Numerical renormalization group

Rok Žitko

Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia and
Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia

(Dated: January 3, 2019)

A quantum impurity is a point-like object with local degrees of freedom, such as spin and orbital angular momentum, that interacts with an environment (continuum of states) in such a way that the values of local variables can change, for instance through the exchange interaction leading to spin-flip scattering. Such models have been studied since the early days of quantum mechanics in the context of decoherence and dissipation, later as effective models for magnetic dopants in metals, and more recently to describe various nanodevices such as quantum dots. Because of their non-perturbative nature, sophisticated theoretical tools are required for their full solution. Here I provide a brief description of the numerical renormalization group (NRG) method, and its “NRG Ljubljana” implementation. I discuss the key ideas of logarithmic discretization, energy-scale separation, and iterative diagonalisation, as well as some recent developments and refinements of the approach. The bibliography includes several key works in the field and hence constitutes a “recommended reading” list.

I. NUMERICAL RENORMALIZATION GROUP

The numerical renormalization group (NRG) is a numerical technique introduced by K. G. Wilson in 1975\(^1\) for solving the Kondo problem, i.e., the non-trivial behaviour of a magnetic impurity embedded in a continuum of itinerant conduction electrons. It had been observed already in 1930s that magnetic dopants in metals lead to anomalous behaviour at low temperatures, in particular to a puzzling resistance minimum: with decreasing temperature the resistivity was at first decreasing, as expected, but then it increased somewhat before saturating at the lowest temperatures. J. Kondo had shown in 1964 that this behaviour is connected to the exchange coupling between the impurity and the host electrons (i.e., spin-flip scattering), and that the perturbation theory breaks down at low temperatures\(^2\). Significant theoretical effort was consequently devoted to developing new many-body techniques beyond the perturbation theory. The Kondo problem hence represented an outstanding problem at that time and its conclusive numerical solution in 1975 marks an important milestone in the field of magnetism. Even more importantly, perhaps, the NRG solution of the Kondo problem was a pioneering implementation of the full framework of the renormalization group (RG) paradigm in the field of condensed-matter physics, which provided a new way of considering many-body problems, introducing a new set of concepts, methods, as well as a new terminology. The approach was soon extended to the single-impurity Anderson model\(^3\)–\(^5\) and other quantum impurity models, while the attempts to generalize it to bulk interacting problems such as the Hubbard model eventually led to the development of the density-matrix renormalization group (DMRG) by S. White\(^6\) which revolutionized the theoretical studies of 1D interacting systems. The DMRG and NRG thus share common origin and they are, in fact, closely related, both being based on the matrix-product-state representation of many-body eigenstates of the Hamiltonian\(^7\)\(^,\)\(^8\). The state-of-the-art of the NRG method as of 2008 was thoroughly reviewed by R. Bulla, T. Costi and...
Th. Pruschke in Ref. 9; that work still constitutes a go-to reference work for the NRG, with the exception of some very recent developments. A key reference for the general field of quantum impurity physics is the book by A. Hewson\textsuperscript{10}. This manuscript is a short introduction to the NRG, prepared for the NRG School at the Universidade Federal de Uberlândia, Brazil, in 2019.

The essence of the NRG approach is a judicious discretization of the continuum of states. Perturbative RG treatment of the Kondo problem by P. W. Anderson\textsuperscript{11–13} had shown that each decade of energy scales contributes equally to the change ("renormalization") of the effective exchange coupling ("running coupling constant"). This observation led Wilson to discretize the continuum into bins with geometrically decreasing width, $[\Lambda^{-(n+1)} : \Lambda^{-n}]$, where $\Lambda > 1$ is some parameter that controls the coarseness of the discretization. Keeping only one state in each such interval, the Hamiltonian can be transformed via a unitary transformation (equivalent to the Gram-Schmidt procedure, or the Lanczos algorithm) to a one-dimensional Hamiltonian with exponentially decreasing hopping constants, $t_n \propto \Lambda^{-n/2}$, also known as the "Wilson chain". Despite a significant simplification of the problem, this representation of the impurity problem still constitutes a challenge for numerical calculations because the Fock space of many-body states grows exponentially with the length of the chain. It turns out, however, that problems of this class have a property of "energy-scale separation", meaning that the high-energy and low-energy states are coupled by very small matrix elements whose amplitudes decrease rapidly with increasing energy separation (on logarithmic scale)\textsuperscript{1}. For this reason, it is possible to perform the diagonalization of the Hamiltonian iteratively: knowing the eigenstates for the chain of length $N$, one adds the $(N+1)$-th site, constructs the corresponding matrix representations of the Hamiltonian, diagonalizes them, then separates the states into two sets: the low-energy kept states that are retained in further steps of the calculation, and the high-energy discarded states which drop out of consideration. The physical quantities are computed with the retained set, and since on a given energy scale the properties are not influenced much by the discarded states at significantly higher energy scales, the error committed is bounded and typically small.

Despite the seeming crudeness of both the discretization and the truncation approximations, it was established that such computational procedure is nevertheless an excellent approximation for all quantum impurity problems considered so far. While the approach is by no means exact, it is frequently possible to obtain results with errors no larger than a few per mil, often at a fraction of the numerical effort required for exact simulations using the quantum Monte Carlo (QMC) techniques. It should be noted that for some particularly simple situations quantum impurity models can be solved analytically\textsuperscript{14–17}. This provides valuable benchmark results for numerical approaches and permits the validation of implementations.

The NRG allows computing thermodynamic properties (impurity free energy, entropy, heat capacity, magnetic susceptibility), expectation values of arbitrary local operators (occupancy, spin), as well as dynamics (impurity spectral function, dynamical charge and spin susceptibilities)\textsuperscript{18–25}. Recent extensions make it possible to also study time-dependent problems such as quantum quenches.

Thermodynamic properties are the easiest to calculate and require only the iterative diagonalisation of the Hamiltonian to be performed. Much information can be extracted simply from analysing the eigenspectra of the finite-length Wilson chains: for instance, one can identify the fixed points, the cross-overs between them, and even establish effective Hamiltonians that govern the physics on different energy scales. This research program was fully expounded already in the early NRG works\textsuperscript{1,4,5}, which still form an essential reading in the field. Expectation values require a straight-forward generalization: in addition to the Hamiltonian, one needs to iterate also the operators of interest. It is possible to write a simple implementation of NRG in no more than a
few tens of lines of Mathematica code, sufficient for computing thermodynamics and expectation values with reasonable efficiency. Dynamic properties are, however, significantly more difficult to compute. Several decades of algorithmic refinements and establishing the know-how about parameter tuning were required to obtain the very reliable high-resolution results that can presently be obtained using the state-of-the-art NRG codes.

II. “NRG LJUBLJANA” PACKAGE

The origin of “NRG Ljubljana” can be traced back to its author’s attempts to study the double quantum dot problem using variational ansatz approaches that (looking retrospectively) didn’t have adequate generality, leading to realization that the use of an unbiased quantum impurity solver such as the NRG is unavoidable. The first implementation was a fairly direct transcription of the ideas in Ref. 4 for the case of conserved charge and total spin quantum numbers, $Q$ and $S$, for the single-impurity Anderson model.

The design goals of NRG Ljubljana were flexibility (very few hard-coded limits) and adaptability (easy to extend). To achieve this, the implementation uses modern high-level programming paradigms (pattern matching and functional programming in Mathematica, object-oriented programming in C++). The requirement to study fairly general multidot configurations with a multitude of coupling topologies and different interaction terms led to parallel development of the package “SNEG” for performing calculations using the second-quantization algebra with the computer-algebra system Mathematica. This turned out to be significantly less error-prone to use compared to alternative approaches where the code needs to be modified and the different Hamiltonians are manually transformed into a matrix representation. SNEG hence provided some key elements: 1) generation of basis states with different symmetries, 2) language for defining the Hamiltonian of the problem and all operators of interest, 3) automatic procedure for generating the matrix representations of all quantities of interest, 4) a reliable tool to determine the transformation coefficients in the NRG iteration for different symmetry types. For this reason, SNEG can be considered as the foundation of NRG Ljubljana, although these are in principle two separate packages (for users’ convenience, a trimmed-down version of SNEG is distributed as part of NRG Ljubljana).

The first public release of the code in January 2007 had support for six symmetry types (for conserved charge or axial charge, full SU(2) or U(1) spin symmetry, and left-right mirror symmetry) for single-channel and two-channel problems. All parameters were configurable at run-time with no need for recompiling the code. To ensure numerical efficiency while maintaining flexibility and ease-of-use, the code was split into an initialisation part written in Mathematica, and a computation-intensive part written in C++ (later optimized so that all numerically intensive parts make BLAS/LAPACK calls that can make full use of multi-core CPUs through multithreaded approaches). Multiple discretization schemes were implemented with support for arbitrary energy-dependence of the conduction-band density of states. For computing dynamical properties, the density-matrix NRG (DM-NRG) algorithm devised by W. Hofstetter was generalized to all symmetry types.

Later refinements included developments in the spectral-function calculations including complete fock space (CFS) and full-density-matrix (FDM) NRG algorithms, improved discretization procedures with elimination of systematic artifacts, and a long list of different symmetry types, recently including e.g. orbital angular momentum $L$ for three-orbital problems and conserved total-angular-momentum $J$ for three-orbital problems with spin-orbit coupling. For the most demanding problems, MPI (message-passing interface) parallelization can be used to spread...
the diagonalisations across several computers.

Recent work consisted in making the package more easily deployable to modern high-performance-computing facilities, for instance by providing an easyconfig file for EasyBuild software build and installation framework, and by preparing a full installation of precompiled and optimized NRG Ljubljana in the form of containers for Docker and Singularity.

III. SNEG

A frequently used approach to write computer codes in science is to hard-code the problem definition and the quantities of interest in the same low-level programming language which is also used to implement the method of solution, i.e., the problem-domain and the solution-domain languages are the same. When the problem under study changes, such codes typically undergo successive modifications and adaptations, often leading to maintainability issues or even bugs. The solution to such difficulties is to use a domain-specific language (DSL), i.e., a specification language adapted to a particular problem domain. If the field of many-body physics such a language already exists: the language of second-quantization operators (particle creation and annihilation operators) in terms of which it is possible to fully express the problem (the Hamiltonian), the quantities of interest (the observables), and the domain of definition (the basis states defined by the creation operators applied to a vacuum state). The Mathematica package SNEG implements a DSL for second-quantization expressions and provides a convenient interface to sophisticated computation backends, such as that in NRG Ljubljana.

Basic elements in SNEG are operators that are concatenated together using non-commutative multiplications and automatically reordered according to the canonical commutation or anticommutation rules; this allows for automatic simplification of expressions and equality testing. For example, \( d^{\uparrow} \) represents \( d^{\dagger} \), i.e., the creation operator for spin-up particle in some local orbital, while \( c^{\downarrow} \) is the annihilation operator for spin-down particle in a conduction band state with momentum \( k \). The non-commutative multiply, denoted by \( \text{nc} \), respects the usual rules: it is linear in all its arguments and it is associative. The occupancy operator can be expressed as \( \text{nc}(d^{\dagger}, c^{\downarrow}) \). For efficiency, SNEG can also work in the occupation-number representation; this is particularly useful for generating matrix representations of abstract operator expressions. Calculations with Dirac bra-ket are also supported and can be intermixed with the second-quantization operators. This is convenient, for example, for describing mixed electron-phonon problems where the fermions can be described using the second-quantization language, and the oscillator using the coherent states.

SNEG includes a number of higher-level functions for generating various physically relevant operators, for example: `number` for the occupancy (charge) operator, `hop` for inter-site hopping, `hubbard` for on-site electron-electron interaction term, `spinx`, `spiny` and `spinz` for spin operators. This allows expressing the Hamiltonian and observables in a particularly compact way. There is also a number of functions for manipulating expressions: `conj` computes the Hermitian conjugate, `commutator` and `anticommutator` calculate the (anti)commutators, `vev` gives the vacuum-expectation-value. Finally, SNEG includes a number of functions for generating the sets of basis states for different symmetries, e.g., well-defined charge and total spin, \( Q \) and \( S \).

Using SNEG, it was possible to achieve a clear separation between the problem domain (coded in Mathematica and SNEG) and the solution domain (coded in C++). This is advantageous not only for reasons of performance, but especially for maintainability of the code. During the lifetime of the project, no major rewrites or design changes were necessary in either part of the code and the development could mostly proceed incrementally without breaking the existing features.
Furthermore, adapting the package to different problems and symmetries, or to calculate new quantities, is rather trivial.

IV. BASIC NOTIONS OF RENORMALIZATION GROUP

When we say “effective Hamiltonian”, what we mean is a description of system properties on the energy/temperature/length scale of our interest. The terms of the effective Hamiltonian are selected based on symmetry considerations and the relevant excitations, while the parameters are very often extracted from the experiment. For example, a magnetic insulator can be described using the Heisenberg model with the exchange coupling constant $J$ determined from the high-temperature Curie’s law behaviour of magnetic susceptibility. One would then apply this effective model to study the low-temperature magnetic phases of the system.

A systematic way of finding relations between the effective descriptions of the same physical system on different scales is known as the renormalization group (RG). The RG provides both an operational procedure for finding such relations and a framework for discussing and describing these relations. The basic idea is that the effective Hamiltonians on two different scales can be related by some transformation $T$, so that

$$H_{N+1} = T[H_N],$$

where $H_N$ is the Hamiltonian at the energy scale indexed by $N$. The transformation $T$ can be iterated, but in most cases it is not invertible, thus the mathematical structure is actually that of a semigroup rather than a group.

The key step in the analysis is the identification of the fixed points of the mapping $T$, i.e., $H^*$ such that $H^* = T[H^*]$. Assuming that a Hamiltonian $H_N$ is close to such a fixed point, we may expand as

$$H_N = H^* + \delta H_N.$$  \hspace{1cm} (2)

Then

$$H_{N+1} = T[H^* + \delta H_N] = H^* + L[H^*] \cdot \delta H_N + O[\delta H_N^2].$$  \hspace{1cm} (3)

Here $L[H^*]$ is the linearization of $T$ around the point $H^*$ in the space of Hamiltonians. This mapping admits eigenoperators $O_m$ and eigenvalues $\lambda_m$:

$$L[H^*] \cdot O_m = \lambda_m O_m.$$  \hspace{1cm} (4)

Thus the generic solution in the vicinity of $H^*$ can be expanded as

$$\delta H_N = \sum_m c_m \lambda_m^N O_m,$$  \hspace{1cm} (5)

with the coefficients $c_m$ determined by the expansion of the initial Hamiltonian $H_1$.

Clearly, if $\lambda_m > 1$, the contribution of the $m$-th term will grow. Such operators $O_m$ are known as relevant operators. They describe the flow away from the fixed point, i.e., the cross-over to a different type of behaviour (provided of course that the $m$-th term is actually present, i.e., $c_m \neq 0$). To the contrary, if $\lambda_m < 1$, the $m$-th term will become small. Such operators are called irrelevant operators. If a fixed point only has irrelevant operators, it is a stable fixed point that has universal behaviour. The special case of $\lambda_m = 1$ requires a more involved analysis based on an expansion to higher order in $\delta H_N$. For example, fixed points may form lines of fixed points if the operator is exactly marginal.
The renormalization procedure can be implemented in different ways. In bulk magnetic systems, one can e.g. perform a coarse graining such that the new spin degree of freedom in the model on a longer spatial scale is equal to the majority value in a cluster of spins in a model on a shorter spatial scale. In the impurity models, one can perform the so-called cutoff renormalization and compute the scaling equation for coupling constants using the perturbation theory. This procedure typically breaks down at low-energy scales. Finally, one can perform the renormalization using a brute-force numerical calculation, as in the NRG.

V. LOGARITHMIC DISCRETIZATION

A number of refinements over Wilson’s original discretization scheme were introduced over the years. In numerical calculations using discretized (finite-size) systems, the results can often be significantly improved by averaging over twisted boundary conditions: the calculation is performed for several realizations of the same Hamiltonian, but with different phase shifts in periodic boundary conditions. The phase shift is namely equivalent to a translation of the discrete mesh of \( k \) points in the reciprocal space, thus using a set of twist parameters is equivalent to performing a better sampling over the Brillouin zone. While this does not recover the thermodynamic limit (which corresponds to a continuous integration over the Brillouin zone), it can to a great extent cancel out some types of the finite-size artifacts.

A similar procedure is possible in the NRG. Instead of a single NRG calculation, one performs a series of calculations with the same model parameters but for a range of different discretization meshes, and averages over these partial results. In this way, the information is sampled from different energy regions in each NRG run. This approach is often referred to as \( z \)-averaging or interleaved method. The twist parameter \( z \) is defined in the interval \( (0 : 1] \), with the standard (Wilson’s) mesh corresponding to \( z = 1 \). Although this method does not restore the continuum limit, \( \Lambda \to 1 \), it is surprisingly successful in removing oscillatory features in the spectra; even averaging over only two values of \( z \) is often very beneficial. The best choice of the number of \( z \) values, \( N_z \), is a power of two, \( N_z = 2^n \): each higher power cancels out further harmonic frequency of oscillating artifacts.

It was noticed in 2008 that the discretization procedures in common use at that time had intrinsic flaws, and that there is a particularly appropriate discretization scheme that largely suppresses such artifacts\(^{38}\). Namely, it is possible to perform the discretization with twist averaging in such a way that for a non-interacting problem, with increasing the number \( N_z \) of \( z \) points, one systematically converges toward the exact continuum result. For interacting problems there is no such a-priori guarantee, but the empirical observations accumulated over the past 10 years indicate that the procedure has excellent convergence properties, and can hence be considered as the best discretization scheme currently known. In particular, this approach has paved the way to high-resolution spectra in the NRG, as well as to more reliable large-\( \Lambda \) calculations unavoidable in complex multi-channel problems. For this reason, in the following I briefly review its implementation.

The discretization mesh is defined by the points

\[
\epsilon^z_j = \begin{cases} D, & j = 1 \\ D\Lambda^{2^{-j-z}} & (j = 2, 3, \ldots), \end{cases}
\]

where \( j \) enumerates the intervals and \( z \) the different meshes. \( D \) is the half-bandwidth of the conduction band, i.e., the continuum of states has support on the interval \([ -D : D ]\). In each interval we introduce a \textit{representative energy} \( \mathcal{E}^z_j \), to be defined in the following. It can be shown\(^{31}\)
that the spectral function on the first site (commonly denoted as $f_0$) of the Wilson chain can be written as

$$A_{f_0}(\omega) = \frac{\epsilon_j^z - \epsilon_{j+1}^z}{2D|d\mathcal{E}_j^z/dz|},$$

(7)

where the arguments $z$ and $j$ on the right hand side need to be chosen such that $\mathcal{E}_j^z = \omega$, which has a unique solution. Now we can require that the discretization fulfills the following requirement:

$$A_{f_0}(\omega) = \rho(\omega),$$

(8)

where $\rho$ is the density of states (DOS) of the conduction band. If this is fulfilled, the $z$-averaging will exactly reproduce the conduction band DOS in non-interacting impurity models.

The constraint leads to a system of equations for $\mathcal{E}_j^z$:

$$\int_{I_j} \rho(\epsilon) d\epsilon \left| d\mathcal{E}_j^z/dz \right| = \rho(\omega).$$

(9)

Here $I_j$ stands for the interval $[\epsilon_{j+1}^z : \epsilon_j^z]$. The true nature of the mathematical problem becomes transparent after introducing continuous indexing as $x = j + z$ with parameter $x$ running from 1 to $+\infty$, so that $\mathcal{E}_j^x$ and $\epsilon_j^x$ both become continuous functions of $x$, $\mathcal{E}(x)$ and $\epsilon(x)$. Then

$$\frac{d\mathcal{E}(x)}{dx} = \frac{\int_{\epsilon(x)}^{\epsilon(x+1)} \rho(\omega) d\omega}{\rho[\mathcal{E}(x)]},$$

(10)

with the initial condition $\mathcal{E}(1) = D$. For the NRG approach to work, we need to impose suitable asymptotic behaviour of the hopping coefficients along the Wilson chain to implement the energy-scale separation. Specifically, we need

$$\mathcal{E}(x) = D f(x) \Lambda^{2-x},$$

(11)

where $f(x)$ has a value of order 1 for most values of $x$ (especially asymptotically for large $x$). We thus need to solve

$$\frac{df(x)}{dx} = \ln \Lambda \ f(x) - \frac{\int_{\epsilon(x+1)}^{\epsilon(x)} \rho(\omega) d\omega}{D \Lambda^{2-x} \rho[\mathcal{E}(x)]},$$

(12)

with the initial condition $f(1) = 1/\Lambda$. This ordinary differential equation can be solved numerically; some care is required since the equation is stiff.

The procedure has also been generalized to channel-mixing problems, which enables to study multi-orbital problems with non-diagonal hybridisation function, as well as superconducting channels with arbitrary energy dependence of the gap function.

VI. NRG TRUNCATION

In addition to the discretization accuracy, the quality of NRG calculations is controlled by the truncation of states during the iteration (without truncation, NRG would be an exact diagonalization of a discretized version of the Hamiltonian). As the number of retained states increases, the result converges toward a constant value (the case of spectral functions is a bit more involved, and there appears to be some degree of irreducible indeterminacy). Depending on the complexity of the problem, this convergence is more or less difficult to achieve.
It is possible to control the truncation in two ways. One can specify either the minimal number of states (symmetry multiplets) retained or the energy cutoff. The energy cutoff has to be expressed in units of the characteristic energy scale at given Wilson chain length, \( \omega_N \propto \Lambda^{-N/2} \). Typically the full convergence is achieved for energy cutoff around \( 10 \omega_N \), but adequate results can be obtained already for \( 7 \omega_N \), and even for a still lower cutoff the results tend to be qualitatively correct and follow the expected trends. Such robustness is surely a distinguishing feature of a good numerical method.

It is important not to truncate inside clusters of nearly degenerate states. The effect of such truncation is equivalent to the presence of an (unphysical) operator, which might be relevant in the RG sense, leading to (unphysical) destabilization of the fixed point of the problem and incorrect results. For the same reason, one can improve the stability of the iteration by correcting for the floating-point round-off errors of nearly degenerate states, thereby restoring the symmetry which would otherwise be lost.

VII. SYMMETRIES

The efficiency of an NRG calculation depends to a great extent on making use of the full symmetry of the problem: The CPU time requirements may differ by orders of magnitude. The symmetry that can be used is the intersection of the symmetries of the Hamiltonian and all operators of interest.

The symmetries are used through the application of the Wigner-Eckart theorem. For this reason, all numerical quantities (matrix representations of operators) in NRG Ljubljana actually consist of irreducible matrix elements. The correct inclusion of suitable Clebsch-Gordan coefficients is done in all parts of the code wherever needed, and the ensuing complexity is hidden from the user.

It is important to express all operators in the spherical tensor operator basis. For SU(2), an operator of rank \( M \) is defined through the following behavior under the action of generators \( J_\nu \) of the symmetry group:

\[
\begin{align*}
[J_z, O^M_\mu] &= \mu O^M_\mu, \\
[J_+, O^M_\mu] &= A(M, \mu) O^{M+1}_{\mu+1}, \\
[J-, O^M_\mu] &= A(M, -\mu) O^{M-1}_{\mu-1},
\end{align*}
\]

with \( A(M, \mu) = \sqrt{(M-\mu)(M+\mu+1)} \).

VIII. QUANTUM IMPURITY PROBLEMS

NRG Ljubljana comes with definitions for some most-commonly used quantum impurity models, such as the Kondo model\(^2\)\(^4\)\(^4\) or the single-impurity Anderson model (SIAM)\(^4\)\(^5\).

The Kondo model (\( \text{model=KONDO} \)) Hamiltonian is

\[
H = \sum_{k\sigma} \epsilon_k c^\dagger_{k\sigma} c_{k\sigma} + JS \cdot s,
\]

where \( S = \frac{1}{2} \sigma \) is a quantum-mechanical spin-1/2 operator, while \( s = (1/N) \sum_{kk'} c^\dagger_k (\sigma/2) c_{k'} \) is the spin density of conduction-band electrons at the position of the impurity. The perturbation theory for this model breaks down at the temperature

\[
T_K \sim D \exp(-1/\rho J),
\]

(15)
where $D$ is the half-bandwidth and $\rho$ is the DOS of the conduction band at the Fermi level. On this scale, known as the Kondo temperature, the local moment becomes screened and the impurity becomes effectively non-magnetic. One says that the conduction band electrons form a Kondo cloud around the impurity spins, forming an over-all spin-singlet many-particle state. The impurity susceptibility saturates to a constant value, $\chi \propto 1/T_K$, rather that behaving according to the Curie’s law for a free local moment, $\chi \propto 1/T$. All low-temperature properties of the system are described by universal functions of the ratio $T/T_K$. This is a manifestation of the universality which results from the flow to the stable fixed-point of the problem; the effect of non-universal irrelevant operators is progressively reduced as the energy/temperature scale is decreased.

The essence of Kondo physics is thus the progressive enhancement of the exchange coupling strength with reducing temperature. At very high temperatures, $T \gg T_K$, the local moments are essentially free (sometimes this is named asymptotic freedom), while for $T \ll T_K$ the local moments are fully screened (sometimes this is named infrared slavery$^{46}$).

More physically realistic model is the single-impurity Anderson model (model=SIAM). It provides, for example, a good description of magnetic ions and quantum dots. Its Hamiltonian is

\[
H = H_{\text{imp}} + H_{\text{band}} + H_{\text{hyb}},
\]

\[
H_{\text{imp}} = \sum_{\sigma} \epsilon n_{\sigma} + U n_{\uparrow} n_{\downarrow},
\]

\[
H_{\text{band}} = \sum_{k\sigma} \epsilon_k c_{k\sigma}^{\dagger} c_{k\sigma},
\]

\[
H_{\text{hyb}} = \sum_{k\sigma} \left( V_k c_{k\sigma}^{\dagger} d_{\sigma} + \text{H.c.} \right).
\]

Here $\epsilon$ is the impurity energy level, $U$ is the electron-electron repulsion, $n_{\sigma} = d_{\sigma}^{\dagger} d_{\sigma}$, $V_k$ are coupling matrix elements. The hybridization between the impurity level and the continuum is characterized by the hybridisation function

\[
\Delta(z) = \sum_k \frac{|V_k|^2}{z - \epsilon_k},
\]

whose imaginary part is also known as the hybridisation strength, $\Gamma(\omega) = \text{Im} |\Delta(\omega + i0^+)|$, and is commonly approximated as a constant (the “flat-band” approximation). The hybridisation function fully describes the effect of the conduction band on the impurity. This is possible because the band is non-interacting. For a flat-band with the half-bandwidth $D = 1$, one has $\Delta(z) = -\log((z - 1)/(z + 1))/(2\pi)$. The effect of external magnetic field can also be easily included in the model$^{47,48}$.

The SIAM has more complex behavior compared to the Kondo model, and a larger number of fixed points. At high temperatures, the orbital is said to be free and all four states (0, 1 spin-up, 1 spin-down, 2 electrons) are equally probable. The model then flows to either the local-moment regime (fluctuations between spin-up and spin-down states states), the valence-fluctuation regime (fluctuations between spin-up, spin-down and one of 0 or 2 states), or the frozen-impurity regime (no fluctuations, impurity in 0 or 2 state). The local-moment regime is followed by the strong-coupling regime where the impurity is fully screened.

One can show that the Kondo model and the SIAM are closely related. In fact, the Schrieffer-Wolff transformation$^{49}$ maps the SIAM onto the Kondo model with $\rho J = 8U/\pi \Gamma$. This mapping is valid around the particle-hole symmetric point of the SIAM, where $\epsilon = -U/2$ and the impurity is half-filled. This corresponds to the local-moment and strong-coupling regimes of the SIAM.
The Kondo impurity model can be generalized to high-spin ($S > 1/2$) and multiple screening channels:

$$H = \sum_i \sum_{k\sigma} \epsilon_{k\sigma} c_{k\sigma,i}^\dagger c_{k\sigma,i} + \sum_i J_i S \cdot s_i,$$

(18)

where $i$ ranges over $k$ channels. It can be shown that a single screening channel can screen one unit ($1/2$) of spin. For this reason, if $2S = k$ the impurity is fully screened, if $2S > k$ the impurity is underscreened and there is residual local moment, while for $2S < k$ the impurity is said to be overscreened and has very particular properties: we say that it is a non-Fermi liquid (NFL)\(^{50}\). Models in the different classes have very different dynamical properties\(^{51–53}\).

For describing impurity atoms adsorbed on metal surfaces, one usually needs to take into account the magnetic anisotropy described by a Hamiltonian term of the type

$$H_{\text{aniso}} = DS_z^2 + E(S_x^2 - S_y^2),$$

(19)

where $D$ and $E$ are known as longitudinal and transverse magnetic anisotropy, respectively. The spectral functions of such models exhibit, in addition to the Kondo peak (when present), spin-flip steps that correspond to the $S_z \rightarrow S_z \pm 1$ transitions (induced, for example, by a tunneling electron when the adsorbate is observed using the tip of a scanning tunneling microscope)\(^{54,55}\).

NRG Ljubljana also comes with predefined two-impurity models, including those that are relevant for studying double quantum dot systems with different coupling topologies (e.g. serial, parallel, side-coupled), as well as multi-orbital dots\(^{56–59}\).

While the NRG has been extended to quantum impurity models with a bosonic bath\(^{60,61}\), such as the paradigmatic spin-boson model (a two-level system coupled to a continuum of harmonic oscillators), NRG Ljubljana currently only supports models with a fermionic bath.

Finally, we mention that it is easy to include local vibration degrees of freedom at the impurity site, such as those in the Anderson-Holstein model\(^{62–64}\).

IX. DYNAMICAL PROPERTIES

The NRG makes it possible to calculate an arbitrary Green’s function,

$$G_{AB}(t) = \langle\langle A; B\rangle\rangle_t := -i\theta(t)\langle[A(t), B(0)]_\pm\rangle,$$

(20)

where $A$ and $B$ are local operators, $A(t) = \exp(iHt)A\exp(-iHt)$, $\langle O \rangle = \text{Tr}[\rho O]$, $\rho = \exp(-\beta H)/Z$, $Z = \text{Tr}[\exp(-\beta H)]$, and one takes $+$ (anticommutator) for fermionic operators and $-$ (commutator) for bosonic operators. We compute the Laplace transform

$$G_{AB}(z) = \langle\langle A; B\rangle\rangle_z = \int_0^\infty dt e^{zt}\langle\langle A; B\rangle\rangle_t,$$

(21)

with $\text{Im} z > 0$. For example, in the SIAM the impurity Green’s function is $G(z) = \langle\langle d; d^\dagger\rangle\rangle_z$, while the spectral function is its imaginary part:

$$A(\omega) = -\frac{1}{\pi}\text{Im} G(\omega + i0^+).$$

(22)

The Green’s function can be decomposed as

$$G(z) = \frac{1}{z - \epsilon - \Delta(z) - \Sigma(z)},$$

(23)
where $\Sigma(z)$ is known as the self-energy. The self-energy function fully describes the effect of the local interaction on the impurity at the level of single-particle excitations (two-particle excitations are described by the vertex function). The asymptotic low-frequency properties of $\Sigma$ are determined by the nature of the low-energy fixed point of the model. For SIAM with a Fermi-liquid ground state, one finds $\text{Im } \Sigma(\omega) \propto \omega^2$ at zero temperature. Checking the fulfilment of this property is a good test of the quality of the numerical results.

In NRG, dynamical quantities are computed by an application of the Lehmann representation (spectral decomposition):

$$G''_{AB}(\omega) = -\pi \sum_{nm} p_n A_{nm} B_{mn} \delta(\omega + E_n - E_m)(1 + e^{-\beta \omega}),$$

where $G''_{AB}$ indicates the spectral function associated with $\langle \langle A; B \rangle \rangle_z$, $p_n = \exp(-\beta E_n)/Z$ with $Z$ the partition function, $E_n$ and $E_m$ are the eigenenergies of eigenstates $|n\rangle$ and $|m\rangle$, and $A_{nm} = \langle n|A|m\rangle$ and $B_{mn} = \langle m|B|n\rangle$. The Green’s function at arbitrary point $z$ in the complex plane is then obtained as

$$G_{AB}(z) = \int_{-\infty}^{\infty} \frac{(-1/\pi)G''_{AB}(\omega)}{z - \omega} d\omega. \quad (25)$$

For example, the impurity spectral function in SIAM is given by

$$A(\omega) = \sum_{nm} |\langle m|d^\dagger|n\rangle|^2 \delta(\omega - E_m - E_n) \frac{e^{-\beta E_m} + e^{-\beta E_n}}{Z}. \quad (26)$$

**X. FINITE-TEMPERATURE ALGORITHMS**

In the original approach to computing dynamical quantities$^{21}$, the higher-frequency parts of the spectral function were calculated without knowing the true ground state of the system that is only obtained in the later stages of the NRG iteration. This leads to severe artifacts in the high-frequency region. The problem was solved by W. Hofstetter in Ref. 24 using the density-matrix NRG (DM-NRG) algorithm. The idea here is to compute the density matrix at the temperature of interest which contains the full information about the state of the system, and then evaluate the spectral function in an additional NRG run using the reduced density matrices instead of the simple Boltzmann weights in the Lehmann representation of the Green’s functions:

$$A(\omega) = \sum_{ijm} \left( \langle j|d^\dagger|m\rangle \langle j|d^\dagger|i\rangle \rho_{im}^{\text{reduced}} + \langle j|d^\dagger|m\rangle \langle i|d^\dagger|m\rangle \rho_{ji}^{\text{reduced}} \right) \delta(\omega - E_j + E_m). \quad (27)$$

Further refinements, the complete Fock space (CFS) and the full density matrix (FDM) algorithms, additionally improve the results of a calculation especially at finite temperatures$^{36,37,65}$. They are based on the observation that a complete basis set for the Wilson chain can be constructed using the discarded states from all steps of the NRG iteration. CFS and FDM are equivalent at $T = 0$. For $T > 0$, FDM is the recommended approach. However, both CFS and FDM are slower than the DM-NRG because the matrix elements for all states, both kept and discarded, need to be determined, and also because the expressions for the spectral functions are more complex.
XI. SPECTRAL BROADENING

The raw output in dynamical-quantity calculations are (binned) $\delta$ peaks that need to be broadened in order to obtain a smooth spectrum. This is performed by convolving the raw results using a broadening kernel. Its shape needs to be adapted to the logarithmic nature of the discretization grid. A suitable function is the log-Gaussian distribution of the form\(^{37}\)

$$ f(\omega, E) = \frac{1}{\alpha \sqrt{\pi}} \exp \left[ - \frac{\left( \log(\omega/E) - \frac{\alpha}{4} \right)^2}{2} \right]. $$

(28)

Here $\alpha$ is the broadening parameter, typically $\alpha > 0.5$ for rough calculations and $\alpha < 0.1$ for high-quality calculations. At finite temperatures, broadening with this log-Gaussian kernel is followed by a further convolution with a regular Gaussian function of width proportional to the temperature.

The choice of the broadening parameter is dictated by the value of $\Lambda$ and the number of $z$ points, $N_z$. The product $\alpha N_z$ needs to remain roughly constant when choosing the compromise between higher energy resolution and lower computational cost.

The most appropriate frequency mesh for the tabulated smooth spectra is a geometric series, $\omega_n = \omega_{\text{max}} \cdot r^{-n}$, with $r > 1$ and $n = 0, 1, \ldots$.

XII. TRANSPORT PROPERTIES OF NANODEVICES

The SIAM can be applied to describe a quantum-dot device (such as semiconductor lateral quantum dot, nanowire, metallic nanoparticle, or even a single molecule or atom) and compute its linear conductance, $G = \frac{dI}{dV}|_{V=0}$, using the Meir-Wingreen formula:

$$ G(T) = G_0 \int_{-\infty}^{\infty} d\omega \left( -\frac{\partial f}{\partial \omega} \right) (\pi \Gamma) A(\omega, T), $$

(29)

where $G_0 = \frac{2e^2}{h} = 1/12.906 \text{ k}\Omega$ is the conductance quantum, $f(\omega) = 1/(1 + e^{\beta \omega})$ is the Fermi function, and $A(\omega)$ is the impurity spectral function. In particular, one can show that at temperatures below the Kondo temperature the spectral function exhibits a peak pinned to the Fermi level with a value of $\approx 1/\pi \Gamma$ for impurity occupancy close to 1. For this reason, at low temperatures a quantum dot exhibits unitary conductance, $G \approx G_0$. For a Fermi liquid ground state, the zero-temperature conductance can also be determined by extracting the quasiparticle scattering phaseshifts directly from the finite-size RG spectra.\(^{66}\) This is a particularly accurate method.

At higher temperatures the Kondo effect is suppressed and one instead observes the physics of Coulomb blockade; the conductance is high only at points of charge degeneracy (i.e., fluctuating occupancies 0-1 and 1-2 in the SIAM).

XIII. SUPERCONDUCTING CHANNELS

The NRG can be applied not only for the case of a normal-metal continuum with a finite DOS at the Fermi level, but also for superconductors. In fact, the Wilson chain representation exists for any non-interacting (quadratic) Hamiltonian, including the reduced BCS Hamiltonian.\(^{67,68}\) This problem has been intensively studied in the context of a quantum dot embedded between two
superconductors and the resulting flow of the Josephson current in the presence of phase difference of order parameters\textsuperscript{69}.

This class of problems recently attracted significant attention due to the search for Majorana zero-energy modes which appear in a very similar setting (semiconductor in contact with superconductors, usually in the presence of an external magnetic field). The NRG is an ideally suited tool for such impurity Hamiltonians because it can access the sub-gap excitation spectrum directly and it can provide detailed information about the Yu-Shiba-Rusinov states (bound states of Bogoliubov quasiparticles trapped by the impurity due to exchange coupling).

The physics of the SIAM with superconducting leads is governed by the ratio of the superconducting gap to the Kondo temperature, $\Delta / T_K$. There are two sub-gap states: a singlet (adiabatically connected with the Kondo state in the limit $\Delta \to 0$) and a doublet (adiabatically connected with the local-moment state in the limit $\Delta \to \infty$). They cross for $\Delta \sim T_K$: this corresponds to a local quantum phase transition whereby the local moment is screened by binding a Bogoliubov quasiparticle, in rough analogy with the formation of the Kondo cloud in the conventional Kondo problem. These theoretical predictions are found to be in an excellent agreement with the experiments\textsuperscript{72–74}. Recently this approach has been successfully applied also to double quantum dot problems\textsuperscript{75–77}.

It may be mentioned in passing that another case of non-trivial continuum is that of ferromagnetic contacts, which can again be studied using the NRG\textsuperscript{78–80}.

**XIV. STRUCTURE OF INPUT FILES**

The input file to NRG Ljubljana is named \texttt{param}. It is a text file file using the keyword=value syntax. A simple example for the SIAM contains the following:

```
[extra]
U=0.01
Gamma=0.0007
delta=0

[param]
symtype=QS
discretization=Z
Lambda=2
Tmin=1e-10
keepenergy=10
keep=5000
model=siam.m
```

The block \texttt{[extra]} contains the model parameters. The block \texttt{[param]} contains the NRG algorithm parameters, specifically the symmetry type, the discretization scheme, the discretization parameter, the lowest temperature (controlling the Wilson chain length), and the truncation parameters.

The Hamiltonian is also specified in a text file. For example, the SIAM can be defined in this way (\texttt{siam.m}):

```
def1ch[1];
H1 = delta (number[d]-1) + U/2 pow[number[d]-1, 2];
```
The first line specifies that this is a 1-channel 1-impurity calculation. \( H_1 \) is the impurity Hamiltonian, \( H_c \) the hybridisation part, while \( H_0 \) is the predefined Hamiltonian of the first site of the Wilson chain.

The calculation is initiated by running the command `nrginit`. This performs the diagonalisation of the Hamiltonian on the initial cluster (typically consisting of the impurity orbitals and the first site of the Wilson chain) and computes the matrix representation of all operators of interest. The result is a file called `data`, which is a text file with tabulated data. The NRG iteration proper is started using the command `nrgrun`.

### XV. TOOLS

The NRG Ljubljana also includes a number of tools, such as a high-quality Kramers-Kronig transform tool (\( \text{kk} \)), Hilbert transform tool (\( \text{hilb} \)), averaging tools (\( \text{intavg}, \text{tdavg} \)), standalone spectral broadening tools (\( \text{broaden}, \text{bw} \)). There is also a number of post-processing scripts (mostly written in perl) for performing simple manipulations of the output from the code. These tools and scripts can be combined to implement, for example, a toolchain for performing dynamical mean-field theory (DMFT) calculations for bulk correlated problems, such as the Hubbard model.

There is also the possibility to drop the requirement of Mathematica in initializing the problem. In fact, NRG Ljubljana makes it possible to produce a “template” file named `data.in`, which is then transformed using a C++ tool to the actual input file `data` for a given set of parameters. Mathematica is hence only required to produce a new template file if the model definition or the set of operators changes.

### XVI. DYNAMICAL MEAN-FIELD THEORY

The dynamical mean-field theory has become an important application area of NRG Ljubljana in recent years, since NRG is unique in being the only impurity solver that 1) provides the spectral function directly on the real-frequency axis, and 2) is reliable both at (arbitrarily) low temperatures and also (somewhat surprisingly) high temperatures, while numerical errors seem to be well controlled in all parameter regimes. While competing continuous-time quantum Monte Carlo (CTQMC) solvers are exact\(^1\), they provide output on the Matsubara axis and require an ill-condition analytical continuation to the real-frequency axis which is highly problematic. In addition, the computation costs of QMC increase with decreasing temperature, often making it difficult to access the asymptotic low-temperature behaviour. Furthermore, the difficulties are exacerbated by the “minus sign” problem for certain models, which leads to much reduced signal-to-noise ratio in the Monte Carlo sampling.

The essence of the dynamical mean-field theory (DMFT) is the mapping of the bulk interacting problem upon an effective impurity model with self-consistently defined hybridisation function\(^2\). For example, the Hubbard model maps upon the SIAM provided that the hybridisation function is given by

\[
\Delta(z) = z + \mu - [G_{\text{loc}}^{-1}(z) + \Sigma(z)],
\]
\[
G_{\text{loc}}(z) = G_0[z + \mu - \Sigma(z)],
\]

(30)
where $\Sigma(z)$ is the impurity self-energy, $G_0$ is the non-interacting Green’s function of the lattice, and $G_{\text{loc}}$ is the local Green’s function of the interacting problem. The approach is exact in the limit of infinite dimensions, and as is a fairly good approximation in three-dimensions. In lower dimensions, non-local extensions of the DMFT become crucial.

The DMFT approach can be used, for example, to study the Mott metal-insulator transition in the Hubbard model. It was shown that the metal becomes insulating by the mechanism of shrinking width of the quasiparticle band as the interaction $U$ increases. Some recent applications of DMFT using the NRG as the impurity solver have identified the existence of resilient quasiparticles at temperatures much about the Fermi-liquid temperature scale\textsuperscript{83}, and unraveled the origin of the linear-in-$T$ resistivity at high temperatures\textsuperscript{84}.

XVII. SELF-ENERGY TRICK

In the DMFT applications, the only output from the impurity solver that is actually needed is the self-energy function $\Sigma$, rather than the impurity spectral function. Using equations of motion, it can be shown that the self-energy can be expressed as a ratio of two Green’s functions:

$$\Sigma_{\sigma}(z) = U \frac{F_{\sigma}(z)}{G_{\sigma}(z)},$$

$$F_{\sigma}(z) = \langle \langle n_{-\sigma} d_{\sigma}; d_{\sigma}^\dagger \rangle \rangle,$$

$$G_{\sigma}(z) = \langle \langle d_{\sigma}; d_{\sigma}^\dagger \rangle \rangle.$$

Interestingly, NRG artifacts usually manifest in both $F$ and $G$, and they tend to cancel out upon taking the ratio of both. Thus the results for $\Sigma$ are typically of significantly better quality than one would naively expect by considering raw $F$ and $G$ alone.

The same approach can be used to reduce the over-broadening artifacts when computing the spectral function in the context of single impurity physics\textsuperscript{85}.

XVIII. PERSPECTIVES

NRG Ljubljana is fairly feature-complete. Possible extensions could go in the direction of time-dependent problems, bosonic baths, and vertex function calculations.

---


46 Infrared and ultraviolet are here just fancy names for the low and high energy/temperature scales, respectively.


