Real-Frequency Correlation Functions from Neural Quantum States via Operator Lanczos Approach

Jonas B. Rigo[1, 2, *] and Markus Schmitt[1, 2, †]

[1]*Forschungszentrum Jülich GmbH, Peter Grünberg Institute, Quantum Control, 52425 Jülich, Germany*
[2]*University of Regensburg*

CONTENTS

I. SPECTRUM INVARIANCE UNDER LINEAR TRANSFORMATIONS

We show the invariance of the spectrum under linear transformation of the operator set by showing the respective transformations of the generalized eigenvalue problem. Let $\{\hat{O}_i\}_{i=1}^{N}$ be a set of operators and define the respective span

$$|\phi_i\rangle = \hat{O}_i |\psi_0\rangle, \qquad \mathcal{S} = \text{span}\{|\phi_i\rangle\}_{i=1}^{N}, \qquad (1)$$

with the seed state $|\psi_0\rangle$. For coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^T$ a state in the span is given as

$$|\psi(\boldsymbol{\theta})\rangle = \sum_{i=1}^{N} \theta_i |\phi_i\rangle \in \mathcal{S}. \qquad (2)$$

To get an eigenstates of the effective Hamiltonian projected in to the span one solves the generalized eigenvalue problem

$$\mathbb{H}\boldsymbol{\theta} = E\,\mathbb{G}\boldsymbol{\theta}, \qquad (3)$$

with matrices

$$\mathbb{H}_{ij} = \langle \phi_i| \hat{H} |\phi_j\rangle, \qquad \mathbb{G}_{ij} = \langle \phi_i|\phi_j\rangle. \qquad (4)$$

Now take any invertible matrix $\mathbb{M} \in \mathbb{C}^{N \times N}$ and define a new operator set

$$\hat{\tilde{O}}_a = \sum_{i=1}^{N} \mathbb{M}_{ai}\hat{O}_i, \qquad |\tilde{\phi}_a\rangle = \hat{\tilde{O}}_a |\psi_0\rangle = \sum_{i=1}^{N} \mathbb{M}_{ai} |\phi_i\rangle. \qquad (5)$$

The corresponding matrices are

$$\tilde{\mathbb{H}}_{ab} = \langle \tilde{\phi}_a| \hat{H} |\tilde{\phi}_b\rangle = \sum_{ij} \mathbb{M}_{ai}\mathbb{H}_{ij}\mathbb{M}_{bj}^* = (\mathbb{M}\mathbb{H}\mathbb{M}^\dagger)_{ab},$$

$$\tilde{\mathbb{G}}_{ab} = \langle \tilde{\phi}_a|\tilde{\phi}_b\rangle = \sum_{ij} \mathbb{M}_{ai}\mathbb{G}_{ij}\mathbb{M}_{bj}^* = (\mathbb{M}\mathbb{G}\mathbb{M}^\dagger)_{ab}.$$

If the solution fo the eigenvalue problem is $\mathbb{H}\boldsymbol{\theta} = E\mathbb{G}\boldsymbol{\theta}$, then set $\tilde{\boldsymbol{\theta}} = (\mathbb{M}^{-1})^\dagger\boldsymbol{\theta}$. With this ansatz we find

$$\tilde{\mathbb{H}}\tilde{\boldsymbol{\theta}} = \mathbb{M}\mathbb{H}\mathbb{M}^\dagger(\mathbb{M}^{-1})^\dagger\boldsymbol{\theta} = \mathbb{M}\mathbb{H}\boldsymbol{\theta} \qquad (6)$$

$$= E\mathbb{M}\mathbb{G}\boldsymbol{\theta} = E\mathbb{M}\mathbb{G}\mathbb{M}^\dagger(\mathbb{M}^{-1})^\dagger\boldsymbol{\theta} = E\tilde{\mathbb{G}}\tilde{\boldsymbol{\theta}}. \qquad (7)$$

Thus E is also an eigenvalue of the transformed problem $\tilde{\mathbb{H}}\tilde{\boldsymbol{\theta}} = E\tilde{\mathbb{G}}\tilde{\boldsymbol{\theta}}$ and the state is unchanged:

$$|\psi(\boldsymbol{\theta})\rangle = \sum_i \theta_i |\phi_i\rangle = \sum_a \tilde{\theta}_a |\tilde{\phi}_a\rangle. \qquad (8)$$

Hence, both the effective energies and the effective states are invariant under linear transformation of the operator set $\{\hat{O}_i\}$.

* j.rigo@fz-juelich.de
† markus.schmitt@ur.de

II. QUANTUM-NUMBER–CONSERVING AUTOREGRESSIVE SAMPLING

In the following, we outline our procedure for exclusively drawing samples with fixed target quantum numbers for charge (Q^0) and magnetization (M^0). The method follows the prescription of Ref. [?] and incorporates operator-sequence constraints at every step of the sampling process, which allow us to apply operators from the SCOL set to the state directly at sampling time, while also patching sites.

1. Local Hilbert space and quantum numbers

The multi-orbital quantum impurity model graph can always be organized into a linear chain, despite some links get non-local connectivity, this is a worthwhile trade-off for the autoregressive process to be feasible. Therefore we consider a one-dimensional lattice of L sites with spinful fermions. At each site $i \in \{1, \ldots, L\}$ we use the local Hilbert alphabet

$$\mathcal{A}_1 = \{ \downarrow,\ \varnothing,\ \uparrow,\ \uparrow\downarrow \} = \{0, 1, 2, 3\}, \tag{9}$$

where the last equality indicates a fixed integer encoding.

Local charge or electron (Q) and magnetization (M) contributions from each local Hilbert space configuration are

$$Q(\downarrow) = 1,\ Q(\varnothing) = 0,\ Q(\uparrow) = 1,\ Q(\uparrow\downarrow) = 2, \tag{10}$$
$$M(\downarrow) = -1,\ M(\varnothing) = 0,\ M(\uparrow) = 1,\ M(\uparrow\downarrow) = 0. \tag{11}$$

For a configuration from the occupation basis $\mathbf{x} = (x_1, \ldots, x_L) \in \mathcal{A}_1^{\times L}$ the total charge and magnetization is

$$Q(\mathbf{x}) = \sum_{i=1}^{L} Q(x_i), \qquad M(\mathbf{x}) = \sum_{i=1}^{L} M(x_i). \tag{12}$$

The target symmetry sector is

$$\mathcal{X}_{Q^0, M^0} = \{ \mathbf{x} \in \mathcal{A}_1^{\times L} : Q(\mathbf{x}) = Q^0,\ M(\mathbf{x}) = M^0 \}. \tag{13}$$

2. Operator-sequence constraints

To efficiently estimate the $\mathbb{G}$ matrix we sample directly $p_a(\mathbf{x}) \propto |\langle \mathbf{x}| \hat{O}_a |\psi_{\boldsymbol{\theta}}\rangle|^2$, where $\hat{O}_i$ is a product of fermionic creation and annihilation operators. This type of operator can directly be applied to an autoregressive $|\psi_{\boldsymbol{\theta}}\rangle$ and sampled, since $\hat{O}_i$ simply enforces local occupation configurations. These per-site restrictions are encoded in an operator sequence

$$\mathrm{op} = \big(\mathrm{op}^{\uparrow}, \mathrm{op}^{\downarrow}\big) \in \{0, 1, 2\}^{\times L} \times \{0, 1, 2\}^{\times L}, \tag{14}$$

where at each site i and spin flavor $\sigma \in \{\uparrow, \downarrow\}$:

- $\mathrm{op}_i^{\sigma} = 0$ enforces the spin-σ orbital to be empty,

- $\mathrm{op}_i^{\sigma} = 1$ enforces it to be occupied,

- $\mathrm{op}_i^{\sigma} = 2$ leaves it unconstrained.

A local state $a \in \mathcal{A}_1$ is allowed at site i if its spin content is compatible with op_i. We denote the allowed set by $\mathcal{C}_{\mathrm{op}}(i) \subseteq \mathcal{A}_1$.

3. Autoregressive factorization and logit masks

The RWKV is an autoregressive network, meaning that its respective Born probability distribution $p_{\boldsymbol{\theta}}(\mathbf{x}) = |\langle \mathbf{x}| |\psi_{\boldsymbol{\theta}}\rangle|^2$ factorizes over configurations as

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{t=1}^{L} p_{\boldsymbol{\theta}}(x_t \mid \mathbf{x}_{<t}), \tag{15}$$

where $\mathbf{x}_{<t} := (x_1, \ldots, x_{t-1})$ and $\boldsymbol{\theta}$ are neural-network parameters. We refer to $\mathbf{x}_{<t}$ as the *prefix* of length $t-1$, that is the partial configuration obtained by fixing all sites strictly to the left of t. The sampling process runs sequentially starting from the empty prefix $\mathbf{x}_{<1} = \varnothing$, at step t the network is evaluated on the current prefix to produce partial probabilities over the local alphabet (which are the configurations of the local Hilbert space of a lattice site), which allows a local state x_t to be drawn from the resulting conditional distribution

$$x_t \sim p_{\boldsymbol{\theta}}(x_t \mid \mathbf{x}_{<t}). \tag{16}$$

Repeating this procedure for $t = 1, \ldots, L$ yields a full configuration $\mathbf{x}$.

In practice the network does not produce probabilities, but at each step t outputs unnormalized values that we consider logarithms of positive numbers, known as *logits* $\ell_{\boldsymbol{\theta}}(t, \cdot \mid \mathbf{x}_{<t}) \in \mathbb{R}^{|\mathcal{A}_1|}$. By applying the *softmax* activation function actual conditioned probabilities can be obtained,

$$p_{\boldsymbol{\theta}}(\cdot \mid \mathbf{x}_{<t}) = \mathrm{softmax}\big(\ell_{\boldsymbol{\theta}}(t, \cdot \mid \mathbf{x}_{<t})\big). \tag{17}$$

Without constraining $\ell_{\boldsymbol{\theta}}$ every configuration $\mathbf{x} \in \mathcal{A}_1^{\times L}$ can be drawn. To impose hard constraints, we add a *logit mask* $m_t(\cdot \mid \mathbf{x}_{<t})$ and instead use

$$\tilde{\ell}_{\boldsymbol{\theta}} = \ell_{\boldsymbol{\theta}} + m_t, \qquad p_{\boldsymbol{\theta}}(\cdot \mid \mathbf{x}_{<t}) = \mathrm{softmax}(\tilde{\ell}_{\boldsymbol{\theta}}). \tag{18}$$

The mask decomposes as

$$m_t(a) = m_t^{\mathrm{op}}(a) + m_t^{\mathrm{QN}}(a), \tag{19}$$

where m_t^{op} enforces the local operator sequence and m_t^{QN} is a *quantum-number mask* that removes choices incompatible with the global target constraints (Q^0, M^0).

The operator mask is purely local and therefore easy to compute

$$m_t^{\mathrm{op}}(a) = \begin{cases} 0, & a \in \mathcal{C}_{\mathrm{op}}(t), \\ -\infty, & \mathrm{otherwise} \end{cases}. \qquad (20)$$

The quantum-number mask is more complex, since the magnetization is not monotonically increasing, but can also decrease throughout the generation process of a sample $\mathbf{x}$.

4. Quantum-number subspace restriction

The quantum-number mask m_t^{QN} is built from two tables, $M_{\min}(t,q)$ and $M_{\max}(t,q)$, which encode the achievable magnetization intervals on suffixes for fixed electron counts.

For $t \in \{0, 1, \ldots, L\}$ let the *suffix* (tail of the chain after position t) starting at t denote the sites $\{t+1, \ldots, L\}$, with $t = L$ corresponding to the empty suffix. For each integer $q \in \{0, 1, \ldots, 2L\}$ we define

$$\begin{aligned} M_{\min}(t,q) = \min\{\, M(y) : y \in \mathcal{A}_1^{L-t}, \; Q(y) = q, \\ y \text{ allowed by op on } \{t+1, \ldots, L\}\,\}, \end{aligned} \qquad (21)$$

$$\begin{aligned} M_{\max}(t,q) = \max\{\, M(y) : y \in \mathcal{A}_1^{L-t}, \; Q(y) = q, \\ y \text{ allowed by op on } \{t+1, \ldots, L\}\,\}. \end{aligned} \qquad (22)$$

If no such configuration y exists, we set $M_{\min}(t,q) = +\infty$ and $M_{\max}(t,q) = -\infty$.

For the empty suffix at $t = L$,

$$M_{\min}(L,0) = M_{\max}(L,0) = 0, \qquad (23)$$

and for $q \neq 0$,

$$M_{\min}(L,q) = +\infty, \qquad M_{\max}(L,q) = -\infty. \qquad (24)$$

The first condition reflects that zero magnetization can be achieved uniquely by the empty configuration with zero electrons and the second ensures that it is impossible to place electrons on an empty suffix.

5. Recursive construction

The tables $M_{\min}(t,q)$ and $M_{\max}(t,q)$ are computed by a backward recursion in t, starting from the boundary conditions at $t = L$. At a given suffix index t, the next physical site to be filled is $t + 1$. For each allowed local state $a \in \mathcal{C}_{\mathrm{op}}(t + 1)$, any suffix configuration y on $\{t+1, \ldots, L\}$ formed by placing a at $t+1$ and an admissible tail y' obeys

$$Q(y) = Q(a) + Q(y'), \qquad M(y) = M(a) + M(y'). \qquad (25)$$

Thus, for fixed total electrons q on the suffix, choosing a and distributing the remaining $q' = q - Q(a)$ electrons on the tail yields magnetizations in

$$\{M(a) + M_{\min}(t+1, q'), \ldots, M(a) + M_{\max}(t+1, q')\}.$$

Taking the lower (upper) envelope over all allowed a gives

$$M_{\min}(t,q) = \min_{a \in \mathcal{C}_{\mathrm{op}}(t+1)} \big[M_{\min}(t+1, q - Q(a)) + M(a) \big], \qquad (26)$$

$$M_{\max}(t,q) = \max_{a \in \mathcal{C}_{\mathrm{op}}(t+1)} \big[M_{\max}(t+1, q - Q(a)) + M(a) \big], \qquad (27)$$

where terms with $q - Q(a) < 0$ or $q - Q(a) > 2L$ are forbidden and therefore treated as $+\infty$ (for $M_{\min}$) and $-\infty$ (for $M_{\max}$). Eqs. (26) and (27) are evaluated for $t = L-1, \ldots, 0$ using the boundary conditions at $t = L$.

6. Quantum-number mask at step t

Assume a prefix $\mathbf{x}_{<t}$ has already been sampled using masked logits. We define the remaining budgets with respect to the target quantum numbers (Q^0, M^0) as

$$q_{\mathrm{left}} = Q^0 - \sum_{i<t} Q(x_i), \qquad m_{\mathrm{left}} = M^0 - \sum_{i<t} M(x_i). \qquad (28)$$

If we choose local state $a \in \mathcal{A}_1$ at site t, the suffix $\{t+1, \ldots, L\}$ must realize

$$q' = q_{\mathrm{left}} - Q(a), \qquad m' = m_{\mathrm{left}} - M(a). \qquad (29)$$

By the definitions (21)–(22), there exists at least one suffix configuration consistent with the op restrictions and achieving (q', m') if and only if

$$0 \leq q' \leq 2L, \qquad M_{\min}(t,q') \leq m' \leq M_{\max}(t,q'). \qquad (30)$$

We therefore define the quantum-number mask as

$$m_t^{\mathrm{QN}}(a) = \begin{cases} 0, & \text{if (30) holds}, \\ -\infty, & \text{otherwise}. \end{cases} \qquad (31)$$

The intuition behind this procedure is as follows. Starting from $t = 1$ with budgets (Q^0, M^0), the sampling procedure at each step t draws x_t from the masked logits $\tilde{\ell}_{\boldsymbol{\theta}} = \ell_{\boldsymbol{\theta}} + m_t^{\mathrm{op}} + m_t^{\mathrm{QN}}$ and updates $(q_{\mathrm{left}}, m_{\mathrm{left}})$. By construction, m_t^{op} enforces local compatibility with op, and m_t^{QN} enforces the necessary and sufficient condition (30) for the existence of a completion.

By induction over t, every prefix sampled from this procedure admits at least one completion in $\mathcal{X}_{Q^0, M^0}$ consistent with op. At the final step the suffix is empty, so a completion exists if and only if the budgets are exhausted exactly, which is guaranteed by the boundary conditions on $M_{\min/\max}$. Hence, every full configuration produced by the masked autoregressive sampler lies in $\mathcal{X}_{Q^0, M^0}$ and respects all operator-sequence constraints.

7. Patching procedure

To reduce the effective sequence length of $\mathbf{x}$ we group

sites into non-overlapping patches of size π by creating larger many-body local Hilbert spaces. Assume π divides L, then the new effective sequence length is $P = L/\pi$. The patches are

$$(1, \ldots, \pi),\ (\pi+1, \ldots, 2\pi),\ \ldots,\ ((P-1)\pi+1, \ldots, P\pi), \tag{32}$$

and the patch alphabet is

$$\mathcal{A}_\pi = \mathcal{A}_1^\pi, \qquad |\mathcal{A}_\pi| = 4^\pi, \tag{33}$$

with the patch state $\boldsymbol{a} = (a_1, \ldots, a_\pi) \in \mathcal{A}_\pi$. We calculate the maps Q and M in the patched case additively,

$$E_\pi(\boldsymbol{a}) = \sum_{k=1}^{\pi} Q(a_k), \qquad M_\pi(\boldsymbol{a}) = \sum_{k=1}^{\pi} M(a_k). \tag{34}$$

The operator sequence induces permissible patched sets

$$\mathcal{C}_{\mathrm{op}}^{(\pi)}(t) = \Big\{\boldsymbol{a} \in \mathcal{A}_\pi : a_k \in \mathcal{C}_{\mathrm{op}}((t-1)\pi + k)\ \forall k\Big\}, \tag{35}$$

for patch index $t \in \{1, \ldots, P\}$.

8. Patched autoregressive model and QN mask

We now express the Born distribution in terms of patches,

$$p_{\boldsymbol{\theta}}(z) = \prod_{t=1}^{P} p_{\boldsymbol{\theta}}(z_t \mid z_{<t}), \tag{36}$$

where $z_t \in \mathcal{A}_\pi$ and $z_{<t}$ denotes the patch prefix. The network outputs logits $\ell_{\boldsymbol{\theta}}(t, \cdot \mid z_{<t}) \in \mathbb{R}^{|\mathcal{A}_\pi|}$, which are masked by

$$\tilde{\ell}_{\boldsymbol{\theta}} = \ell_{\boldsymbol{\theta}} + m_t^{\mathrm{op}} + m_t^{\mathrm{QN}}, \tag{37}$$

now defined over the patch alphabet. The patched operator mask is

$$m_t^{\mathrm{op}}(\boldsymbol{a}) = \begin{cases} 0, & \boldsymbol{a} \in \mathcal{C}_{\mathrm{op}}^{(\pi)}(t), \\ -\infty, & \text{otherwise.} \end{cases} \tag{38}$$

The patched quantum-number mask m_t^{QN} is constructed from patch suffix bounds $M_{\mathrm{min/max}}^{(\pi)}$, defined in analogy to Eqs. (21)–(22) with $\mathcal{A}_1$ replaced by $\mathcal{A}_\pi$ and (Q, M) by (E_π, M_π). This goes as follows: for $t \in \{0, \ldots, P\}$ and $q \in \{0, \ldots, 2L\}$ we define

$$M_{\mathrm{min}}^{(\pi)}(t, q) = \min\{\, M_\pi(y) : y \in \mathcal{A}_\pi^{P-t},\ E_\pi(y) = q, \\ y \text{ obeys } \mathcal{C}_{\mathrm{op}}^{(\pi)} \,\}, \tag{39}$$

$$M_{\mathrm{max}}^{(\pi)}(t, q) = \max\{\, M_\pi(y) : y \in \mathcal{A}_\pi^{P-t},\ E_\pi(y) = q, \\ y \text{ obeys } \mathcal{C}_{\mathrm{op}}^{(\pi)} \,\}. \tag{40}$$

The recursion is identical in structure to Eqs. (26)–(27):

$$M_{\mathrm{min}}^{(\pi)}(t, q) = \min_{\boldsymbol{a} \in \mathcal{C}_{\mathrm{op}}^{(\pi)}(t+1)} \Big[M_{\mathrm{min}}^{(\pi)}(t+1, q - E_\pi(\boldsymbol{a})) + M_\pi(\boldsymbol{a}) \Big], \tag{41}$$

$$M_{\mathrm{max}}^{(\pi)}(t, q) = \max_{\boldsymbol{a} \in \mathcal{C}_{\mathrm{op}}^{(\pi)}(t+1)} \Big[M_{\mathrm{max}}^{(\pi)}(t+1, q - E_\pi(\boldsymbol{a})) + M_\pi(\boldsymbol{a}) \Big], \tag{42}$$

with $M_{\mathrm{min}}^{(\pi)}(P, 0) = M_{\mathrm{max}}^{(\pi)}(P, 0) = 0$ and $M_{\mathrm{min}}^{(\pi)}(P, q) = +\infty$, $M_{\mathrm{max}}^{(\pi)}(P, q) = -\infty$ for $q \neq 0$.

At patch step t, with patch prefix $z_{<t}$, the remaining budgets with respect to the target quantum numbers are

$$q_{\mathrm{left}} = Q^0 - \sum_{u<t} E_\pi(z_u), \qquad m_{\mathrm{left}} = M^0 - \sum_{u<t} M_\pi(z_u). \tag{43}$$

For a candidate patch $\boldsymbol{a}$, the suffix must realize

$$q' = q_{\mathrm{left}} - E_\pi(\boldsymbol{a}), \qquad m' = m_{\mathrm{left}} - M_\pi(\boldsymbol{a}). \tag{44}$$

The patched quantum-number feasibility condition is

$$0 \le q' \le 2L, \qquad M_{\mathrm{min}}^{(\pi)}(t, q') \le m' \le M_{\mathrm{max}}^{(\pi)}(t, q'), \tag{45}$$

and we set

$$m_t^{\mathrm{QN}}(\boldsymbol{a}) = \begin{cases} 0, & \text{if this condition holds,} \\ -\infty, & \text{otherwise.} \end{cases} \tag{46}$$

Since patches are non-overlapping and the additive quantum numbers $Q(\cdot)$ and $M(\cdot)$ are extensive, there is a one-to-one correspondence between site-level and patched configurations that preserves $(Q(\mathbf{x}), M(\mathbf{x}))$ and operator feasibility. As a consequence, the reachable (q, m) pairs on any suffix are the same whether computed sitewise or patchwise, and the patched m^{QN} is equivalent to a site-level m^{QN} evaluated at patch boundaries.

III. DETAILS OF NUMERICAL RENORMALIZATION GROUP SIMULATION

For the single- and two-orbital impurity problems the reference data were obtained using full-density matrix NRG [?] for a Wilson-chain discretization parameter of $\Lambda = 3$, keeping up to 4000 states per iteration and a maximum Wilson chain length of $L_c = 35$. The dynamical correlators were computed at temperature $T = 10^{-6}$.

For the three-orbital impurity problem, we constructed an effective NRG reference by attaching three additional Wilson-chain bath sites with $\Lambda = 12$ to the third impurity orbital and then performing a standard two-channel NRG calculation with 8000 kept states. The Green's function was converged with respect to the discretization parameter Λ and the number of kept states. Since we do not perform a dynamical mean-field self-consistency here, we only require the spectral function of a single bath leg as a reference for the NQS+SCOL results.