

## Automatic differentiable numerical renormalization group

Jonas B. Rigo<sup>\*</sup> and Andrew K. Mitchell<sup>†</sup>*School of Physics, University College Dublin, Belfield, Dublin 4, Ireland and Centre for Quantum Engineering, Science, and Technology, University College Dublin, Belfield, Dublin 4, Ireland*

(Received 31 August 2021; accepted 24 February 2022; published 25 March 2022)

Machine learning techniques have recently gained prominence in physics, yielding a host of new results and insights. One key concept is that of backpropagation, which computes the exact gradient of any output of a program with respect to any input. This is achieved efficiently within the differentiable programming paradigm, which utilizes automatic differentiation (AD) of each step of a computer program and the chain rule. A classic application is in training neural networks. Here, we apply this methodology instead to the numerical renormalization group (NRG), a powerful technique in computational quantum many-body physics. We demonstrate how derivatives of NRG outputs with respect to Hamiltonian parameters can be accurately and efficiently obtained. Physical properties can be calculated using this differentiable NRG scheme—for example, thermodynamic observables from derivatives of the free energy. Susceptibilities can be computed by adding source terms to the Hamiltonian, but still evaluated with AD at precisely zero field. As an outlook, we briefly discuss the derivatives of dynamical quantities and a possible route to the vertex.

DOI: [10.1103/PhysRevResearch.4.013227](https://doi.org/10.1103/PhysRevResearch.4.013227)

## I. INTRODUCTION

The numerical renormalization group (NRG) [1,2] is a standard tool in computational physics for solving a certain class of quantum many-body problem, known as *quantum impurity models*. These comprise a few interacting quantum degrees of freedom, coupled to one or more noninteracting baths. The most famous example is the Kondo model, which involves a single spin- $\frac{1}{2}$  “impurity” coupled to a single fermionic bath [3]. The low-energy physics is nonperturbative and controlled by a single emergent energy scale [the Kondo temperature ( $T_K$ )], below which the impurity is dynamically screened by a spatially extended entanglement “Kondo cloud” of surrounding conduction electrons [4]. Generalized quantum impurity models describe the scattering from magnetic impurities in metals [3,5], semiconductor quantum dot devices [6–9], and single-molecule transistors [10,11] and are the local effective models within dynamical mean field theory (DMFT) for correlated materials [12–14].

NRG is a tensor network method [15] that exploits the renormalization group structure of such quantum impurity models [1]. At its core, the method involves the iterative numerical diagonalization of a set of renormalized effective Hamiltonians. There is no statistical element, as with quantum Monte Carlo impurity solvers [16].

In this paper, we apply the “differentiable programming” methodology to NRG and demonstrate some advantages of its

use for practical physics applications. We show that the core eigensolver routine in NRG is well suited to automatic differentiation (AD), an emerging programming paradigm that has its origin in deep learning (DL) [17]. In DL, hugely complicated neural networks (NNs) feed their output into a loss function, which must be optimized with respect to the NN parameters (weights). This can be accomplished efficiently via gradient descent using AD programming: The gradient of the loss function with respect to the NN weights can be computed at about the cost of the evaluation of the NN [18]. The remarkable aspect of this technique is that the computed gradient is not approximated, but is numerically exact. Without AD, derivatives can of course be approximated using finite-difference (FD) derivatives. However, without care, FD derivatives can be inaccurate; when treated rigorously to guarantee convergence, FD can become computationally expensive. The efficiency and accuracy of AD are owed to the fact that if each step of a computer program can be differentiated, then outputs can be differentiated with respect to inputs via the chain rule. Evaluating derivatives is trivial and cheap since the program itself is differentiated.

AD has been made more accessible through the introduction of AD libraries such as JAX [19], TENSORFLOW [20], or PYTORCH [21]. The advantage for computational physics is that when formulated using an AD library, exact derivatives can be obtained “for free” [22–28]. Recent applications in physics include optimal control in quantum systems [29], mitigating the sign problem in Monte Carlo [30], and optimizing tensor networks [31]. In addition to its use for optimization problems, AD has particular appeal in physics because physical quantities are related to derivatives of generating functionals [32].

Here, we formulate a differentiable NRG ( $\partial$ NRG) scheme. The key step is an explicit expression for the derivative of NRG eigenvalues and eigenvectors: The exact gradient

<sup>\*</sup>jonas.rigo@ucdconnect.ie<sup>†</sup>andrew.mitchell@ucd.ie

of the entire eigensolver routine at any given NRG step is determined symbolically. This eliminates the need for full backpropagation and, when combined with AD for the other program elements, provides highly efficient access to derivatives of program outputs with respect to model parameters. For example, thermodynamic observables may be obtained from derivatives of the free energy, and the physical response of a system to perturbations can be simply and cheaply determined.

## II. FORWARD- AND BACKWARD-MODE AD

The AD approach is similar to symbolic differentiation, with standard derivative rules applied algorithmically [17,33]. Regarding a program as a function  $f$  that maps a given input  $x$  to an output  $y$ , we may write  $f(x) = y$ . This is typically achieved in practice by concatenating several primitive functions  $f^{(i)}$ , viz.,

$$f = f^{(n)} \circ f^{(n-1)} \circ \dots \circ f^{(1)}, \quad (1)$$

where “ $\circ$ ” denotes the composition of two functions. We denote the number primitive functions comprising  $f$  as  $\text{ops}(f)$ . The flow of data between primitive functions can be expressed as a compute graph [34]; for simplicity, we consider here only programs whose graph has a chain topology as in Eq. (1). To compute the derivative of  $f$  with respect to  $x$ , one can apply the chain rule,

$$\frac{\partial f}{\partial x} = \frac{\partial f^{(n)}}{\partial f^{(n-1)}} \frac{\partial f^{(n-1)}}{\partial f^{(n-2)}} \dots \frac{\partial f^{(1)}}{\partial x}. \quad (2)$$

Equation (2) can be evaluated algorithmically in either the *forward mode* or the *backward mode*. In the forward mode, we apply the recursion

$$\frac{\partial f^{(i)}}{\partial x} = \frac{\partial f^{(i)}}{\partial f^{(i-1)}} \delta_{i-1} \equiv \delta_i, \quad \text{with } \delta_0 = 1, f^{(0)} = x, \quad (3)$$

starting with  $i = 1$  and increasing up to  $i = n$ , with  $f^{(n)} \equiv f$  such that  $\delta_n = \partial f / \partial x$  is the desired derivative. In backward mode (also known as *backpropagation*) we use the recursion

$$\frac{\partial f}{\partial f^{(i)}} = \frac{\partial f^{(i+1)}}{\partial f^{(i)}} \bar{\delta}_{i+1} \equiv \bar{\delta}_i, \quad \text{with } \bar{\delta}_n = 1, f^{(0)} = x, \quad (4)$$

starting from  $i = n - 1$  and decreasing down to  $i = 0$ , with  $\bar{\delta}_0 = \partial f / \partial x$  being the desired derivative.

These forward- and backward-mode AD recursions are symbolically equivalent methods for calculating derivatives. In the practical implementation of AD the intermediate derivatives  $\delta_i$  or  $\bar{\delta}_i$  are evaluated numerically and stored in memory. This provides a way to compute numerically exact derivatives, with the true analytical result recovered when the value of  $\delta_i$  is not truncated [35].

By contrast, the FD approximation reads

$$\frac{\partial f}{\partial x} \approx \frac{f(x+h) - f(x)}{h} \equiv D_h[f](x), \quad (5)$$

with  $h$  being the FD value. The exact limit  $h \rightarrow 0^\pm$  cannot be taken numerically, and so  $\partial_x f$  must be approximated for finite  $h$  and the convergence  $D_h[f](x) \rightarrow \partial_x f$  checked explicitly. This may require many function evaluations (runs of the whole program) since for very small  $h$ ,  $f(x)$  and  $f(x+h)$  may be numerically indistinguishable (truncation error). Furthermore, convergence should be checked for both the

forwards ( $h > 0$ ) and backwards ( $h < 0$ ) difference quotient. The situation gets worse for higher-order derivatives, and numerical convergence of FD approximations can be costly and/or fraught [36].

Above we considered scalar functions of the form  $f : \mathbb{R} \rightarrow \mathbb{R}$ . However, both AD and FD approaches can be straightforwardly generalized to compute the Jacobian of vector-valued functions defined over a vector space  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . The FD and forward- or backward-mode AD scale differently with respect to the input dimension  $n$  and output dimension  $m$ . The FD exhibits the most expensive scaling of computational cost with  $\sim 2nm$  function evaluations required to compute the Jacobian of  $f$  in the single-shot case. By contrast, AD methods do not require the function to be evaluated in order to compute the derivative, as with the symbolic approach (although since many derivatives are related to their antiderivative, function evaluations can be reused to reduce computational cost at the expense of increased memory cost [37]). Instead, the computational cost of AD methods is controlled by  $\text{ops}(f)$ . Specifically, the forward mode scales as  $\sim c_F n \text{ops}(f)$ , while the backward mode scales as  $\sim c_B m \text{ops}(f)$ , with  $c_F, c_B < 6$  (although the forward mode typically outperforms the backward mode unless  $m \ll n$  [37]). The computational cost of the FD approach is therefore typically much higher than either of the AD modes; in this paper we use forward-mode AD unless stated otherwise. A comparison of the performance and precision of the different methods, as applied to NRG, is presented in Appendix A.

Through AD one can obtain the numerically exact derivative of any program with a single run, as long as all operations comprising the program are themselves differentiable. Thus one can differentiate a given physics solver program simply by utilizing known derivatives of its constituent operations.

On the other hand, in certain cases it may be possible to find the symbolic derivative of the entire solver with respect to certain input parameters. In the following we show how the latter can be achieved for NRG through analysis of the backpropagation chain.

## III. NRG

Wilson’s NRG is a numerical solver for quantum impurity models, of the type  $\hat{H} = \hat{H}_{\text{imp}} + \hat{H}_{\text{bath}} + \hat{H}_{\text{hyb}}$ , where  $\hat{H}_{\text{imp}}$  is the “impurity” Hamiltonian describing a few local, interacting quantum degrees of freedom, while  $\hat{H}_{\text{bath}} = \sum_{\sigma, \mathbf{k}} \epsilon_{\mathbf{k}} \hat{c}_{\sigma \mathbf{k}}^\dagger \hat{c}_{\sigma \mathbf{k}}$  describes a noninteracting fermionic bath. The coupling between them is given by  $\hat{H}_{\text{hyb}} = \sum_{\sigma, \mathbf{k}} V_{\mathbf{k}} (\hat{d}_\sigma^\dagger \hat{c}_{\sigma \mathbf{k}} + \hat{c}_{\sigma \mathbf{k}}^\dagger \hat{d}_\sigma)$ , where we have assumed for simplicity here that a single impurity orbital  $\hat{d}_\sigma$  couples to the bath.

Due to the impurity interactions, such a model is a genuine quantum many-body problem and in general has no exact solution [3]. NRG treats instead a discretized version of the model, which becomes computationally tractable through a process of iterative diagonalization and truncation. This constitutes a renormalization group (RG) procedure, in which useful physical information can be extracted at each step, as progressively lower energy scales are probed. Full details can be found in Refs. [1,2]; here, we introduce only the elements necessary to formulate  $\partial$ NRG.

The first step is the logarithmic discretization of the bath and mapping to a Wilson chain. The continuous local density of states of the uncoupled bath at the impurity position is divided up into intervals of decreasing width, defined by the points  $x_n^\pm = \pm D\Lambda^{-n}$ , where  $D$  is the half bandwidth,  $\Lambda > 1$  is the discretization parameter, and  $n = 0, 1, 2, \dots$ . The spectrum is discretized by replacing the continuous density in each interval by a single pole of the same weight and at the average position. A semi-infinite tight-binding chain (the Wilson chain) is then defined such that the spectral function at the end is the same as the discretized bath spectrum (this is achieved in practice by the Lanczos algorithm). The discretized bath Hamiltonian reads

$$\hat{H}_{\text{bath}} \mapsto \hat{H}_{\text{bath}}^{\text{disc}} = \sum_{\sigma} \sum_{n=0}^{\infty} t_n (\hat{c}_{\sigma n+1}^\dagger \hat{c}_{\sigma n} + \hat{c}_{\sigma n}^\dagger \hat{c}_{\sigma n+1}), \quad (6)$$

where we have assumed particle-hole symmetry here for simplicity. The original model is recovered as  $\Lambda \rightarrow 1$ . The specific form of the Wilson chain hopping parameters  $t_n$  depends on details of the dispersion relation  $\epsilon_{\mathbf{k}}$ . However, for a metallic system,  $t_n \sim D\Lambda^{-n/2}$  for large  $n$ .

For a Wilson chain of total length  $N$ , adding an extra site  $N+1$  is therefore always a small perturbation. The NRG exploits this energy scale separation down the chain through the iterative diagonalization scheme. Starting from the impurity, the chain is built up by successively adding new Wilson chain sites. At each step, the system is diagonalized, and high-energy states are discarded. One defines a sequence of rescaled Hamiltonians, comprising the impurity and the first  $N$  sites of the Wilson chain,

$$\hat{H}_N = \Lambda^{(N-1)/2} \left[ \hat{H}_{\text{imp}} + V \sum_{\sigma} (\hat{d}_{\sigma}^\dagger \hat{c}_{\sigma 0} + \hat{c}_{\sigma 0}^\dagger \hat{d}_{\sigma}) + \sum_{\sigma} \sum_{n=0}^{N-1} t_n (\hat{c}_{\sigma n+1}^\dagger \hat{c}_{\sigma n} + \hat{c}_{\sigma n}^\dagger \hat{c}_{\sigma n+1}) \right], \quad (7)$$

with  $V$  being the effective impurity-bath coupling. The full (discretized) Hamiltonian is recovered in the limit  $\hat{H} = \lim_{N \rightarrow \infty} \Lambda^{-(N-1)/2} \hat{H}_N$ . The sequence of Hamiltonians  $\hat{H}_N$  satisfy the recursion relation

$$\hat{H}_{N+1} = \sqrt{\Lambda} \hat{H}_N + \Lambda^{N/2} \sum_{\sigma} t_N (\hat{c}_{\sigma N+1}^\dagger \hat{c}_{\sigma N} + \hat{c}_{\sigma N}^\dagger \hat{c}_{\sigma N+1}). \quad (8)$$

At each step  $N$ , the Hamiltonian  $\hat{H}_N$  is diagonalized to find the eigenvectors (states)  $\{|i\rangle_N\}$  and eigenvalues (energies)  $\{E_{N;i}\}$  that satisfy the Schrödinger equation

$$\hat{H}_N |i\rangle_N = E_{N;i} |i\rangle_N. \quad (9)$$

We denote excitation energies relative to the ground state as  $\Delta E_{N;i} = E_{N;i} - E_{N;0}$ , and  $\hat{H}_{\text{imp}} \equiv \hat{H}_{-1}$ . To construct  $\hat{H}_{N+1}$ , we add another Wilson chain site. The Fock space of  $\hat{H}_{N+1}$  is spanned by basis states,

$$|i; s\rangle_{N+1} = |i\rangle_N \otimes |s\rangle, \quad (10)$$

comprising the tensor product of eigenstates  $|i\rangle_N$  of  $\hat{H}_N$ , and the added chain site  $N+1$  denoted as  $|s\rangle$ . Matrix elements of  $\hat{H}_{N+1}$  in this basis read

$$\hat{H}_{N+1}(is; i's') = {}_N \langle i; s | \hat{H}_{N+1} | i'; s' \rangle_{N+1}. \quad (11)$$

Finally,  $\hat{H}_{N+1}$  is diagonalized to find the new eigenbasis at step  $N+1$ , viz.,

$$|j\rangle_{N+1} = \sum_{is} U_{N+1}(j, is) |i; s\rangle_{N+1}. \quad (12)$$

Equation (11) can be simplified using the energies  $E_{N;i}$  of  $\hat{H}_N$  and the tensor  $\eta_{ss'\sigma} = \langle s' | \hat{c}_{\sigma N+1} | s \rangle$  (which is independent of  $N$ ), viz.,

$$\hat{H}_{N+1}(is; i's') = \sqrt{\Lambda} \delta_{ss'} \delta_{ii'} \times E_{N;i} + (-1)^s \Lambda^{N/2} t_N \sum_{\sigma} \eta_{ss'\sigma} \eta_{N;i i'\sigma} + \text{H.c.}, \quad (13)$$

where  $\eta_{N;j j'\sigma} = \sum_{\bar{i}\bar{s}, \bar{i}'\bar{s}'} U_N^\dagger(j, \bar{i}\bar{s}) U_N(j', \bar{i}'\bar{s}') \times \delta_{\bar{i}\bar{i}'} \eta_{\bar{s}\bar{s}'\sigma}$  and  $s = \{0, -1, +1, 2\}$  for  $|s\rangle = \{|0\rangle, |\downarrow\rangle, |\uparrow\rangle, |\uparrow\downarrow\rangle\}$ .

When following these steps, the dimension of the Fock space grows exponentially with the length of the chain. This is avoided in NRG by truncating the Fock space at each step, discarding high-energy states. In practice, one retains  $M_K$  of the lowest-energy eigenstates of  $\hat{H}_N$  at each step. The *NRG approximation* [1,38] is that the states and energies of  $\hat{H}_N$  approximate those of the full  $\hat{H}$ . This approximation is justified by the exponential decay of the Wilson chain coefficients  $t_n$ , which means that high-energy states discarded at a given iteration do not become low-lying states at a later iteration. Convergence of the NRG calculation can be checked *post hoc* by increasing  $M_K$ . Hereinafter it is to be understood that the condition  $i \leq M_K$  applies to Eq. (10) (with the eigenvalues sorted from lowest to highest). The dimensionality of the Fock space is therefore constant at each step, and the NRG calculation scales *linearly* with chain length  $N$ . With increasing  $N$  the physics on successively lower energy scales is uncovered.

The entire process of going from one iteration to the next can be summarized as an RG transformation,

$$\hat{H}_{N+1} = R[\hat{H}_N]. \quad (14)$$

The full (discretized, renormalized) Hamiltonian  $\hat{H}_N$  can therefore be constructed iteratively starting from the impurity Hamiltonian  $\hat{H}_{\text{imp}}$ . Regarding the latter as a function of  $n$  physical model parameters  $\{\theta\}$ , the entire NRG algorithm can be viewed as a function,

$$f(\{\theta\}) \equiv R^{\circ N+1}[\hat{H}_{\text{imp}}(\{\theta\})], \quad f: \mathbb{R}^n \rightarrow \mathbb{C}^{M_K \times M_K}, \quad (15)$$

where we have used the shorthand notation for an  $(N+1)$ -fold concatenation of functions  $R^{\circ N+1} = R \circ R \circ \dots \circ R$ . All physical quantities of interest for the original quantum impurity model are obtained from the eigenvalues and eigenvectors of the set of  $\{\hat{H}_N\}$  [1,2,38].

Note that an operator  $\hat{O}$ , acting only on impurity degrees of freedom, can be expressed in the eigenbasis of  $\hat{H}_N$ . We denote this matrix as  $\hat{O}_N(i; i') = {}_N \langle i | \hat{O} | i' \rangle_N$ . From Eq. (12) it follows that

$$\begin{aligned} \hat{O}_{N+1}(j; j') &= \sum_{is, i's'} U_{N+1}^\dagger(j, is) U_{N+1}(j', i's') \times \hat{O}_{N+1}(is; i's') \\ &= \sum_{is, i's'} U_{N+1}^\dagger(j, is) U_{N+1}(j', i's') \times \hat{O}_N(i; i') \delta_{s,s'}, \end{aligned} \quad (16)$$

where the primed sum implies that  $i, i' \leq M_K$ , and the second line follows from the fact that  $\hat{O}$  does not act on states  $|s\rangle$  of site  $N + 1$ . Equation (16) can be used iteratively to transform  $\hat{O}_{-1}(i; i')$ , evaluated explicitly in the impurity basis, into the eigenbasis of any later iteration.

Symbolically, we denote this process

$$\hat{O}_{N+1} = P[\hat{O}_N], \tag{17}$$

which we refer to as *propagating forward* the operator.

#### IV. DERIVATIVE OF EIGENVALUES AND EIGENVECTORS IN NRG

As outlined in the previous section, NRG outputs the eigenvalues and eigenvectors for the set of  $\hat{H}_N$ . Since all physical quantities for the quantum impurity model are obtained from these, calculating their derivatives using Eq. (2) requires the derivatives of the eigenvalues and eigenvectors of  $\hat{H}_N$ . From Eq. (15), we regard NRG as a function defined over an  $n$ -dimensional domain, with an  $M_K \times M_K$ -dimensional image. Since  $n$  denotes the number of model parameters in  $\hat{H}_{\text{imp}}$  and  $M_K \times M_K$  is the Fock space dimension of  $\hat{H}_N$ , we have  $M_K \times M_K \gg n$ , and therefore forward-mode AD is significantly faster [35]. In the following we consider only forward-mode AD.

For a Hermitian matrix  $\hat{A}$  (such as the Hamiltonian  $\hat{H}_N$ ) with nondegenerate eigenvectors  $|i\rangle$  and distinct eigenvalues  $\lambda_i$  satisfying  $\hat{A}|i\rangle = \lambda_i|i\rangle$ , we may express the differentials through [39,40]

$$d\lambda_i = \langle i | d\hat{A} | i \rangle, \tag{18a}$$

$$d|i\rangle = \sum_{i \neq j} \frac{\langle j | d\hat{A} | i \rangle}{\lambda_i - \lambda_j} |j\rangle. \tag{18b}$$

These expressions are of course well known in the context of nondegenerate perturbation theory [41].

However, the assumption implicit in Eq. (18b) that  $\hat{A}$  must be free of eigenvalue degeneracies is rather restrictive. Methods treating the general degenerate case are much more computationally involved and may suffer from numerical instabilities [40,42]. Of course in physics, where  $\hat{A}$  is the Hamiltonian of some system, energy eigenvalue degeneracies are common. This may be because of underlying non-Abelian symmetries [for example, SU(2) spin symmetry], which endow the spectrum of the Hamiltonian with a multiplet structure. In this case, the Hamiltonian becomes block diagonal in the associated conserved quantum numbers, and the diagonalization equation (12) can be performed separately in each block. Alternatively, NRG can be formulated directly in multiplet space by using the Wigner-Eckart theorem [43]. Either approach removes the problem of such degeneracies in Eq. (18b). However, this may not fully solve the problem, since there could be accidental degeneracies or emergent symmetries [1,3,7] that lead to additional energy eigenvalue degeneracies in physical systems.

To overcome this, we make a simple approximation. To the diagonal entries of  $\hat{H}_N$  we add noise, with random variables drawn from a normal distribution of width  $\sigma$ . The noise lifts the eigenvalue degeneracy, meaning that derivatives can be obtained using Eqs. (18a) and (18b). Note that the smallest

terms in the rescaled Hamiltonian  $\hat{H}_N$  are  $O(1)$  [1,2], so  $\sigma \ll 1$  constitutes a small perturbation. Care should be taken to add the noise in such a way as to respect bare symmetries.

With a straight application of AD via Eq. (3), the derivatives of eigenvalues and eigenvectors for  $\hat{H}_N$  require the evaluation of Eqs. (18a) and (18b) at every NRG step (and hence noise must be added at every NRG step). For small enough noise width  $\sigma$ , physical properties at a given iteration should be unaffected. Indeed, in Appendix A we show for the Anderson impurity model (AIM) that NRG remains highly accurate when  $\sigma \lesssim 10^{-6}$ . At larger noise levels, errors may propagate through the iterative process (and will snowball if they introduce RG-relevant perturbations), so care must be taken to avoid this with such an AD approach. On the other hand, forward-mode AD for the AIM requires  $\sigma \gtrsim 10^{-6}$  to stabilize the calculation of derivatives using Eqs. (18a) and (18b). Smaller noise levels introduce more severe numerical instabilities because of the recursive nature of the derivative calculation in AD via Eq. (3): Derivatives at one iteration depend on those of previous iterations. This stability-accuracy trade-off in AD is analyzed in Appendix A.

Below, we derive an alternative formulation used in  $\partial$ NRG, which utilizes the symbolic derivative of the entire eigensolver for  $\hat{H}_N$ . This has the advantage that Eqs. (18a) and (18b) need only be applied once, at the NRG step for which derivatives are required. Not only is this far less computationally costly, but also it avoids accumulating errors introduced by noise, which can then be added at a much lower level.

#### Formulation in $\partial$ NRG

In the context of NRG, one can formulate a differential recursion relation based on the RG transformation equation (14),

$$d\hat{H}_N = d\hat{H}_{N-1} \frac{dR(\hat{H}_{N-1})}{d\hat{H}_{N-1}}. \tag{19}$$

This allows derivatives of  $\hat{H}_N$  to be computed in AD forward or backward mode. However, one can also reformulate the problem in a much more simple and efficient way. Taking derivatives with respect to model parameters  $\theta$  of  $\hat{H}_{\text{imp}}$ , we may write from Eq. (7) the operator identity  $\partial_\theta \hat{H}_N \simeq \Lambda^{(N-1)/2} \partial_\theta \hat{H}_{\text{imp}}$  (where  $\simeq$  follows from the NRG approximation). Since  $\partial_\theta \hat{H}_{\text{imp}}$  only involves impurity operators, it can be trivially evaluated in the impurity basis,  $[\partial_\theta \hat{H}_{\text{imp}}]_{-1}(k; k') = -_1 \langle k | \partial_\theta \hat{H}_{\text{imp}} | k' \rangle_{-1}$ . Equation (17) can then be used to propagate this operator forward into the basis of  $\hat{H}_N$ . It therefore follows that

$$[\partial_\theta \hat{H}_N]_N \simeq \Lambda^{(N-1)/2} P^{\circ N+1} ([\partial_\theta \hat{H}_{\text{imp}}]_{-1}). \tag{20}$$

From Eqs. (18a) and (18b) we then obtain

$$\partial_\theta E_{N;i} \simeq \Lambda^{(N-1)/2} P^{\circ N+1} ([\partial_\theta \hat{H}_{\text{imp}}]_{-1})(i; i), \tag{21a}$$

$$\partial_\theta |i\rangle_N \simeq \Lambda^{(N-1)/2} \sum_{i \neq j} \frac{P^{\circ N+1} ([\partial_\theta \hat{H}_{\text{imp}}]_{-1})(j; i)}{E_{N;i} - E_{N;j}} |j\rangle_N. \tag{21b}$$

Equations (21a) and (21b) are the main result of this work. These equations show how the derivative of energies and

states of the NRG Hamiltonians  $\hat{H}_N$  can be obtained by simply forward-propagating the impurity operators  $\partial_\theta \hat{H}_{\text{imp}}$ . Derivatives of physical quantities can then be related to those obtained via Eqs. (21a) and (21b). The above can also be generalized to deal with derivatives  $\partial_\phi$  of other Hamiltonian parameters  $\phi$  [for example,  $V$  or  $t_n$  in Eq. (7)] and higher-order derivatives (see Appendix B).

To calculate first-order derivatives of eigenvalues and eigenvectors in AD forward mode (assuming the simplest NRG implementation), one must perform seven additional operations for every NRG step. The  $\partial$ NRG approach involves only one additional operation per NRG step and does not require any propagation of derivatives related to the chain rule (see Appendix C). Hence  $\partial$ NRG, utilizing Eqs. (21a) and (21b), is far more efficient than standard forward-mode AD. Performance benchmarking is demonstrated in Appendix A.

Derivatives with respect to multiple parameters can be obtained straightforwardly by propagating forward any  $[\partial_\theta \hat{H}_{\text{imp}}]_{-1}$  of interest using Eq. (16) in  $\partial$ NRG. Therefore  $\partial$ NRG scales linearly with the number of derivatives  $n$  but is independent of  $\text{ops}(f)$ , unlike forward-mode AD.

In practice, the use of Eq. (21b) again requires lifting eigenvalue degeneracies by adding a small diagonal noise term. We found in numerical tests for the AIM that  $\sigma \simeq 10^{-11}$  is sufficient for stabilizing the numerics in  $\partial$ NRG, without noticeably impacting any measurable physical properties; see Appendix A.

If derivatives at only step  $N$  are required, such a noise term need only be introduced at that step (rather than adding noise at each step), and the RG flow is unaffected. This is typically the situation for ground-state properties or low-temperature thermodynamics. On the other hand, if information is required from every NRG step (for example, in the calculation of dynamical quantities [38]), noise must be introduced at each iteration. Unlike with the straight AD implementation,  $\partial$ NRG allows us to mitigate the possibility of snowballing errors introduced by propagating the noise terms—if this level of accuracy should be required (e.g., in the vicinity of a quantum critical point). This is because the eigensystem derivatives at different  $N$  via Eq. (21b) are independent in  $\partial$ NRG. Noise may be added to  $\hat{H}_N$  for the purpose of evaluating Eq. (21b), but the pristine  $\hat{H}_N$  without noise can be used for the main NRG recursion. This gives the most accurate and reliable results, at the cost of an additional matrix diagonalization. See Appendixes A and C for performance comparisons (and note that this more rigorous approach is still faster than FD and straight AD in many circumstances, and certainly more accurate).

## V. APPLICATION TO ANDERSON IMPURITY MODEL

We illustrate the use of Eqs. (21a) and (21b) by applying the  $\partial$ NRG scheme to the paradigmatic Anderson impurity model (AIM), for which

$$\hat{H}_{\text{imp}} = \epsilon \hat{n} + U \hat{n}_\uparrow \hat{n}_\downarrow, \quad (22)$$

where  $\hat{n} = \sum_\sigma \hat{n}_\sigma$  and  $\hat{n}_\sigma = \hat{d}_\sigma^\dagger \hat{d}_\sigma$ . With  $\hat{H}_{\text{imp}}$  so defined, the exact (discretized)  $\hat{H}_N$  is given by Eq. (7). In NRG,  $\hat{H}_N$  is

approximated through the RG procedure equation (14),  $\hat{H}_N = R^{\circ N+1}(\hat{H}_{\text{imp}})$ .

For a given  $\hat{H}_N$ , we obtain the partition function  $\mathcal{Z}_N(\bar{\beta}) = \sum_i e^{-\bar{\beta} E_{N;i}}$  and free energy  $F_N(\bar{\beta}) = -\frac{1}{\bar{\beta}} \ln[\mathcal{Z}_N(\bar{\beta})]$ . In the original Wilsonian formulation [1,2], the effective inverse temperature  $\beta \equiv 1/k_B T$  is related to the NRG iteration number (Wilson chain length)  $N$  via  $\beta = \Lambda^{(N-1)/2} \bar{\beta}$ , with  $\bar{\beta} = O(1)$  in practice. With this definition, the NRG free energy at inverse temperature  $\bar{\beta}$  is a good approximation to the true free energy of the original (undiscretized) model at inverse temperature  $\beta$  [1,2],  $F(\beta) \simeq F_N(\bar{\beta})$ . The corresponding differential follows as

$$dF_N = \frac{\Lambda^{-(N-1)/2}}{\mathcal{Z}_N} \sum_i e^{-\bar{\beta} E_{N;i}} dE_{N;i}. \quad (23)$$

For the AIM, we may use Eq. (21a) to obtain derivatives of the free energy with respect to the impurity parameters  $\epsilon$  and  $U$ . For example,

$$\partial_\epsilon F_N = \frac{1}{\mathcal{Z}_N} \sum_i e^{-\bar{\beta} E_{N;i}} P^{\circ N+1}([\hat{n}]_{-1})(i; i). \quad (24)$$

Since  $\partial_\epsilon F_N(\bar{\beta}) = \langle \hat{n} \rangle_{H_N, \bar{\beta}} \simeq \langle \hat{n} \rangle_{H, \beta}$ , Eq. (24) is precisely equivalent to the standard Wilsonian approach to calculating local thermodynamic expectation values in NRG [2]. The above illustration demonstrates that  $\partial$ NRG is analytically equivalent to the well-known result for such local thermodynamic quantities.

Another commonly computed quantity for such models is the *local* impurity magnetic susceptibility at zero field,

$$\chi(T) = \int_0^\beta d\tau \langle \hat{S}_z(\tau) \hat{S}_z \rangle - \beta \langle \hat{S}_z \rangle^2,$$

where  $\hat{S}_z = \frac{1}{2}(\hat{n}_\uparrow - \hat{n}_\downarrow)$  is the impurity spin operator and  $\tau$  is imaginary time. This can be alternatively obtained by adding a source term  $B\hat{S}_z$  to the Hamiltonian equation (22) and then taking the second-order derivative of the free energy with respect to  $B$ , evaluated at  $B = 0$ . That is, we can write  $\chi(T) = \frac{\partial^2}{\partial B^2} F|_{B=0}$ . Since  $\partial$ NRG is able to deal with derivatives of second (and higher) order (see Appendix B), we calculate  $\chi_{H_N, \bar{\beta}} = \frac{\partial^2}{\partial B^2} F_N(\bar{\beta})|_{B=0}$ . This is obtained automatically in  $\partial$ NRG, but from Eq. (B2b) it can also be expressed as

$$\begin{aligned} \chi_{H_N, \bar{\beta}} = & -\frac{\bar{\beta}}{\mathcal{Z}_N^2} \left( \sum_i e^{-\bar{\beta} E_{i;N}} P^{\circ N+1}([\hat{S}_z]_{-1})(i; i) \right)^2 \\ & + \frac{\bar{\beta}}{\mathcal{Z}_N} \sum_i e^{-\bar{\beta} E_{i;N}} |P^{\circ N+1}([\hat{S}_z]_{-1})(i; i)|^2 \\ & - \frac{2}{\mathcal{Z}_N} \sum_i e^{-\bar{\beta} E_{i;N}} \sum_{i \neq j} \frac{|P^{\circ N+1}([\hat{S}_z]_{-1})(j; i)|^2}{E_{N;i} - E_{N;j}}. \quad (25) \end{aligned}$$

This form of  $\chi_{H_N, \bar{\beta}}$  is equivalent to the Lehmann representation of  $\chi(T)$  evaluated in NRG Hamiltonian  $\hat{H}_N$  at effective inverse temperature  $\bar{\beta} = \beta \Lambda^{-(N-1)/2}$  [2]. Since all degeneracies are lifted in  $\partial$ NRG, convergence factors typically used in the Lehmann representation of dynamical quantities are not needed here. Note that  $\chi_{H_N, \bar{\beta}}$  can be obtained at exactly zero

magnetic field, so no symmetries are broken and no limit is taken numerically. Since the derivatives of primitive program functions are symbolic, source terms added to  $\hat{H}_{\text{imp}}$  may be evaluated at zero coupling constant and still yield finite derivatives. However, we note that such symmetry-breaking source terms cannot be added to the Hamiltonian if non-Abelian quantum numbers are implemented from the outset. In the case of the magnetic susceptibility, we can therefore utilize conserved total  $S_z$ , but not conserved total  $S$ , when adding the source term  $B\hat{S}_z$  (even though in the end we set  $B = 0$ ). Numerical results are presented in Appendix A and reproduce precisely the results of standard NRG using Eq. (25).

Similarly,  $\partial$ NRG may be used to obtain the charge susceptibility, which in the wide-band limit is given simply by  $\chi_C(T) \simeq \partial^2 F_N(\beta)/\partial \epsilon^2 = \partial_\epsilon \langle \hat{n} \rangle_{H_N, \beta}$  [and can be obtained by replacing  $\hat{S}_z$  by  $\hat{n}$  in Eq. (25)]. Indeed, Maxwell relations provide nontrivial connections between physical quantities obtained as derivatives. This was exploited recently in Ref. [44] to extract the fractional entropy of multichannel Kondo states in quantum dot experiments, comparing with NRG calculations of charge derivatives.

The above examples for the AIM demonstrate that the  $\partial$ NRG algorithm is equivalent to known expressions for certain derived quantities. However, the power of  $\partial$ NRG is that it does this automatically within a generalized framework and works equally well for any derivative in any quantum impurity model.

## VI. NUMERICAL RESULTS

For the following numerical demonstrations, we implemented a basic NRG code in JAX [19], using the included AD routines to obtain derivatives via Eqs. (21a) and (21b). For simplicity we did not exploit quantum numbers here, and so eigenstate degeneracies were removed by adding Gaussian noise with a small variance  $\sigma \simeq 10^{-11}$  (FD derivatives were calculated for comparison without noise,  $\sigma = 0$ ). In the following we set the conduction electron half bandwidth to  $D = 1$  and use an NRG discretization parameter  $\Lambda = 3$ . Further details on the numerical calculations and the finite-difference derivatives can be found in Appendixes D and E, respectively.

First we use  $\partial$ NRG to compute the derivative of the ground-state energy  $E_{N,0}$  of the AIM with respect to the interaction  $U$ . AD results in Fig. 1 (blue lines) are compared with FD approximations (circles), as a function of  $U$  at iteration  $N = 5$  (upper panel) and  $N = 40$  (lower panel), normalized by their respective  $U \rightarrow 0$  derivatives. The green lines show the variation of  $E_{N,0}$  itself. The results show the nontrivial effect of renormalization going from  $N = 5$  to  $N = 40$  at intermediate  $U$ , as well as the saturation of the ground-state derivatives at both large and small  $U$ . In this case we see excellent agreement between AD and FD results (although the former are far less computationally expensive to obtain).

Figure 2 demonstrates the use of  $\partial$ NRG to obtain thermodynamic quantities from derivatives of the NRG free energy. The inset shows the impurity occupation  $\langle \hat{n} \rangle \simeq \partial_\epsilon F_N$  at a temperature  $T/D \sim 10^{-5}$  (corresponding to  $N = 20$  and  $\beta = 0.9$ ) for the same systems as in Fig. 1. Since  $\epsilon = -U/2$ , the AIM possesses an exact particle-hole symmetry and hence is at half filling,  $\langle \hat{n} \rangle = 1$ . The AD results (blue line) sat-

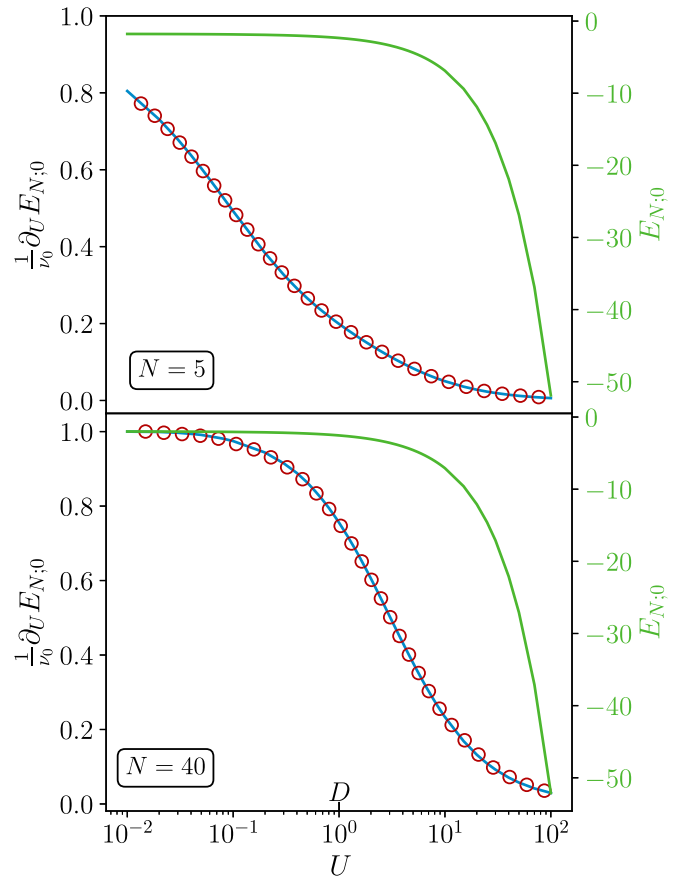


FIG. 1. Ground-state energy  $E_{N,0}$  of the AIM (green line) and its normalized derivative  $\partial_U E_{N,0}/\nu_0$  obtained by  $\partial$ NRG (blue line) and by FD (red circles) for iteration  $N = 5$  (top panel) and  $N = 40$  (bottom panel). Results plotted as a function of  $U$ , with  $\epsilon = -U/2$  and constant  $J \equiv 8V^2/U = 0.3$ . We define  $\nu_0 = \partial_U E_{N,0}|_{U \rightarrow 0}$  as the derivative in the limit  $U \rightarrow 0$ .

isfy this exact result precisely, while the FD results (circles) show some numerical error. More interestingly, the  $\partial$ NRG framework allows us to obtain higher derivatives with equal ease [Eqs. (B2a) and (B2b) are used instead of Eqs. (18a) and (18b)], as shown in the main panel of Fig. 2. Here, we calculate the corresponding second derivatives  $\partial^2 F_N/\partial U \partial \epsilon \simeq \partial_U \langle \hat{n} \rangle$ , which again show nontrivial behavior as a function of interaction strength  $U$ . The FD approximations agree well but are much more costly to obtain [36], requiring for every point five executions of the entire NRG code per second derivative, and an expensive convergence test.  $\partial$ NRG requires only a single function evaluation (see Appendix C).

As a final application of  $\partial$ NRG for the AIM, we turn to the RG energy level flow diagram shown in the top panel of Fig. 3. The excitation spectrum of the effective rescaled  $\hat{H}_N$  plotted against iteration number  $N$  shows the well-known flow between fixed points [1–3]. In the example shown, the crossover scales are well separated, such that we see distinct level structures associated with the free-orbital (FO), local-moment (LM), and strong-coupling (SC) fixed points as marked on the diagram. At a fixed point, the level structure does not change with  $N$ . Indeed, the RG structure of the problem and resulting universality imply that the fixed point level structure

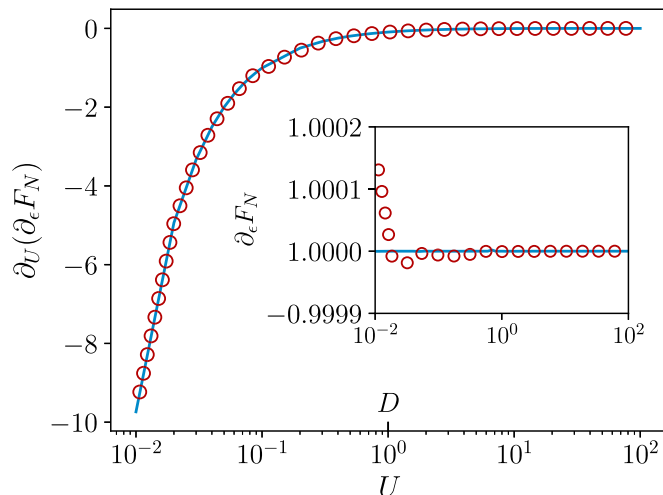


FIG. 2. Free-energy derivatives of the AIM obtained by  $\partial$ NRG. The inset shows  $\partial_\epsilon F_N \simeq \langle \hat{n} \rangle_T$ ; the main panel shows  $\partial_U(\partial_\epsilon F_N) \simeq \partial_U \langle \hat{n} \rangle_T$ , comparing AD (blue lines) with FD (red circles). The same parameters are used as for Fig. 1, but with  $N = 20$  and  $\bar{\beta} = 0.9$  yielding an effective temperature  $T/D \simeq 3 \times 10^{-5}$ .

is the same, independent of the microscopic model parameters  $U$  ( $\equiv -2\epsilon$ ) and  $V$ : Only the crossover scales between fixed points are affected.

This is demonstrated in the bottom panel of Fig. 3, where we plot the derivatives with respect to  $U$  of the NRG energy levels  $E_{N;i}$ , again as a function of iteration number  $N$ . As expected, the derivatives vanish at the fixed points (the level structure does not depend on  $U$  at the fixed points); however, there is a strong dependence along the crossovers, since the crossover scales depend on  $U$ .

## VII. OUTLOOK: DYNAMICS AND THE VERTEX

The above numerical results for the AIM are provided as a demonstration proof of principle. Future useful applications exploiting the full power of  $\partial$ NRG may be found for more complex models, situations involving higher-order derivatives, optimization techniques requiring exact gradients, or cases where partial derivatives of NRG outputs with respect to inputs are difficult to obtain by standard FD means. An example of the latter is the derivative of frequency-dependent dynamical quantities, such as impurity Green's functions or the conduction electron scattering T matrix, with respect to bare model parameters.

In the context of dynamical mean field theory (DMFT) for correlated materials [12,13,45], model machine learning techniques [46] could be applied to optimize simpler effective models with respect to more complicated microscopic ones, by comparing their Green's functions. Given the nontrivial dependence of such dynamical quantities at different energies on model parameters, the exact gradient within the AD framework becomes an essential ingredient for gradient descent optimization.

References [47,48] recently uncovered the analytic structure of the full local vertex and proposed a scheme to compute it within NRG. The vertex is an important object, entering, for example, in extensions of DMFT beyond the local limit [49].

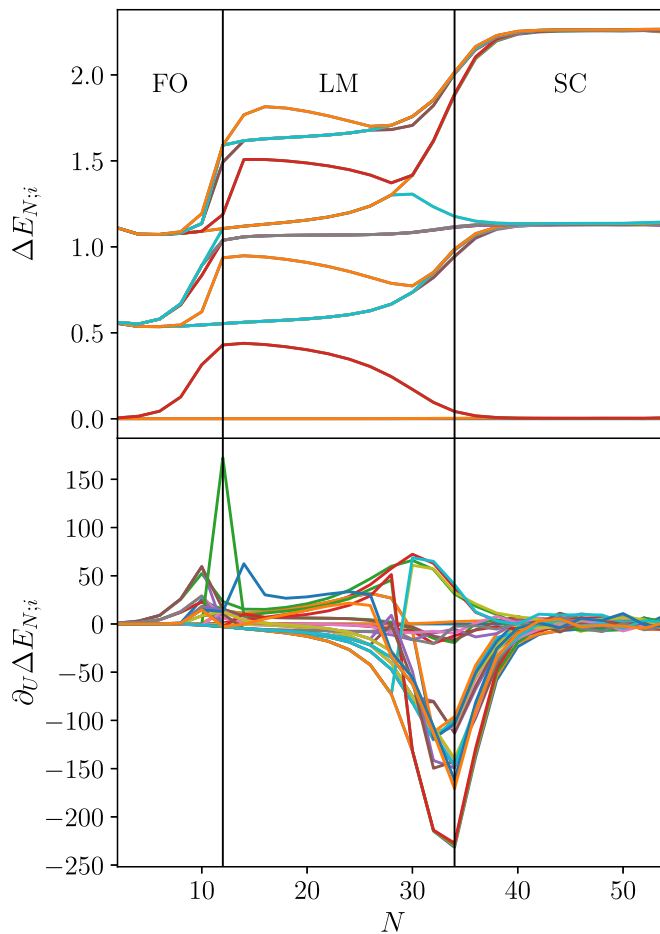


FIG. 3. Top: NRG energy level flow diagram for the AIM. The lowest 32 rescaled excitation energies  $\Delta E_{N;i}$  above the ground state at iteration  $N$  are plotted as a function of  $N$  (for even  $N$  only). Plotted for model parameters  $U = 0.01$ ,  $\epsilon = -U/2$ , and  $J \equiv 8V^2/U = 0.2$ , with the free-orbital (FO), local-moment (LM), and strong-coupling (SC) regimes marked. Bottom: the corresponding derivatives with respect to  $U$  obtained by  $\partial$ NRG.

It is possible that some partial information on the vertex could also be obtained by  $\partial$ NRG. This is inspired by the result in Ref. [50] for the AIM which, within the Matsubara formalism, connects the vertex at zero bosonic frequency  $F_{\nu\nu'}^{\text{loc}}(\omega = 0)$  to the functional derivative of the interaction self-energy  $\Sigma_\nu$  with respect to the hybridization  $\Delta_{\nu'}$ , viz.,

$$\frac{\delta \Sigma_\nu}{\delta \Delta_{\nu'}} = T[G_{\nu'}]^2 F_{\nu\nu'}^{\text{loc}}(\omega = 0), \quad (26)$$

where  $G_{\nu'}$  is the single-particle impurity Matsubara Green's function. For a precise definition and discussion of  $F_{\nu\nu'}^{\text{loc}}(\omega = 0)$ , see Refs. [47,50]. Since  $G_{\nu'}$  and  $\Sigma_\nu$  can be calculated within standard NRG [51], we argue that such an object is obtainable within  $\partial$ NRG. We speculate that a Keldysh version of Eq. (26) may similarly provide access to certain information about the real-frequency vertex at finite temperatures.

In order to calculate such derivatives of dynamical quantities, further code development is required, since the full-density-matrix NRG method would need to be implemented using AD libraries such as JAX [19] and integrated

within the  $\partial$ NRG scheme described in this paper. We leave this for future work.

### VIII. CONCLUSION

NRG is the gold-standard method of choice for solving generalized quantum impurity models [1,2]. In this paper, we introduce a variant of the standard algorithm— $\partial$ NRG—which makes use of the differential programming paradigm to automatically and efficiently obtain derivatives of NRG outputs with respect to input model parameters.

We make use of the AD JAX library [19], together with a bespoke routine based on Eqs. (21a) and (21b) which allows the derivatives of Hamiltonian eigenvalues and eigenvectors to be obtained at any iteration in an accurate and inexpensive way.

We demonstrated the use of  $\partial$ NRG by application to the Anderson impurity model, for which we obtained the derivative of NRG energy levels with respect to model parameters, calculated thermodynamic quantities from derivatives of the NRG free energy, and susceptibilities from derivatives of Hamiltonian source terms.  $\partial$ NRG may be useful for machine learning applications involving NRG [46] for which gradient descent optimization requires derivatives of a loss function; for adaptive broadening schemes [52]; or for optimal control protocols [29]. Richer physical information may be obtained from derivatives of dynamical quantities. This also opens the door to automatic differentiable DMFT, with  $\partial$ NRG as the impurity solver. Finally, we note that the  $\partial$ NRG methodology is compatible with interleaved-NRG (iNRG) [53,54] and generalizations utilizing full symmetries [43]. This is left for future work.

Our fully commented code is available as open-source software [55].

### ACKNOWLEDGMENTS

The authors thank L. Wang, L. Coopmans, J. von Delft, S.-S. Lee, and E. Minarelli for helpful discussions. We acknowledge funding from the Irish Research Council Laureate Awards 2017/2018 through Grant No. IRCLA/2017/169.

### APPENDIX A: ACCURACY AND SPEEDUP NUMERICALLY BENCHMARKED

In this Appendix we demonstrate that (i) NRG and  $\partial$ NRG are not affected by the addition of Gaussian noise with a small variance and (ii) the real-world performance of  $\partial$ NRG exceeds that of basic AD and FD in terms of both accuracy and speed.

A degeneracy-free NRG Hamiltonian is required for application of Eqs. (18a) and (18b). However, physical systems often do have eigenvalue degeneracies, arising, for example, from underlying symmetries. Although one can partially mitigate this problem by exploiting Abelian and non-Abelian quantum numbers, some degeneracies may remain. Here, we consider the “worst-case” scenario, in which no symmetries are exploited.

First we examine the effect of adding Gaussian noise with mean  $\mu = 0$  and variance  $\sigma$  to the diagonal elements of  $\hat{H}_N$ .

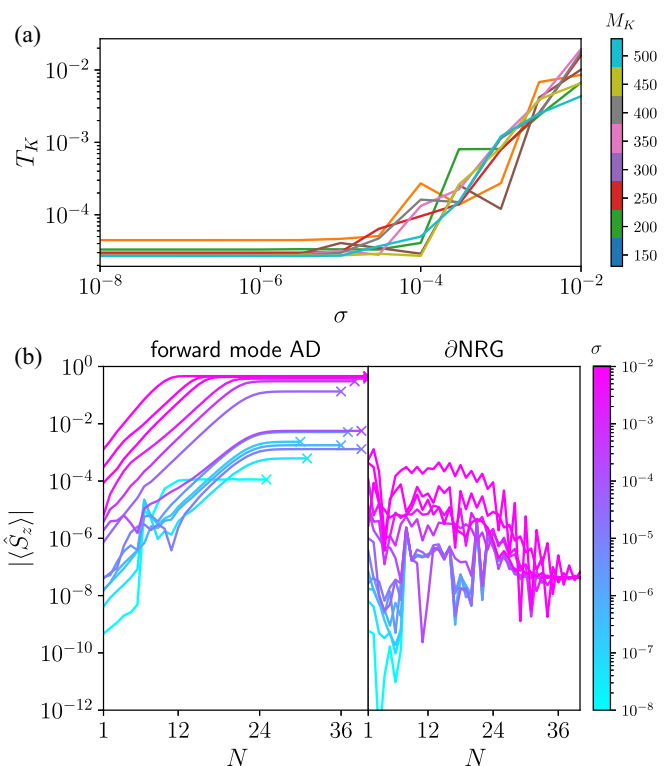


FIG. 4. (a) Dependence of  $T_K$  on the noise variance  $\sigma$  at different  $M_K$ . The number of kept states ranges from 150 to 500 in steps of 50. The AIM parameters used are  $\epsilon = -0.15$ ,  $U = 0.3$ ,  $V = 0.1$ , and  $B = 0$ . (b) Zero-field impurity magnetization  $\langle \hat{S}_z \rangle$  computed via derivatives of the free energy using  $\partial$ NRG and forward-mode AD, as a function of Wilson chain length  $N$  (for  $M_K = 500$ ). The noise variance is  $\sigma = 1 \times 10^{-8}$ ,  $3 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $3 \times 10^{-7}$ , ...,  $1 \times 10^{-2}$ . The model parameters are the same as in (a).

We use the Kondo temperature  $T_K$  as one figure of merit for assessing the effect of the noise term. For simplicity we define the Kondo temperature through the impurity entropy via  $S_{\text{imp}}(T = T_K) = 0.5$  (suitably between the local-moment and strong-coupling fixed point values). We compute  $T_K$  for  $10^{-8} \leq \sigma \leq 10^{-2}$  and different numbers of kept states  $100 \leq M_K \leq 500$  (using fixed  $\Lambda = 3$ ). Figure 4(a) shows clearly that for  $\sigma \leq 10^{-6}$  the Kondo temperature does not noticeably depend on the noise level (we have confirmed explicitly down to  $\sigma = 10^{-15}$  that the results are fully converged for each  $M_K$ ). Other physical quantities computed in standard NRG show similar behavior.

However, the effect of adding noise is more pronounced in derivatives. In Fig. 4(b) we compare  $\partial_B F_N(\vec{\beta})|_{B=0} = \langle \hat{S}_z \rangle_{H_N, \vec{\beta}}$  as a function of Wilson chain length  $N$ , as computed with  $\partial$ NRG and straight AD for different  $\sigma$ . Since the impurity magnetization is evaluated at zero field, SU(2) spin symmetry implies the exact result  $\langle \hat{S}_z \rangle_{H_N, \vec{\beta}} = 0$ . However, this spin symmetry is not enforced, and so we see deviations due to the noise. As Fig. 4(b) shows,  $\partial$ NRG accurately approximates the exact result even at relatively high  $\sigma$ . By contrast, AD yields derivatives that strongly depend on  $\sigma$  and have much higher error than  $\partial$ NRG at a given  $\sigma$ . Indeed, derivatives are numerically not computable for all  $N$  in standard forward-mode AD for  $\sigma \gtrsim 10^{-6}$ ; at larger noise levels, AD derivatives



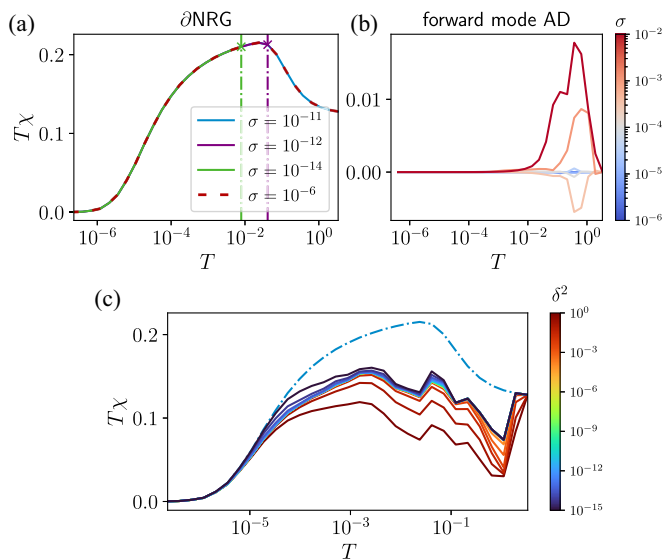


FIG. 5. (a) Magnetic susceptibility  $\chi$  calculated with  $\partial$ NRG for different noise variances  $\sigma$ . (b)  $\chi$  computed using forward-mode AD and  $10^{-8} \leq \sigma \leq 10^{-2}$ . (c)  $\chi$  computed with  $\partial$ NRG using the pole broadening approach, with  $10^{-15} \leq \delta^2 \leq 10^0$ . All calculations are performed with  $M_K = 500$  and with the same AIM parameters as in Fig. 4.

are computable but the accuracy can become poor, especially at later iterations due to the propagation and accumulation of errors through the derivative chain rule.

Therefore, although NRG for the AIM is insensitive to noise for  $\sigma \leq 10^{-6}$ , the AD approach is not stable at these noise levels. At higher noise levels, AD is stabilized, but the accuracy suffers. On the other hand,  $\partial$ NRG avoids such problems: A lower noise level can be used since each derivative calculation is independent, and highly accurate results can be obtained.

Within  $\partial$ NRG, the magnetization calculation via the analog of Eq. (24) shown in Fig. 4(b) does not involve eigenvector derivatives and is therefore particularly stable. By contrast, the AD calculation of any derivative is built up recursively via Eq. (3) and therefore does involve the computation of eigenvector derivatives at each and every step. This contributes to the relative performance gain in  $\partial$ NRG.

In Fig. 5 we examine the magnetic susceptibility  $\chi(T)$ . The calculation of this quantity, which is obtained automatically in  $\partial$ NRG via the second derivative of the NRG free energy, is formally equivalent to the analytic expression, Eq. (25) (and *does* involve eigenvector derivatives). Gaussian noise of width  $\sigma$  is added to the diagonal of  $\hat{H}_N$  at iteration  $N$ . The calculated  $\chi_{H_N, \bar{\beta}}$  at  $\bar{\beta} = 0.9$  (used hereinafter) is then interpreted as the true  $\chi(T)$  at inverse temperature  $\beta = \Lambda^{(N-1)/2} \bar{\beta}$ . The numerical results from  $\partial$ NRG show that  $\chi_{H_N, \bar{\beta}}$  is obtained reliably at all  $N$  (and hence all  $T$ ) for  $\sigma$  as low as  $10^{-11}$ . Only at smaller  $\sigma$  does the method break down: Derivatives are then not computable for earlier iterations or higher temperatures (the correct low- $T$  behavior is, however, still captured). As demonstrated in Fig. 4 and confirmed in Fig. 5(a), highly accurate results are obtained for  $\sigma < 10^{-6}$ . Therefore in  $\partial$ NRG we have a wide window

$10^{-11} < \sigma < 10^{-6}$  over which numerical results are fully converged and stable.

By contrast, in Fig. 5(b) we show results for the same quantity obtained by straight AD. The calculation is not numerically stable for small  $\sigma$ , but very inaccurate for large  $\sigma$  (typical of the breakdown for higher-order derivatives). As such there is no reliable regime for which robust results can be obtained by straight AD.

In Fig. 5(c) we examine the feasibility of an alternative approach to the eigenvalue-degeneracy problem that avoids adding Gaussian noise. The method, proposed by Liao *et al.* in Ref. [31], consists of reformulating Eq. (18b),

$$d|i\rangle = \sum_{i \neq j} \frac{\langle j|d\hat{A}|i\rangle}{\lambda_i - \lambda_j} |j\rangle \equiv \sum_{i \neq j} \langle j|d\hat{A}|i\rangle f(\lambda_i - \lambda_j) |j\rangle, \quad (\text{A1})$$

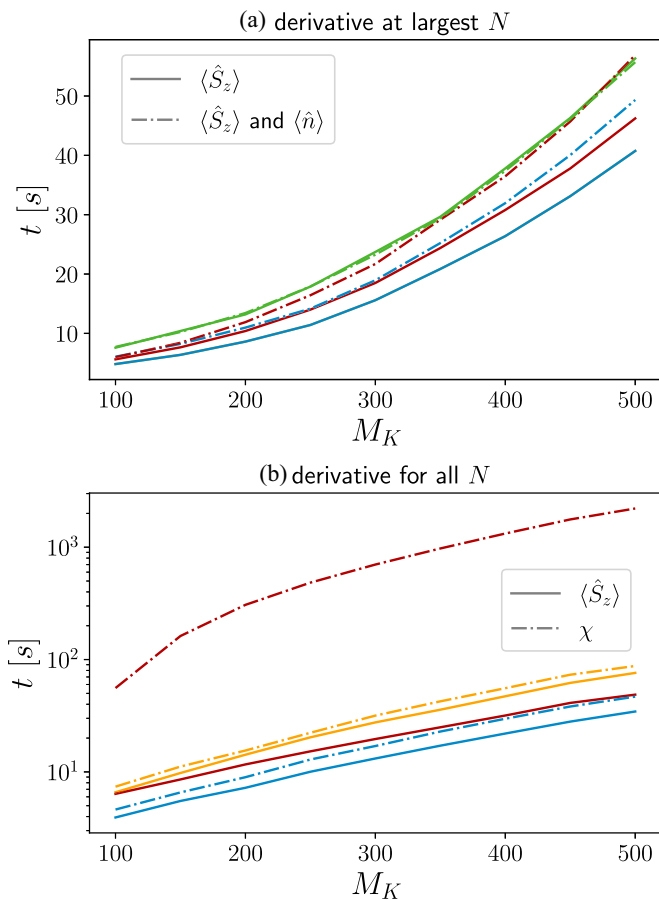


FIG. 6. Representative calculation runtime in seconds, vs number of kept states  $M_K$ , comparing different automatic derivative techniques and computed quantities. In all cases shown, blue lines correspond to  $\partial$ NRG with single diagonalization, orange lines are for  $\partial$ NRG with double diagonalization, red lines are for forward-mode AD, and green lines are for backward-mode AD. (a) Calculation at  $N = 40$  of  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  (solid lines) compared with the calculation of both  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  and  $\langle \hat{n} \rangle_{H_N, \bar{\beta}}$  (dash-dotted lines), obtained via the first derivatives of the NRG free energy. (b) Calculation for all  $N \leq 40$  of  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  requiring first derivatives (solid lines), and  $\chi_{H_N, \bar{\beta}}$  requiring second derivatives (dash-dotted lines). Model parameters are as in Fig. 4.

with  $f(x) = 1/x$ . Divergences induced by eigenvalue degeneracies can be avoided by replacing the function  $f(x)$  with the approximate broadened form  $f(x) \approx x/(x^2 + \delta^2)$  with  $\delta \ll 1$ . This broadening approximation is known to distort somewhat the overall results but has the advantage that divergences are removed without the need to lift degeneracies. Figure 5(c) shows numerical results for  $\chi(T)$  obtained in this way, as a function of  $T$  for different broadenings  $\delta$ . Although the method always yields a computable result, the true result (dash-dotted line) is only obtained in the  $\delta \rightarrow 0$  limit at the lowest temperature scales. Finite  $\delta$  at intermediate  $T$  appears to yield rather poor results. We conclude that  $\partial$ NRG is indeed the method of choice in this context, in terms of accuracy.

We now turn to an analysis of the real-world performance of  $\partial$ NRG in terms of calculation time, comparing with straight AD implemented in JAX [19]. The figure of merit is the runtime measured in seconds of both algorithms run on the same machine (in this case an AMD Threadripper 2950X platform running PYTHON 3.8.10 with the OMP\_NUM\_THREAD = 16 flag; further package specifications can be found in Ref. [56]). We do not utilize any quantum numbers here.

In Fig. 6(a) we compare the runtime for the calculation of (i)  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  (solid lines), and (ii)  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  and  $\langle \hat{n} \rangle_{H_N, \bar{\beta}}$  together (dash-dotted lines); using  $\partial$ NRG (blue), forward-mode AD (red), and backward-mode AD (green), at the last NRG iteration,  $N = 40$ . We see that for all  $M_K$ ,  $\partial$ NRG performs best (with the relative speedup becoming more pronounced at larger  $M_K$ ). For straight AD, we see that forward mode beats backward mode for the calculation of a single derivative; however, since backward-mode AD scales with the number of inputs rather than outputs, it will overtake the forward mode when many derivatives are calculated.

TABLE I. Wengert list for NRG with forward-mode AD. Left column: the primal trace prescribes all steps to be performed to execute an NRG calculation with  $N$  iterations. Right column: the tangent trace prescribes all steps to compute the derivative of the NRG Hamiltonian  $\hat{H}_N$ .

Forward primal trace	Forward tangent trace
$v_{-3} = \Lambda$	$dv_{-3} = 0$
$v_{-2} = \{t_n\}$	$dv_{-2} = 0$
$v_{-1} = \eta$	$dv_{-1} = 0$
$v_0 = H_{\text{imp}}$	$dv_0 = dH_{\text{imp}}$
$v_1 = \text{EIGENVALUES}(v_0)$	$[dv_1]_i = [v_2]_i^\dagger \cdot dv_0 \cdot [v_2]_i$
$v_2 = \text{EIGENVECTORS}(v_0)$	$[dv_2]_i = \sum_{j \neq i} \frac{[v_2]_j^\dagger \cdot dv_0 \cdot [v_2]_i}{[v_1]_i - [v_1]_j} [v_2]_j$
$v_3 = v_{-1}^T$	$dv_3 = 0$
FOR $n = 0$ to $N$	FOR $n = 0$ to $N$
$v_{7n+4} = v_{7n}$	$dv_{7n+4} = dv_{7n}$
$[v_{7n+5}]_{iiss} = v_{-3}^{1/2} [v_{7n+1}]_l$	$[dv_{7n+5}]_{iiss} = v_{-3}^{1/2} [dv_{7n+2}]_l$
$[v_{7n+6}]_{ii'ss'} = (-1)^j [v_{-2}]_n \sum_\sigma [v_{7n+3}]_{\sigma ii'} [v_{-1}]_{\sigma ss'}$	$[dv_{7n+6}]_{ii'ss'} = (-1)^j [v_{-2}]_n \sum_\sigma [dv_{7n+3}]_{\sigma ii'} [v_{-1}]_{\sigma ss'}$
$[v_{7n+7}]_{ii'ss'} = [v_{7n+5}]_{ii'ss'} + [v_{7n+6}]_{ii'ss'} + [v_{7n+6}]_{ii'ss'}^\dagger$	$[dv_{7n+8}]_i = [v_{7n+9}]_i^\dagger \cdot dv_{7n} \cdot [v_{7n+9}]_i$
$v_{7n+8} = \text{EIGENVALUES}(v_{7n+7})$	$[dv_{7n+8}]_i = [v_{7n+9}]_i^\dagger \cdot dv_{7n} \cdot [v_{7n+9}]_i$
$v_{7n+9} = \text{EIGENVECTORS}(v_{7n+7})$	$[dv_{7n+9}]_i = \sum_{j \neq i} \frac{[v_{7n+9}]_j^\dagger \cdot dv_{7n} \cdot [v_{7n+9}]_i}{[v_{7n+8}]_i - [v_{7n+8}]_j} [v_{7n+9}]_j$
$[v_{7n+10}]_{\sigma ii'} = \sum_{ss'} \sum_{jj'} [v_{7n+9}]_{i'sj}^\dagger [v_{7n+9}]_{i's'j'} [v_{-1}]_{\sigma ss'}$	$[dv_{7n+10}]_{\sigma ii'} = \sum_{ss'} \sum_{jj'} ([dv_{7n+9}]_{i'sj}^\dagger [v_{7n+9}]_{i's'j'} + [v_{7n+9}]_{i'sj}^\dagger [dv_{7n+9}]_{i's'j'}) [v_{-1}]_{\sigma ss'}$
$y_0 = v_{7N+7}$	$dy_0 = dv_{7N+7}$

In Fig. 6(b) we compare runtimes using  $\partial$ NRG with a single diagonalization of  $\hat{H}_N$  (blue),  $\partial$ NRG with double diagonalization of  $\hat{H}_N$  (orange), and forward-mode AD (red). We calculate  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  (solid lines), and  $\chi_{H_N, \bar{\beta}}$  (dash-dotted lines) at all iterations  $N \leq 40$ . For the orange lines, two diagonalizations of  $\hat{H}_N$  are performed at each step (with and without noise), which thereby eliminates error propagation from the added noise and provides the most accurate calculation (this may or may not be needed in practice, depending on the situation). We see that  $\partial$ NRG with a single diagonalization per step is the fastest in all cases. For the simpler quantity  $\langle \hat{S}_z \rangle_{H_N, \bar{\beta}}$  (which involves only a first derivative), the  $\partial$ NRG using two diagonalizations and the forward-mode AD have similar runtimes, although AD is slightly faster. However, for  $\chi_{H_N, \bar{\beta}}$ , both versions of  $\partial$ NRG are considerably faster. This is because the calculation of  $\chi_{H_N, \bar{\beta}}$  requires a second derivative, which is much more computationally expensive in AD, but almost as cheap in  $\partial$ NRG. The relative performance gain for  $\partial$ NRG also increases if several derivatives are computed.

In conclusion, in a typical setting,  $\partial$ NRG is considerably more efficient than AD (often by orders of magnitude)—while at the same time being more accurate.

## APPENDIX B: HIGHER-ORDER DERIVATIVES

Equations (18a) and (18b) allow us to take the derivatives of eigenvalues and eigenvectors of some Hermitian, nondegenerate Hamiltonian  $\hat{H}$ . Consider now a Hamiltonian that can be linearly decomposed as

$$\hat{H}(\{\theta\}) = \sum_i \theta_i \hat{h}_i, \quad (\text{B1})$$

where  $\hat{h}_i$  is an operator defined over the same Hilbert space as  $\hat{H}$ , and with the corresponding coupling constant  $\theta_i$  being a real scalar. In this case, all second-order derivatives of the Hamiltonian must vanish,  $\frac{\partial^2 \hat{H}}{\partial \theta_i \partial \theta_j} = 0$ . It is then possible to derive a closed-form expression for the second-order derivatives of eigenvalues and eigenvectors with respect to the couplings  $\theta_1, \theta_2 \in \{\theta\}$ , viz.,

$$\partial_{\theta_1} \partial_{\theta_2} E_i = 2 \sum_{i \neq j} \frac{\langle i | \partial_{\theta_2} \hat{H} | j \rangle \langle j | \partial_{\theta_1} \hat{H} | i \rangle}{E_i - E_j}, \quad (\text{B2a})$$

$$\partial_{\theta_1} \partial_{\theta_2} |i\rangle = \sum_{i \neq j} \left[ \Pi_{ij} |j\rangle + \frac{\langle j | \partial_{\theta_1} \hat{H} | i \rangle}{\lambda_i - \lambda_j} \sum_{j \neq k} \frac{\langle k | \partial_{\theta_2} \hat{H} | j \rangle}{\lambda_j - \lambda_k} |k\rangle \right], \quad (\text{B2b})$$

where

$$\begin{aligned} \Pi_{ij} = & - \frac{\langle i | \partial_{\theta_2} \hat{H} | i \rangle - \langle j | \partial_{\theta_2} \hat{H} | j \rangle}{(\lambda_i - \lambda_j)^2} \langle j | \partial_{\theta_1} \hat{H} | i \rangle \\ & + \sum_{j \neq k} \frac{\langle j | \partial_{\theta_2} \hat{H} | k \rangle}{\lambda_j - \lambda_k} \langle k | \partial_{\theta_1} \hat{H} | i \rangle \\ & + \sum_{k \neq i} \frac{\langle k | \partial_{\theta_2} \hat{H} | i \rangle}{\lambda_i - \lambda_k} \langle j | \partial_{\theta_1} \hat{H} | k \rangle. \end{aligned} \quad (\text{B3})$$

With these formulas one can compute physical observables depending on second derivatives, such as susceptibilities, using  $\partial$ NRG.

### APPENDIX C: COMPARING DIFFERENTIATION METHODS

The basic NRG algorithm is described in the main text, with the key steps contained in Eqs. (6)–(13). For more details, see Ref. [2]. Assuming that degeneracies in the NRG Hamiltonian  $\hat{H}_N$  are lifted, we can use Eqs. (18a) and (18b) to compute the derivatives of eigenvectors and eigenvalues of  $\hat{H}_N$ . The Wengert list [37,57] with the forward primal trace (function evaluation) and the forward tangent trace (function differentiation) can then be compiled, as shown in Table I. Similarly, we compile the forward primal trace and forward tangent trace for the  $\partial$ NRG algorithm; see Table II. For  $\partial$ NRG, we have one additional step in the primal trace for the main algorithm [corresponding to Eq. (16)] but a trivial tangent trace.

We can now compare the efficiency of the two approaches. For forward-mode AD (fAD) we have  $\text{ops}(\text{fAD}) = 7N$ , and to compute forward primal and tangent traces,  $2 \times \text{ops}(\text{fAD})$  operations are required. For  $\partial$ NRG, by contrast,  $\text{ops}(\partial\text{NRG}) = 8N$ , but no other step is required to compute the derivative of  $\hat{H}_N$ . However, this does not mean that  $\partial$ NRG is twice as fast as fAD because the bottleneck eigensolver routines [58] appear only in the forward primal trace. Nonetheless,  $\partial$ NRG still

TABLE II. Wengert list for  $\partial$ NRG. Left column: the forward primal trace for  $\partial$ NRG outputs the NRG Hamiltonian  $\hat{H}_N$  and its derivative  $d\hat{H}_N$ . Since derivatives are computed via Eqs. (21a) and (21b) in the main algorithm, there is no corresponding forward tangent trace in the right column.

Forward primal trace	Forward tangent trace
$v_{-4} = dH_{\text{imp}}$	$dv_{-4} = dH_{\text{imp}}$
$v_{-3} = \Lambda$	$dv_{-3} = 0$
$v_{-2} = \{t_n\}$	$dv_{-2} = 0$
$v_{-1} = \eta$	$dv_{-1} = 0$
$v_0 = H_{\text{imp}}$	$dv_{-4} = 0$
$v_1 = \text{EIGENVALUES}(v_0)$	
$v_2 = \text{EIGENVECTORS}(v_0)$	
$v_3 = v_{-1}^T$	
$[v_4]_{i'j'} = \sum_{ss'} \sum_{jj'} [v_2]_{i;sj}^\dagger [v_2]_{i;s'j'} [v_{-4}]_{j'j} \delta_{ss'}$	
FOR $n = 0$ to $N$	
$v_{8n+5} = v_{8n}$	
$[v_{8n+6}]_{iiss} = v_{-3}^{1/2} [v_{8n+1}]_i$	
$[v_{8n+7}]_{i'iss'} = (-1)^{i'} [v_{-2}]_n \sum_{\sigma} [v_{8n+3}]_{\sigma i' i'} [v_{-1}]_{\sigma ss'}$	
$[v_{8n+8}]_{i'iss'} = [v_{8n+6}]_{i'iss'} + [v_{8n+7}]_{i'iss'} + [v_{8n+7}]_{i'iss'}^\dagger$	
$v_{8n+9} = \text{EIGENVALUES}(v_{8n+8})$	
$v_{8n+10} = \text{EIGENVECTORS}(v_{8n+8})$	
$[v_{8n+11}]_{\sigma i' i'} = \sum_{ss'} \sum_{jj'} [v_{8n+10}]_{i;sj}^\dagger [v_{8n+10}]_{i;s'j'} [v_{-1}]_{\sigma ss'}$	
$[v_{8n+12}]_{i' i'} = \sum_{ss'} \sum_{jj'} [v_{8n+10}]_{i;sj}^\dagger [v_{8n+10}]_{i;s'j'} [v_{8n+4}]_{j'j} \delta_{ss'}$	
$y_0 = v_{8N+8}$	$dy_0 = v_{8N+12}$

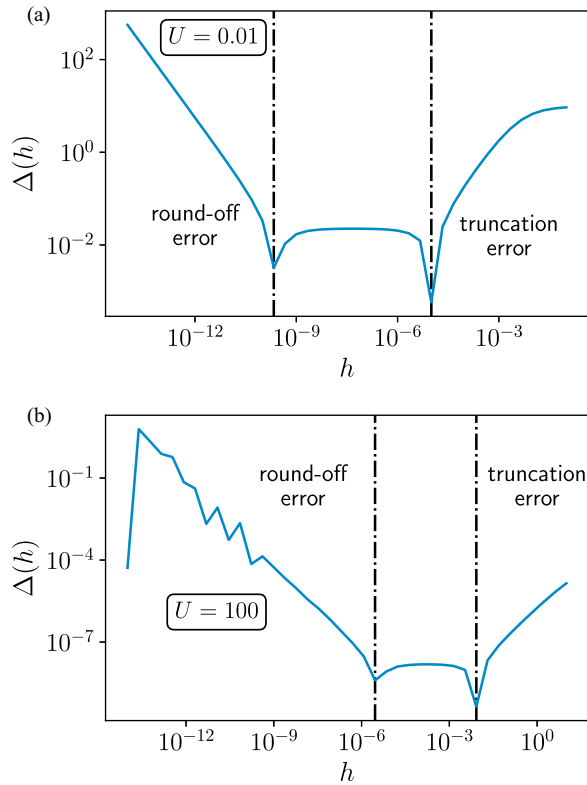


FIG. 7. FD derivative errors  $\Delta(h)$  vs FD value  $h$  for the calculation of  $\partial_U \langle \hat{n} \rangle_{H_N, \beta}$  at  $N = 40$ . Shown for AIM parameters  $\epsilon = -U/2$ ,  $V = 0.1$ , and  $B = 0$  with (a)  $U = 0.01$  and (b)  $U = 100$ . In all cases,  $M_K = 500$ .

provides a considerable performance advantage, as established by explicit benchmarks in Appendix A.

#### APPENDIX D: DETAILS OF NRG CALCULATIONS

We used standard thermodynamic NRG [1] to calculate the free energy, impurity occupation, and eigenvalues of  $\hat{H}_N$ . An NRG discretization parameter  $\Lambda = 3$  was used, and  $M_K = 200$  states were retained at each step for Figs. 1 and 3, while  $M_K = 600$  states were retained for Fig. 2. No quantum numbers were used in this demonstration calculation.

#### APPENDIX E: DETAILS OF FINITE-DIFFERENCE CALCULATIONS

We calculate the finite-difference (FD) derivative of a function  $f$  via

$$D_h[f](x) = \frac{f(x+h) - f(x)}{h}, \quad (\text{E1})$$

where  $h$  is the FD value. The FD derivative is connected to the definition of the derivative by taking the limit,  $\frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} D_h[f](x)$ .

In practice, we compute FD derivatives for a set of finite  $\{h_i\}$  and check for convergence. However, care must be taken due to the trade-off between *round-off* and *truncation* errors [37]. The truncation error arises due to the finite  $h > 0$ , which is required for the numerical evaluation of  $D_h[f](x)$ , and which would diminish in the formal limit  $h \rightarrow 0$ . However, for very small  $h$ , the difference in the function evaluations for  $f(x)$  and  $f(x+h)$  cannot be distinguished numerically due to inevitable round-off errors in any finite precision arithmetic. This leads to increasing errors as  $h \rightarrow 0$ .

We quantify the FD error as

$$\Delta(h) = \left| D_h[f](x) - \frac{\partial f}{\partial x} \right|. \quad (\text{E2})$$

This quantity is plotted in Fig. 7 for the derivative of the AIM impurity occupation with respect to the impurity interaction,  $\partial_U \langle \hat{n} \rangle_{H_N, \beta}$ . We have used the  $\partial$ NRG derivative as the true value of  $\frac{\partial f}{\partial x}$  in Eq. (E2). Figure 7 shows rather typical behavior, with truncation errors dominating at large  $h$  and round-off errors dominating at small  $h$ , with a stability plateau in between where the derivative should be evaluated. However, comparison of Figs. 7(a) and 7(b) (which correspond to model parameters  $U = 10^{-2}$  and  $U = 10^2$ , respectively) demonstrates an important limitation of FD differentiation: The optimal  $h$  is not fixed, but depends on input model parameters. Reliable results therefore require such a convergence analysis for each new set of model parameters and for each new derivative. This becomes computationally costly. The situation becomes significantly worse when considering higher-order derivatives.

- [1] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, *Rev. Mod. Phys.* **47**, 773 (1975).
- [2] R. Bulla, T. A. Costi, and T. Pruschke, Numerical renormalization group method for quantum impurity systems, *Rev. Mod. Phys.* **80**, 395 (2008).
- [3] A. C. Hewson, *The Kondo Problem to Heavy Fermions*, Cambridge Studies in Magnetism (Cambridge University Press, Cambridge, UK, 1997), Vol. 2.
- [4] A. K. Mitchell, M. Becker, and R. Bulla, Real-space renormalization group flow in quantum impurity systems: Local moment formation and the Kondo screening cloud, *Phys. Rev. B* **84**, 115120 (2011).
- [5] T. A. Costi, L. Bergqvist, A. Weichselbaum, J. von Delft, T. Micklitz, A. Rosch, P. Mavropoulos, P. H. Dederichs, F. Mallet,

L. Saminadayar, and C. Bauerle, Kondo Decoherence: Finding the Right Spin Model for Iron Impurities in Gold and Silver, *Phys. Rev. Lett.* **102**, 056802 (2009).

- [6] D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, and M. A. Kastner, Kondo effect in a single-electron transistor, *Nature (London)* **391**, 156 (1998).
- [7] A. J. Keller, S. Amasha, I. Weymann, C. P. Moca, I. G. Rau, J. A. Katine, H. Shtrikman, G. Zaránd, and D. Goldhaber-Gordon, Emergent SU(4) Kondo physics in a spin-charge-entangled double quantum dot, *Nat. Phys.* **10**, 145 (2014).
- [8] Z. Iftikhar, S. Jezouin, A. Anthore, U. Gennser, F. D. Parmentier, A. Cavanna, and F. Pierre, Two-channel Kondo effect and renormalization flow with macroscopic quantum charge states, *Nature (London)* **526**, 233 (2015); A. K. Mitchell,

- L. A. Landau, L. Fritz, and E. Sela, Universality and Scaling in a Charge Two-Channel Kondo Device, *Phys. Rev. Lett.* **116**, 157202 (2016); Z. Iftikhar, A. Anthore, A. K. Mitchell, F. D. Parmentier, U. Gennser, A. Ouerghi, A. Cavanna, C. Mora, P. Simon, and F. Pierre, Tunable quantum criticality and super-ballistic transport in a “charge” Kondo circuit, *Science* **360**, 1315 (2018).
- [9] W. Pouse, L. Peeters, C. L. Hsueh, U. Gennser, A. Cavanna, M. A. Kastner, A. K. Mitchell, and D. Goldhaber-Gordon, Exotic quantum critical point in a two-site charge Kondo circuit, [arXiv:2108.12691](https://arxiv.org/abs/2108.12691).
- [10] W. Liang, M. P. Shores, M. Bockrath, J. R. Long, and H. Park, Kondo resonance in a single-molecule transistor, *Nature (London)* **417**, 725 (2002).
- [11] A. K. Mitchell, K. G. L. Pedersen, P. Hedegård, and J. Paaske, Kondo blockade due to quantum interference in single-molecule junctions, *Nat. Commun.* **8**, 15210 (2017).
- [12] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, *Rev. Mod. Phys.* **78**, 865 (2006).
- [13] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [14] K. M. Stadler, Z. P. Yin, J. von Delft, G. Kotliar, and A. Weichselbaum, Dynamical Mean-Field Theory Plus Numerical Renormalization-Group Study of Spin-Orbital Separation in a Three-Band Hund Metal, *Phys. Rev. Lett.* **115**, 136401 (2015).
- [15] A. Weichselbaum, F. Verstraete, U. Schollwöck, J. I. Cirac, and J. von Delft, Variational matrix-product-state approach to quantum impurity models, *Phys. Rev. B* **80**, 165117 (2009).
- [16] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, *Rev. Mod. Phys.* **83**, 349 (2011).
- [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [18] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, Automatic differentiation of algorithms, *J. Comput. Appl. Math.* **124**, 171 (2000).
- [19] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs, 2018, <https://github.com/google/jax>.
- [20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, <https://tensorflow.org>.
- [21] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems* 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Red Hook, NY, 2019), pp. 8024–8035.
- [22] T. Tamayo-Mendoza, C. Kreisbeck, R. Lindh, and A. Aspuru-Guzik, Automatic differentiation in quantum chemistry with applications to fully variational Hartree–Fock, *ACS Cent. Sci.* **4**, 559 (2018).
- [23] S.-X. Zhang, Z.-Q. Wan, and H. Yao, Automatic differentiable Monte Carlo: Theory and application, [arXiv:1911.09117](https://arxiv.org/abs/1911.09117).
- [24] H. Xie, J.-G. Liu, and L. Wang, Automatic differentiation of dominant eigensolver and its applications in quantum physics, *Phys. Rev. B* **101**, 245139 (2020).
- [25] N. McGreivy, S. R. Hudson, and C. Zhu, Optimized finite-build stellarator coils using automatic differentiation, *Nucl. Fusion* **61**, 026020 (2021).
- [26] B. Ponsioen, F. F. Assaad, and P. Corboz, Automatic differentiation applied to excitations with projected entangled pair states, *SciPost Phys.* **12**, 006 (2022).
- [27] S. Dick and M. Fernandez-Serra, Using differentiable programming to obtain an energy and density-optimized exchange-correlation functional, [arXiv:2106.04481](https://arxiv.org/abs/2106.04481).
- [28] L. Coopmans, D. Luo, G. Kells, B. K. Clark, and J. Carrasquilla, Protocol Discovery for the Quantum Control of Majoranas by Differentiable Programming and Natural Evolution Strategies, *PRX Quantum* **2**, 020332 (2021).
- [29] R. A. Vargas-Hernández, R. T. Q. Chen, K. A. Jung, and P. Brumer, Fully differentiable optimization protocols for non-equilibrium steady states, *New J. Phys.* **23**, 123006 (2021).
- [30] Z.-Q. Wan, S.-X. Zhang, and H. Yao, Mitigating sign problem by automatic differentiation, [arXiv:2010.01141](https://arxiv.org/abs/2010.01141).
- [31] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, Differentiable Programming Tensor Networks, *Phys. Rev. X* **9**, 031041 (2019).
- [32] W.-L. Tu, H.-K. Wu, N. Schuch, N. Kawashima, and J.-Y. Chen, Generating function for tensor network diagrammatic summation, *Phys. Rev. B* **103**, 205155 (2021).
- [33] S. Linnainmaa, Taylor expansion of the accumulated rounding error, *BIT Numer. Math.* **16**, 146 (1976).
- [34] G. Tinhofer, R. Albrecht, E. Mayr, H. Noltemeier, and M. M. Syslo, *Computational Graph Theory*, Computing Supplementum, Vol. 7 (Springer, New York, 2012).
- [35] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation* (SIAM, Philadelphia, PA, 2008).
- [36] C. Grossmann, H.-G. Roos, and M. Stynes, *Numerical Treatment of Partial Differential Equations*, Universitext (Springer, Berlin, 2007), Vol. 154.
- [37] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, Automatic differentiation in machine learning: a survey, *J. Mach. Learn. Res.* **18**(153), 1 (2018).
- [38] A. Weichselbaum and J. von Delft, Sum-Rule Conserving Spectral Functions from the Numerical Renormalization Group, *Phys. Rev. Lett.* **99**, 076402 (2007).
- [39] M. Seeger, A. Hetzel, Z. Dai, E. Meissner, and N. D. Lawrence, Auto-differentiating linear algebra, [arXiv:1710.08717](https://arxiv.org/abs/1710.08717).
- [40] M. B. Giles, Collected matrix derivative results for forward and reverse mode automatic differentiation, in *Advances in Automatic Differentiation*, Lecture Notes in Computational Science and Engineering, edited by C. H. Bischof, H. M. Bücker, P. Hovland, U. Naumann, and J. Utke (Springer, Berlin, 2008), Vol. 64.
- [41] R. P. Feynman, Forces in Molecules, *Phys. Rev.* **56**, 340 (1939).
- [42] R. L. Dailey, Eigenvector derivatives with repeated eigenvalues, *AIAA J.* **27**, 486 (1989).

- [43] A. Weichselbaum, Non-abelian symmetries in tensor networks: A quantum symmetry space approach, *Ann. Phys. (Amsterdam)* **327**, 2972 (2012).
- [44] C. Han, A. K. Mitchell, Z. Iftikhar, Y. Kleeorin, A. Anthore, F. Pierre, Y. Meir, and E. Sela, Extracting entropy of exotic quasiparticles from conductance measurements, [arXiv:2108.12878](https://arxiv.org/abs/2108.12878).
- [45] E. Sheridan, C. Rhodes, F. Jamet, I. Rungger, and C. Weber, Data-driven dynamical mean-field theory: An error-correction approach to solve the quantum many-body problem using machine learning, *Phys. Rev. B* **104**, 205120 (2021).
- [46] J. B. Rigo and A. K. Mitchell, Machine learning effective models for quantum systems, *Phys. Rev. B* **101**, 241105(R) (2020).
- [47] F. B. Kugler, S.-S. B. Lee, and J. von Delft, Multipoint Correlation Functions: Spectral Representation and Numerical Evaluation, *Phys. Rev. X* **11**, 041006 (2021).
- [48] S.-S. B. Lee, F. B. Kugler, and J. von Delft, Computing Local Multipoint Correlators Using the Numerical Renormalization Group, *Phys. Rev. X* **11**, 041007 (2021).
- [49] G. Rohringer, H. Hafermann, A. Toschi, A. A. Katanin, A. E. Antipov, M. I. Katsnelson, A. I. Lichtenstein, A. N. Rubtsov, and K. Held, Diagrammatic routes to nonlocal correlations beyond dynamical mean field theory, *Rev. Mod. Phys.* **90**, 025003 (2018).
- [50] E. G. C. P. van Loon, F. Krien, and A. A. Katanin, Bethe-Salpeter Equation at the Critical End Point of the Mott Transition, *Phys. Rev. Lett.* **125**, 136402 (2020).
- [51] Ž. Osolin and R. Zitko, Padé approximant approach for obtaining finite-temperature spectral functions of quantum impurity models using the numerical renormalization group technique, *Phys. Rev. B* **87**, 245135 (2013).
- [52] S.-S. B. Lee and A. Weichselbaum, Adaptive broadening to improve spectral resolution in the numerical renormalization group, *Phys. Rev. B* **94**, 235127 (2016).
- [53] A. K. Mitchell, M. R. Galpin, S. Wilson-Fletcher, D. E. Logan, and R. Bulla, Generalized Wilson chain for solving multichannel quantum impurity problems, *Phys. Rev. B* **89**, 121105(R) (2014).
- [54] K. M. Stadler, A. K. Mitchell, J. von Delft, and A. Weichselbaum, Interleaved numerical renormalization group as an efficient multiband impurity solver, *Phys. Rev. B* **93**, 235101 (2016).
- [55] J. Rigo, AD-numerical-renormalization-group, 2021, <https://github.com/JonasRigo/AD-Numerical-Renormalization-Group>.
- [56] J. Rigo, Build BLAS on a TR2 NUMA platform, 2019, <https://github.com/JonasRigo/Summer-Wine/wiki/Build-BLAS-on-a-TR2-NUMA-platform>.
- [57] R. E. Wengert, A simple automatic derivative evaluation program, *Commun. ACM* **7**, 463 (1964).
- [58] J. W. Demmel, O. A. Marques, B. N. Parlett, and C. Vömel, Performance and accuracy of LAPACK's symmetric tridiagonal eigensolvers, *SIAM J. Sci. Comput.* **30**, 1508 (2008).