a modified Hamiltonian

$$H_g := H - gV, \tag{11}$$

where V is a perturbation operator and g a small parameter. While ρ_0 in (9) amounts to a thermal equilibrium state with respect to the modified Hamiltonian H_g, it is a nonequilibrium state with respect to the actual Hamiltonian H of the system we are considering. Accordingly, this initial state then evolves for $t > 0$ into the state

$$\rho(t) = e^{-iHt/\hbar}\rho_0\, e^{iHt/\hbar}, \tag{12}$$

yielding for an observable A the time-dependent expectation values:

$$\langle A\rangle_t := \mathrm{Tr}\{\rho(t)A\}. \tag{13}$$

Equivalently, they can be rewritten in the Heisenberg picture by means of (6) and (12) as

$$\langle A\rangle_t = \mathrm{Tr}\{\rho_0 A(t)\}. \tag{14}$$

In terms of the so-defined quantities, Onsager's regression hypothesis assumes the form

$$\langle A\rangle_t - A_{\mathrm{th}} = g\beta\, R_{VA}(t) \tag{15}$$

for sufficiently weak perturbations, i.e., up to corrections of order g^2. As announced, it relates the temporal correlations at thermal equilibrium to the time-dependent expectation values of a system close to equilibrium (usually exhibiting some kind of relaxation).

For quantum systems as we consider them here, Onsager's regression hypothesis (15) is known to be incorrect [3–5]

(a simple analytical example will be provided in Sec. V). However, an appropriately adjusted version of such a relation has to our knowledge never been proposed until now. This is the main objective of our present paper, namely, we will derive for sufficiently small g the following modified relation:

$$\langle A\rangle_t - A_{\rm th} = g\beta \sum_{k=0}^{\infty} \frac{(i\tau_B)^k}{(k+1)!}\, C_{VA}^{(k)}(t), \tag{16}$$

where $f^{(k)}(t)$ indicates the kth derivative of any given function $f(t)$, and where

$$\tau_B := \hbar\beta \tag{17}$$

is the so-called Boltzmann time. Furthermore, we will show that the sum on the right-hand side of (16) is a purely real function of t (even though the single summands may possibly be complex). Exploiting the definitions in (7) and (8), one can thus conclude that

$$\langle A\rangle_t - A_{\rm th} = g\beta \sum_{k=0}^{\infty} \frac{\tau_B^k}{(k+1)!}\, S_k(t), \tag{18}$$

where

$$S_k(t) := \begin{cases} (-1)^{k/2}\, R_{VA}^{(k)}(t) & \text{for even } k \\ (-1)^{\frac{k+1}{2}}\, I_{VA}^{(k)}(t) & \text{for odd } k. \end{cases} \tag{19}$$

In particular, Onsager's hypothesis (15) is asymptotically recovered for small values of τ_B, commonly considered as corresponding to the classical limit in view of (17).

The rest of the paper is organized as follows. The derivation of our main results (16)–(19) is provided in Sec. II. The validity range of these results is established in Sec. III, complemented by the more rigorous details in Appendix A. Some general implications of physical interest are discussed in Sec. IV, followed by various numerical examples in Sec. V, a modification and extension of our main results in Sec. VI and Appendix B, and some concluding remarks in Sec. VII.

II. DERIVATION OF THE MAIN RESULTS

In essence, the derivation of our main results (16)–(19) is relatively easy. Additional details are provided in Sec. III and Appendix A.

Employing the definitions

$$\psi(\lambda) := e^{\lambda H} e^{-\lambda H_g}, \tag{20}$$

$$q := {\rm Tr}\{\rho\, \psi(\beta)\}, \tag{21}$$

it follows with (1) and (9) that

$$q = Z_g/Z, \tag{22}$$

and with (14) that

$$\langle A\rangle_t = q^{-1}\, {\rm Tr}\{\rho\, \psi(\beta) A(t)\}. \tag{23}$$

Furthermore, we can conclude from (20) and (11) that

$$\begin{aligned} \psi'(\lambda) &:= d\psi(\lambda)/d\lambda = e^{\lambda H} H e^{-\lambda H_g} - e^{\lambda H} H_g e^{-\lambda H_g} \\ &= e^{\lambda H} g V e^{-\lambda H_g} = g V_\lambda \psi(\lambda), \end{aligned} \tag{24}$$

$$V_\lambda := e^{\lambda H} V e^{-\lambda H}. \tag{25}$$

Integrating (24) and exploiting that $\psi(0) = 1$ yields

$$\psi(\lambda) = 1 + g \int_0^\lambda dx\, V_x\, \psi(x). \tag{26}$$

Upon iteration, we thus obtain

$$\psi(\lambda) = 1 + g \int_0^\lambda dx\, V_x + g^2 \int_0^\lambda dx\, V_x \int_0^x dy\, V_y \psi(y). \tag{27}$$

Together with (1), (21), and (25), this implies

$$q = 1 + g\,\beta\, {\rm Tr}\{\rho V\} + g^2 Q, \tag{28}$$

$$Q := \int_0^\beta dx \int_0^x dy\, {\rm Tr}\{\rho V_x V_y \psi(y)\}. \tag{29}$$

Likewise, one finds that

$${\rm Tr}\{\rho\, \psi(\beta) A(t)\} = {\rm Tr}\{\rho A(t)\} + g \int_0^\beta d\lambda\, Y_t(\lambda) + g^2 R_t, \tag{30}$$

where

$$Y_t(\lambda) := {\rm Tr}\{\rho\, V_\lambda A(t)\}, \tag{31}$$

$$R_t := \int_0^\beta dx \int_0^x dy\, {\rm Tr}\{\rho V_x V_y \psi(y) A(t)\}. \tag{32}$$

Assuming temporarily that

$$V_{\rm th} := {\rm Tr}\{\rho V\} = 0 \tag{33}$$

[see also (4)], and omitting terms of order g^2, we thus can rewrite (23) by means of (28) and (30) as

$$\langle A\rangle_t = {\rm Tr}\{\rho A(t)\} + g \int_0^\beta d\lambda\, Y_t(\lambda). \tag{34}$$

One readily infers from (1), (4), and (6) that

$${\rm Tr}\{\rho A(t)\} = {\rm Tr}\{\rho A\} = A_{\rm th}, \tag{35}$$

yielding

$$\Delta(t) := \langle A\rangle_t - A_{\rm th} = g \int_0^\beta d\lambda\, Y_t(\lambda) \tag{36}$$

up to corrections of order g^2. Since the left-hand side is real, one also expects that the right-hand side is real separately in every order of g. In accordance with this expectation, one can explicitly verify that the leading order term on the right-hand side of (36) is indeed real by exploiting that (31) implies $Y_t^*(\lambda) = Y_t(\beta - \lambda)$.

Denoting the eigenvalues and eigenvectors of H as E_ν and $|\nu\rangle$, respectively, the corresponding matrix elements, for instance, of V are abbreviated as

$$V_{\nu\mu} := \langle \nu | V | \mu \rangle. \tag{37}$$

Hence, ρ from (1) becomes a diagonal matrix,

$$\rho_{\nu\mu} = \delta_{\nu\mu} p_\nu, \tag{38}$$

with diagonal elements

$$p_\nu := Z^{-1} e^{-\beta E_\nu}. \tag{39}$$

Employing (1), (6), and (25), we thus can rewrite (31) as

$$\begin{aligned} Y_t(\lambda) &= \mathrm{Tr}\{\rho\, e^{\lambda H} V e^{-\lambda H} e^{iHt/\hbar} A e^{-iHt/\hbar}\} \\ &= \sum_{\nu\mu} p_\nu\, e^{\lambda E_\nu} V_{\nu\mu} e^{-\lambda E_\mu} e^{iE_\mu t/\hbar} A_{\mu\nu} e^{-iE_\nu t/\hbar} \\ &= \sum_{\nu\mu} p_\nu\, V_{\nu\mu} A_{\mu\nu} e^{E_{\mu\nu}(it/\hbar-\lambda)}, \end{aligned} \tag{40}$$

where

$$E_{\mu\nu} := E_\mu - E_\nu. \tag{41}$$

The last factor in (40) can be rewritten as

$$\begin{aligned} e^{E_{\mu\nu}(it/\hbar-\lambda)} &= e^{iE_{\mu\nu}t/\hbar} \sum_{k=0}^{\infty} \frac{(-\lambda E_{\mu\nu})^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} (-i\hbar)^k \frac{d^k}{dt^k} e^{iE_{\mu\nu}t/\hbar}, \end{aligned} \tag{42}$$

implying

$$Y_t(\lambda) = \sum_{k=0}^{\infty} \frac{(i\lambda\hbar)^k}{k!} \frac{d^k}{dt^k} \sum_{\nu\mu} p_\nu\, V_{\nu\mu} A_{\mu\nu} e^{iE_{\mu\nu}t/\hbar}. \tag{43}$$

Similarly as in (40), the last double sum can be identified with $\mathrm{Tr}\{\rho\, VA(t)\}$, while the remaining integral over λ in (36) can now be carried out, yielding

$$\Delta(t) = g\beta \sum_{k=0}^{\infty} \frac{(i\tau_B)^k}{(k+1)!} \frac{d^k}{dt^k} \mathrm{Tr}\{\rho\, VA(t)\}. \tag{44}$$

So far, our derivation only applies for V's with the property $V_{\mathrm{th}} = 0$; see (33). The generalization to arbitrary V's is straightforward: To begin, we define $\tilde{V} := V - V_{\mathrm{th}}$ and observe that employing $\tilde{V}$ instead of V in (11) does not affect any physically relevant system properties such as the relations (1)–(10) or (12)–(14). In particular, $\Delta(t)$ as defined in (36), remains unchanged. On the other hand, $\tilde{V}$ now exhibits the property $\tilde{V}_{\mathrm{th}} = 0$ and hence (44) must apply with $\tilde{V}$ instead of V. Altogether, the generalization to arbitrary V's thus simply amounts to replacing V in (44) by $V - V_{\mathrm{th}}$. Exploiting (5) and (35), this finally yields (16).

Moreover, according to the discussion below Eq. (36), the right-hand side of (44) and thus of (16) must be a purely real function of t, as claimed below Eq. (17).

III. VALIDITY OF THE APPROXIMATION

Combining (23), (28), (30), (33), and (35) yields

$$\langle A\rangle_t = \frac{A_{\mathrm{th}} + g\int_0^\beta d\lambda \mathrm{Tr}\{\rho V_\lambda A(t)\} + g^2 R_t}{1 + g^2 Q}. \tag{45}$$

Our main (and only) approximation in Sec. II was to neglect the terms of order g^2 in (45), resulting in (34). However, since Q and R_t in (45) actually still depend on g themselves via (29), (32), and (20), one may feel that a more careful justification of such an approximation would be desirable. This is accomplished in Appendix A in the form of rigorous upper bounds for $|Q|$ and $|R_t|$. Most importantly, and in contrast to ordinary time-dependent perturbation theory, those bounds are not restricted to sufficiently small times but rather apply uniformly in t.

The derivation of those bounds is based on the following two premises: The system Hamiltonian H must be a sum of local operators, and the perturbation V must be a local operator; see Appendix A for further details. From now on, these two premises are thus tacitly taken for granted.

Roughly speaking, the relevant small parameter in the detailed analytical considerations in Appendix A is given, as one might have intuitively expected, by $g\beta\|V\|_{\mathrm{op}}$ (operator norm): If this parameter is small, the neglected corrections (of higher order in g) on the right-hand side of (16) and (18) can be shown to be small. More precisely speaking, the system Hamiltonian H must also, of course, somehow enter the game, giving rise to an extra factor $f(\beta)$ [see Eq. (A33)]. The latter accounts for the specific properties of H as detailed below Eq. (A33), and is often expected to be of the order of unity for small-to-moderate β values when working in natural units.

Another very important question is how Q and R_t in (45) depend on the size of the considered system and, in particular, how they behave in the thermodynamic limit [6]. This issue is clearly far from obvious in view of (29), (32), and the exponential dependence on the system Hamiltonians in (6), (20), and (25). From the rigorous considerations in Appendix A, one can infer the following quite strong conclusion regarding this issue: If $\|V\|_{\mathrm{op}}$ and $\|A\|_{\mathrm{op}}$ remain bounded in the thermodynamic limit, then the same must apply to $|Q|$ and $\max_t |R_t|$. Therefore, Q and R_t are indeed negligible in (45) for sufficiently small g even after taking the thermodynamic limit.

Due to our above premise that V is a local operator, the requirement that $\|V\|_{\mathrm{op}}$ must remain bounded in the thermodynamic limit is (almost) automatically fulfilled. Similarly, since the entire formalism is linear in A, and since most physically relevant observables are (sums of) local operators, the requirement that $\|A\|_{\mathrm{op}}$ must remain bounded in the thermodynamic limit does not amount to any substantial loss of generality either.

We finally note that the extension of our present analytical results to nonlocal perturbations V amounts to a very challenging unsolved problem, as can be understood by the following nonrigorous arguments. Let us, for instance, consider a local observable A and an extensive perturbation V (meaning that V consists of a sum of local operators which grows extensively with the system size [6]). For physical reasons, it then seems intuitively reasonable to expect that the quantities $\langle A\rangle_t$ and A_{th} in (45) exhibit a well-defined (finite) thermodynamic limit [7]. Turning to the second term in the numerator of (45), we observe that this term coincides with the right-hand side of (36) and thus with (44) and (16). As before, it is reasonable to expect that the correlations $C_{VA}(t)$ in (5) exhibit a well-defined (finite) thermodynamic limit [8]. Hence, (16) is also expected to remain finite in the thermodynamic limit. Finally, by means of numerical explorations and of similar analytical considerations as in Appendix A, we found that both Q and R_t generally seem to diverge in the thermodynamic limit. [Incidentally, R_t/Q must still converge to $\langle A\rangle_t$ according to (45).] Thus, neglecting Q and R_t in (45) can only be justified for smaller and smaller g values as the system size increases. In other words, the main approximation

of our present approach breaks down for any finite g in the thermodynamic limit.

IV. GENERAL REMARKS

Although systems with many degrees of freedom (DOFs) are usually of foremost interest (see also Sec. III), it is nevertheless remarkable that our main results actually apply to systems with an arbitrary number of DOFs, including very small systems.

Similarly, it is remarkable that issues like integrability, ergodicity, or many-body localization (MBL) do not play any role.

In most previous theoretical works in this context, the physical interest of such temporal correlations as in (5), (7), (8) is considered self-evident. Here, we adopt the same viewpoint that relations like (15)–(18) are of considerable theoretical interest in themselves without any further discussion of whether and how such temporal correlations may be experimentally measurable. (The latter is quite obvious in the special case $t = 0$, which we address in more detail below, but not any more for $t \neq 0$.) For a more detailed discussion of such issues, we refer to Refs. [3,5,11] and further references therein.

As already mentioned below (19), Onsager's original hypothesis (15) is recovered for asymptotically small values of τ_B in (18). Moreover, one can infer from (18) that the deviations from Onsager's hypothesis remain small as long as the characteristic timescale of the correlations in (5) is large compared to the Boltzmann time τ_B [9]. Experimentally, this will often be the case unless we are dealing with rather low temperatures. On the other hand, in numerical studies, where choosing units with $\hbar = 1$ is quite natural, those corrections are expected to be non-negligible in many cases (see also Sec. V below).

For sufficiently large systems, one often expects, observes, or can even show [10] that the time-dependent expectation values in (13) approach a (nearly) constant value after initial transients have died out, and analogously for the the temporal correlations in (5) [11]. Quantitatively, these constant values can be determined by taking the long-time average on both sides of (18). One readily sees that this average is zero for each summand with $k > 0$ on the right-hand side of (18). Remarkably, the dependence on τ_B thus disappears altogether and we recover once again the same result as in Onsager's original relation (15) as far as long-time averages are concerned.

On the other hand, for large but finite systems one expects that both sides in (18) still exhibit (even for arbitrarily late times) some small temporal fluctuations around their long-time averages [10,11]. A quite nontrivial prediction of our present paper is that those long-time fluctuations are again connected via the relation (18). Furthermore, it is not obvious at all how significant the finite τ_B corrections of Onsager's hypothesis (15) will be with respect to those long-time fluctuations.

More generally speaking, our present theory must remain valid even in such unusual cases where the system does not exhibit any kind of relaxation, equilibration, or thermalization in the long run. For instance, this may be the case for systems with few DOFs or due to some special symmetries and conservation laws.

Yet another interesting special case arises for $t = 0$, namely, the left-hand side in the relations (15), (16), and (18) then amounts to the difference between the thermal expectation values for two sightly different systems with Hamiltonians H_g and H [see Eqs. (1), (4), and (9)–(13)]. In other words, all quantities appearing in (15), (16), and (18) now solely refer to thermal equilibrium properties. In the context of equilibrium statistical mechanics, such equations are sometimes also denoted, for instance, as fluctuation-response relations, though the choice of this and other names tends to be somewhat unfortunate [5]. While Eq. (15) with $t = 0$ amounts to a textbook relation of this kind, our present results (16) and (18) amount to a nontrivial improvement for quantum systems, which to our knowledge has not been previously known.

V. QUANTITATIVE EXAMPLES

Here, we will quantitatively illustrate some of the general issues addressed in the previous sections by means of specific examples. A systematic exploration of all those various analytical predictions is clearly impossible, hence we will confine ourselves to just a few instances.

In particular, we will mainly focus on one-dimensional Heisenberg-type Hamiltonians (XYZ models) with open boundary conditions,

$$H = -\sum_{a\in\{x,y,z\}} J_a \sum_{l=1}^{L-1} s_{l+1}^a s_l^a + h\sum_{l=1}^{L} s_l^x, \tag{46}$$

where $s_l^{x,y,z}$ are spin-1/2 operators at the lattice sites $l \in \{1, ..., L\}$, the $J_{x,y,z}$ quantify the coupling strength of the nearest-neighbor interactions, and h may be considered to account for an external magnetic field. Focusing on nonvanishing and nonidentical couplings $J_{x,y,z}$, this model is nonintegrable for $h \neq 0$ [12], and integrable (but still of a nontrivial, so-called interacting type) for $h = 0$ [13–15]. Moreover, such models are commonly expected to exhibit a well-defined thermodynamic limit when $L \to \infty$ (see Sec. III).

As mentioned in Sec. III, the perturbation V in (11) is assumed to be a local operator. As a particularly simple choice, we will mainly consider examples of the form $V = s_l^z$ for some $l \in \{1, ..., L\}$. Physically, it is reasonable to expect (and numerically seen) that such a local perturbation will only lead to a notable response for observables which are not too far away from the perturbation, and will be particularly well visible if the observable is in some sense similar to the perturbation. Therefore, it is natural to focus on observables of the form $A = V$.

Yet another appealing feature of the above choice $A = V = s_l^z$ is that the corresponding thermal expectation values $A_{\rm th}$ in (5) can be shown analytically to vanish for symmetry reasons [16]. Moreover, one readily concludes that $VA(0) = (s_l^z)^2 = 1/4$. Together with (5), it follows that $C_{VA}(0) = 1/4$. Exploiting (7), Onsager's regression hypothesis at the time

point $t = 0$ thus assumes the simple form

$$\langle A\rangle_{t=0} = g\beta/4, \tag{47}$$

independent of any further details of the considered system. Since the left-hand side in (47) is bounded by the largest and smallest eigenvalues of $A = s_l^z$, that is, by $\pm 1/2$, the prediction (47) will certainly be wrong when $|g\beta| > 2$. To the best of our knowledge, this is the simplest analytical example for the failure of Onsager's regression hypothesis in the quantum regime.

Our next observation is that a spin-1/2 model like in (46) does not admit a physically meaningful classical limit. The basic reason is that the classical limit requires the emergence of an asymptotically continuous level density for $\hbar \to 0$ (the simplest example is a harmonic oscillator), while the number of energy levels in (46) is finite and independent of $\hbar$. As usual, in numerical explorations of such models, we thus work in units with

$$\hbar = 1, \tag{48}$$

implying that τ_B in (17) formally coincides with β.

The general strategy in our subsequent numerical explorations will be to compute the left-hand side in (18) and compare it with the numerically evaluated functions

$$P_K(t) := g \sum_{k=0}^{K} \frac{\beta^{k+1}}{(k+1)!}\, S_k(t), \tag{49}$$

converging towards the right-hand side of (18) for large values of K, while $K = 0$ corresponds to Onsager's hypothesis (15).

In passing, we note that the dependence of (49) on β is not as simple as it might appear at first glance since the functions $S_k(t)$ also depend on β according to (1)–(8) and (19). Nevertheless, it is reasonable to expect, and will be later confirmed numerically, that the factor β^{K+1} is asymptotically dominating in (49) for large β.

In the following subsections, we will present our results for the left-hand side of (18) and for the functions $P_K(t)$ in (49). These results have been obtained by the numerically exact diagonalization of the pertinent Hamiltonian H in (46). Moreover, we have circumvented the numerical evaluation of the kth derivative appearing in (19) as follows: As a first step, we observe that (1) and (6) imply

$$\mathrm{Tr}\{\rho\, V(s)A(t)\} = \mathrm{Tr}\{\rho\, VA(t-s)\} \tag{50}$$

for arbitrary $t, s \in \mathbb{R}$, and hence

$$\begin{aligned}\mathrm{Tr}\{\rho\, V^{(k)}(0)A(t)\} &= \frac{d^k}{ds^k}\mathrm{Tr}\{\rho\, V(s)A(t)\}|_{s=0}\\ &= \frac{d^k}{ds^k}\mathrm{Tr}\{\rho\, VA(t-s)\}|_{s=0}\\ &= (-1)^k\,\mathrm{Tr}\{\rho\, VA^{(k)}(t)\}.\end{aligned} \tag{51}$$

As a second step, we define (as usual) $\dot V := i\,[H, V]$ [see also (48)], while the higher derivatives $V^{(k)}$ then follow recursively as

$$V^{(k+1)} := i\,[H, V^{(k)}]. \tag{52}$$

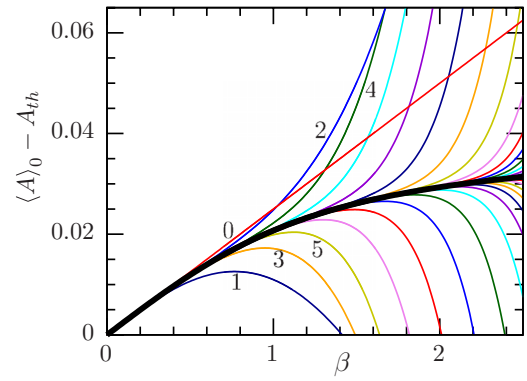

FIG. 1. Numerical results for the nonintegrable spin model from (49) with $L = 16$, $J_x = 1$, $J_y = 1.2$, $J_z = 1.5$, $h = 1$, and a local perturbation in (11) with $g = 0.1$ and $V = s^z_{L/2}$. Bold black line: Left-hand side of (18) versus β for $A = s^z_{L/2}$ and $t = 0$. Colored lines: Numerical results for $P_K(0)$ from (49) with $K = 0, 1, ..., 30$. The first few K values are indicated as numbers close to the corresponding lines. The continuation for all other K is obvious. In particular, the straight red line ($K = 0$) represents Onsager's regression hypothesis from (15).

Alternatively, if we set $V^{(0)} := V$ then (52) is valid for all $k \in \mathbb{N}_0$. As expected, it follows by exploiting (6) that

$$V^{(k)}(0) = V^{(k)}, \tag{53}$$

and thus with (5) and (51) that

$$C^{(k)}_{VA}(t) = (-1)^k\,\mathrm{Tr}\{\rho\, V^{(k)}A(t)\} \tag{54}$$

for all $k \geqslant 1$. Analogous relations apply to the derivatives of the real and imaginary parts in (7) and (8), which arise in (19) and are thus needed in (49). The main point is that the commutators, appearing on the right-hand side of (54) via (52), are numerically much more convenient and accurate than directly evaluating the kth derivative on the left-hand side of (54).

A. Numerical results for $t = 0$

As mentioned at the end of Sec. IV, in the special case $t = 0$ our findings are essentially tantamount to a generalization of the so-called fluctuation-response relation in the context of equilibrium statistical mechanics.

A first example is depicted in Fig. 1. The main conclusion is that Onsager's regression hypothesis (straight red line, $K = 0$) indeed becomes quite bad for large β (low temperatures), while our improved prediction from (18) (corresponding to $K \to \infty$) works very well. Furthermore, and as announced below Eq. (49), for any given (finite) K value, the function $P_K(0)$ indeed seems to diverge approximately as β^{K+1} for large β. Some interesting additional features, which are numerically observed, and which can also be to some extent understood as a consequence of this divergence and the fact that the left-hand side of (18) is bounded (uniformly in β), are as follows: (i) Upon increasing K, the functions $P_K(0)$ from (49) diverge for asymptotically large β alternatingly towards plus and minus infinity. (ii) The convergence of the finite sums $P_K(0)$ from (49) towards the infinite sum in (18) cannot be uniform in β: The larger the value of β,

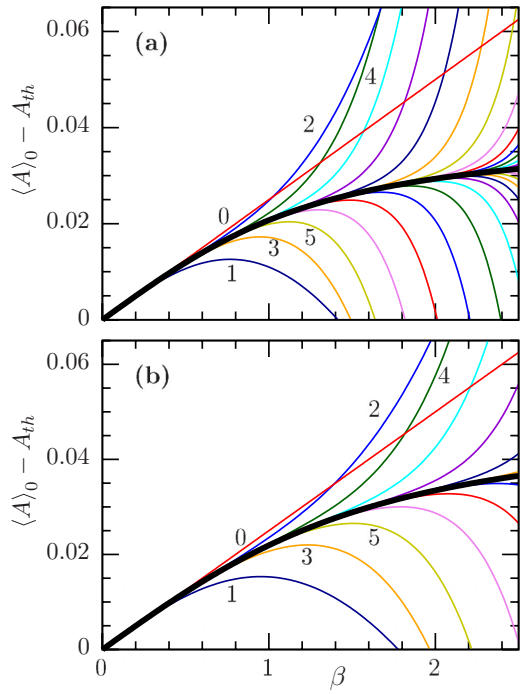

FIG. 2. Same as in Fig. 1 but for smaller systems with $L = 8$ in (a) and $L = 2$ in (b).

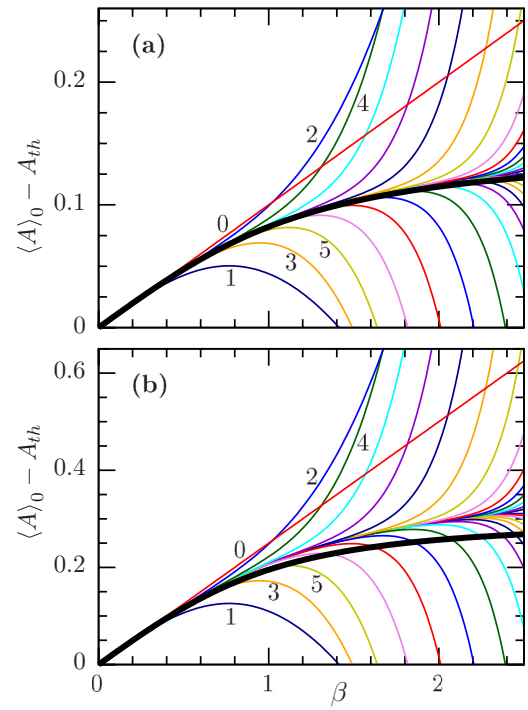

FIG. 3. Same as in Fig. 1 but for stronger perturbations with $g = 0.4$ in (a) and $g = 1$ in (b).

the larger values of K are needed until convergence sets in. (iii) For sufficiently large but fixed β, the approximations $P_K(0)$ considered as a function of K will initially (for small-to-moderate K) even become worse upon increasing K, and only later (for sufficiently large K) converge towards a really good approximation of the infinite sum in (18).

We also remark that the perfectly straight numerical lines for $K = 0$ in Figs. 1–5 are in agreement with the analytical prediction from (47). An exception is Fig. 4(b), where the requirement $A = V = \sigma_l^z$ above Eq. (47) is not fulfilled, and hence no straight line arises for $K = 0$.

Turning to Fig. 2, its main message (together with Fig. 1) is that the dependence on the system size L is remarkably weak. In other words, the thermodynamic limit is approached very quickly upon increasing L. We conjecture that our employing canonical ensembles as initial states in (9) may play an important role in this context, since similar observations have also been reported in various other numerical explorations in the literature. We also see that, as predicted in Sec. III, our analytical theory indeed works very well independently of whether the considered system is small or large.

Figure 3 exemplifies the impact of the neglected higher order terms in g. In agreement with the analytics at the end of Appendix A (see also Sec. III), the essential quality parameter for the reliability of our present linear response type approximation is the product $g\beta\|V\|_{\mathrm{op}}$, where $\|V\|_{\mathrm{op}} = 1/2$ in all our numerical examples. Quantitatively, the differences between the numerically exact results (bold black lines) and the analytical linear response theory (colored lines with sufficiently large K) indeed become visible on the scale of these plots when β is comparable or larger than $1/g$.

Figure 4 illustrates the behavior for a different choice of the perturbation V and the observable A. Particularly noteworthy are the following features of Fig. 4(b): (i) Quantitatively, the observed effects are generally weaker than in the other examples (note the y-axis scales). (ii) The relative deviations

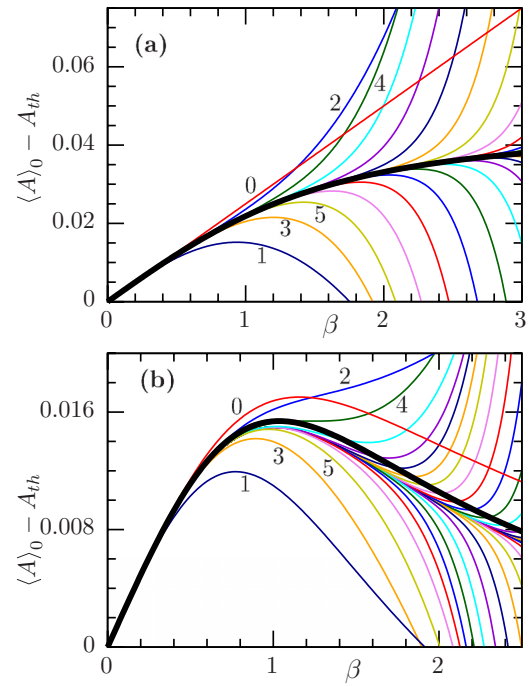

FIG. 4. Same as in Fig. 1 but for a different perturbation and observable, namely, $V = A = s_1^z$ in (a) and $V = A = s_{L/2}^x$ in (b). Moreover, in (a) the depicted range of β values is somewhat larger than in Fig. 1.

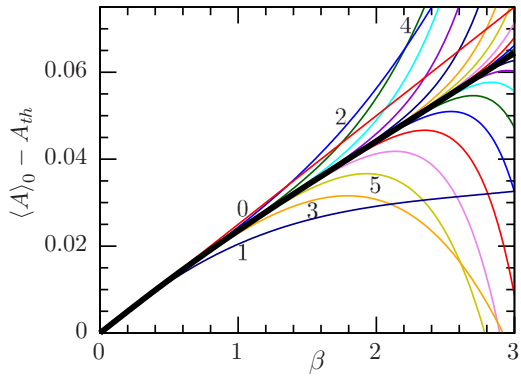

FIG. 5. Same as in Fig. 1 but for an integrable model with $h = 0$ in (49). Moreover, the depicted range of β values is somewhat larger than in Fig. 1.

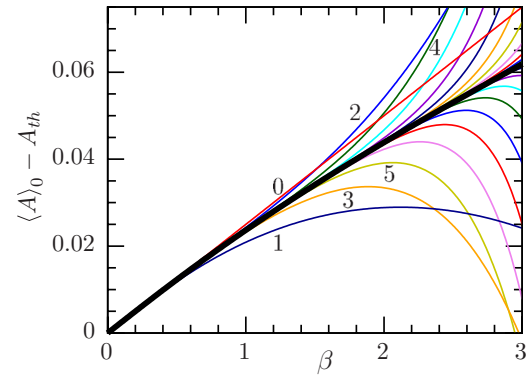

FIG. 6. Same as in Fig. 1 but for the frustrated spin chain model from (55) with $J = 0.25$. Moreover, the depicted range of β-values is somewhat larger than in Fig. 1.

(finite-g effects) between the numerically exact results (black) and our analytics (corresponding to $K \to \infty$) are larger than in Figs. 1, 2, and 4(a), but the absolute deviations remain comparable. (iii) Unlike in Fig. 3, those deviations hardly grow with increasing β as far as their absolute values are concerned, but the relative deviations still grow in nearly the same way. (iv) Similar considerations apply to the differences between Onsagers's hypothesis ($K = 0$, red) and the numerically exact results. (v) As stated, Onsager's hypothesis no longer amounts to a straight line.

For the rest, it does not seem to us of great interest to present here further results for still other examples of V and A; see also the general considerations above Eq. (47).

Turning to Fig. 5, a comparison of these results (for $h = 0$) with those in the other examples (with $h = 1$) confirms the prediction from Sec. IV that our analytical theory works very well for nonintegrable ($h \neq 0$) as well as for integrable ($h = 0$) models.

As yet another, somewhat more exotic example, we finally consider a so-called frustrated ferromagnetic Heisenberg spin-1/2 chain of the form [17,18]

$$H = -\sum_{l=1}^{L-1} \vec{s}_{l+1} \cdot \vec{s}_l + J \sum_{l=1}^{L-2} \vec{s}_{l+2} \cdot \vec{s}_l, \tag{55}$$

with $J > 0$ (frustration) and $\vec{s}_l := (s_l^x, s_l^y, s_l^z)$; see also below Eq. (46). Besides the special feature of frustration, this model is also known to exhibit numerous degeneracies [due to its SU(2) symmetry] and to be nonintegrable. The concomitant numerical findings in Fig. 6 are quite similar to those in (5), indicating that frustration and/or degeneracies do not seem to play an important role with respect to the questions in which we are interested here.

B. Time-dependent numerical results

Figure 7 depicts the time-dependent comparison of the left-hand side of Eq. (18) with the approximations on the right-hand side of Eq. (49) for two representative K values. As before, Onsager's regression hypothesis (red lines, $K = 0$) quite notably deviates from the numerically exact results (black). The approximation for $K = 10$ (blue) is sometimes better and sometimes worse than that for $K = 0$. This is in accordance with the fact that the lines for $K = 0$ and $K = 10$ in Fig. 1 cross each other near $\beta = 2$ [see also item (iii) in Sec. V A]. For $K \geqslant 20$, our numerical data were found to practically coincide with the black lines in Fig. 7. In other words, our analytical theory (corresponding to $K \to \infty$) indeed agrees very well with the numerically exact results for all times t which we actually explored.

We also observe that the finite-K effects in Fig. 7 are particularly pronounced when the numerically exact curves (black)

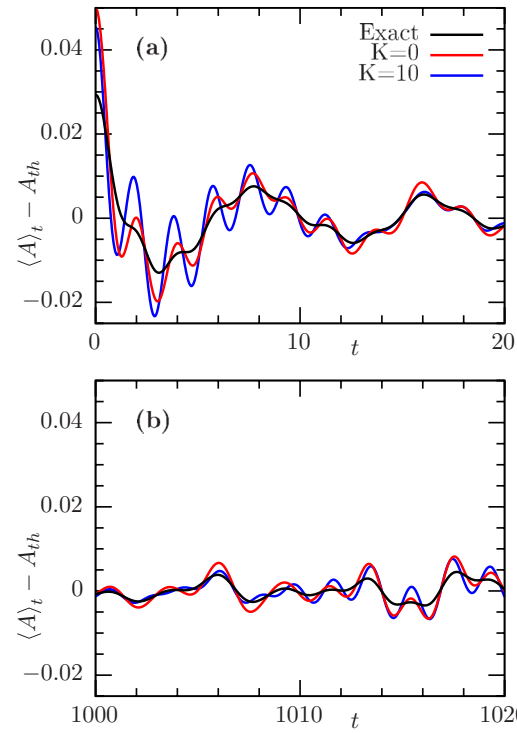

FIG. 7. Numerical results for the same model as in Fig. 1 with $\beta = 2$. Black lines: Left-hand side of (18) versus t for (a) $t \in [0, 20]$ (initial behavior) and (b) $t \in [1000, 1020]$ (long-time behavior). Colored lines: Numerical results for $P_K(t)$ from (49) with $K = 0$ (red) and $K = 10$ (blue). The corresponding results $K = 20$ and $K = 30$ were found to be indistinguishable from the black lines and are therefore not shown.

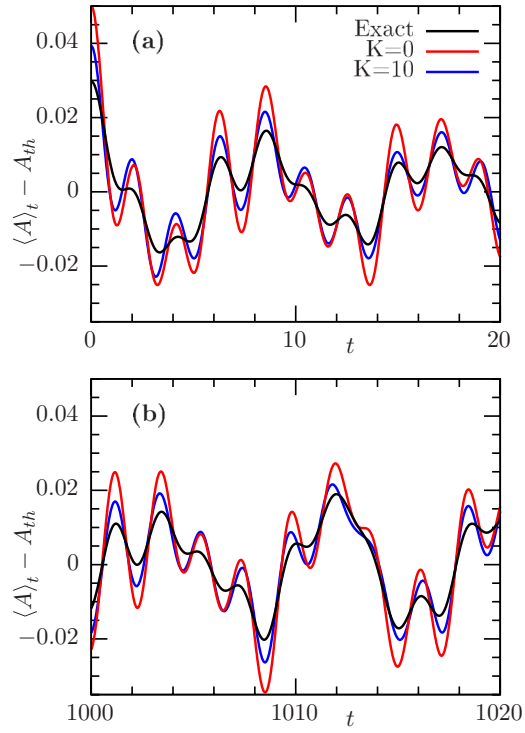

FIG. 8. Same as in Fig. 7 but for a smaller system with $L = 4$.

exhibit large curvatures, and are comparatively small near turning points. Unfortunately, we did not succeed in coming up with a simple intuitive explanation of this observation.

While Fig. 7(a) illustrates the initial relaxation behavior for $t \in [0, 20]$, the everlasting long-time fluctuations discussed at the end of Sec. IV are exemplified by the behavior for $t \in [1000, 1020]$ in Fig. 7(b). Though the detailed behavior for those large times is quite nontrivial, the agreement with our analytical theory is still nearly perfect for sufficiently large K, while the deviations for $K = 0$ (Onsager's hypothesis) are still quite notable.

Finally, Fig. 8 confirms (upon comparison with Fig. 7) the theoretical prediction [10,11] (see also Sec. IV) that these long-time fluctuations quickly decrease as the system size is increased. For the rest (initial relaxation behavior, long-time average), the dependence on the system size is rather weak, as already observed in the previous subsection. Moreover, Fig. 8 once more confirms the prediction that our analytical theory works very well even for small systems and/or in the absence of a distinct initial relaxation.

VI. GENERALIZATION

Our goal is to extend the results from Sec. I to more general density operators than in (1) and (9), provided the system satisfies certain additional assumptions, namely, we assume that there exists a projector P onto some sub-Hilbert space $\tilde{\mathcal{H}}$ with the property

$$[H, P] = [V, P] = 0. \tag{56}$$

The most important example arises when the system Hamiltonian H as well as the perturbed Hamiltonian H_g from (11) exhibit a common conserved quantity S, i.e.,

$$[H, S] = [V, S] = 0, \tag{57}$$

and if we identify $\tilde{\mathcal{H}}$ with one of the eigenspaces of S. [One readily verifies that the property (56) is indeed fulfilled in such a case.] More generally, if there are several (commuting) conserved quantities of H and H_g, the subspace $\tilde{\mathcal{H}}$ may be chosen as an eigenspace of one of those conserved quantities, but also as a common eigenspace of several conserved quantities.

Given that (56) is fulfilled, we show in Appendix B that our main results from (16)–(19) are still valid if the canonical ensembles in (1) and (9) are replaced by

$$\rho := Pe^{-\beta H} / \operatorname{Tr}\{Pe^{-\beta H}\}, \tag{58}$$

$$\rho_0 := Pe^{-\beta H_g} / \operatorname{Tr}\{Pe^{-\beta H_g}\}. \tag{59}$$

In other words, we can replace the original density operators from (1) and (9) by their projections (or reductions) onto any invariant subspace $\tilde{\mathcal{H}}$, whereas the observable A and system Hamiltonian H still remain entirely unchanged, for instance, in (4)–(6) and in (12)–(14).

While one may come up with various intuitive arguments of why these findings might not be entirely unexpected, the details of a more rigorous line of reasoning are not obvious at all; see Appendix B.

If relations as in (56) apply simultaneously to two different projectors P_1 and P_2, our main results from (16)–(19) are valid for each of the two corresponding ensembles of the forms (58) and (59). It thus might seem tempting to conjecture that linear combinations thereof are also still be admissible. However, this cannot be true in view of the fact that the left-hand side of (16) is linear in ρ, while the last term in (5) and thus the right-hand side of (16) is nonlinear in ρ.

To numerically illustrate these analytical findings, we consider a random field Heisenberg spin-1/2 chain of the form

$$H = \sum_{l=1}^{L-1} \vec{s}_{l+1} \cdot \vec{s}_l + \sum_{l=1}^{L} h_l s_l^x, \tag{60}$$

where $\vec{s}_l := (s_l^x, s_l^y, s_l^z)$; see also below Eq. (46). Moreover, the h_l are independent random variables, uniformly distributed within the interval $[-h_{\max}, h_{\max}]$ for some arbitrary but fixed $h_{\max} \geqslant 0$. This model is thus quite similar to (46), except that all nearest-neighbor coupling strengths J_a are now chosen equal to -1 and the external magnetic field is now randomized.

The model (60) is one of the standard examples considered in the context of MBL, see, for instance, Refs. [19,20] and further references therein. Regarding the disorder strength $h_{\max}$, we focus on the choice $h_{\max} = 5$, for which the model has been reported to exhibit MBL in Ref. [19]. We remark that whether and in which sense this finding from Ref. [19] is actually correct is still debated in the literature. This controversy is in itself not of immediate relevance in our present paper: For us, the model (60) just serves the purpose to further enlarge the diversity of different models we are numerically exploring.

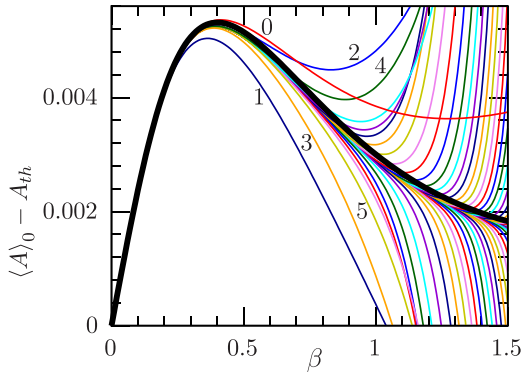

FIG. 9. Numerical results for the random field spin chain from (60) with disorder strength $h_{\max} = 5$ for an observable A and a perturbation in (11) with $A = V = s^x_{L/2}$ and $g = 0.1$. Furthermore, Eqs. (58) and (59) have been employed with a projector P onto the eigenspace of the conserved quantity from (61) with vanishing eigenvalue (subsector with zero total magnetization). All further details are as in Fig. 1 except that curves up to $K = 50$ are shown.

Particularly important in our present context is the well-known fact [19] that the total magnetization

$$S := \sum_{l=1}^{L} s^x_l \tag{61}$$

is a conserved quantity of the model from (60). Similarly as in Ref. [19], we thus restrict ourselves to the eigenspace $\tilde{\mathcal{H}}$ of S with eigenvalue zero, and thus with the corresponding projector P onto $\tilde{\mathcal{H}}$ appearing in Eqs. (58) and (59). Moreover, we focus on perturbations of the form $V = s^x_l$ to fulfill the second condition in (56).

The main qualitative features of our numerical findings in Fig. 9 are quite similar to those in Fig. 4(b), suggesting that neither the disorder nor the restriction to a subspace of the conserved quantity seems to have a major impact with respect to the main issues of our present paper.

VII. CONCLUSIONS

Our main result consists of a modification of Onsager's regression hypothesis [see Eq. (18)], which is analytically shown to be correct in linear order of the perturbation strength g in (11). In contrast to most other perturbative approaches in similar contexts, the neglected nonlinear corrections can be rigorously bounded for a large class of physically relevant model systems and for arbitrarily large times t (see Sec. III). Further noteworthy general features of the theory are collected in Secs. IV and VI, while more detailed quantitative examples are provided in Secs. V and VI.

Since Onsager's regression hypothesis is known to fail in the quantum regime [3–5] [see also around Eq. (47)], a common proposal has been that the adequate quantum version of the hypothesis should be nothing else than the well-established fluctuation-dissipation theorem (FDT) [4,21,22]. This proposal is incorrect for the following reasons. With respect to the specific questions in which we are interested here, a very basic feature of the FDT is that it does not admit any non-trivial prediction concerning the initial time point $t = 0$. On the other hand, Onsager's regression hypothesis from (15) does make a nontrivial prediction for $t = 0$. While the latter prediction is still not quantitatively right, it is appropriately corrected by our present theory (see Sec. V A).

Formally speaking, Onsager's original regression hypothesis works well in the combined linear response and near-classical regime, i.e., when both the perturbation parameter $g\beta\|V\|_{\mathrm{op}}$ (see Sec. III) and the Boltzmann time $\tau_{\mathcal{B}}$ in (17) are simultaneously small. Or, when working in natural units with $\hbar = 1$ and $\|V\|_{\mathrm{op}} = 1$, both $g\beta$ and β must be simultaneously small. In contrast, our present theory works well as long as the product $g\beta$ is small, while β may not be small. On the other hand, to explore very low temperatures (large β) and not too small perturbations, one necessarily must go beyond the present linear response regime. This will be the subject of a separate publication.

A quite common practice in cases like ours is to devise some sort of expansion in the perturbation parameter, and then to simply neglect all nonlinear terms without any further consideration of their actual effect. Such a strategy has been severely criticized in a hardly accessible but nevertheless highly cited paper by van Kampen [23]; see also Sec. 2 in Ref. [21]. Within our present approach, those nonlinear terms can be rigorously bounded for a large class of physically relevant model systems. On the other hand, we pointed out in Sec. III that the higher order terms are seen to diverge in the thermodynamic limit for other model classes, most prominently when the perturbation operator is an extensive quantity. In other words, the widespread opinion that van Kampen's criticism is irrelevant for all practical purposes may not be appropriate. A more detailed account of our ongoing explorations in the case of extensive perturbations will be given elsewhere.

ACKNOWLEDGMENTS

We are indebted to A. Dymarsky for informing us about his important rigorous results in Ref. [24] and about unpublished generalizations thereof. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 355031190 within the Research Unit FOR 2692 and under Grant No. 502254252, and by the Paderborn Center for Parallel Computing (PC^2) within the project HPC-PRF-UBI2.

APPENDIX A: RIGOROUS JUSTIFICATION OF THE APPROXIMATION (34)

As usual, the Hilbert space of the considered model is denoted as $\mathcal{H}$, and the norm $\|\phi\|$ of any vector $|\phi\rangle \in \mathcal{H}$ is defined as $\langle\phi|\phi\rangle^{1/2}$. Furthermore, for any linear (but not necessarily Hermitian) operator $B : \mathcal{H} \to \mathcal{H}$, the operator norm is defined as

$$\|B\|_{\mathrm{op}} := \sup_{\|\phi\|=1} \||B|\phi\rangle\|. \tag{A1}$$

If B is Hermitian, this is equal to the maximum (supremum) of all its eigenvalues in modulus. Some well-established

relations for arbitrary linear (but not necessarily Hermitian) operators B and C are

$$\|BC\|_{\rm op} \leqslant \|B\|_{\rm op}\|C\|_{\rm op}, \tag{A2}$$

$$\|B+C\|_{\rm op} \leqslant \|B\|_{\rm op} + \|C\|_{\rm op}, \tag{A3}$$

$$\|B^\dagger\|_{\rm op} = \|B\|_{\rm op}, \tag{A4}$$

$$\|B^\dagger B\|_{\rm op} = \|B\|_{\rm op}^2. \tag{A5}$$

For example, $\psi(\lambda)$ in (20) is such a linear but generally not Hermitian operator.

Our first main goal is to derive bounds for the quantity R_t from Eq. (32), which we rewrite as

$$R_t = \int_0^\beta dx \int_0^x dy\, r_t(x,y), \tag{A6}$$

$$r_t(x,y) := {\rm Tr}\{\rho V_x V_y \psi(y) A(t)\}. \tag{A7}$$

Considering ${\rm Tr}\{\rho B^\dagger C\}$ as a scalar product of two arbitrary linear (but not necessarily Hermitian) operators B and C, the Cauchy-Schwarz inequality takes the form

$$|{\rm Tr}\{\rho\, B^\dagger C\}|^2 \leqslant {\rm Tr}\{\rho\, B^\dagger B\}{\rm Tr}\{\rho\, C^\dagger C\}. \tag{A8}$$

Moreover, for an arbitrary Hermitian, non-negative operator D, one readily verifies by evaluating the trace by means of the eigenbasis of D that $|{\rm Tr}\{\rho D\}| \leqslant \|D\|_{\rm op}{\rm Tr}\{\rho\} = \|D\|_{\rm op}$. Since $B^\dagger B$ and $C^\dagger C$ in (A8) are Hermitian and non-negative, it follows with (A5) and (A8) that

$$|{\rm Tr}\{\rho B^\dagger C\}|^2 \leqslant \|B\|_{\rm op}^2\|C\|_{\rm op}^2. \tag{A9}$$

By exploiting this result and (A2), we can infer from (A7) that

$$|r_t(x,y)| \leqslant \|V_y\|_{\rm op}\|V_x\|_{\rm op}\|\psi(x)\|_{\rm op}\|A\|_{\rm op}. \tag{A10}$$

Without any significant loss of generality, we henceforth restrict ourselves to non-negative values of β (negative β can be readily accounted for by changing the sign of H, H_g, and V in (11)). Equation (A6) together with (A10) then yields

$$|R_t| \leqslant \int_0^\beta dx \int_0^x dy\, |r_t(x,y)| \leqslant \frac{\beta^2}{2} M_V^2 M_\psi \, \|A\|_{\rm op}, \tag{A11}$$

$$M_V := \max_{y\in[0,\beta]} \|V_y\|_{\rm op}, \tag{A12}$$

$$M_\psi := \max_{y\in[0,\beta]} \|\psi(y)\|_{\rm op}. \tag{A13}$$

Exploiting (24) and (A2), we can conclude that

$$\|\psi'(y)\|_{\rm op} \leqslant g\,\|V_y\|_{\rm op}\|\psi(y)\|_{\rm op}. \tag{A14}$$

To deduce an upper bound for $\|\psi(y)\|_{\rm op}$ from this relation, we proceed similarly as in Ref. [25]: Observing that $|\|B\|_{\rm op} - \|C\|_{\rm op}| \leqslant \|B-C\|_{\rm op}$ for arbitrary B and C [see (A3)], and choosing $B = \psi(y+dy)$ and $C = \psi(y)$, it readily follows that

$$\left|\frac{d\|\psi(y)\|_{\rm op}}{dy}\right| \leqslant \|\psi'(y)\|_{\rm op}. \tag{A15}$$

With (A14) and (A12), this implies for all $y \in [0,\beta]$ that

$$\left|\frac{d\|\psi(y)\|_{\rm op}}{dy}\right| \leqslant g\,\|V_y\|_{\rm op}\|\psi(y)\|_{\rm op} \leqslant g M_V \|\psi(y)\|_{\rm op}. \tag{A16}$$

Focusing on $y \in [0,\beta]$ and observing that the real valued function $\|\psi(y)\|_{\rm op}$ of y is non-negative and that its growth is upper bounded by the right-hand side of (A16), it must be upper bounded by a function $f(y)$ which satisfies $f'(y) = M_V f(y)$ and $f(0) = \|\psi(0)\|_{\rm op}$. Upon integrating this equation and exploiting that $\psi(0) = 1$ according to (20), we thus obtain

$$\|\psi(y)\|_{\rm op} \leqslant e^{ygM_V}. \tag{A17}$$

With (A13), this yields

$$M_\psi \leqslant e^{g\beta M_V}. \tag{A18}$$

In view of (A11), (A12), (A18), the main remaining task is to upper bound the operator norm of $V_y := e^{yH}Ve^{-yH}$ [see Eq. (25)]. Such bounds have been previously obtained for a considerable variety of Hamiltonians H and perturbations V; see, for instance, Refs. [24,26–28] and further references therein. Focusing on Ref. [24], Eq. (4) therein can be rewritten in our present notation as

$$\|V_y\|_{\rm op} = \|e^{yH}Ve^{-yH}\|_{\rm op} \leqslant \|V\|_{\rm op} f(y). \tag{A19}$$

Concerning $f(y)$, let us first consider lattice systems in one dimension with nearest-neighbor interactions and Hamiltonians of the form

$$H = \sum_{l=1}^L h_l. \tag{A20}$$

See Eq. (5) in Ref. [24] and our present Eq. (46) for an explicit example. Similarly as below Eq. (5) and in Eq. (14) of Ref. [24], we employ the definitions

$$\gamma := \max_l \|h_l\|_{\rm op}, \tag{A21}$$

$$q_y := e^{2|y|\gamma} - 1. \tag{A22}$$

The first main result in Ref. [24] then takes the form [see Eq. (19) therein]

$$f(y) = e^{2q_y}. \tag{A23}$$

Analogously, Eq. (39) in Ref. [24] implies the following bound for lattice systems in arbitrary dimensions:

$$f(y) \leqslant \frac{9}{1-2\gamma|y|z}, \tag{A24}$$

where z is the coordination number of the considered lattice, which means that each vertex is attached or adjacent to at most z bonds (see above Eq. (38) and beginning of Appendix B in Ref. [24]). It thus seems that this result, in particular, is not restricted to cases with nearest-neighbor interactions. The main point is that only β values with $\beta \leqslant 1/2\gamma z$ are now admitted. In any case, the Hamiltonian H must be of the form (A20) with local operators h_l (hence H itself is often denoted as a local Hamiltonian).

More precisely speaking, the results in (A22)–(A24) actually only apply if V is a local operator which acts on a single site of the lattice; see below Eq. (5) in Ref. [24]. By linearity, one readily infers that the same results remain valid if V is a sum of single site operators and $\|V\|_{\rm op}$ in (A19) is replaced by the corresponding sum of operator norms. Moreover, for

one-dimensional lattice Hamiltonians (A20) with nearest-neighbor interactions, the bound (A19) still remains valid for operators V which act on N_s consecutive sites (the letter s stands for support), while (A23) must now be generalized as follows [29]:

$$f(y) = \frac{1-q_y^{N_s}}{1-q_y}\, e^{2q_y}. \tag{A25}$$

The generalization to sums of such operators is again obvious. For lattice Hamiltonians (A20) with more general interactions and in more than one dimension, analogous generalizations are expected to be feasible as well [29].

Next, we can conclude from (25), (A12), and (A19) that

$$M_V \leqslant \|V\|_{\mathrm{op}} f(\beta) \tag{A26}$$

and with (A13), (A18) that

$$|R_t| \leqslant \|A\|_{\mathrm{op}}\, M^2 e^{gM}/2, \tag{A27}$$

$$Mf(\beta) := \beta\, f(\beta)\, \|V\|_{\mathrm{op}}. \tag{A28}$$

Finally, we turn to the quantity Q from (29). Obviously, this quantity is recovered by choosing $A = 1$ in (32) and thus in (A27), yielding

$$|Q| \leqslant M^2 e^{gM}/2. \tag{A29}$$

It is instructive to rewrite (A27)-(A29) as

$$g^2|R_t| \leqslant \|A\|_{\mathrm{op}}\, F(\epsilon), \tag{A30}$$

$$g^2|Q| \leqslant F(\epsilon), \tag{A31}$$

$$F(x) := x^2 e^x/2, \tag{A32}$$

$$\epsilon := g\beta f(\beta)\|V\|_{\mathrm{op}}. \tag{A33}$$

We also recall that $g\beta\|V\|_{\mathrm{op}}$ quantifies the perturbation strength in (11) in units of the thermal energy, and that $f(\beta)$ is of order unity at least for small-to-moderate values of β if we work in natural energy units with $J = 1$ in (A21).

The main conclusion is that the right-hand side of (A30) is independent of t. Moreover, g^2R_t in (A30) and g^2Q in (A31) scale asymptotically (at least) quadratically with g, β, and V.

Altogether, (A30)–(A33) thus amount to a rigorous justification of the approximation (34) in the main text (see also discussion at the beginning of Sec. III).

APPENDIX B: DERIVATION OF THE GENERALIZATION FROM SEC. VI

As in Sec. VI, we take for granted that the Hamiltonian H and the perturbation V in (11) obey the relations (56), where P is a projector onto some subspace $\tilde{\mathcal{H}}$.

By exploiting (56) and utilizing the common eigenbasis of H and P, one readily verifies that

$$Pe^{zH} = e^{zH}P = Pe^{zH}P \tag{B1}$$

for any $z \in \mathbb{C}$, and similarly for H_g from (11):

$$Pe^{zH_g} = e^{zH_g}P = Pe^{zH_g}P. \tag{B2}$$

Defining the projected (or reduced) Hamiltonians as

$$\tilde{H} := PHP,\ \tilde{H}_g := PH_gP, \tag{B3}$$

one furthermore can infer that

$$Pe^{zH} = Pe^{z\tilde{H}} = e^{z\tilde{H}}P = Pe^{z\tilde{H}}P, \tag{B4}$$

and likewise for $\tilde{H}_g$. On the other hand, relations such as, for instance, $Pe^{zH}P = e^{zPHP}$ are generally *not* valid ! (This is particularly obvious for $z = 0$.)

Similarly as in (B3), we define the following projected (or reduced) counterparts of the original quantities introduced in Sec. I:

$$\tilde{A} := PAP, \quad \tilde{V} := PVP, \tag{B5}$$

$$\tilde{\rho} := Pe^{-\beta\tilde{H}}P/\operatorname{Tr}\{Pe^{-\beta\tilde{H}}P\}, \tag{B6}$$

$$\tilde{A}_{\mathrm{th}} := \operatorname{Tr}\{\tilde{\rho}\tilde{A}\}, \tag{B7}$$

$$\tilde{A}(t) := Pe^{i\tilde{H}t/\hbar}\tilde{A}e^{-i\tilde{H}t/\hbar}P, \tag{B8}$$

$$\tilde{C}_{VA}(t) := \operatorname{Tr}\{\tilde{\rho}\ \tilde{V}\tilde{A}(t)\} - \tilde{V}_{\mathrm{th}}\tilde{A}_{\mathrm{th}}, \tag{B9}$$

$$\tilde{\rho}_0 := Pe^{-\beta\tilde{H}_g}P/\operatorname{Tr}\{Pe^{-\beta\tilde{H}_g}P\}, \tag{B10}$$

$$\widetilde{\langle A\rangle}_t := \operatorname{Tr}\{\tilde{\rho}_0\tilde{A}(t)\}. \tag{B11}$$

Essentially, everything is thus projected (or reduced) onto the subspace $\tilde{\mathcal{H}}$.

The key point of this Appendix consists of the claim that our main result (16) still remains valid when working from the outset within the subspace $\tilde{\mathcal{H}}$, and assumes the form

$$\widetilde{\langle A\rangle}_t - \tilde{A}_{\mathrm{th}} = g\beta \sum_{k=0}^{\infty} \frac{(i\tau_B)^k}{(k+1)!}\, \tilde{C}_{VA}^{(k)}(t). \tag{B12}$$

To substantiate this claim, we simply must replace all the quantities in Sec. I and Appendix A by their above defined reduced counterparts. The only major problem is that the reduced Hamiltonian $\tilde{H}$ and perturbation $\tilde{V}$ generally no longer exhibit the specific properties of their original counterparts H and V which are required, for instance, around (A20), (A24), and (A25). This problem can be solved by noting that the actual remaining task in our present case is, analogously to the sentence below (A18), to upper bound the operator norm of

$$\tilde{V}_y := Pe^{y\tilde{H}}\tilde{V}e^{-y\tilde{H}}P. \tag{B13}$$

By exploiting (B4) and (B5), it follows that

$$\tilde{V}_y = Pe^{yH}Ve^{-yH}P, \tag{B14}$$

and with (25), (A2) that

$$\|\tilde{V}_y\|_{\mathrm{op}} \leqslant \|V_y\|_{\mathrm{op}}\|P\|_{\mathrm{op}}^2. \tag{B15}$$

Recalling the definition (A1) and the fact that P is a projector, it follows that $\|P\|_{\mathrm{op}} = 1$ and hence

$$\|\tilde{V}_y\|_{\mathrm{op}} \leqslant \|V_y\|_{\mathrm{op}}. \tag{B16}$$

Accordingly, we can again exploit (A19) and all the results for $f(y)$ as detailed below (A19). Altogether, this completes the justification (A12).

Finally, we note that some of the projectors P and tilde symbols in (B6), (B8), (B10) are actually superfluous according to (B1)–(B4). In particular, one readily confirms that

$$\tilde{\rho} = Pe^{-\beta H}/\operatorname{Tr}\{Pe^{-\beta H}\}, \tag{B17}$$

$$\tilde{A}(t) = Pe^{iHt/\hbar}Ae^{-iHt/\hbar}P = PA(t)P, \tag{B18}$$

$$\tilde{\rho}_0 = Pe^{-\beta H_g}/\operatorname{Tr}\{Pe^{-\beta H_g}\}, \tag{B19}$$

where $A(t)$ is defined in (6). Moreover, (B7), (B9), (B11) can be rewritten as

$$\tilde{A}_{\rm th} := \operatorname{Tr}\{\tilde{\rho}A\}. \tag{B20}$$

$$\tilde{C}_{VA}(t) = \operatorname{Tr}\{\tilde{\rho}\,VA(t)\} - \tilde{V}_{\rm th}\tilde{A}_{\rm th}. \tag{B21}$$

$$\widetilde{\langle A\rangle}_t = \operatorname{Tr}\{\tilde{\rho}_0 A(t)\}. \tag{B22}$$

It follows that (B12) assumes the same form as (16) if the canonical ensembles ρ and ρ_g from (1) and (9) are replaced by their counterparts $\tilde{\rho}$ and $\tilde{\rho}_g$ from (B17) and (B19). Analogous modifications of (7), (8), (18), (19), as well as of Onsager's hypothesis (15), are obvious and therefore omitted. Upon dropping the tilde symbols in (B17) and (B19), our verification of the statement above Eq. (58) is thus completed.

[1] L. Onsager, Reciprocal relations in irreversible processes II., Phys. Rev. **37**, 405 (1931); Reciprocal relations in irreversible processes I, **38**, 2265 (1931).

[2] H. Takahasi, Generalized theory of thermal fluctuations, J. Phys. Soc. Jpn. **7**, 439 (1952).

[3] P. Talkner, The failure of the quantum regression hypothesis, Ann. Phys. **167**, 390 (1986).

[4] G. W. Ford and R. F. O'Connell, There is no quantum regression hypothesis, Phys. Rev. Lett. **77**, 798 (1996).

[5] A. Shimizu and K. Fujikura, Quantum violation of fluctuation-dissipation theorem, J. Stat. Mech. (2017) 024004.

[6] We tacitly restrict ourselves to model systems which exhibit a well-defined thermodynamic limit.

[7] Rigorously speaking, one can readily upper and lower bound the quantities $\langle A\rangle_t$ and $A_{\rm th}$ in terms of the operator norm of A, which is independent of the system size. The remaining possibility that these quantities nevertheless do not converge in the thermodynamic limit appears quite unlikely.

[8] An upper bound for $|C_{VA}(t)|$ in (5), which is moreover independent of the system size, can be deduced at least for $t = 0$ from the so-called cluster decomposition property. The latter is known to apply to the canonical ensembles of a large class of system Hamiltonians H in (1). For finite t, similar conclusions may possibly be inferred by invoking Lieb-Robinson bounds. In view of such system size-independent bounds for $|C_{VA}(t)|$, it seems reasonable to expect that the quantity $C_{VA}(t)$ itself asymptotically approaches a well-defined (finite) thermodynamic limit.

[9] In many cases, one such important characteristic timescale is related to the relaxation behavior (often an exponential decay) for sufficiently late times.

[10] P. Reimann, Foundation of statistical mechanics under experimentally realistic conditions, Phys. Rev. Lett. **101**, 190403 (2008); N. Linden, S. Popescu, A. J. Short, and A. Winter, Quantum mechanical evolution towards equilibrium, Phys. Rev. E **79**, 061103 (2009); A. J. Short, Equilibration of quantum systems and subsystems, New J. Phys. **13**, 053009 (2011); P. Reimann and M. Kastner, Equilibration of macroscopic quantum systems, *ibid.* **14**, 043020 (2012); A. J. Short and T. C. Farrelly, Quantum equilibration in finite time, *ibid.* **14**, 013063 (2012); B. N. Balz and P. Reimann, Equilibration of isolated many-body quantum systems with respect to general distinguishability measures, Phys. Rev. E **93**, 062107 (2016).

[11] A. M. Alhambra, J. Riddell, and L. P. Garcia-Pintos, Time evolution of correlation functions in quantum many-body systems, Phys. Rev. Lett. **124**, 110605 (2020).

[12] D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov, Gap generation in the XXZ model in a transverse magnetic field, Phys. Rev. B **65**, 172409 (2002).

[13] E. K. Sklyanin, Boundary conditions for integrable quantum systems, J. Phys. A **21**, 2375 (1988).

[14] F. C. Alcaraz, M. N. Barber, M. T. Batchelor, R. J. Baxter, and G. R. W. Quispel, Surface exponents of the quantum XXZ, Ashkin-Teller and Potts models, J. Phys. A **20**, 6397 (1987).

[15] N. Beisert, L. Fievet, M. de Leeuw, and F. Loebbert, Integrable deformations of the XXZ spin chain, J. Stat. Mech. (2013) P09028.

[16] C. Eidecker-Dunkel and P. Reimann, Allosteric impurity effects in long spin chains, Phys. Rev. B **108**, 054407 (2023)

[17] R. Roth and U. Schollwöck, Frustrated antiferromagnetic quantum spin chains for spin length $S > 1$, Phys. Rev. B **58**, 9264 (1998).

[18] J. Sudan, A. Lüscher, and A. M. Läuchli, Emergent multipolar spin correlations in a fluctuating spiral: The frustrated ferromagnetic spin-1/2 Heisenberg chain in a magnetic field, Phys. Rev. B **80**, 140402(R) (2009).

[19] A. Pal and D, A. Huse, Many-body localization phase transition, Phys. Rev. B **82**, 174411 (2010).

[20] R. Nandkishore and D. A. Huse, Many-body localization and thermalization in quantum statistical mechanics, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).

[21] M. V. S. Bonança, Relaxation in finite and isolated classical systems: An extension of Onsager's regression hypothesis, Phys. Rev. E **85**, 031122 (2012).

[22] T. A. Elsayed and B. V. Fine, Regression relation for pure quantum states and its implications for efficient computing, Phys. Rev. Lett. **110**, 070404 (2013).

[23] N. G. van Kampen, The case against linear response theory, Phys. Norv. **5**, 279 (1971).

[24] A. Avdoshkin and A. Dymarsky, Euclidean operator growth and quantum chaos, Phys. Rev. Res. **2**, 043234 (2020).
[25] L. Dabelow, P. Vorndamme, and P. Reimann, Thermalization of locally perturbed many-body quantum systems, Phys. Rev. B **105**, 024310 (2022).
[26] D. A. Abanin, W. De Roeck, and F. Huveneers, Exponentially slow heating in periodically driven many-body systems, Phys. Rev. Lett. **115**, 256803 (2015).
[27] I. Arad, T. Kuwahara, and Z. Landau, Connecting global and local energy distributions in quantum spin models an a lattice, J. Stat. Mech. (2016) 033301.
[28] T. R. de Oliveira, C. Charalambous, D. Jonathan, M. Lewenstein, and A. Riera, Equilibration time scales in closed many-body quantum systems, New J. Phys. **20**, 033032 (2018).
[29] A. Dymarsky (private communication).