
Machine Learning Schrieffer–Wolff Transformation and Its Applications

Yuhao Liu

Chair of Theoretical Solid State Physics
Faculty of Physics
Ludwig-Maximilians-Universität München

Yuhao Liu
Bachelor thesis

Supervisor:
Prof. Dr. Jan von Delft

Munich, 20.07.2021

Maschinelles Lernen Schrieffer–Wolff Transformation und ihre Anwendung

Yuhao Liu

Lehrstuhl für theoretische Festkörperphysik
Fakultät für Physik
Ludwig-Maximilians-Universität München

Yuhao Liu
Bachelorarbeit

Betreuer:
Prof. Dr. Jan von Delft

München, den 20.07.2021

Contents

1	Introduction and motivation	1
2	Anderson Model, Kondo Model and Schrieffer–Wolff transformation	3
2.1	Single Impurity Anderson Model (SIAM)	3
2.2	Single Impurity Kondo Model (SIKM) and SWT	5
2.3	Double Quantum Dot system (DQD)	5
3	Numerical Renormalization Group	7
3.1	Logarithmic discretization	7
3.2	From star-geometry to Wilson chain	9
3.3	Iterative diagonalization	11
3.3.1	Renormalization group transformation	11
3.3.2	Iterative diagonalization	12
3.4	Anders-Schiller basis and full density matrix (fdm)	14
4	Machine learning effective models for quantum impurity systems	16
4.1	Cost function and gradient descent	16
4.1.1	Cost function	16
4.1.2	Gradient descent	17
4.2	General procedure	17
4.3	Effective model of SIAM	18
4.3.1	Procedure	18
4.3.2	Results	19
4.3.3	Temperature for calculating partitions function	23
4.4	Effective model of 2IAM for DQD	24
4.4.1	SIKM with spin-1 as an effective model	24
4.4.2	2IKM for DQD as an effective model	24
4.4.3	Choosing the ML temperature for the next regime	28
4.5	Adopting Wilson’s truncation scheme in Model ML	29
5	Conclusion and outlook	32

Chapter 1

Introduction and motivation

Effective models are often used when only a specific part, e.g. low-temperature physics, is required so that one does not need the complicated bare model. So how to find an effective model for a given bare model is an interesting topic in physics. Traditionally, by Schrieffer–Wolff transformation (SWT) [1] one can derive a low-energy effective Hamiltonian from an original Hamiltonian of a given system perturbatively, normally up to the second order. This may lead to imperfectness of the effective Hamiltonian which could potentially cause errors between bare and effective models.

Numerical renormalization group (NRG) firstly introduced by Kenneth Wilson [2] in 1975 is an important tool to solve many-body systems in which impurities play a role. However, the NRG method was not popularly used for several reasons, especially due to the limited computational ability. In the last and this decade, the resonance of NRG takes place combining with the tensor networks techniques. Since energies can be resolved logarithmically in different scales, NRG applies a coarse-graining process and brings a high resolution.

Recently, J. B. Rigo and A. K. Mitchell [3] proposed a model machine learning approach to find effective models that can in principle be derived by a continuous renormalization group (RG) transformation from the bare model. They begin with a known effective model with unknown parameters and optimize the effective parameters by minimizing the cost function based on the partition function.

With this combination of NRG and model machine learning, Rigo and Mitchell provide two examples as proof of principle. The first bare model is single impurity Anderson model (SIAM) whose effective model is single impurity Kondo model (SIKM) with an impurity of spin-1/2. The second bare model is a two-impurity Anderson model for a double quantum dot (2IAM for DQD) system and its effective model is a SIKM with an impurity of spin-1. With the calculation of some physical properties of the effective models with optimized parameters, they show in their paper that the resulting effective parameters indeed fit the bare model perfectly for the low-temperature physics.

In this thesis, some basic information of the Anderson model and Kondo model will be introduced in Chapter 2. Then, a short discussion of NRG is presented in Chapter 3. By NRG, the bath should be firstly discretized logarithmically, then tridiagonalized from a star-geometry to a chain-geometry, and in the end diagonalized iteratively. In Chapter 4, the general procedure of the model machine learning is provided, and the two examples in Rigo and Mitchell's paper will be reproduced. Also, another effective model of 2IAM for DQD will be found so that more low-energy physics can be covered. This would be more challenging because more than one effective parameter must be optimized. It will also be shown that by the single-shell approximation the computational cost of the model machine learning can be reduced remarkably. Chapter 5 will give a summary of all the findings, problems, and a future outlook of the model machine learning.

Chapter 2

Anderson Model, Kondo Model and Schrieffer–Wolff transformation

In this thesis, single impurity models and double quantum dot (DQD) systems are going to be considered. As mentioned before, the Kondo model is an effective model of the Anderson model to describe the same low-energy physics. In this chapter, the models in the table below and the Schrieffer–Wolff transformation (SWT) will be introduced.

Bare model	Effective model
Single impurity Anderson model (SIAM) 2IAM for DQD	Single impurity Kondo model (SIKM) 2IKM for DQD

Table 2.1: Effective models of Anderson models.

In the following, the impurity of a spinful fermionic level and the bath consisting of non-interacting spinful fermions are considered. The quantum impurity has only a small number of degrees of freedom, so the impurity Hamiltonian can be diagonalized exactly. The bath Hamiltonian can be diagonalized exactly as well since the bath is non-interacting.

2.1 Single Impurity Anderson Model (SIAM)

The Hamiltonian of a general quantum impurity model is given by

$$H = H_{\text{imp}}[d_s, d_s^\dagger] + H_{\text{bath}}[c_{ks}, c_{ks}^\dagger] + H_{\text{hyb}}[d_s, d_s^\dagger, c_{ks}, c_{ks}^\dagger], \quad (2.1)$$

where d_s is the annihilation operator for the impurity Hamiltonian, c_{ks} is the annihilation operator for the bath. Here, $s = \uparrow, \downarrow$ stands for spin, and k is the momentum of bath fermion. The bath Hamiltonian H_{bath} and the hybridization Hamiltonian H_{hyb} can be expressed further as

$$H_{\text{bath}} = \sum_k \sum_{s=\uparrow, \downarrow} \epsilon_k c_{ks}^\dagger c_{ks}, \quad (2.2)$$

$$H_{\text{hyb}} = \sum_k \sum_{s=\uparrow,\downarrow} \nu_k (d_s^\dagger c_{ks} + c_{ks}^\dagger d_s) \quad (2.3)$$

with

ϵ_k : the energy of bath fermion of momentum k ,

ν_k : the coupling amplitude between the impurity level and the bath level of momentum k .

The coupling between the impurity and the bath is characterized by the hybridization function

$$\Delta(\omega) = \sum_k \nu_k^2 \delta(\omega - \epsilon_k), \quad (2.4)$$

For simplicity, a box-shaped hybridization function is chosen:

$$\Delta(\omega) = \frac{\Gamma}{\pi} \Theta(D - |\omega|). \quad (2.5)$$

The impurity Hamiltonian of SIAM has the following form:

$$H_{\text{imp}} = U n_{d\uparrow} n_{d\downarrow} + \epsilon_d (n_{d\uparrow} + n_{d\downarrow}), \quad (2.6)$$

where $n_{ds} = d_s^\dagger d_s$ is a number operator at the impurity, $U > 0$ is the local Coulomb interaction (plays a role only when double occupied), $\epsilon_d < 0$ is the energy level, Γ is hybridization strength and D is the half-bandwidth of the bath. Without loss of generality, $D = 1$ is chosen for the whole thesis. Due to Eq.(2.6), the states and the corresponding energies at impurity can be listed as follow:

State	Local energy
$ 0\rangle$	0
$ \uparrow\rangle$	ϵ_d
$ \downarrow\rangle$	ϵ_d
$ \uparrow\downarrow\rangle$	$2\epsilon_d + U$

Table 2.2: States at local Hilbert space of impurity.

Here, the particle-hole symmetry is considered so that the energy levels of empty state and double occupied state are the same.

$$0 = 2\epsilon_d + U \implies \epsilon_d = -U/2. \quad (2.7)$$

For the purpose of this thesis, only the low-energy physics part is cared about, which means only the singly occupied states are crucial, since their energy levels lay far below the empty and double occupied states. Hence, the average local occupancy of local level is $\langle n_d \rangle \approx 1$.

2.2 Single Impurity Kondo Model (SIKM) and SWT

In 1964, Jun Kondo described the Kondo effect by using the third-order perturbation theory that is an unusual low-temperature behavior for dilute magnetic alloys.[4] The Kondo model, also the so-called s-d interaction model in Kondo's original formulation, is named after him as well.

SIKM is an effective model of SIAM which only describes the singly occupied states. The general form of the Kondo model is

$$H = H_{\text{bath}} + H_{\text{exc}}. \quad (2.8)$$

The bath Hamiltonian for SIKM is the same as for SIAM in Eq.(2.2), while

$$H_{\text{hyb}} = 2J\vec{S}_d \cdot \vec{S}_0. \quad (2.9)$$

Here \vec{S}_d is the local spin operator for the impurity and

$$\vec{S}_0 = \sum_{k,k'} \sum_{s,s'} c_{ks}^\dagger \frac{1}{2} \vec{\sigma}_{ss'} c_{k's'} \quad (2.10)$$

is the spin operator acting on the first bath site with the vector of spin-1/2 Pauli matrices $\vec{\sigma} = [\sigma_x, \sigma_y, \sigma_z]$. J is the Kondo coupling strength.

By Schrieffer-Wolff transformation[1], the SIAM can be projected onto SIKM where the empty state and the double occupied state are filtered out. The Kondo coupling strength J can be expressed by parameters from SIAM. Considering the particle-hole symmetry, one gets

$$J_{\text{SW}} = \frac{8\Gamma D}{\pi U}. \quad (2.11)$$

However, as one will see later in this thesis, SIKM with J_{SW} is not the optimal parameter to describe the original low-energy regime of SIAM. By the model machine learning approach, one will find the optimal J .

2.3 Double Quantum Dot system (DQD)

In this section, the Hamiltonians of both 2IAM and 2IKM for DQD are going to be introduced.

The only difference compared with single impurity models is that now two impurities interact with the bath instead of one. So the bath Hamiltonian stays unchanged. The Hamiltonian of 2IAM for DQD has the following general form

$$H = H_{\text{bath}} + H_{\text{dots}} + H_{\text{hyb}}, \quad (2.12)$$

where H_{bath} is the same conduction band Hamiltonian as in (2.2) and $H_{\text{dots}} = \sum_{i=1,2} H_{\text{dot},i}$, $H_{\text{hyb}} = \sum_{i=1,2} H_{\text{hyb},i}$. Corresponding Eq.(2.6) and Eq.(2.3),

$$H_{\text{dot},i} = U n_{d\uparrow i} n_{d\downarrow i} + \epsilon_d (n_{d\uparrow i} + n_{d\downarrow i}) \quad (2.13)$$

$$H_{\text{hyb},i} = \sum_k \sum_{s=\uparrow,\downarrow} \nu_k (d_{is}^\dagger c_{ks} + c_{ks}^\dagger d_{is}). \quad (2.14)$$

In 2IKM for DQD, two quantum dots (two spin-1/2 as impurities) can also interact with each other. Similarly, the Hamiltonian of 2IKM for DQD can be expressed as

$$H = H_{\text{bath}} + \sum_{i=1,2} 2J_i \vec{S}_{d,i} \cdot \vec{S}_0 + J_s \vec{S}_{d,1} \cdot \vec{S}_{d,2}. \quad (2.15)$$

Due to the symmetry, the two quantum dots should be aligned in the same direction with the same strength. So one can set $J_1 = J_2$. Actually, the inter-dot spin-spin coupling is a special case of Ruderman-Kittel-Kasuya-Yosida(RKKY) [5] interaction in bulk system and J_s can be approximately given by

$$J_{\text{RKKY}} \propto \frac{64 \Gamma^2}{\pi^2 U}. \quad (2.16)$$

Same as J_{SW} , J_{RKKY} can still be optimized by machine learning approaches. See Ref. 6 for further insight of DQD.

Chapter 3

Numerical Renormalization Group

The numerical Renormalization Group (NRG) method is a tool to solve quantum impurity models non-perturbatively in which a quantum mechanical impurity with a small number of degrees of freedom couples to a non-interacting bath of fermions or bosons. This method was firstly invented by Kenneth Wilson in 1975 [2] as he was trying to solve the Kondo model. In 1982, Wilson was awarded the Nobel Prize in Physics and it was part of his Nobel Prize citation for development for the RG concept. A detailed review of this method can be found in [7]. The codes used in this thesis have been developed by Andreas Weichselbaum and Seung-Sup Lee. The lecture materials of tensor networks are also utilized. [8]

In Wilson's scheme of NRG method, the band of a bath should be discretized logarithmically to resolve the low-energy regime. By tridiagonalization, the Hamiltonian with a star-geometry can be mapped on a tight-binding Wilson chain. Then this Hamiltonian should be diagonalized iteratively to solve the Wilson chain numerically. In the following sections, this process will be discussed in detail with the example of SIAM from Sec. 2.1.

3.1 Logarithmic discretization

Recall that the hybridization function is box-shaped which should be now discretized logarithmically. Choose a logarithmic discretization parameter $\Lambda > 1$ and separate the band into the intervals $I_{\pm n}$ with $n \in \mathbb{N}$:

$$I_{+n} = [\Lambda^{-n}, \Lambda^{-n+1}] \quad I_{-n} = [-\Lambda^{-n+1}, \Lambda^{-n}]. \quad (3.1)$$

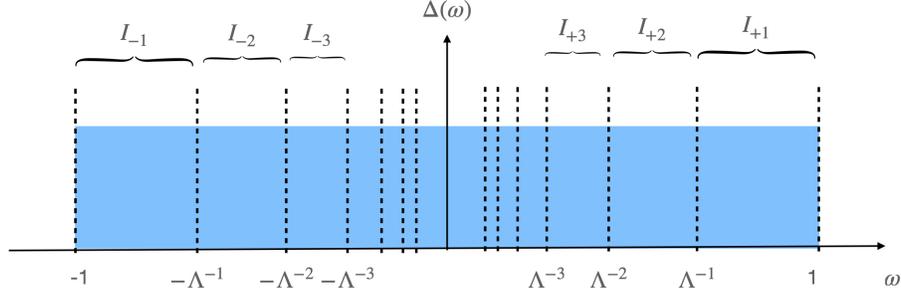


Figure 3.1: Discretized intervals of a box-shaped hybridization function, modified from Fig. 1a) in Ref. 7

In each interval, there can be found a representative state $|\pm n\rangle$ with its energy $\xi_{\pm n}$ as representative energy of this interval. Hence, the Hamiltonian of SIAM can be rewritten approximately as

$$H_{\text{disc}} = H_{\text{imp}}[d_s, d_s^\dagger] + H_{\text{bath}}^{\text{star}}[a_{\pm n, s}, a_{\pm n, s}^\dagger] + H_{\text{hyb}}^{\text{star}}[d_s, d_s^\dagger, a_{\pm n, s}, a_{\pm n, s}^\dagger], \quad (3.2)$$

$$H_{\text{bath}}^{\text{star}} = \sum_{\pm n} \sum_{s=\uparrow, \downarrow} \xi_{\pm n} a_{\pm n, s}^\dagger a_{\pm n, s}, \quad (3.3)$$

$$H_{\text{hyb}}^{\text{star}} = \sum_{\pm n} \sum_{s=\uparrow, \downarrow} \gamma_{\pm n} (d_s^\dagger a_{\pm n, s} + h.c.). \quad (3.4)$$

Here, $a_{\pm n, s}$ is the annihilation operator[7] for the representative state $|\pm n\rangle$ in the interval $I_{\pm n}$. More precisely, $a_{+n, s}^\dagger$ can be regarded as a particle-like excitation while $a_{-n, s}^\dagger$ as a hole-like excitation. The continuous hybridization function (2.4) becomes also discrete with the new discrete coupling constant $\gamma_{\pm n}$ for the whole interval $I_{\pm n}$:

$$\Delta(\omega) \approx \sum_{\pm n} \gamma_{\pm n}^2 \delta(\omega - \gamma_{\pm n}). \quad (3.5)$$

$$\gamma_{\pm n}^2 = \int_{I_{\pm n}} d\omega \Delta(\omega) \quad (3.6)$$

Using the Campo-Oliveira scheme[9], the representative energies can be defined as

$$\xi_{\pm n} = \frac{\int_{I_{\pm n}} d\omega \Delta(\omega)}{\int_{I_{\pm n}} d\omega \frac{\Delta(\omega)}{\omega}}. \quad (3.7)$$

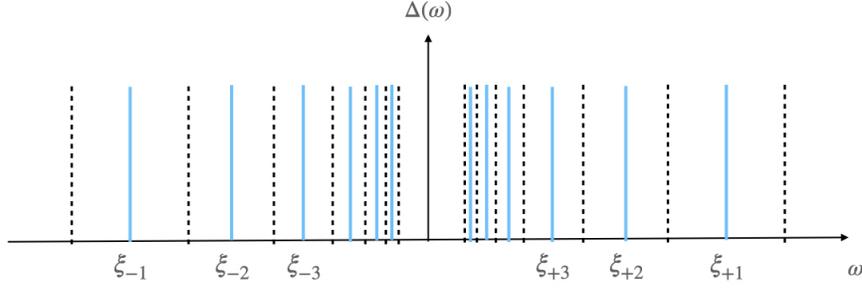


Figure 3.2: Representative states with energies $\xi_{\pm n}$ in $I_{\pm n}$, modified from Fig. 1b) in Ref. 7

According to the box-shaped hybridization function (2.4), the integrals can be easily calculated, which results

$$\gamma_{\pm n}^2 = \Lambda^{-n}(1 - \Lambda^{-1}) \sim \Lambda^{-n} \quad (3.8)$$

$$\xi_{\pm n} = \pm \frac{1}{2} \Lambda^{-n}(1 + \Lambda^{-1}) \sim \Lambda^{-n}. \quad (3.9)$$

One can see, both the coupling constants and the representative energies decrease exponentially.

3.2 From star-geometry to Wilson chain

The discrete hybridization Hamiltonian $H_{\text{hyb}}^{\text{star}}$ (see Eq.(3.4)) means that each representative state can interact with the impurity. This can be seen as a star-geometry.

$$H_{\text{star}} = H_{\text{hyb}}^{\text{star}} + H_{\text{bath}}^{\text{star}} = \begin{pmatrix} d & a_{+1} & \dots & a_{+N/2} & a_{-1} & \dots & a_{-N/2} \\ d^\dagger & 0 & \gamma_{+1} & \dots & \gamma_{+N/2} & \gamma_{-1} & \dots & \gamma_{-N/2} \\ a_{+1}^\dagger & \gamma_{+1} & \xi_{+1} & & & & & \\ \vdots & \vdots & & \ddots & & & & \\ a_{+N/2}^\dagger & \gamma_{+N/2} & & & \xi_{+N/2} & & & \\ a_{-1}^\dagger & \gamma_{-1} & & & & \xi_{-1} & & \\ \vdots & \vdots & & & & & \ddots & \\ a_{-N/2}^\dagger & \gamma_{-N/2} & & & & & & \xi_{-N/2} \end{pmatrix}.$$

By tridiagonalization using Lanczos's method[10], this star-geometry Hamiltonian can be transformed into a semi-finite tight-binding Wilson chain Hamiltonian with

$$H_{\text{chain}} = H_{\text{bath}}^{\text{chain}} + H_{\text{hyb}}^{\text{chain}} = \begin{matrix} & d & f_0 & f_1 & f_2 & \dots & f_N \\ \begin{matrix} d^\dagger \\ f_0^\dagger \\ f_1^\dagger \\ f_2^\dagger \\ \vdots \\ f_N^\dagger \end{matrix} & \begin{pmatrix} 0 & t_{\text{imp}} & & & & & \\ t_{\text{imp}} & \epsilon_0 & t_0 & & & & \\ & t_0 & \epsilon_1 & t_1 & & & \\ & & t_1 & \epsilon_2 & \ddots & & \\ & & & \ddots & \ddots & & \\ & & & & & & \epsilon_N \end{pmatrix} & \cdot \end{matrix}.$$

Here, all the interactions with the impurity from the star-geometry are summarized into the interaction between the impurity and the first bath site (with index 0):

$$t_{\text{imp}} f_{0s} \equiv \sum_{n=1}^{N/2} \gamma_{+n} a_{+n,s} + \gamma_{-n} a_{-n,s} \quad (3.10)$$

with normalization constraint

$$\{f_{ls}^\dagger, f_{l's'}\} = \delta_{ll'} \delta_{ss'} \text{ for } l = 0, \dots, N. \quad (3.11)$$

Indeed, due to the particle-hole symmetry, the on-site energies ϵ_l are all zeros, since they are all non-interacting. Hence, the bath and hybridization Hamiltonian of the chain can be rewritten as

$$H_{\text{bath}}^{\text{chain}} = \sum_{l \geq 0} \sum_{s=\uparrow, \downarrow} t_l (f_{l,s}^\dagger f_{l+1,s} + h.c.), \quad (3.12)$$

$$H_{\text{hyb}}^{\text{chain}} = \sum_s t_{\text{imp}} (d_s^\dagger f_{0,s} + h.c.). \quad (3.13)$$

Note that the discretization and tridiagonalization act only on bath and hybridization Hamiltonian, but do not act on impurity Hamiltonian. Adding the impurity at site $l = -1$, the whole Wilson chain Hamiltonian can be expressed by

$$H_{\text{Wilson}} = H_{\text{imp}} + H_{\text{chain}}. \quad (3.14)$$

A remarkable result is that the coupling strength between nearest-neighbouring sites decay exponentially along the chain

$$t_l \sim \Lambda^{-l/2}. \quad (3.15)$$

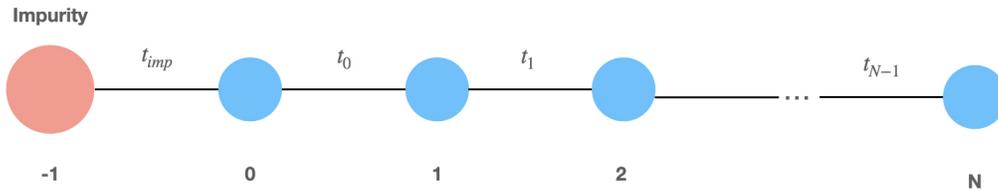


Figure 3.3: Semi-finite tight-binding Wilson chain with an impurity at site $l = -1$, modified from Fig. 3 in Ref. 7

Hence, all energy scales are equally important, though the contribution of perturbations of later sites becomes weaker and weaker.

3.3 Iterative diagonalization

In the next step, the Wilson chain should be diagonalized iteratively and the energies can be resolved in a coarse-graining way. The center scheme of the iterative diagonalization in the renormalization group (RG) context is going to be discussed. In this and the following sections, the matrix product state (MPS) method will be used. A mathematical introduction in MPS [11] is recommended for further reading.

In the following, the RG steps are going to proceed on the Hamiltonian, then the rescaled Hamiltonian will be iteratively diagonalized.

3.3.1 Renormalization group transformation

Recall the Wilson chain Hamiltonian with chain length N :

$$H^l = H_{\text{imp}} + \sum_{l \geq 0} \sum_{s=\uparrow, \downarrow} t_l (f_{l,s}^\dagger f_{l+1,s} + h.c.) + \sum_s t_{\text{imp}} (d_s^\dagger f_{0,s} + h.c.) \quad (3.16)$$

and that the coupling strength t_l decays exponentially as shown in Eq.(3.15). Now define the rescaled coupling strength

$$\tilde{t}_l = \Lambda^{l/2} t_l, \quad (3.17)$$

so that all the rescaled coupling strengths \tilde{t}_l are in the order of $\mathcal{O}(1)$. Therefore, lowest energy splitting of H^l is in $\mathcal{O}(\Lambda^{-(l-1)/2})$. To continue to resolve this lowest energy splitting, rescale the Hamiltonians and shift the ground state energies equal to zeros:

$$\tilde{H}^l = \Lambda^{(l-1)/2} (H^l - E_g^l). \quad (3.18)$$

This is the standard RG procedure: set the ground state energy to zero, then rescale the whole Hamiltonian to $\mathcal{O}(1)$. Adding one more site, the rescaled Hamiltonian \tilde{H}^{l+1} can be expressed by \tilde{H}^l :

$$\begin{aligned} \tilde{H}^{l+1} &= \Lambda^{l/2} (H^{N+1} - E_g^{l+1}) \\ &= \Lambda^{l/2} (H^l - E_g^l) + \Lambda^{l/2} t_l \sum_s (f_{l,s}^\dagger f_{l+1,s} + h.c.) - \Lambda^{l/2} (E_g^{l+1} - E_g^l) \\ &= \Lambda^{1/2} \tilde{H}^l + \sum_s \tilde{t}_l (f_{l,s}^\dagger f_{l+1,s} + h.c.) - \delta E^{l+1}. \end{aligned} \quad (3.19)$$

This can be viewed as a so-called RG transformation T of \tilde{H}^l :

$$\tilde{H}^{l+1} = T(\tilde{H}^l). \quad (3.20)$$

3.3.2 Iterative diagonalization

Finally, the scheme of iterative diagonalization is to be introduced. Assume that the chain until site l is already diagonalized numerically with eigenstates $|\alpha\rangle_l$:

$$\tilde{H}^l |\alpha\rangle_l = \tilde{E}_\alpha^l |\alpha\rangle_l. \quad (3.21)$$

Here \tilde{H}^l and \tilde{E}_α^l are the rescaled Hamiltonian and eigenenergy. According to Eq.(3.20), multiply a factor $\Lambda^{1/2}$ to \tilde{H}^l , then add the next site $|\sigma_{l+1}\rangle$ with bond dimension d and get the enlarged Hilbert space $\mathcal{H}^{l+1} = \text{span}\{|\alpha\rangle \otimes |\sigma_{l+1}\rangle\}$. Now, our new Hamiltonian looks almost the same as Eq.(3.20) until the ground state energy shift.

$$\tilde{H}^{l+1} = \Lambda^{1/2} \tilde{H}^l + \sum_s \tilde{t}_l (f_{l,s}^\dagger f_{l+1,s} + h.c.) + \Lambda^{1/2} E_g^l = \tilde{H}^{l+1} + \Lambda^{1/2} E_g^{l+1} \quad (3.22)$$

After diagonalizing this new rescaled Hamiltonian \tilde{H}^{l+1} (s. Eq.(3.20)), the new eigenstate $|\beta\rangle$ can be written as a linear combination of this basis:

$$|\beta\rangle_{l+1} = \sum_{s, \sigma_{l+1}} |\sigma_{l+1}\rangle |\alpha\rangle_l A_\beta^{\alpha, \sigma_{l+1}} \quad (3.23)$$

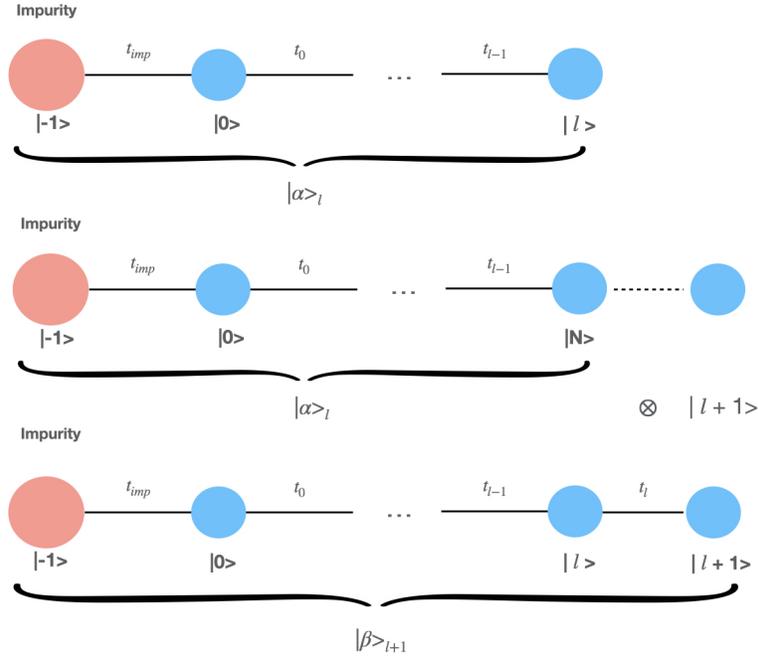


Figure 3.4: One site of the chain is added in each iteration, modified from Fig. 2 in Ref. 7

Then the ground state energy of this iteration should be shifted to zero. Obviously, the dimension of the Hilbert space becomes larger every time when one more site is added

to the Wilson chain. Due to the limited computational ability, in each iteration only the lowest N_{keep} states are kept, all others should be discarded:

$$|\beta\rangle_{l+1} = \begin{cases} |\beta\rangle_{l+1}^D, & \text{lowest } N_{\text{keep}} \text{ states} \\ |\beta\rangle_{l+1}^K, & \text{otherwise} \end{cases} \quad (3.24)$$

A short summary of the iteration diagonalization scheme is shown below (also see fig.(3.5)):

- (a) Use the kept rescaled eigenenergies \widetilde{E}_l from last iteration and make sure that the ground state energy is zero.
- (b) Scale the eigenenergies \widetilde{E}_l by a factor $\Lambda^{1/2}$.
- (c) Enlarge the Hilbert space by adding a new site and diagonalize the new Hamiltonian.
- (d) Shift the new ground state energy to zero and keep only the lowest N_{keep} states. Then repeat this procedure starting from step (a).

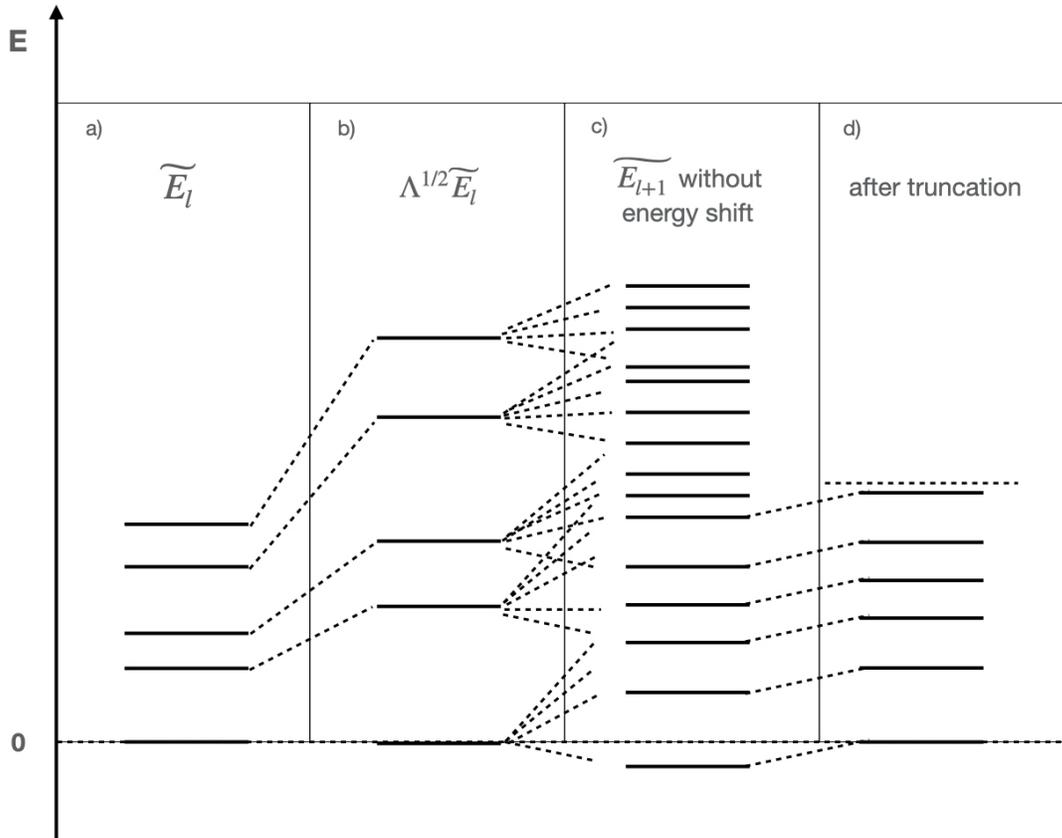


Figure 3.5: Modified from Fig. 3 of Ref. 7: Procedure of iterative diagonalization.

One might ask why the standard diagonalization (not iteratively) would not work here. The reason is that the dimension of the whole Wilson chain scales with the length of the chain, namely proportional to d^N . Hence, the Hamiltonian is too large to diagonalize at once. By the iterative diagonalization with truncation, the number of states are manageable. There are some more advantages of this method. Thank the logarithmic discretization, information from all energy scales is obtained. Also, the small energies can be resolved very well.

However, there is also a problem with this method: since all the discarded states are thrown away in each iteration, the kept states do not build a complete basis, which causes the failure when the spectral function is calculated using Lehmann's representation. Hence, it would be helpful, if there is a scheme to get a complete basis set. Indeed, this will be introduced in the next section.

3.4 Anders-Schiller basis and full density matrix (fdm)

In 2005, Frithjof Anders and Avraham Schiller introduced a brilliant idea [12][13] to build a complete basis set from the discarded states which were thrown away by Wilson in the iterative diagonalization. With this approach, one can construct the full density matrix. More importantly for this thesis, one can construct the partition function of the system under a specific temperature.

Define the environmental states $|e_l\rangle$ at the N th. iteration of NRG:

$$|e_l\rangle = \bigotimes_{n=l+1}^N |\sigma_n\rangle, \quad (3.25)$$

which can be combined with the shell state $|\alpha\rangle_l^X$ with $X \in \{K, D\}$ onto the Hilbert space of the whole Wilson chain:

$$|\alpha, e\rangle_l^X \equiv |e_l\rangle |\alpha\rangle_l^X. \quad (3.26)$$

In the references mentioned above, it is proved that

$$B = \{|\alpha, e\rangle_l^D\} \quad (3.27)$$

forms a complete basis set, known as 'Anders-Schiller (AS) basis'. In this thesis, the proof is spared. By construction, the states $|\alpha, e\rangle_l^D$ are orthogonal to each other since they are in different iterations. These basis states are approximately eigenstates of the Hamiltonian of the whole chain:

$$H^N |\alpha, e\rangle_l \approx H^l |\alpha, e\rangle_l = E_\alpha^l |\alpha, e\rangle_l. \quad (3.28)$$

Here the 'NRG approximation' is made: when acting on states from shell l , H^N is approximated by H^l . Neglecting the later sites is acceptable since they describe finer structures which are small contributions compared with shell l . With this approach, the full density

matrix can be constructed by the basis states and their approximated eigenenergies in each shell:

$$\rho_{[l]} \equiv \frac{\exp(-\beta H^N)}{Z} \approx \sum_l \sum_{\alpha e} |\alpha, e\rangle_l^D \frac{\exp(-\beta E_\alpha^l)}{Z} {}^D_l \langle \alpha, e|. \quad (3.29)$$

For the purpose of this thesis, it is important to know how the partition function Z results from this construction. The partition function is crucial for the model machine learning approach since the partition functions of bare and effective models are going to be matched later and therefore the cost function depends on the partition function Z . The density matrix ρ , as shown in Eq.(3.29), can be simplified to a reduced density matrix $\bar{\rho}$ where the environmental states are contracted. Since for shell l nothing happens yet in the later shell, one can trace out all the later shells to define the reduced density matrix for shell l :

$$\bar{\rho}_{[l]X}^{X'} = \text{Tr} \left\{ \rho_{[l]X}^{X'} \right\}_{\text{shell} > l}. \quad (3.30)$$

Here $\rho_{[l]X}^{X'} = P_l^{X'} \rho P_l^X$ is sector projection of ρ for shell l . Hence, Eq.(3.29) can be rewritten as

$$\rho_{[l]} \approx \sum_l \omega_n \rho_{[l]D}^D, \quad (3.31)$$

where Z is the partition function, $\omega \equiv d^{N-l} Z_l^D / Z$ and $Z_l^D \equiv \sum_\alpha^D \exp(-\beta E_\alpha^l)$. See [14] for more details about the full density matrix method in NRG (fmdNRG).

Chapter 4

Machine learning effective models for quantum impurity systems

In this chapter, the method of model machine learning in Ref. 3 is going to be implemented. A machine learning approach that optimizes an effective impurity model based on partition functions will be introduced. The effective impurity model will be explored to describe the low-energy physics of the bare model. As proof of principle, SIAM and 2IAM for DQD will play the role of bare models and the effective models of them will be found.

4.1 Cost function and gradient descent

In this section, two fundamental concepts in machine learning will be discussed, namely the cost function and gradient descent[15]. These are crucial for the model machine learning approach in this thesis.

4.1.1 Cost function

A cost function (also called a loss function) is used to learn parameters that explain the data well and defines how costly the mistakes are[15]. What machine learning steadily does after choosing the proper cost function, is trying to minimize the cost function. Ideally, a cost function should be a convex function to make sure that a local minimum of the cost function is also the only global minimum. If a cost function is not convex, then one has to pay attention to check whether this local minimum is the global minimum, which means it is smaller than all other local minima. An RG-derivable effective model has a Hamiltonian of the following form as shown in Ref. 3:

$$H_{\text{eff}} = \sum_i \theta_i h_i, \quad (4.1)$$

where θ_i 's are the effective parameters and h_i 's are the effective operators. In this thesis, the effective operators are taken as granted and the job is to find the optimal effective

parameter by tuning them iteratively. As shown in Ref. 3, the cost function is chosen to be dependent on the partition functions of the bare and effective models:

$$L_Z = [\log(Z_{\text{eff}}) - \log(Z_{\text{bare}})]^2. \quad (4.2)$$

The free energy F is convex and defined by

$$F = -kT \log(Z), \quad (4.3)$$

which ensures the cost function is always convex in each effective parameter effectively. Thus, this cost function has only one local minimum.

4.1.2 Gradient descent

The gradient of a multi-variable function points to the direction of the steepest ascent where the gradient is evaluated. So if this point goes in the opposite direction of the gradient (adding a minus sign in front of the gradient) for the same step size (also called learning step) $\alpha > 0$, the cost function will decrease. Iteratively applying this procedure, the minimal value of the cost function will be achieved. The gradient of the cost function in Eq.4.2 can be calculated as in Ref.[3]:

$$\frac{\partial L_Z}{\partial \theta_i} = 2 \left[\log(Z_{\text{eff}}) - \log(Z_{\text{bare}}) \right] \cdot \frac{\frac{\partial Z_{\text{eff}}}{\partial \theta_i}}{Z_{\text{eff}}} = 2 \cdot \left[\log(Z_{\text{eff}}) - \log(Z_{\text{bare}}) \right] \cdot (-\beta \langle h_i \rangle), \quad (4.4)$$

where $\langle h_i \rangle$ is the expectation value of h_i in eigenbasis of the effective Hamiltonian. Note that the proportionality constant is not important for the gradient descent, since it can be absorbed into the learning step. Hence, the new parameter after each gradient descent iteration is

$$\theta_{i,\text{new}} = \theta_i - \tilde{\alpha} \cdot \frac{\partial L_Z}{\partial \theta_i} = \theta_i + \alpha \cdot \left[\log(Z_{\text{eff}}) - \log(Z_{\text{bare}}) \right] \cdot \frac{\langle h_i \rangle}{T}. \quad (4.5)$$

4.2 General procedure

In this section, a general procedure to learn an effective model for a quantum impurity system will be given. In specific examples introduced later, the procedure may differ a little, but the main steps do hold.

First, one has to generate the bare model, iteratively diagonalize it via NRG and calculate its partition function for a proper low temperature¹ by NRG as shown in Sec. 3.4. Then beginning with some random numbers for the effective parameters, tune them by applying the gradient descent (see Eq.(4.5)) iteratively until the cost function arrives its only minimum. Alternatively, one can try to get ‘approximately optimal’ effective parameters and start with them, then proceed with the gradient descent. Actually, this alternative is

¹How to choose this low temperature depends on specific problems which are introduced in Sec. 4.3.3

better than the previous one. Since the starting parameters are already somehow physically meaningful, the chance to get nonsensical results is much lower. Also, the speed of convergence to the minimum is normally faster. Hence, one should at first try to use this alternative. If one cannot find such ‘approximately optimal’ parameters, then just use the random numbers. After the optimal parameters are achieved, it is meaningful to calculate some physical properties of both effective and bare models and compare them. The effective model should correspond to the low-energy part of the bare model. In the end, it is worth thinking about whether there is any improvement for the algorithm or some other better ideas for some specific problem. To summarize, the steps above are listed below:

1. Generate the bare model and calculate its partition function at a proper low temperature.
2. Begin with physically meaningful parameters for the effective model. Or just begin with some random parameters.
3. Repeat gradient descent to find the optimal parameters by using Eq. (4.5).
4. Compare the physical properties of bare and effective model.
5. Improvement and discussion.

4.3 Effective model of SIAM

In this section, as proof of principle, the effective model of a given SIAM will be found.

4.3.1 Procedure

Go through the steps mentioned in the last section (see Sec. 4.2):

1. In Chapter 3, it was in detail introduced how to solve a SIAM by logarithmic discretization and map it from star-geometry to Wilson chain which should be iteratively diagonalized using NRG. With the help of the Ander-Schiller basis as shown in Sec. 3.4, the partition function can also be calculated by using the fdmNRG approach. As mentioned before, a proper low temperature should be chosen. Here, we choose the Kondo temperature, since at the Kondo temperature the crossover between the strong-coupling and local-moment regimes takes place. By Bethe ansatz, the Kondo temperature can be calculated as in Ref. 16

$$T = T_K = \sqrt{\frac{U\Gamma}{2}} \exp\left(-\frac{\pi U}{8\Gamma} + \frac{\pi\Gamma}{2U}\right) \quad (4.6)$$

2. It is known that the SIKM is the effective model of SIAM if both models share the same bath as shown in Ref. 3

$$H_{\text{eff}} = \sum_i \theta_i h_i = 2J\vec{S}_d \cdot \vec{S}_0 \quad (4.7)$$

with $i = 1$, $\theta_i = 2J$ and $h_i = \vec{S}_d \cdot \vec{S}_0$. Using Schrieffer–Wolff transformation[1] (see Section 2.2) the SIAM can be approximately mapped to a Kondo model at low temperature with $J_{\text{SW}} = \frac{8\Gamma D}{\pi U}$ that is a good starting point for J .

3. In SIKM, the expectation value of effective operator can be calculated with the help of the reduced density matrix within the Hilbert space of the impurity ρ_{imp} , where the bath sites are all traced out. As shown in Ref. 3,

$$\langle h_i \rangle = \langle \vec{S}_d \cdot \vec{S}_0 \rangle = \text{Tr} \left(\rho_{\text{imp}} (\vec{S}_d \cdot \vec{S}_0) \right). \quad (4.8)$$

Rewrite Eq.(4.5)

$$J_{n+1} = J_n - \alpha \cdot \frac{\partial L_Z}{\partial J_n} = J_n + \alpha \cdot \left[\log(Z_{\text{SIKM}}|_{J=J_n}) - \log(Z_{\text{SIAM}}) \right] \cdot \frac{\langle \vec{S}_d \cdot \vec{S}_0 \rangle|_{J=J_n}}{T} \quad (4.9)$$

and repeat the gradient descent step until the minimum of the cost function (Eq. 4.2) is reached².

4. Calculate the impurity entropy of SIAM and SIKM as a function of temperature and see whether the low-energy part matches, which means the strong-coupling and local-moment regimes should be the same.

4.3.2 Results

In the following, two examples of SIAM will be shown as results.

Example 1

Example 1 is a reproduction of an example in Ref. 3. Fix some parameters for the SIAM: local Coulomb interaction $U = 0.3$, the half-bandwidth of the bath $D = 1$, the coupling strength of the box-shaped hybridization function $\Gamma = \pi \cdot 3 \times 10^{-3}$, discretization parameter $\Lambda = 2.5$, Wilson chain length $N = 50$, number of states to be kept $N_{\text{keep}} = 300$ and $\epsilon_d = -U/2$ due to particle-hole symmetry.

One can see in Fig.4.1(a) below that the cost function decreases with the iteration step and arrives at zero in the end while the cost gradient also becomes zero. To extend the range of J , cost functions and cost gradients of some extra points are also calculated and plotted to visualize the minimum as shown in Fig.4.1(b). The cost gradient perfectly crosses the zero-value at J_{ML} , since the cost gradient equal to zero corresponds to the minimum of the cost function.

² α should be positive. However, because of an unknown reason, α used in this thesis is always negative. Probably there is a mistake in the derivative, but we did not find it.

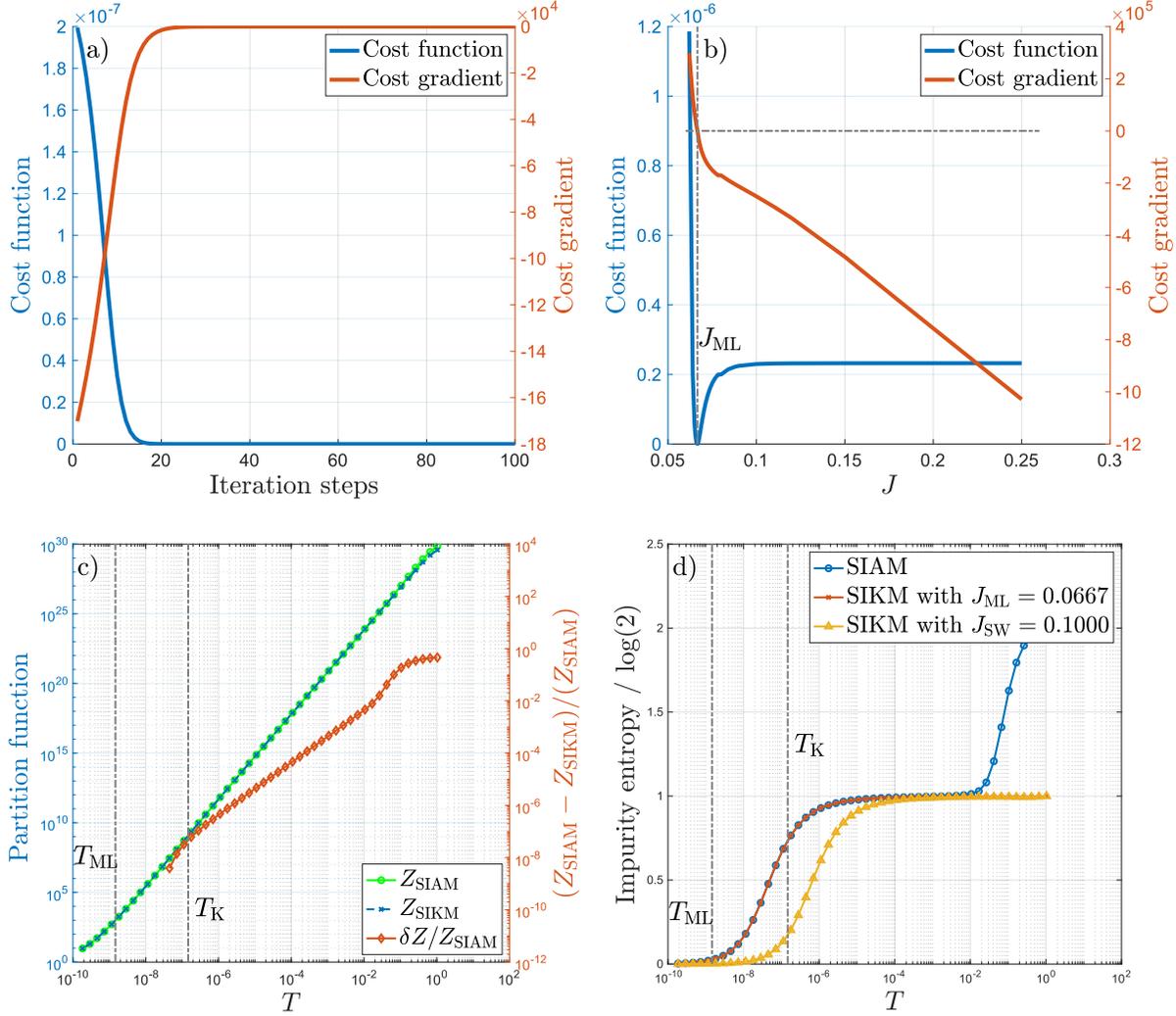


Figure 4.1: Model machine learning process of SIKM from the SIAM with $U = 0.3$ and $\Gamma = \pi \cdot 3 \times 10^{-3}$. The effective parameter results as $J_{ML} = 0.0667$. (a) Cost function and cost gradient during the learning process. Both of them go towards zero. (b) Cost function and cost gradient as functions of J . The minimum of the cost function corresponds to the cost gradient of zero. (c) The partition functions of SIAM and SIKM with optimal $J = J_{ML}$ as well as the relative difference of partition functions. (d) Impurity contribution to the entropy of SIAM and SIKM with J_{ML} as well as with J_{SW} , modified from Fig. 2a) in Ref. 3. SIKM with J_{ML} overlaps perfectly on SIAM for the first two regimes from the left.

The partition functions of SIAM and SIKM with J_{ML} match for low temperature as shown in Fig.4.1(c) very well. There is no difference between them around T_{ML} and first from about $T = 10^{-7}$ the difference is not zero. Nevertheless, the relative difference is still very small and increases with the temperature. In Fig. 4.1(d) the impurity contribution to

the entropy is plotted. One can see SIKM with J_{SW} has approximately the same low-energy physics while the impurity entropy is perfectly reproduced by J_{ML} .

Example 2

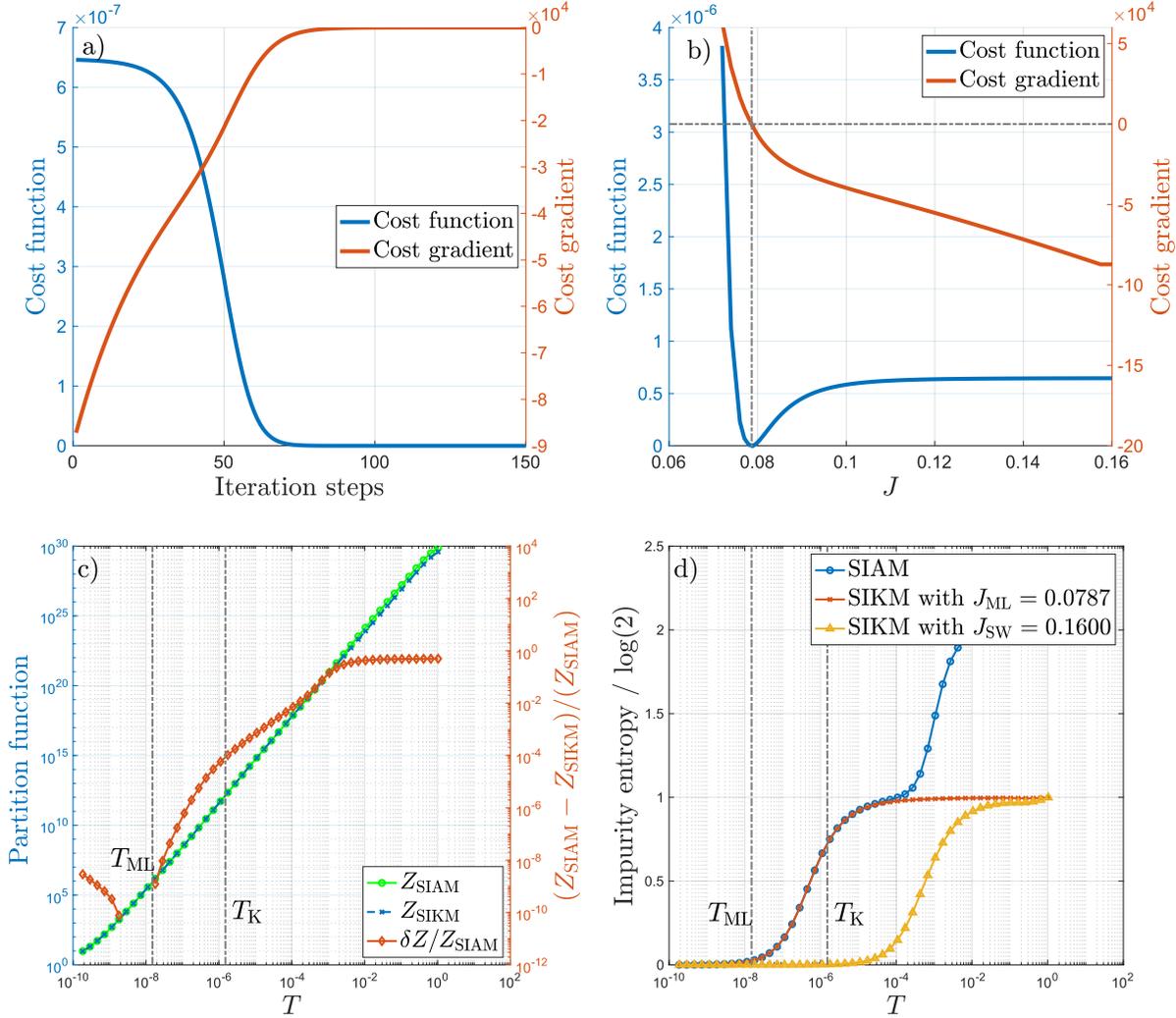


Figure 4.2: Model machine learning process of SIKM from the SIAM with $U = 0.004$ and $\Gamma = \pi \cdot 8 \times 10^{-5}$. The effective parameter results as $J_{\text{ML}} = 0.787$. (a) Cost function and cost gradient during the learning process. Both of them go towards zero. (b) Cost function and cost gradient dependent on J with some extra values of J . The minimum of J corresponds the cost gradient of zero. (c) The partition functions of SIAM and SIKM with optimal $J = J_{\text{ML}}$. They are almost the same for low temperatures. (d) Impurity contribution to the entropy of SIAM and SIKM with J_{ML} as well as with J_{SW} . SIKM with J_{ML} matches perfectly on SIAM for the first two regimes from left.

Using another set of parameters, the algorithm works as well. Choose $U = 0.004, \Gamma = \pi \cdot 8 \times 10^{-5}$, all other parameters stay the same as in Example 1. One gets results shown in Fig.4.2 above. They are similar to the results in Example 1.

Spectral functions

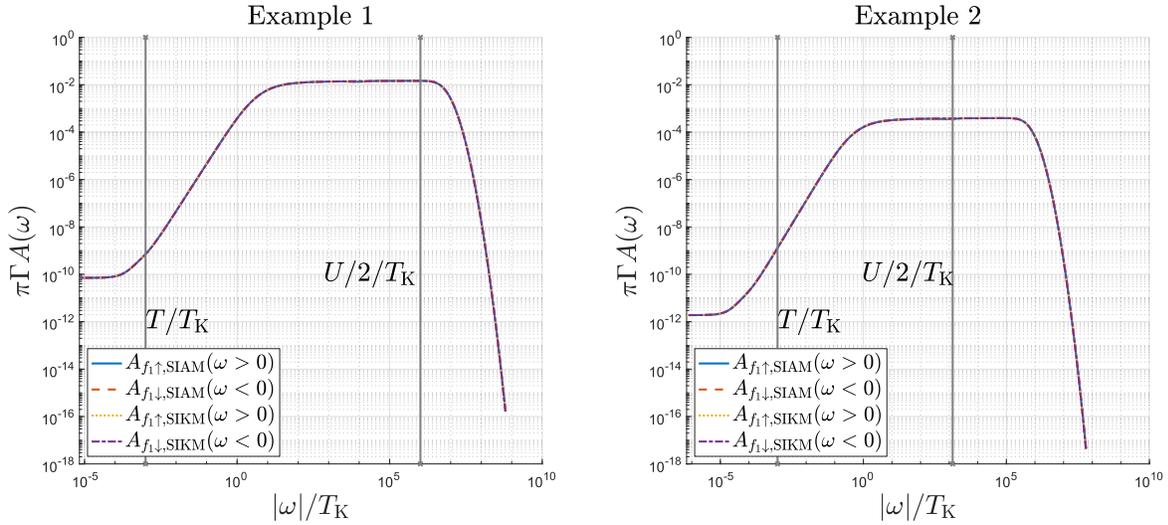


Figure 4.3: Spectral function of the annihilation operator $f_{1,\uparrow}$ for a spin-up electron at the first bath site.

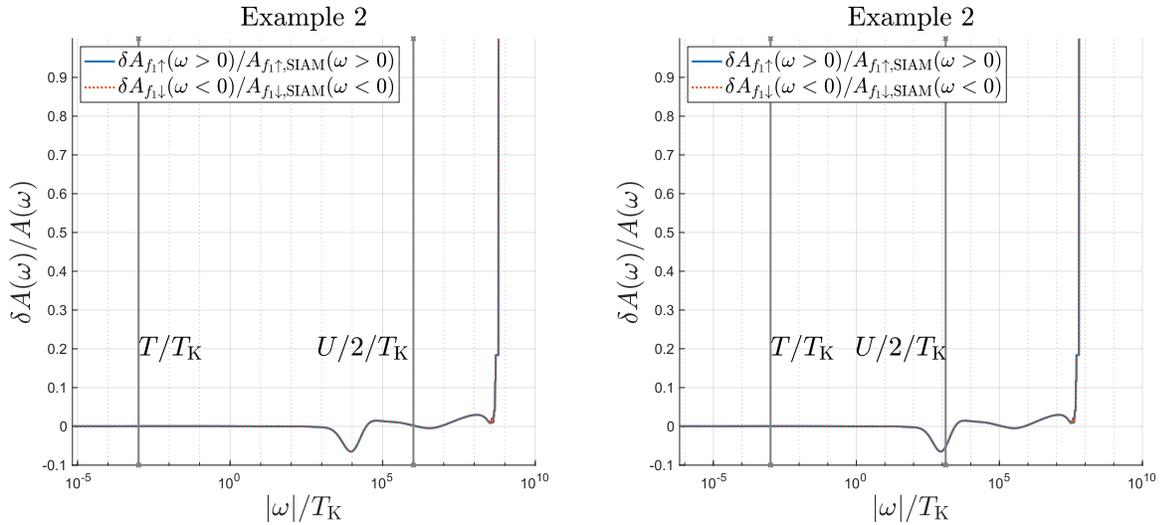


Figure 4.4: Relative error of spectral functions of the annihilation operator $f_{1,\uparrow}$ for spin-up electron at the first bath site .

Not only the static properties, but also dynamics properties can be optimized perfectly. In Fig. 4.3 above, the spectral functions of the annihilation operator $f_{1,\uparrow}$ for a spin-up electron at the first bath site is plotted. The only difference is there is an anti-resonance for SIAM at about $\omega = \frac{U}{2}$. The relative error $\delta A_{f_{1,\uparrow}}(\omega > 0)/A_{f_{1,\uparrow},\text{SIAM}}(\omega > 0)$ is shown in Fig. 4.4.

4.3.3 Temperature for calculating partitions function

In the following, the dependence of the machine learning temperature is going to be discussed about.

Since the low-temperature physics must be optimized, the machine learning temperature should be small, e.g. $T \leq T_K$. Such choices are valid because of the physical meaning of Kondo temperature T_K : the crossover takes place around Kondo temperature. So the question now is how large temperature T can be chosen so that the temperature is still small enough to describe the low-temperature physics. In Fig. 4.5, the temperature dependence of J for two SIAM with the same Kondo temperature T_K is shown. One can see, even when the temperature is a little bit higher than Kondo temperature T_K , J can still be optimized correctly. Depending on different SIAM, the tolerance is different. So for safety, one can always choose T smaller than or equal to T_K .

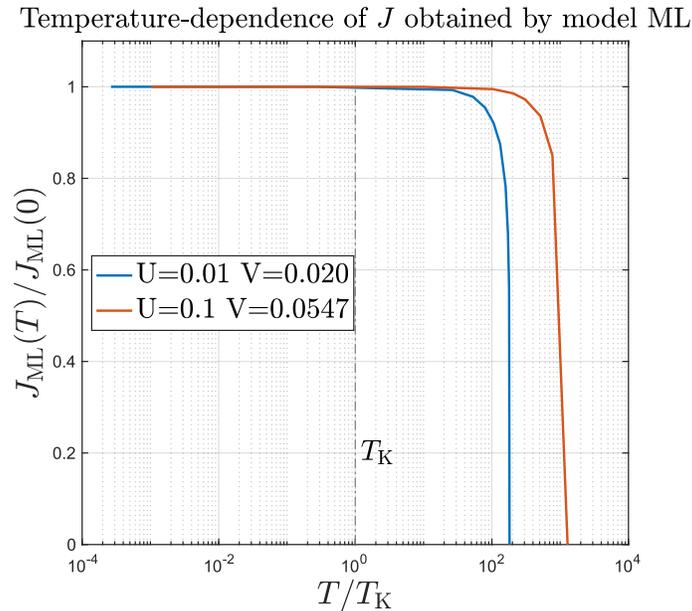


Figure 4.5: Temperature dependence of J obtained by model ML for two SIAM with the same T_K but varying U, V , modified from Fig. 3c) in Ref. 3. Even when the temperature a little bit higher than Kondo temperature T_K , J can still be optimized correctly.

4.4 Effective model of 2IAM for DQD

Using the idea above, one can also find effective models of 2IAM for DQD. Now there are two options for us to set the Hamiltonian of effective model H_{eff} . The first option is to choose a SIKM with spin-1. Alternatively and also more complicatedly, one can choose a 2IKM for DQD as an effective model. This makes sense since a 2IKM for DQD can be mapped to a SIKM with spin-1. It would be more challenging to deal with a 2IKM for DQD as an effective model because there are then two parameters to optimize, which could lead to several local minimums. But first of all, the first option is considered that is an example in Ref. 3.

4.4.1 SIKM with spin-1 as an effective model

A SIKM with spin-1 is similar to the normal SIKM with spin-1/2. The only difference is that spin-1 instead of spin-1/2 Pauli-matrices should be used (s. Sec. 2.2). Hence, the procedure for SIKM (s. Sec. 4.3.1) can be recycled and the following results can be achieved: with the increasing iteration steps, both cost function and cost gradient go to zero as shown in Fig. 4.6(a). After adding some extra values of J and the corresponding cost functions and cost gradients, one can see that at $J = J_{\text{ML}}$ the cost function is indeed the minimum and the cost gradient is zero in Fig. 4.6(b).

One can see, the partition function can be fitted until T in $\mathcal{O}(10^{-3})$ and the first crossover from left in Fig.4.6(d) can be covered perfectly. Depending on the demand or the purpose, if only the most lowest-energy physics is required, then SIKM with spin-1 with J_{ML} is a proper effective model of 2IAM for DQD. However, if one is more ambitious and wants to get a richer structure of low-temperature physics (e.g. covering the second crossover from left as well), a 2IKM for DQD can be applied as an effective model for a larger range of temperatures.

4.4.2 2IKM for DQD as an effective model

With even more details, one can use 2IKM for DQD as an effective model. Recall the Hamiltonian of a 2IKM for DQD (Eq.(2.15)):

$$H = H_{\text{bath}} + 2J(\vec{S}_{d,1} \cdot \vec{S}_0 + \vec{S}_{d,2} \cdot \vec{S}_0) + J_s \vec{S}_{d,1} \cdot \vec{S}_{d,2}. \quad (4.10)$$

Now there are two parameters to optimize. Directly applying exactly the same procedure for SIAM may not lead to the desired result. Nevertheless, go through the first three steps of general procedure quickly and see what kind of changes can be made for the current situation:

1. 2IAM for DQD can be easily generated. But for which temperature the partition function should be calculated is not clear. As known from Sec. 4.3.3, choosing temperatures lower than Kondo temperature T_K can lead to the correct parameter for SIKM. Choosing a much higher temperature than T_K , the low-energy physics cannot be reproduced properly. So for now, it makes sense to choose T around T_K .

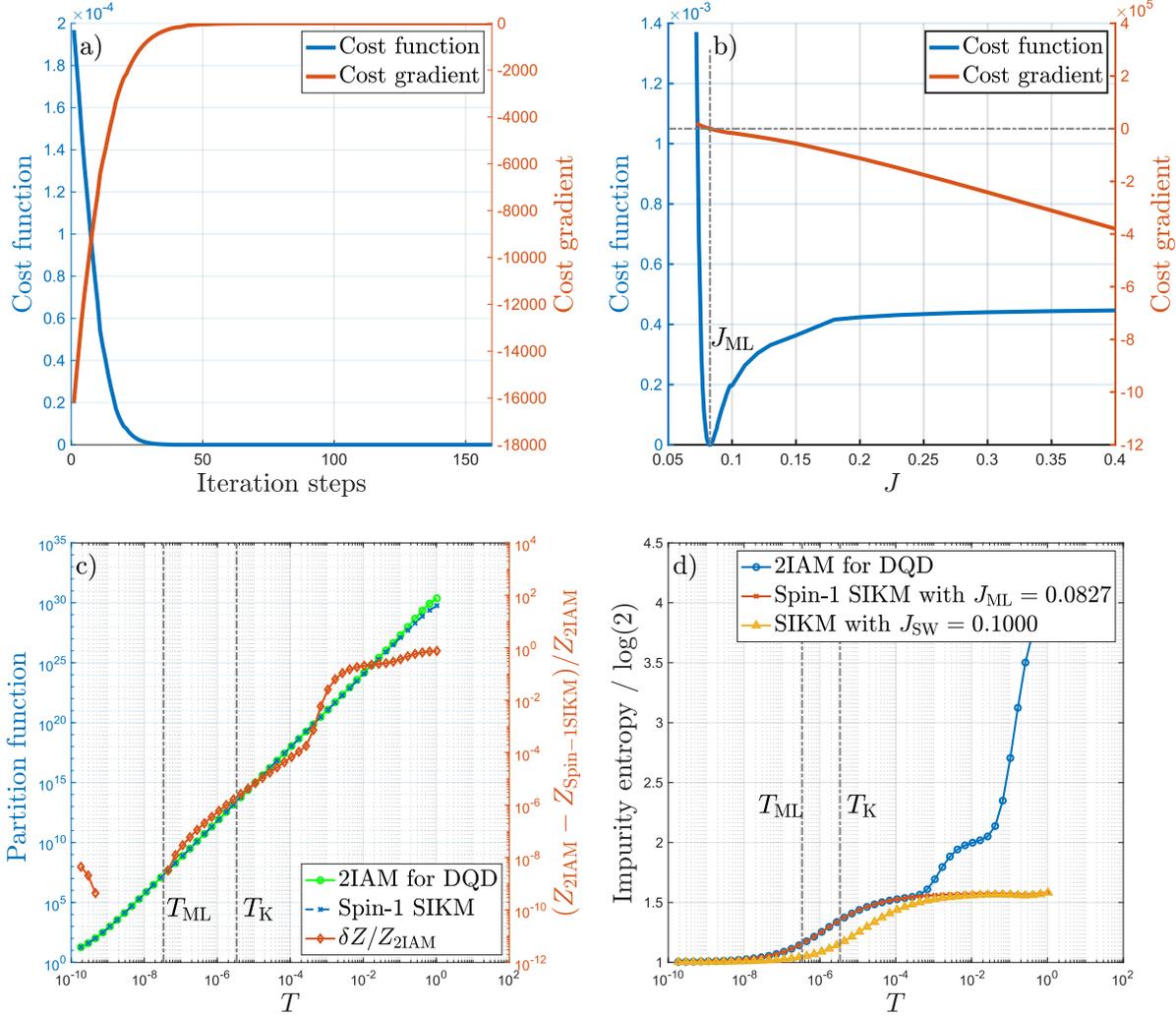


Figure 4.6: Model machine learning process of spin-1 SIKM from 2IAM for DQD with $U = 0.5$ and $\Gamma = 0.0196$. The effective parameter results as $J_{ML} = 0.827$. (a) Cost function and cost gradient during the learning process. Both of them go towards zero. (b) Cost function and cost gradient dependent on J with some extra values of J . The minimum of J corresponds to the cost gradient of zero. (c) The partition functions of 2IAM for DQD and spin-1 SIKM with optimal $J = J_{ML}$ and their relative differences. The relative differences are zero around T_{ML} and increase with the temperature. (d) Impurity contribution to the entropy of SIAM and SIKM with J_{ML} as well as with J_{SW} , modified from Fig. 4 in Ref. 3. Spin-1 SIKM with J_{ML} matches perfectly on 2IAM for DQD for the first two regimes from left.

2. As starting parameters one can choose $J = J_{\text{SW}}$ and $J_s = J_{\text{RKKY}}$ as shown in Eq.(2.16).
3. Now the expectation values of the effective operators can be calculated using the same method as Eq.(4.8):

$$\langle h_1 \rangle = \langle \vec{S}_{d,1} \cdot \vec{S}_0 + \vec{S}_{d,2} \cdot \vec{S}_0 \rangle = \text{Tr} \left(\rho_{\text{imp}}(\vec{S}_{d,1} \cdot \vec{S}_0 + \vec{S}_{d,2} \cdot \vec{S}_0) \right), \quad (4.11)$$

$$\langle h_2 \rangle = \langle \vec{S}_{d,1} \cdot \vec{S}_{d,2} \rangle = \text{Tr} \left(\rho_{\text{imp}}(\vec{S}_{d,1} \cdot \vec{S}_{d,2}) \right). \quad (4.12)$$

Hence, the parameters J and J_s can be updated through

$$J_{n+1} = J_n - \alpha \cdot \frac{\partial L_Z}{\partial J_n} = J_n + \alpha \cdot \left[\log(Z_{\text{DQD,Kondo}}|_{J=J_n}) - \log(Z_{\text{DQD,AIM}}) \right] \cdot \frac{\langle h_1 \rangle}{T}, \quad (4.13)$$

$$J_{s,n+1} = J_{s,n} - \beta \cdot \frac{\partial L_Z}{\partial J_{s,n}} = J_{s,n} + \beta \cdot \left[\log(Z_{\text{DQD,Kondo}}|_{J_s=J_{s,n}}) - \log(Z_{\text{DQD,AIM}}) \right] \cdot \frac{\langle h_2 \rangle}{T}. \quad (4.14)$$

Here, another learning step parameter β can be defined in dependence on α , which can speed up the optimization with a proper choice of β , since J and J_s are in different scales possibly. Choose $\beta = \alpha$ and $\beta = \alpha \cdot J_{\text{RKKY}}/J_{\text{SW}}$ separately and get two different results as shown in Fig. 4.7.

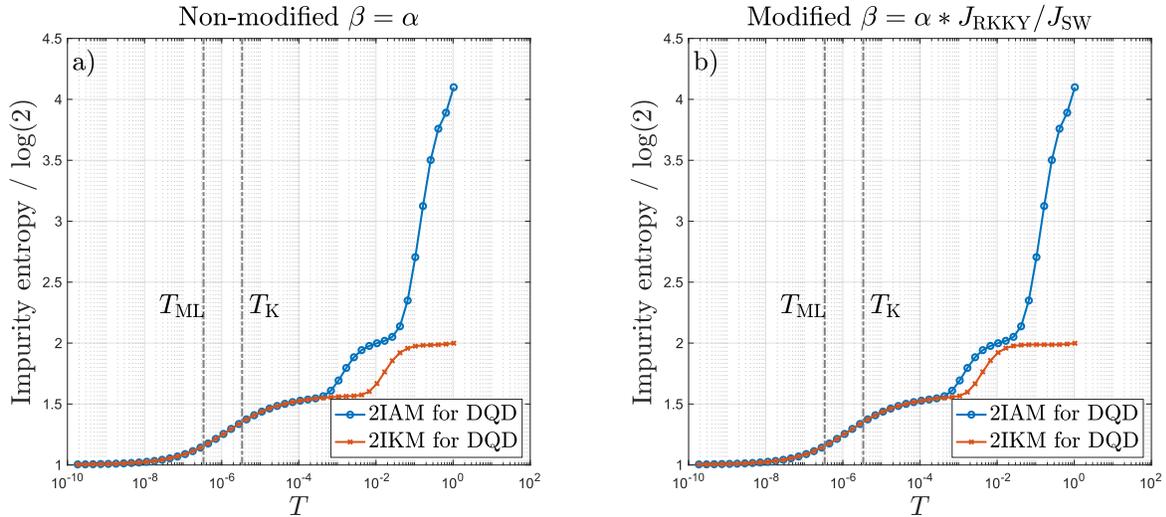


Figure 4.7: Model machine learning results of 2IKM for DQD from 2IAM for DQD with $U = 0.5$ and $\Gamma = 0.0196$. (a) Impurity entropy with modified $\beta = \alpha \cdot \frac{J_{\text{RKKY}}}{J_{\text{SW}}}$, resulting $J = 0.0827$, $J_s = 0.0037$. (b) Impurity entropy with $\beta = \alpha$, resulting $J = 0.0827$, $J_s = -0.0199$.

Actually, as mentioned above, different choices of β only change the speed of convergence to the optimal values. However, two different sets of parameters have been resulted. The reason for this unusual outcome is that there is more than one local minimum for the cost function because the partition function is only a scalar (namely there is only one feature for optimization), which should control two parameters J and J_s . Nevertheless, almost the same J is achieved, since the partition function is calculated at a temperature under T_K . So the partition function encodes mainly the feature of the most low-energy part. By using this property, one can optimize the parameters in such a fashion: fix $J_s = J_{\text{RKKY}}$ in the very beginning and optimize J . With optimized $J = J_{\text{ML}}$ one can then optimize J_s by choosing another higher temperature that corresponds to the second crossover (from the left) of the impurity entropy. How to choose this temperature will be talked about in the next section. Assuming that this temperature is known as $T_{\text{ML},2\text{nd}}$. Then β can also scale with the temperature, so set $\beta = \alpha \cdot \frac{J_{\text{RKKY}}}{J_{\text{SW}}} \cdot \frac{T_{\text{ML},2\text{nd}}}{T_{\text{ML},1\text{st}}}$. With these constructions, one gets an optimized J after 43 iteration steps and then changes the machine learning temperature as shown in Fig. 4.8. The partition functions can be fitted very well, as shown in Fig. 4.9. The relative differences of partition functions are zero around $T_{\text{ML},1\text{st}}$ and $T_{\text{ML},2\text{nd}}$, and increase with the temperature.³

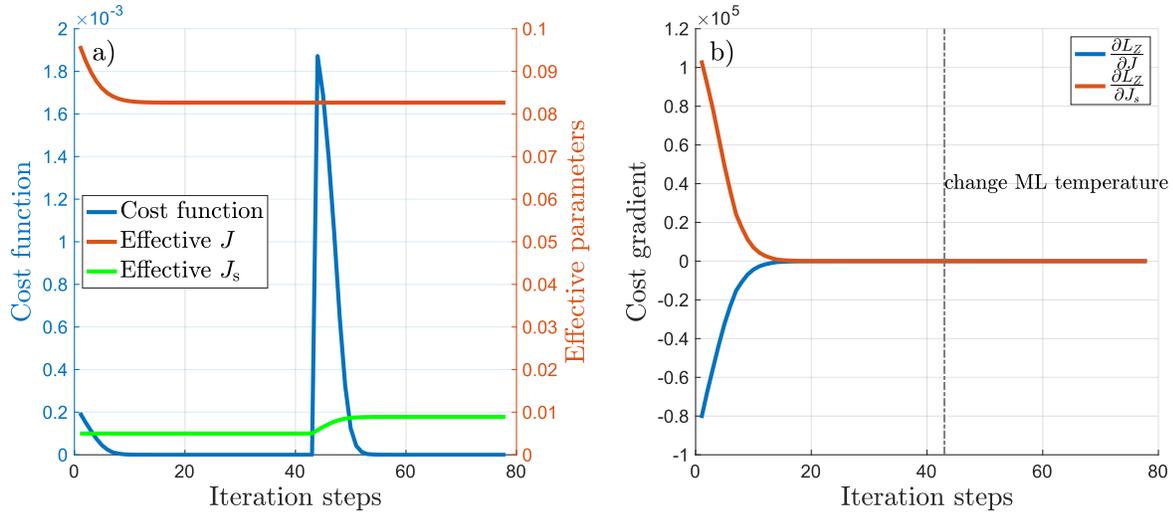


Figure 4.8: Model machine learning process of 2IKM for DQD from 2IAM for DQD with $U = 0.5$ and $\Gamma = 0.0196$. Firstly optimize J , then J_s . The ML temperature is changed at the 43th iteration step. (a) Cost function and effective parameters, (b) cost gradient with increasing iteration steps.

³The relative differences of the partition functions are zero at $T_{\text{ML},2\text{nd}}$! The temperature next to $T_{\text{ML},2\text{nd}}$ is larger than $T_{\text{ML},2\text{nd}}$, but not the same.

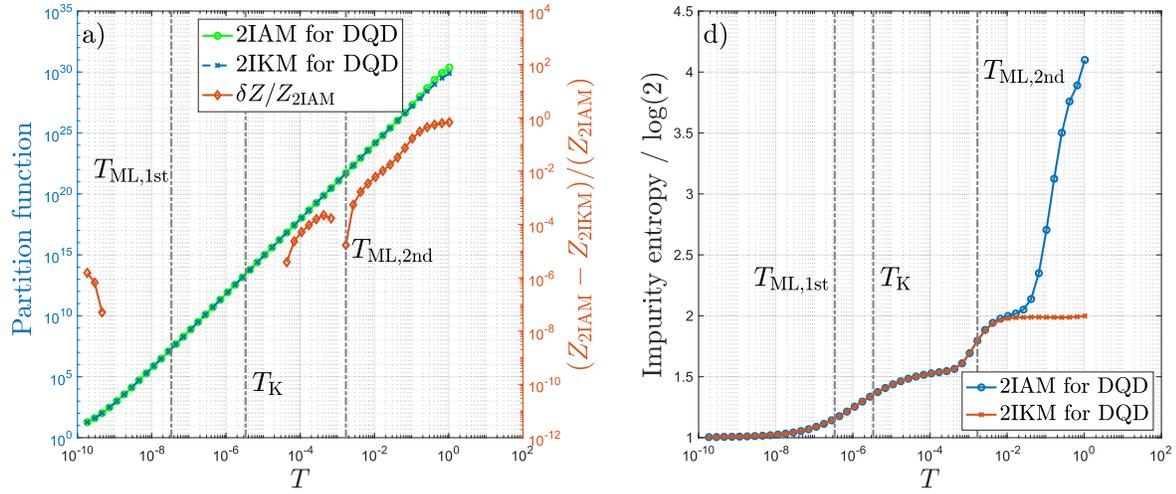


Figure 4.9: Model machine learning results of 2IKM for DQD from DOD for AIM with $U = 0.5$ and $\Gamma = 0.0196$. Firstly optimize J , then J_s . (a) Partition function of 2IAM for DQD and 2IKM for DQD with $J_{\text{ML}} = 0.0827$ and $J_{s,\text{ML}} = 0.0089$ as well as the relative difference of the partition functions. Around $T_{\text{ML},1\text{st}}$ and $T_{\text{ML},2\text{nd}}$ the relative difference is zero. (b) Impurity contribution to the entropy of 2IAM for DQD and 2IKM for DQD with optimized parameters.

4.4.3 Choosing the ML temperature for the next regime

Now let's go back to the question of how to choose the machine learning temperature to optimize the second crossover (second regime).

After optimizing J , the very low-temperature part has been already optimized. So differences of the partition functions between bare and effective models should be almost zero. To cancel the exponential influence resulting from the degeneracy, one can calculate the logarithmic differences (see Fig. (4.10)) of partition functions between AIM and Kondo model:

$$\delta = \log(Z_{\text{bare}}) - \log(\tilde{Z}_{\text{eff}}|_{J=J_{\text{ML}}}), \quad (4.15)$$

where \tilde{Z}_{eff} is the partition function of the effective model with already optimized J_{ML} . One can choose the temperature for the second ML temperature $T_{\text{ML},2\text{nd}}$ at which δ changes most rapidly, namely, choose the local extremum⁴ of $\frac{\partial \delta}{\partial T}$ (see Fig.(4.10)). This temperature lies in the beginning of the second regime.

⁴Actually, one finds the maximum of the absolute value of the derivative. In this case, it is a maximum.

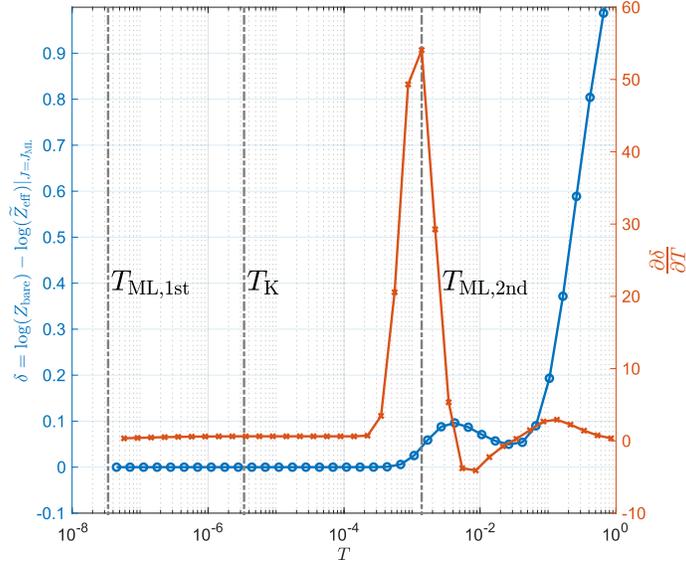


Figure 4.10: Logarithmic differences of partition functions between AIM and Kondo model and its derivative over temperature. The part with $T_{\text{ML},1\text{st}}$ is not shown, since that part is not the interest for now.

With $T_{\text{ML},2\text{nd}}$ one can reproduce the second crossover of 2IAM for DQD from 2IKM for DQD. The scheme described above can actually be generalized if there are more than two regimes to be optimized. One just needs to calculate the logarithmic differences again after the optimization of the second regime, find the local maximal derivative and repeat this procedure.

4.5 Adopting Wilson’s truncation scheme in Model ML

Until now, the full density matrix method is used to calculate the partition functions and expectation values with NRG, which works very well but has a high computational cost, since the Wilson chain Hamiltonian with chain length N has to be fully iteratively diagonalized. One can actually adopt Wilson’s truncation scheme, use single-shell approximation and speed up the model machine learning remarkably.

Explanation

Choosing 2IAM for DQD as a bare model, there are two machine learning temperatures $T_{\text{ML},1\text{st}}$ and $T_{\text{ML},2\text{nd}}$ for calculating the partition functions to optimize the effective parameters. So by coarse-graining in Wilson’s scheme, one only needs the first l_T sites to describe the physics at temperature T with $E_\alpha^l \propto T$ and then compare the partition functions in

shell l_T :

$$Z^l = \sum_{\alpha} e^{-\beta E_{\alpha}^l} \quad (4.16)$$

with eigenenergies E_{α}^l in shell l . Hence, one only needs to iteratively diagonalize until site L_T and the rest of the sites plays a role in perturbations that are small enough to be ignored. This can be done for both $T_{\text{ML},1\text{st}}$ and $T_{\text{ML},2\text{nd}}$. For $T_{\text{ML},1\text{st}}$, $l_{\text{ML},1\text{st}} = 26$ and for $T_{\text{ML},2\text{nd}}$, $l_{\text{ML},2\text{nd}} = 15$ for the same example in Sec. (4.4). Indeed, one gets almost exactly the same results. The relative difference of J is 0.0049% and the relative difference of J_s is 0.00045%, which hardly changes the physical features of effective models. In this approach, the computational cost is reduced from $\mathcal{O}(N_{\text{keep}}^N)$ to $\mathcal{O}(N_{\text{keep}}^l)$ and in the example above is from $\mathcal{O}(N_{\text{keep}}^{50})$ to $\mathcal{O}(N_{\text{keep}}^{26})$ for the optimization of J and from $\mathcal{O}(N_{\text{keep}}^{50})$ to $\mathcal{O}(N_{\text{keep}}^{15})$ for the optimization of J_s .

An interesting plot

From Sec. 4.3.3 one knows there is a range of tolerance of temperatures that can be chosen to calculate the partition functions. Hence, one can actually choose different shells corresponding to the temperatures. Observing a SIAM and its effective model and varying the shells for the single-shell approximation, one gets an oscillation for temperatures higher than Kondo temperature (s. Fig. 4.11). The reason for this ‘even-odd oscillation’ is unclear, but it is still interesting to show this plot.

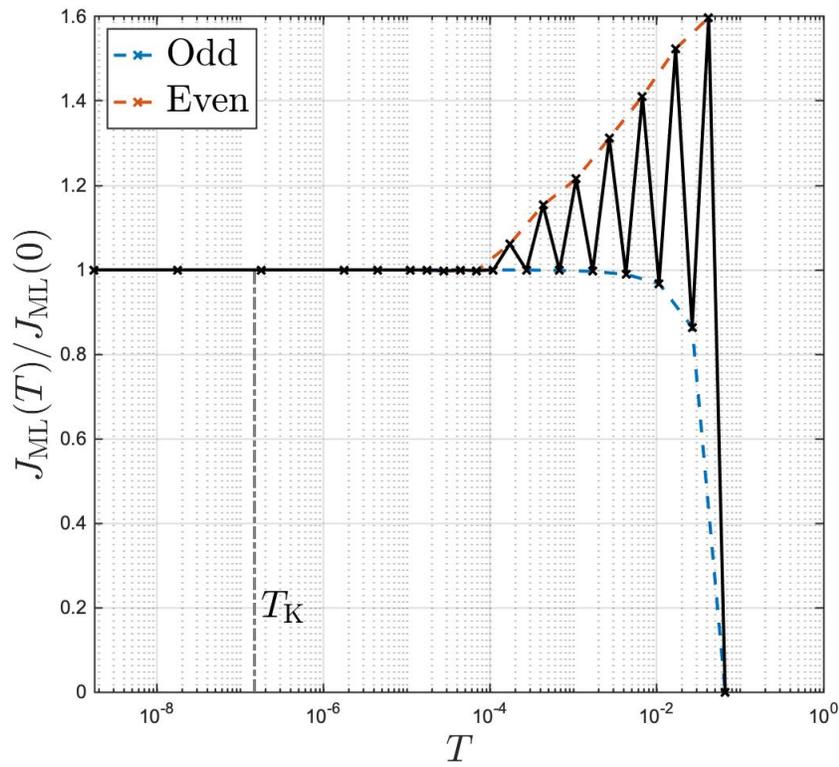


Figure 4.11: Temperature dependence for calculating the partitions functions with single-shell approximations. The same J is resulting in around Kondo temperature. For even higher temperatures, there exists an ‘even-odd oscillation’.

Chapter 5

Conclusion and outlook

In this thesis, the effective models of SIAM and 2IAM for DQD are reproduced as in Rigo and Mitchell's paper. To check the result, models with different parameters are also shown. The machine learning process, for example, the cost function and cost gradient, are plotted to verify that the cost function with the optimized effective parameters is indeed the minimum of all. By the impurity contribution of entropy, one can see that the static low-energy physical properties of bare models are indeed reproduced by the effective model. However, the spectral functions of the annihilation operator for the spin-up electron at the first bath sites are not always the same. For the single impurity model, the spectral functions of bare and effective models are almost exactly the same. Also, the temperature dependence of effective parameter J is discussed. For SIAM, one can always choose the temperature under or a little bit higher than Kondo temperature to calculate the partition functions for the cost function. For the effective models with more than one parameter, e.g. 2IKM for DQD, one must optimize the effective parameters separately, from very-low temperature to low temperature. Every time when one parameter has been optimized for one regime, one can choose the locally largest absolute value of derivative for the logarithmic difference of the partition functions. This is indeed a variant of the utilized cost function and can be calculated numerically. In the end, by adopting Wilson's truncation scheme and using the single-shell approximation, one can get almost the same effective parameters as before up to a $\mathcal{O}(10^{-5})$ accuracy, while the computational cost is reduced exponentially. There is still an unsolved fundamental problem in this thesis: the learning step α should be positive. As mentioned before, a negative α is used for an unknown reason.

To conclude, the model machine learning method reproduced from Ref. 3 and further developed in this thesis works perfectly for the physical properties of SIAM and 2IAM for DQD, which is based on the assumption that the effective operators are already known and one only needs to optimize the effective parameters. However, if the effective parameters are also unknown, then one cannot apply the method in this thesis. Also, the partition function is only a number. Thus, two unrelated, arbitrary partition functions are the same by accident. To avoid this, one could add more constraints to the cost function.

Bibliography

- [1] J. R. Schrieffer and P. A. Wolff. Relation between the anderson and kondo hamiltonians. *Phys. Rev.*, 149(2):491–492, Sep 1966.
- [2] Kenneth G. Wilson. The renormalization group: Critical phenomena and the kondo problem. *Rev. Mod. Phys.*, 47(4):773–840, Oct 1975.
- [3] Jonas B. Rigo and Andrew K. Mitchell. Machine learning effective models for quantum systems. *Phys. Rev. B*, 101:241105(R), 2020.
- [4] Jun Kondo. Resistance minimum in dilute magnetic alloys. *Prog. Theor. Phys.*, 32:37, 1964.
- [5] M. A. Ruderman and C. Kittel. Indirect exchange coupling of nuclear magnetic moments by conduction electrons. *Phys. Rev.*, 96(1):99, Oct 1954.
- [6] R. Žitko and J. Bonča. Multiple-impurity anderson model for quantum dots coupled in parallel. *Phys. Rev. B*, 74(16):045312, Jul 2006.
- [7] Ralf Bulla, Theo A. Costi, and Thomas Pruschke. Numerical renormalization group method for quantum impurity systems. *Rev. Mod. Phys.*, 80(2):395, 2008. arXiv:cond-mat/0701105 [cond-mat.str-el].
- [8] J. von Delft. Tensor networks 2020. LMU Munich.
- [9] Jr. Vivaldo L. Campo and Luiz N. Oliveira. Alternative discretization in the numerical renormalization-group method. *Phys. Rev. B*, 72(23):104432, Sep 2005.
- [10] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat'l. Bur. Std.*, 45:255–282, 1950.
- [11] U. Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326:96–192, 2011. arXiv:1008.3477 [cond-mat.str-el].
- [12] Frithjof B. Anders and Avraham Schiller. Real-time dynamics in quantum-impurity systems: A time-dependent numerical renormalization-group approach. *Phys. Rev. Lett.*, 95(19):196801, 2005.

-
- [13] Frithjof B. Anders and Avraham Schiller. Spin precession and real-time dynamics in the kondo model: Time-dependent numerical renormalization-group study. *Phys. Rev. B*, 74(24):245113, 2006.
- [14] Andreas Weichselbaum and Jan von Delft. Sum-rule conserving spectral functions from the numerical renormalization group. *Phys. Rev. Lett.*, 99(7):076402, 2007. arXiv:cond-mat/0607497.
- [15] Andrew Ng. Machine learning by standford university at coursera.org.
- [16] H. R. Krishna-murthy, J. W. Wilkins, and K. G. Wilson. Renormalization-group approach to the anderson model of dilute magnetic alloys. ii. static properties for the asymmetric case. *Phys. Rev. B*, 21(3):1044–1083, Feb 1980.

Selbständigkeitserklärung

Ich versichere hiermit, die vorliegende Arbeit mit dem Titel

Maschinelles Lernen Schrieffer–Wolff Transformation und ihre Anwendung

selbständig verfasst zu haben und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

Yuhao Liu

München, den 20. Juli 2021