Maximum quantum entropy method

Jae-Hoon Sim and Myung Joon Han*

Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea

(Received 4 April 2018; revised manuscript received 16 September 2018; published 1 November 2018)

A maximum entropy method for analytic continuation is extended by introducing quantum relative entropy. This method is formulated in terms of matrix-valued functions and therefore invariant under arbitrary unitary transformation of input matrix. As a result, the continuation of off-diagonal elements becomes straightforward. Without introducing any further ambiguity, the Bayesian probabilistic interpretation is maintained just as in the conventional maximum entropy method. The applications of our generalized formalism to a model spectrum and a real material demonstrate its usefulness and superiority.

DOI: 10.1103/PhysRevB.98.205102

I. INTRODUCTION

An imaginary time Green's function method such as quantum Monte Carlo (QMC) is a main workhorse for various finite temperature many-body problems [1–3]. While it has been successful for both impurity and periodic systems, the output data on the imaginary axis [e.g., Matsubara Green's function $G(i\omega_n)$] should be transformed to the real-frequency spectrum $A(\omega)$ in order to be compared with experimental results. Namely, physical observables can only be accessed indirectly via analytic continuation. While the calculation of $G(i\omega_n)$ from $A(\omega)$ is straightforward, its inverse is an ill-posed problem due to the large conditional number of the kernel matrix. The small noise in $G(i\omega_n)$ can lead to large fluctuations in $A(\omega)$, and the double precision is far from being enough [4]. Among well-established methods, such as Pade [5,6], stochastic method [7], and others [8–13], the maximum entropy method (MEM) is one of the most widely used [14–17].

One obvious limitation of conventional MEM is about the nondiagonal components of Matsubara functions. Since the conventional formalism is rigorous only for non-negative and additive functions [18], it has been a challenge to make the continuation of off-diagonal matrix elements which can be negative or complex. This limitation becomes particularly serious when one tries to understand the real materials based on, for example, dynamical mean-field theory (DMFT) [19–21] combined with QMC impurity solver [22–26]. No matter how correctly the Matsubara function or self-energy is computed, severely limited is to understand the electronic property, especially the effect of spin-orbit coupling (SOC), crystal-field effect, or any other factors that can generate nondiagonal parts of the functions [27,28].

One possible way to overcome this limitation is to transform the imaginary frequency data to a "good basis set" on which the Green's function can be represented as diagonal as possible and the off-diagonal elements be neglected. Another approach tries to relax the non-negativity conditions or to construct the auxiliary functions with positive definite property [29–31]. Recently, a notable idea has been suggested [32]. In order to apply MEM to the off-diagonal elements of spectral function $A^{\text{off}}(\omega)$, Kraberger *et al.* decomposed $A^{\text{off}}(\omega)$ so that the positive definite condition is satisfied. It is noted, however, that in this element-wise approach SU(N) invariance is hardly preserved in the sense that the resulting spectrum is dependent on the basis choice.

In the current study we generalize MEM by reformulating it with quantum relative entropy. It can be regarded as a quantum version of MEM. Within this maximum quantum entropy method (MQEM), the matrix is not decomposed nor treated element wise, but is directly continued as a single object. Thus our formalism is guaranteed to be basis independent. This outstanding feature enables us to perform the analytic continuation for the off-diagonal parts. We apply MQEM to a model spectrum, whose ideal spectrum can be known by construction, and to a realistic material example of Sr_2IrO_4, for which the full matrix information is essential due to strong SOC and structural distortion. The results demonstrate the usefulness and superiority of our formulation.

II. FORMALISM

A. Quantum entropy

Matsubara frequency Green's function $G(i\omega_n)$ [or self-energy $\Sigma(i\omega_n)$] is analytically continued to real-frequency $G(\omega)$ [or $\Sigma(\omega)$]. For a given $G(i\omega_n)$, spectral function $A(\omega) = -\frac{1}{\pi}\text{Im}G(\omega + i0^+)$ is obtained by inverting the integral equation

$$G(i\omega_n) = \int d\omega \frac{A(\omega)}{i\omega_n - \omega} \tag{1}$$

$$= \int d\omega \mathbf{K}(i\omega_n, \omega)A(\omega). \tag{2}$$

Note that both $G(i\omega_n)$ and A are in general matrix-valued functions. A kernel $\boldsymbol{K}(i\omega_n, \omega)$ is ill-conditioned and the direct inverse of $A = \mathbf{K}^{-1}G$ is not quite feasible, due to the large conditional number (i.e., the large ratio between minimum and maximum eigenvalues), likely leading to the violation of

*mj.han@kaist.ac.kr

non-negativity condition $[A(\omega) > 0]$ and sum rule $[\int d\omega A(\omega) = 1]$.

To solve this ill-posed problem, MEM introduces entropy S for diagonal components,

$$S[A(\omega)||D(\omega)] = \int d\omega A(\omega) \ln \frac{A(\omega)}{D(\omega)}, \tag{3}$$

which is also known as Kullback-Leibler distance [33]. $D(\omega)$ is a default model providing the essential features of spectra, which can be determined by the "annealing" procedure [15,33] or by making use of high-frequency behavior of input data [15]. The latter is used in the current study. In MEM, it is the "free energy" $F = \chi^2 + \alpha S$ that is minimized with a fitting parameter α, the goodness-of-fit functional

$$\chi^2[A;G] = \frac{1}{2}\sum_n \frac{1}{\sigma_n^2}\left\| G(i\omega_n) - \int d\omega \mathbf{K}(i\omega_n,\omega)A(\omega)\right\|^2, \tag{4}$$

and the variance σ_n^2. In the case of nondiagonal covariance, one can introduce the uncorrelated variables $\tilde{G}_i = \sum_n \mathbf{U}_{in}^\dagger G(i\omega_n)$, where $\mathbf{U}$ is the eigenvectors of the covariance matrix, such that $\tilde{C} = \mathbf{U}^\dagger C \mathbf{U}$ is diagonal. We stress that the entropy S requires the positiveness of the spectrum [32].

Hereafter, we use a hat ($\hat{}$) notation to emphasize the matrix values. In order to consider the whole matrix continuation (not element decomposed) and not to lose any off-diagonal information, we first renormalize Matsubara function by $\hat{G} \to \hat{G}/\mathrm{Tr}[\hat{M}_0]$. Here $\hat{M}_0 = \int d\omega \hat{A}(\omega)$ is zeroth moments of the spectrum. With a renormalized Matsubara function, the asymptotic behavior can be written as $\hat{G}(i\omega_n \to \infty) = \hat{M}_0/i\omega_n$ with $\mathrm{Tr}\hat{M}_0 = 1$. Second, we divide $\hat{A}(\omega)$ into two parts, $\hat{A}(\omega) = P(\omega)\hat{\rho}(\omega)$ and $P(\omega) = \mathrm{Tr}\hat{A}(\omega)$. Here the sum rule is written as

$$\int d\omega \mathrm{Tr}\hat{A} = \int d\omega P(\omega) = \mathrm{Tr}\hat{M}_0 = 1. \tag{5}$$

Thus, within this formalism, $P(\omega)$ can be interpreted as a classical probability distribution and $\hat{\rho}(\omega)$ as a density matrix.

Now we extend the entropy of Eq. (3) to quantum relative entropy, which is widely used in the nonequilibrium thermodynamics [34,35] as well as the information science [36]:

$$\begin{aligned} S^Q(\hat{A}||\hat{D}) &= \int d\omega \mathrm{Tr}\hat{A}(\omega)[\ln \hat{A}(\omega) - \ln \hat{D}(\omega)] \\ &= S[P(\omega)||D(\omega)] + \int d\omega P(\omega) S^Q[\hat{\rho}(\omega)||\hat{\sigma}(\omega)]. \end{aligned} \tag{6}$$

Here the default model is further decomposed into $D(\omega) = \mathrm{Tr}\hat{D}(\omega)$ and $\hat{\sigma}(\omega) = \hat{D}(\omega)/D(\omega)$; $\hat{D}(\omega) = D(\omega)\hat{\sigma}(\omega)$. While the first term in the second line, $S[P(\omega)||D(\omega)]$, is the classical entropy used in MEM [see Eq. (3)], the second term is introduced to regularize matrix elements. In our formalism, the free energy functional to be minimized is defined by the matrix-valued functions $F[\hat{A};\hat{G}] = \chi^2 - \alpha S^Q(\hat{A}||\hat{D})$. We stress that this free energy functional is SU(N) invariant, i.e.,

$$F[U\hat{A}U^\dagger; U\hat{G}U^\dagger] = F[\hat{A};\hat{G}] \tag{7}$$

for unitary matrix U.

While our formalism assumes that the spectrum $\hat{A}$ is Hermitian, any non-Hermitian spectrum $\hat{A}_{\mathrm{nonH}}$ can be divided into two Hermitian matrices:

$$\hat{A}^R_{\mathrm{nonH}} = (\hat{A}_{\mathrm{nonH}} + \hat{A}^\dagger_{\mathrm{nonH}})/2, \tag{8}$$

$$\hat{A}^I_{\mathrm{nonH}} = (\hat{A}_{\mathrm{nonH}} - \hat{A}^\dagger_{\mathrm{nonH}})/2i. \tag{9}$$

And therefore, $\hat{A}^R_{\mathrm{nonH}}$ and $\hat{A}^I_{\mathrm{nonH}}$ can be dealt with separately.

B. Iterative equation

The key task is to minimize free energy F:

$$\begin{aligned} \Phi &= \min_{A(\omega)} F[\hat{A}(\omega)] \\ &= \min_{A(\omega)} \left[\tfrac{1}{2}\chi^2 + \alpha S^Q(\hat{A}||\hat{D})\right]. \end{aligned} \tag{10}$$

This minimization can be conducted by using the stationary condition $\delta F/\delta \hat{A} = 0$. With a trace norm $\|\hat{M}\| = \mathrm{Tr}(\hat{M}^\dagger\hat{M})$ for Eq. (4), a set of self-consistent equations is given as follows:

$$\hat{H}_\omega|\psi_\omega\rangle = \epsilon_\omega|\psi_\omega\rangle, \tag{11}$$

where

$$\begin{aligned} &\hat{H}_\omega[\hat{A}(\omega)] \\ &= \frac{1}{2}\sum_{i\omega} K^*(i\omega_n,\omega)\left(\hat{G}(i\omega_n) - \int d\omega' K(i\omega_n,\omega')\hat{A}(\omega')\right) \\ &\quad + \frac{\alpha}{2}\ln \hat{D}(\omega) + \mathrm{H.c.} \end{aligned} \tag{12}$$

and

$$\hat{A}(\omega) = \int d\epsilon \frac{1}{Z} e^{-\epsilon_w/\alpha}|\psi_\omega\rangle\langle\psi_\omega|. \tag{13}$$

Then Eq. (11) can be solved iteratively. For more details, see Appendix A. Note that these equations represent a quantum system described by Hamiltonian $\hat{H}_\omega[\hat{A}(\omega)]$, and its spectrum is given by a density matrix of the canonical ensemble with temperature α. We would like to note that the MEM is closely related to stochastic approximation (SA) [37]. Also, the self-consistent form of Eq. (11) is not surprising since MEM can be regarded as a mean-field realization of SA [37].

Equation (11) represents the matrix diagonalization procedure for all ω-grid points. This is the only part that requires the additional computation compared to the conventional MEM. Considering that the size of the Hamiltonian matrix is typically less than 10×10, this extra cost is not significant.

We used Pulay mixing scheme [38] and its generalization [39] to achieve the stable convergence. The results were compared to the solution of the reduced independent variables in the singular space [32,40]. To minimize the real-frequency grid size, cubic splines in combination with nonuniform real-frequency grids have been adopted [15]. For more details, see Appendix B.

C. Default model

We take a Gaussian shape of the default model to avoid the data noise. The asymptotic behavior of high-frequency data

determines the first a few moments of spectra [15]:

$$\hat{G}(i\omega_n) = \frac{\hat{M}_0}{i\omega_n} + \frac{\hat{M}_1}{(i\omega_n)^2} + \frac{\hat{M}_2}{(i\omega_n)^3} + \cdots, \quad (14)$$

where $\hat{M}_j = \int \omega^j \hat{A}(\omega) d\omega$ is the jth moment of $\hat{A}(\omega)$. To define Gaussian curves for given moments M_j $(j = 0, 1, 2)$ are straightforward in the scalar version of MEM. In the matrix formalism of our MQEM, on the other hand, finding out the analytic solution is not quite feasible due to the fact that $\hat{M}_j$ does not commute in general with each other; $[M_j, M_{j'}] \neq 0$.

Here we propose a way to find out the "featureless" default models for a few given moments, M_j $(j = 0, 1, 2)$. Recalling that Gaussian curve has the maximum entropy among the distributions with a specified variance, we define a default model that maximizes

$$S_D = \int d\omega \hat{D}(\omega) \ln \hat{D}(\omega) + \sum_{j=0}^{2} \mathrm{Tr} \hat{\mu}_j \int \omega^j \hat{D}(\omega) d\omega, \quad (15)$$

where $\hat{\mu}_j$ is Lagrange multiplier introduced by the constraint $\int \omega^j \hat{D}(\omega) d\omega = M_j$. The stationarity condition $\partial S_D / \partial \hat{D}(\omega) = 0$ reads

$$\hat{D}(\omega) = \exp\left(\sum_{j=0}^{2} \hat{\mu}_j \omega^j \right). \quad (16)$$

D. Fitting parameter α

A popular approach to calculate spectral functions is to optimize the parameter α by a statistical method within the probabilistic interpretation of MEM [16]. Alternatively, an average value of the spectra calculated by many different α values can be taken [40]. Recently, a different approach has been suggested [15]. In this approach $\log \chi^2$ is computed as a function of $\log \alpha$, and two different regions (namely, "information-fitting" and "noise-fitting" region) are considered. The optimal α is then determined at the maximum curvature of $\log \chi^2[\log(\alpha)]$. We used a similar approach in our MQEM implementation. We fit the $\log \chi^2[\log(\alpha)]$ curve by Fermi-Dirac function as shown in Fig. 1. The optimal α is determined by the maximum second deviation of the fitting function. A clear advantage of this technique is the numerical stability against the grid changes.

III. RESULT AND DISCUSSION

A. Simple model spectrum

As the first example, we apply our method to a simple model system. The Green's function $\hat{G}^{\mathrm{in}}(i\omega_n)$ is obtained from a model spectral function which is given by a 2×2 matrix:

$$\hat{A}(\omega) = \hat{R}\left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} (e^{-\frac{1}{2}(\frac{\omega+\omega_0}{2})^2} + e^{-\frac{1}{2}(\frac{\omega-\omega_0}{2})^2}) \right] \hat{R}^{\dagger}. \quad (17)$$

Note that obtaining $\hat{G}^{\mathrm{in}}(i\omega_n)$ from $\hat{A}(\omega)$ is not ill-conditioned. Here the two-peak Gaussian spectrum centered at $\omega_0 = \pm 1.5$ is rotated by a rotation matrix $\hat{R} = \begin{pmatrix} \cos(\theta) & i\sin(\theta) \\ i\sin(\theta) & \cos(\theta) \end{pmatrix}$ with

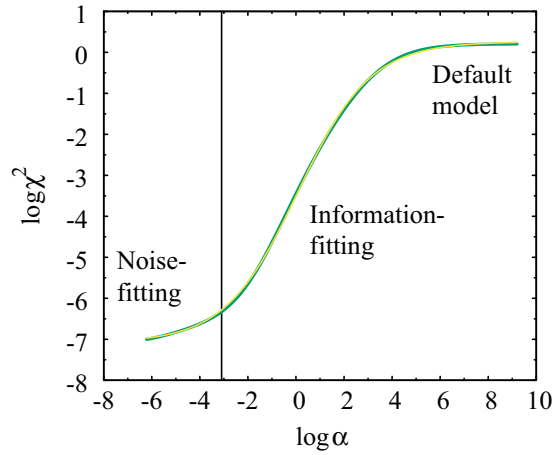

FIG. 1. The green line presents $\log \chi^2$ as a function of $\log \alpha$ in the case of model spectrum discussed in Sec. III A. Gaussian noise of $\sigma = 10^{-3}$ is introduced to each element of Green's function matrix. Yellow line is the fitting curve with optimal $\log \alpha_{\mathrm{opt}} = -3.09$ (black vertical line).

$\theta = (2\pi\omega/T_\omega)^2$. For $T_\omega = \infty$ $(\theta = 0)$, the spectral function corresponds to the trivial case that off-diagonal elements are all zero. At finite T_ω, $\hat{A}(\omega)$ has nonzero off-diagonal values. In performing MEM continuation, we also introduced random Gaussian noises to the Green's functions with a standard deviation of $\sigma = 10^{-4}$ in order to mimic a realistic QMC situation.

Figure 2 shows the calculated spectra from the input of Eq. (17). The conventional MEM and the generalized MQEM results are presented in magenta and blue lines, respectively, along with the ideal spectrum (green) from which the input Green's function $\hat{G}^{\mathrm{in}}(i\omega_n)$ is generated. It is noted that the conventional MEM does not well reproduce the off-diagonal part of spectral function [Fig. 2(b)] while the diagonal part is in good agreement with the ideal spectrum [Fig. 2(a)]. This is a well-known limitation of MEM. Here the results of conventional MEM are obtained from the properly chosen basis set in which the off-diagonal components of $\hat{G}^{\mathrm{in}}(i\omega_n)$ are minimized; i.e., min $\sum_{i\omega_n} \sum_{i \neq j} |\hat{G}^{\mathrm{in}}_{ij}(i\omega_n)|^2$ $(\theta = 0.972$ rad). Note that, even with this "best" basis, the off-diagonal elements are significantly deviated from the ideal result as shown in Fig. 2(b). The same feature is also observed in $G(i\omega_n)$, see Fig. 2(c). The conventional MEM result shows the noticeable deviation from the ideal (or original) curve especially for the off-diagonal part. Note that, in this example, there is no unitary transformation for the basis set on which the matrix-valued $\hat{A}(\omega)$ [or equivalently $\hat{G}^{\mathrm{in}}(i\omega_n)$] is diagonalized at all frequencies, and therefore the conventional MEM has no way to be satisfactory.

A remarkable improvement is clearly noticed in our result of MQEM. Even for the off-diagonal components, the generalized MQEM results are in good agreement with the ideal spectrum; see Figs. 2(a) and 2(b). The excellent agreement is also found for $G(i\omega_n)$ as shown in Fig. 2(c). This result of simple model spectrum demonstrates the capability of MQEM for the continuation of matrix-valued functions.

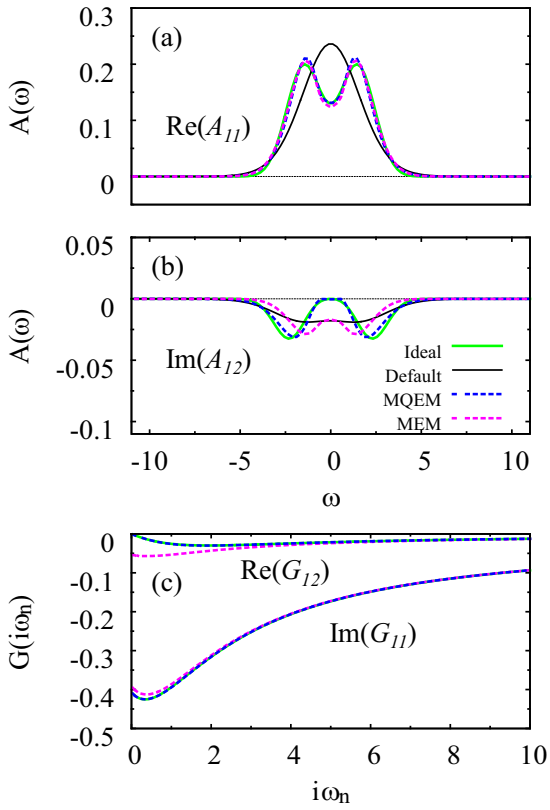

FIG. 2. (a) and (b) The calculated default model $\hat{D}(\omega)$ (black solid) and spectral function $\hat{A}(\omega)$ by using the conventional MEM (magenta) and our method (blue). The ideal spectra, from which $G(i\omega_n)$ is calculated, are also presented for comparison (green). The diagonal and off-diagonal components are presented in (a) and (b), respectively. The off-diagonal components are dominated by imaginary part due to the form of the rotation matrix $\hat{R}$. (c) The input Green's function in Matsubara frequency axis (green) and the Green's function reconstructed from $\hat{A}(\omega)$ using Eq. (1) (blue and magenta). The solid (below) and dashed lines (above) correspond to $\mathrm{Im}[G_{11}(i\omega_n)]$ and $\mathrm{Re}[G_{12}(i\omega_n)]$, respectively.

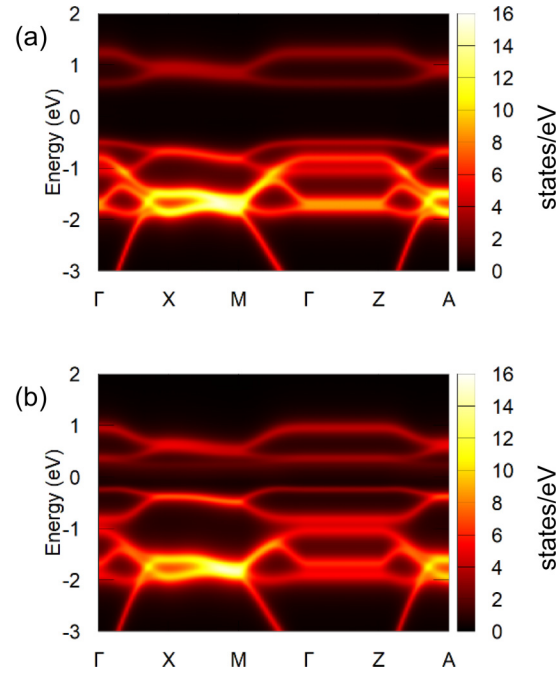

FIG. 3. The calculated DMFT spectral function of Ir-t_{2g} states obtained by (a) our generalized MEM and (b) the conventional MEM. The chemical potential is set to be zero energy.

B. Real material example: Sr_2IrO_4

As a real material example, we consider Sr_2IrO_4. The local Green's function and self-energy of this material are featured by the significant off-diagonal components caused by strong SOC and structural distortions. Thus, dealing properly the off-diagonal elements is of crucial importance to describe its electronic structure. We calculate Matsubara functions by LDA+DMFT (local density approximation plus dynamical mean-field theory) method based on Wannier-projected t_{2g} orbitals [41,42]. The interaction parameter of $U = 2.2$ eV is adopted [28]. Further computation details can be found in Appendix C. Analytic continuation of impurity self-energy $\Sigma(i\omega_n)$ is conducted to obtain $\mathrm{Im}\Sigma(\omega)$, and the real part is obtained by Kramers-Kronig transformation.

Real frequency self-energy can be obtained via the analytic continuation of Weiss field $\mathcal{G}_0(i\omega_n)$ and impurity Green's function $G_{\mathrm{imp}}(i\omega_n)$. Self-energy on the real-frequency axis is then given by Dyson's equation $\Sigma(\omega) = \mathcal{G}_0^{-1}(\omega) - G_{\mathrm{imp}}^{-1}(\omega)$ [43]. In practice, widely used to perform the continuation of auxiliary Green's functions which are constructed from the self-energy [28,32,43–46]. While there are many different ways to construct the auxiliary Green's functions, we perform the continuation of $\hat{\Sigma}^{\mathrm{dyn}}(i\omega_n) = \hat{\Sigma}(i\omega_n) - \hat{\Sigma}(i\infty)$ [43], assuming diagonal covariance with constant noise. It is noted that the element-wise MEM is not quite feasible for $\hat{\Sigma}^{\mathrm{dyn}}$ due to the fact that the high-frequency behavior of the off-diagonal components of $\hat{\Sigma}^{\mathrm{dyn}}$ is proportional to $1/i\omega_n$ with the finite norm of the spectral function [32]. We emphasize that our formalism is free from this deficiency and provides the full matrix information of high-frequency coefficients, see Eqs. (5) and (14).

The result of MQEM is presented in Fig. 3(a). The calculated spectral function $A(\mathbf{k}, \omega)$ is in reasonable agreement with the well-known features of this material including the relative position of so-called $j_{\mathrm{eff}} = 1/2$ and $j_{\mathrm{eff}} = 3/2$ bands [47–49].

MQEM result is significantly different from that of conventional MEM. By comparing Figs. 3(a) and 3(b), the differences are clearly noticed. For example, the separation between the conduction and valence band states is markedly enhanced in MQEM [Fig. 3(a)] and therefore the band gap becomes larger. While $j_{\mathrm{eff}} = 1/2$ states (upper and lower Hubbard band) move away from the Fermi level, the $j_{\mathrm{eff}} = 3/2$ states do not show a significant change. It is likely due to that Σ^{dyn} is dominated by the static Hartree terms in these fully occupied states.

Our results show that taking the full account of off-diagonal matrix elements is important to correctly describe the electronic structure. While the effect of ignoring the off-diagonal part can be minimized by taking a better basis set rather than Ir-t_{2g} [28,44], it is not always straightforward to make the right choice. In many different situations and due to many different reasons, the off-diagonal elements can become

non-negligible. Therefore it is important to take all matrix information through the continuation process.

It can be an interesting future direction to further extend the idea of MQEM. Introducing quantum entropy can extend the physical implication and the applicability of currently available methods or techniques, especially in our case for dealing with the off-diagonal information. The similar idea might be applicable to the other non-Hermitian matrix-valued functions such as Gorkov's Green's function for superconducting order parameter [29,45].

IV. SUMMARY

By introducing quantum relative entropy we resolve a long-standing issue of analytic continuation, namely, the continuation of off-diagonal matrix elements. Based on quantum relative entropy, the functions are treated as being matrix valued and the non-negativity condition as well as the sum rule are extended. The invariance under unitary transformation and the Hermiticity of spectral function $\hat{A}(\omega)$ are inherently satisfied in the general context. As a result, it becomes straightforward to perform analytic continuation of the off-diagonal as well as diagonal components without any further approximation or ad-hoc treatment. The capability and usefulness of our method is demonstrated by two examples. In both of model spectrum and a real material example of Sr_2IrO_4, our MQEM provides a reliable description of off-diagonal elements which cannot be well treated within the conventional schemes.

ACKNOWLEDGMENTS

We thank Junya Otsuki, Hongkee Yoon, and Hunpyo Lee for useful comments and discussions. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A2B2005204).

APPENDIX A: DERIVATION OF EQ. (11)

The minimization of $F[\hat{A}]$ can be achieved by solving the variational equation

$$\frac{\delta}{\delta\hat{A}(\omega)}\left\{\frac{1}{2}\chi^2+\alpha S^Q(\hat{A}||\hat{D})-\alpha\mu\int d\omega[\hat{A}(\omega)]\right\}=0, \tag{A1}$$

where Lagrange multiplier $\alpha\mu$ is required to ensure the normalization of the spectrum $\int d\omega \mathrm{Tr}\hat{A}(\omega)=1$. First, let us consider the differentiation:

$$\delta\chi^2=\chi^2[\hat{A}+\delta\hat{A}]-\chi^2[\hat{A}] \tag{A2}$$

$$=\frac{1}{2}\sum_n\frac{1}{\sigma_n^2}\left\|\hat{G}(i\omega_n)-\int d\omega\mathbf{K}(i\omega_n,\omega)[\hat{A}(\omega)+\delta\hat{A}(\omega)]\right\|^2-\chi^2[\hat{A}] \tag{A3}$$

$$=\frac{1}{2}\sum_n\frac{1}{\sigma_n^2}\mathrm{Tr}\left[\int d\omega\mathbf{K}^*(i\omega_n,\omega)\delta\hat{A}^\dagger(\omega)\left\{\hat{G}(i\omega_n)-\int d\omega\mathbf{K}(i\omega_n,\omega)\hat{A}(\omega)\right\}\right]+\mathrm{H.c.}+O(\delta A^2) \tag{A4}$$

$$=\int d\omega\sum_{ij}\delta\hat{A}_{ij}(\omega)\hat{h}_\omega[\hat{A}]_{ji}+O(\delta A^2), \tag{A5}$$

where $\|\cdot\|$ refers to the trace norm. From the last equality we have

$$\frac{\delta\chi^2}{\delta\hat{A}(\omega)}=\hat{h}_\omega[\hat{A}], \tag{A6}$$

with

$$\hat{h}_w[\hat{A}]=\sum_n\frac{1}{\sigma_n^2}\mathrm{Tr}\left[\int d\omega\mathbf{K}^*(i\omega_n,\omega)\left\{G(i\omega_n)-\int d\omega\mathbf{K}(i\omega_n,\omega)A(\omega)\right\}\right]. \tag{A7}$$

Similarly, the derivative of entropy terms is

$$\frac{\delta}{\delta\hat{A}(\omega)}S^Q(\hat{A}||\hat{D})=\ln\hat{A}(\omega)-\ln\hat{D}(\omega)+\hat{I}. \tag{A8}$$

Now the variational equation (A1) can be rewritten as

$$\hat{h}_\omega[\hat{A}]+\alpha\ln\hat{A}(\omega)-\alpha\ln\hat{D}(\omega)+\alpha(1-\mu)\hat{I}=0 \tag{A9}$$

from Eqs. (A6) and (A8). Solving Eq. (A9) about $\hat{A}(\omega)$, we arrive at

$$\hat{A}(\omega)=\exp-\hat{H}_\omega[\hat{A}]/\alpha+f. \tag{A10}$$

Here $f=1-\mu=\ln Z$ is adjusted to satisfying the normalization condition, i.e., Z is the partition function. Diagonalizing $\hat{H}_\omega[\hat{A}]$, it can be shown that the Eq. (A10) is equivalent to Eqs. (11), (12), and (13).

APPENDIX B: IMPLEMENTATION DETAILS

1. Kernel matrix for cubic spline spectral function

For a given discrete grid set of frequencies $\omega_j\in[-W,W]$, Eq. (1) reads

$$G(i\omega_n)\approx\sum_{j=1}^{N_\omega+1}\frac{A(\omega_j)}{i\omega_n-\omega_j}\Delta\omega_j. \tag{B1}$$

In order to achieve a high accuracy with a decent number of grids, we adopted the cubic spline interpolation. The coefficients of cubic polynomials $S_j(\omega) = a_j(\omega-\omega_j)^3 + b_j(\omega-\omega_j)^2 + c_j(\omega-\omega_j) + d_j$ (for $\omega_j < \omega < \omega_{j+1}$) are the solution of the linear equations:

$$S_j(\omega_j) = d_j = A(\omega_j), \tag{B2}$$

$$S_j(\omega_{j+1}) = a_j\Delta\omega_j^3 + b_j\Delta\omega_j^2 + c_j\Delta\omega_j + d_j = A(\omega_{j+1}), \tag{B3}$$

$$S_j'(\omega_{j+1}) - S_{j+1}'(\omega_{j+1}) = 3a_j\Delta\omega_j^2 + 2b_j\Delta\omega_j + c_j - c_{j+1} = 0, \tag{B4}$$

$$S_j''(\omega_{j+1}) - S_{j+1}''(\omega_{j+1}) = 6a_j\Delta\omega_j + 2b_j - 2b_j = 0. \tag{B5}$$

Here $j = 1, \ldots, N_\omega$ in (A2) and (A3), and $j = 1, \ldots, N_\omega - 1$ in (A4) and (A5), providing $(4N_\omega - 2)$ equations. Two more equations are from boundary conditions:

$$S_1''(\omega_1) = S_{N_\omega}''(\omega_{N_\omega+1}) = 0. \tag{B6}$$

Thus the transformation matrix T is obtained and it gives rise to a vector $\Gamma = TA$ for the spline coefficients in terms of the spectral function A at the grid points [15]; $A_i = A(\omega_i)$ and $\Gamma = (a_1, b_1, c_1, d_1, \ldots, d_{N_\omega})$.

With the known coefficients, Eq. (1) can be rewritten as

$$G(i\omega_n) = \sum_{j=1}^{N_\omega}\int_{\omega_j}^{\omega_{j+1}} d\omega \frac{S_j(\omega)}{i\omega_n - \omega} \tag{B7}$$

$$= K\Gamma. \tag{B8}$$

Here K is the matrix obtained by integrating Eq. (B7) [15]. Finally, we have

$$G = \boldsymbol{K}A = KTA. \tag{B9}$$

This procedure provides us the more stable and efficient numerics for MQEM.

2. Nonuniform real-frequency gird

To reduce the number of grid points, a nonuniform real-frequency grid technique [15,25,50] can be used. For the details of implementation, we follow Ref. [15] in which three different regions are considered as the grid sections: $W_L = [\omega_{\min}, w_l)$, $W_C = [w_l, \omega_r]$, and $W_R = (\omega_r, \omega_{\max}]$. For the central region W_C, we take a regular grid spacing of $\Delta\omega$. For W_L, on the other hand, the grid is defined by

$$\omega_j = \frac{1}{u_j} + \omega_{0l} \in W_L, \tag{B10}$$

where $j = 1, \ldots, N_L$. The free parameters N_L, u_j, and ω_{0l} are to be determined. By assuming a constant step Δu, $\omega_l = \frac{1}{u_{N_L+1}} + \omega_{0l}$, and $\omega_{N_L} = \omega_l - \Delta\omega$, we have

$$\Delta u = u_{N_{L+1}} - u_{N_L} \tag{B11}$$

$$= \frac{1}{\omega_l - \omega_{0l}} - \frac{1}{\omega_l - \Delta\omega - \omega_{0l}}, \tag{B12}$$

$u_j = j\Delta u$, and $N_L\Delta u = \frac{1}{\omega_l - \Delta\omega - \omega_{0l}}$. With a given Δu,

$$\omega_{\min} = \frac{1}{\Delta u} + \omega_{0l} \tag{B13}$$

$$= -\frac{(\omega_l - \Delta\omega - \omega_{0l})(\omega_l - \omega_{0l})}{\Delta\omega} + \omega_{0l}, \tag{B14}$$

and $\omega_{0l} = \omega_l + \sqrt{\Delta\omega(\omega_l - \omega_{\min})}$. Since we have an integer value of

$$N_L = \mathrm{ceil}\left(\frac{1}{(\omega_l - \Delta\omega - \omega_{0l})\Delta u}\right) = \mathrm{ceil}\left(\frac{\omega_{0l} - \omega_l}{\Delta\omega}\right), \tag{B15}$$

we redefine $\omega_{0l} = \omega_l + N_L\Delta\omega_l$ and $\omega_{\min}$. The same numerical approach is also used for W_R.

APPENDIX C: CALCULATIONAL DETAILS OF LDA+DMFT

First-principles electronic structure calculations have been carried out based on DFT (density functional theory) within LDA (local density approximation) [51]. We used our DFT software package OpenMX [52–55] for Sr_2IrO_4. $8 \times 8 \times 1$ **k** points for the slab geometry have been taken. SOC is treated within a fully relativistic j-dependent formalism [56]. To describe the electronic correlation, single-site DMFT has been adopted [20,21]. The correlated subspace was constructed by maximally localized Wannier functions starting from the initial projections onto the atomic Ir-t_{2g} orbitals [41,42]. This Hamiltonian serves as the noninteracting H_0 for the multiband Hubbard Hamiltonian $H = H_0 + H_{\mathrm{int}}$. The interaction part is expressed in the Slater-Kanamori form of $H_{\mathrm{int}} = \sum_i h_{i,\mathrm{int}}$:

$$h_{i,\mathrm{int}} = \sum_\alpha U n_{i\alpha\uparrow}n_{i\alpha\downarrow} + \sum_{\alpha\neq\beta} U' n_{i\alpha\uparrow}n_{i\beta\downarrow}, \tag{C1}$$

where U and U' refers to the intraorbital and interorbital interaction, respectively, and $U' = U = 2.2$ eV. Hund interaction J_H is set to zero which does not change any of our main conclusions. The Hamiltonian is solved within single-site DMFT (dynamical mean-field theory) by employing a hybridization expansion continuous-time quantum Monte Carlo (CT-QMC) [23,24] as implemented in Ref. [26] with 2.24×10^8 measurements. In this procedure, local Green's functions are calculated using momentum-independent self-energy:

$$G_{\mathrm{loc}}(i\omega_n) = \frac{1}{N_k}\sum_{\mathbf{k}} \frac{1}{i\omega_n + \mu - H_0(\mathbf{k}) - \Sigma(i\omega_n)}, \tag{C2}$$

where $H_0(\mathbf{k})$ and $\Sigma(i\omega_n)$ are given by 12×12 matrices. Self-energy is decomposed into 6×6 matrices corresponding to two Ir sites, $\Sigma(i\omega_n) = \Sigma_{\mathrm{Ir}(1)}(i\omega_n) \oplus \Sigma_{\mathrm{Ir}(2)}(i\omega_n)$. We would like to note that the self-energy was sampled in the "intermediate representation" [57] to reduce the fluctuation in the high-frequency region.

[1] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011).
[2] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[3] J. E. Hirsch and R. M. Fye, Phys. Rev. Lett. **56**, 2521 (1986).
[4] J. Schött, I. L. M. Locht, E. Lundin, O. Grånäs, O. Eriksson, and I. Di Marco, Phys. Rev. B **93**, 075104 (2016).
[5] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).
[6] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni, Phys. Rev. B **82**, 165125 (2010).
[7] A. W. Sandvik, Phys. Rev. B **57**, 10287 (1998).
[8] J. Otsuki, M. Ohzeki, H. Shinaoka, and K. Yoshimi, Phys. Rev. E **95**, 061302 (2017).
[9] P. Staar, B. Ydens, A. Kozhevnikov, J.-P. Locquet, and T. Schulthess, Phys. Rev. B **89**, 245114 (2014).
[10] L.-F. Arsenault, R. Neuberg, L. A. Hannah, and A. J. Millis, Inverse Probl. **33**, 115007 (2017).
[11] K. Haule, C.-H. Yee, and K. Kim, Phys. Rev. B **81**, 195107 (2010).
[12] O. Goulko, A. S. Mishchenko, L. Pollet, N. Prokof'ev, and B. Svistunov, Phys. Rev. B **95**, 014102 (2017).
[13] N. V. Prokof'ev and B. V. Svistunov, JETP Lett. **97**, 649 (2013).
[14] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990).
[15] D. Bergeron and A.-M. S. Tremblay, Phys. Rev. E **94**, 023303 (2016).
[16] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).
[17] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni, Phys. Rev. B **81**, 155107 (2010).
[18] E. Laue, J. Skilling, and J. Staunton, J. Magn. Reson. **63**, 418 (1985).
[19] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[20] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
[21] X. Y. Zhang, M. J. Rozenberg, and G. Kotliar, Phys. Rev. Lett. **70**, 1666 (1993).
[22] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005).
[23] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, Phys. Rev. Lett. **97**, 076405 (2006).
[24] P. Werner and A. J. Millis, Phys. Rev. B **74**, 155107 (2006).
[25] E. Gull, P. Werner, S. Fuchs, B. Surer, T. Pruschke, and M. Troyer, Comput. Phys. Commun. **182**, 1078 (2011).
[26] K. Haule, Phys. Rev. B **75**, 155113 (2007).
[27] E. Pavarini, S. Biermann, A. Poteryaev, A. I. Lichtenstein, A. Georges, and O. K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004).
[28] C. Martins, M. Aichhorn, L. Vaugier, and S. Biermann, Phys. Rev. Lett. **107**, 266404 (2011).
[29] A. Reymbaut, D. Bergeron, and A.-M. S. Tremblay, Phys. Rev. B **92**, 060509 (2015).
[30] J. M. Tomczak and S. Biermann, J. Phys.: Condens. Matter **19**, 365206 (2007).
[31] J. Yoshitake, J. Nasu, Y. Kato, and Y. Motome, Phys. Rev. B **96**, 024438 (2017).
[32] G. J. Kraberger, R. Triebl, M. Zingl, and M. Aichhorn, Phys. Rev. B **96**, 155128 (2017).
[33] M. Jarrell, *Correlated Electrons: From Models to Materials*, edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (Verlag des Forschungszentrum, Jülich, 2012).
[34] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[35] K.-H. Kim and S. W. Kim, Phys. Rev. E **84**, 012101 (2011).
[36] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, 2010).
[37] K. S. D. Beach, arXiv:cond-mat/0403055.
[38] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[39] A. S. Banerjee, P. Suryanarayana, and J. E. Pask, Chem. Phys. Lett. **647**, 31 (2016).
[40] R. K. Bryan, Eur. Biophys. J. **18**, 165 (1990).
[41] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
[42] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
[43] X. Wang, E. Gull, L. de' Medici, M. Capone, and A. J. Millis, Phys. Rev. B **80**, 045101 (2009).
[44] H. Zhang, K. Haule, and D. Vanderbilt, Phys. Rev. Lett. **111**, 246402 (2013).
[45] E. Gull and A. J. Millis, Phys. Rev. B **90**, 041110 (2014).
[46] J. Mravlje and A. Georges, Phys. Rev. Lett. **117**, 036401 (2016).
[47] Q. Wang, Y. Cao, J. A. Waugh, S. R. Park, T. F. Qi, O. B. Korneta, G. Cao, and D. S. Dessau, Phys. Rev. B **87**, 245109 (2013).
[48] S. J. Moon, H. Jin, K. W. Kim, W. S. Choi, Y. S. Lee, J. Yu, G. Cao, A. Sumi, H. Funakubo, C. Bernhard, and T. W. Noh, Phys. Rev. Lett. **101**, 226402 (2008).
[49] B. J. Kim, H. Jin, S. J. Moon, J.-Y. Kim, B.-G. Park, C. S. Leem, J. Yu, T. W. Noh, C. Kim, S.-J. Oh, J.-H. Park, V. Durairaj, G. Cao, and E. Rotenberg, Phys. Rev. Lett. **101**, 076402 (2008).
[50] R. Levy, J. P. LeBlanc, and E. Gull, Comput. Phys. Commun. **215**, 149 (2017).
[51] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[52] M. J. Han, T. Ozaki, and J. Yu, Phys. Rev. B **73**, 045110 (2006).
[53] T. Ozaki and H. Kino, Phys. Rev. B **69**, 195113 (2004).
[54] T. Ozaki, Phys. Rev. B **67**, 155108 (2003).
[55] http://www.openmx-square.org.
[56] A. H. MacDonald and S. H. Vosko, J. Phys. C: Solid State Phys. **12**, 2977 (1979).
[57] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, Phys. Rev. B **96**, 035147 (2017).