Sparse modeling approach to analytical continuation of imaginary-time quantum Monte Carlo data

Junya Otsuki,[1] Masayuki Ohzeki,[2] Hiroshi Shinaoka,[3] and Kazuyoshi Yoshimi[4]

[1]*Department of Physics, Tohoku University, Sendai 980-8578, Japan*
[2]*Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan*
[3]*Department of Physics, Saitama University, Saitama 338-8570, Japan*
[4]*Institute for Solid State Physics, University of Tokyo, Chiba 277-8581, Japan*

A data-science approach to solving the ill-conditioned inverse problem for analytical continuation is proposed. The root of the problem lies in the fact that even tiny noise of imaginary-time input data has a serious impact on the inferred real-frequency spectra. By means of a modern regularization technique, we eliminate redundant degrees of freedom that essentially carry the noise, leaving only relevant information unaffected by the noise. The resultant spectrum is represented with minimal bases and thus a stable analytical continuation is achieved. This framework further provides a tool for analyzing to what extent the Monte Carlo data need to be accurate to resolve details of an expected spectral function.

I. INTRODUCTION

Numerical and analytical calculations in quantum many-body systems are in most cases performed with the imaginary-time framework. Diagrammatic perturbation theory [1] and variants of quantum Monte Carlo (QMC) simulations [2,3] take full advantage of imaginary-time descriptions of statistical averages. One needs, however, to perform analytical continuation to transform calculated imaginary-time quantity $G(\tau)$ to real-frequency spectra $\rho(\omega)$, which can be compared directly to experimental results. This becomes problematic particularly when handling QMC data, because analytical continuation is extremely sensitive to noise.

Due to the sensitivity to noise, the standard Padé approximation [4] often yields unphysical spectra that even break preconditions such as the sum rule and causality. In order to stably obtain a physically reasonable spectrum, various numerical algorithms have been developed such as the maximum entropy method (MaxEnt) [5–7] and stochastic method [8–12]. With the recent progress in computational and information theories, there are yet growing attempts to settle this long-standing issue [13–21]. Despite extensive efforts from many different angles, a fundamental question still remains: to what extent imaginary-time data with statistical errors have relevant information in the first place, and in other words, how much we can, in principle, reconstruct the fine structure of real-frequency spectra.

In this Rapid Communication, we address this fundamental issue by presenting an alternative approach based on the concept of sparse modeling (SpM), which has been developed in the context of data-driven science. Technically, SpM provides ways to extract relevant variables for representing high-dimensional data, eliminating redundant variables that potentially cause overfitting. Our idea is that, by *enforcing sparseness on imaginary-time data represented in a properly constructed basis set*, we are able to extract relevant information to perform stable analytical continuations against noise. We shall demonstrate that this idea does work and further reveals the accuracy of imaginary-time QMC data required for reproducing the structure of spectral functions.

II. FORMALISM

A. Descriptions of the problem

The input of analytical continuation is the imaginary-time Green's function $G(\tau)$ or its Fourier transform $G(i\omega_n)$, where ω_n denotes a Matsubara frequency. If the analytical expression of $G(i\omega_n)$ is known, one can readily obtain the spectral function $\rho(\omega)$ using the relation $\rho(\omega) = (1/\pi)\mathrm{Im}G(\omega + i0)$ by replacing $i\omega_n$ with $\omega + i0$. For numerical data, one may use the exact integral equation between $\rho(\omega)$ and $G(\tau)$ [22],

$$G(\tau) = \int_{-\infty}^{\infty} d\omega \, K_{\pm}(\tau,\omega)\rho(\omega), \tag{1}$$

where $0 \leqslant \tau \leqslant \beta \equiv 1/T$ and the kernel $K_{\pm}$ is given by

$$K_{\pm}(\tau,\omega) = \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}}. \tag{2}$$

Here, the $+$ $(-)$ sign is for fermionic (bosonic) correlation functions. In this representation, the analytical continuation may be read as an inverse problem in which one infers $\rho(\omega)$ from its integrated values $G(\tau)$ in the presence of noise. This constitutes an ill-posed problem: the kernel K is exponentially small at large ω, and so the right-hand side of Eq. (1) is insensitive to variations of $\rho(\omega)$. Therefore, tiny noise in the left-hand side can considerably affect the "best" inference of $\rho(\omega)$. In other words, there are enormous numbers of plausible solutions that satisfy Eq. (1) within a given accuracy. Our objective is to *select* a reasonable solution which is independent of noise.

For convenience, we recast Eq. (1) into a conventional linear equation with dimensionless quantities as

$$\boldsymbol{G} = K\boldsymbol{\rho}. \tag{3}$$

Here, the vector $\boldsymbol{G}$ is defined by $G_i \equiv G(\tau_i)$ with τ_i being M division of $[0 : \beta]$. In the fermionic cases, the quantities on the right-hand side are defined by $K_{ij} \equiv K_+(\tau_i,\omega_j)$ and $\rho_j \equiv \rho(\omega_j)\Delta\omega$ [23], which are obtained after replacing the integral over ω with N-point finite differences in the range $[-\omega_{\max} : \omega_{\max}]$. One may use a nonlinear mesh for better efficiency, but the formulation below does not change. When the input $\boldsymbol{G}$ has noise, deviation from Eq. (3) needs to be taken into account.

For this reason, we consider the square error

$$\chi^2(\boldsymbol{\rho}) = \tfrac{1}{2}\|\boldsymbol{G} - K\boldsymbol{\rho}\|_2^2, \qquad (4)$$

and find $\boldsymbol{\rho}$ such that $\chi^2(\boldsymbol{\rho}) < \eta$ with η being a small constant depending on the magnitude of noise. Here, $\|\cdot\|_2$ stands for the L_2 norm defined by $\|\boldsymbol{\rho}\|_2 \equiv (\sum_j \rho_j^2)^{1/2}$.

The solution must hold two conditions: non-negativity $\rho(\omega) \geqslant 0$ and the sum rule $\int_{-\infty}^{\infty} \rho(\omega)d\omega = 1$. They are expressed in terms of the vector $\boldsymbol{\rho}$ as

$$\rho_j \geqslant 0, \qquad \sum_j \rho_j = 1. \qquad (5)$$

These constraints are applied to diagonal components of Green's functions. For off-diagonal components, the non-negativity is not applied, while the sum rule always exists [24]. Our algorithm presented below works both with and without those constraints. This is a technical advantage over MaxEnt, in which the entropy term requires the positiveness, $\rho(\omega) > 0$ [25].

B. Efficient basis set

We discuss what basis best describes spectral functions $\rho(\omega)$. Here, "best" means the ability to reproduce the correct $\rho(\omega)$ (i) with a small number of bases (ii) for wide models independent of details of interactions or parameters. For this purpose, we focus on the fact that the matrix K is ill-conditioned. To see this, we use the singular value decomposition (SVD) of the matrix K:

$$K = USV^{\mathrm{t}}, \qquad (6)$$

where S is an $M \times N$ diagonal matrix, and U and V are orthogonal matrices of size $M \times M$ and $N \times N$, respectively. It should be noted that the singular values s_l ($l = 0, 1, 2, \ldots$) decay exponentially or even faster [see Fig. 3(a1)]. This makes a numerical optimization of $\chi^2(\boldsymbol{\rho})$ unstable. A standard recipe for avoiding this difficulty is to drop vectors corresponding to small s_l below a certain threshold [5,26]. Although this yields some definite solution, the result depends totally on the threshold. We make another use of SVD of the ill-conditioned matrix in modern perspective of SpM.

C. L_1 regularization

We reconsider the expression of $\chi^2(\boldsymbol{\rho})$ in Eq. (4). Introducing new vectors

$$\boldsymbol{\rho}' \equiv V^{\mathrm{t}}\boldsymbol{\rho}, \quad \boldsymbol{G}' \equiv U^{\mathrm{t}}\boldsymbol{G}, \qquad (7)$$

we obtain

$$\chi^2(\boldsymbol{\rho}') = \frac{1}{2}\|\boldsymbol{G}' - S\boldsymbol{\rho}'\|_2^2 = \frac{1}{2}\sum_l (G'_l - s_l\rho'_l)^2. \qquad (8)$$

It turns out that the contribution of ρ'_l to $\chi^2(\boldsymbol{\rho}')$ is weighted by the corresponding s_l. Since s_l decays exponentially, most elements of $\boldsymbol{\rho}'$ give only negligible contribution to $\chi^2(\boldsymbol{\rho}')$. Hence, such elements are essentially indefinite as far as $\chi^2(\boldsymbol{\rho}')$ is concerned, making a naive analytical continuation quite sensitive to noise.

The above consideration brings us to the idea that, by imposing sparseness on $\boldsymbol{\rho}'$, we can find a stable solution
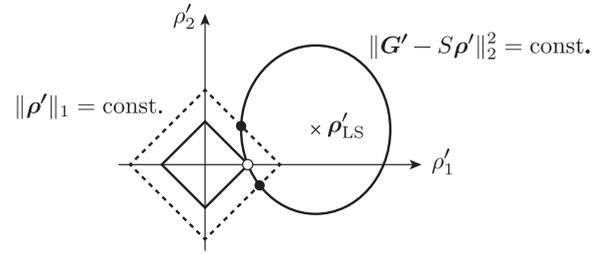

FIG. 1. Explanation for the mechanism that a sparse solution is chosen by the L_1 regularization. For details, see the paragraph below Eq. (9).

which is robust against noise. To this end, we consider the cost function including an L_1 regularization term

$$F(\boldsymbol{\rho}') \equiv \tfrac{1}{2}\|\boldsymbol{G}' - S\boldsymbol{\rho}'\|_2^2 + \lambda\|\boldsymbol{\rho}'\|_1, \qquad (9)$$

where λ is a positive constant and $\|\cdot\|_1$ denotes the L_1 norm defined by $\|\boldsymbol{\rho}'\|_1 \equiv \sum_l |\rho'_l|$. This form of optimization problems is referred to as LASSO (least absolute shrinkage and selection operators) [27].

The role of the L_1 term may be explained as follows. We consider a two-dimensional vector $\boldsymbol{\rho}' = (\rho'_1, \rho'_2)$ as the simplest example (Fig. 1). We suppose that the minimum of $\chi^2(\boldsymbol{\rho}')$, namely, the solution of the least-square method, is located at $\boldsymbol{\rho}' = \boldsymbol{\rho}'_{\mathrm{LS}}$. Equal-value contours of $\chi^2(\boldsymbol{\rho}')$ are elliptic centered at $\boldsymbol{\rho}'_{\mathrm{LS}}$, and all points on this line, e.g., open and closed circles in Fig. 1, are equally feasible when a certain extent of errors are allowed. When the L_1 term is included, the open circle becomes the most favorable, since the contours of the L_1 norm exhibit a cusp on each axis. The L_1 term thus selects a sparse solution out of an infinite number of plausible solutions of least squares. Furthermore, since the LASSO is a convex optimization, the global minimum can be obtained regardless of initial conditions [28]. Therefore, the present scheme is computationally inexpensive.

Our task now is to find $\boldsymbol{\rho}'$ that minimizes Eq. (9) subject to the constraints in Eq. (5). We applied an algorithm named alternating direction method of multipliers (ADMM) developed by Boyd et al. [29]. See the Appendix for detailed explanations for this algorithm. How to choose a reasonable value of λ will be discussed later.

III. DEMONSTRATIVE RESULTS

We present demonstrative results of our SpM scheme. Imaginary-time input data were prepared as follows. We construct a model spectrum $\rho^{\mathrm{exact}}(\omega)$ with three Gaussians that imitate a typical spectrum in the single-impurity Anderson model, as shown in Fig. 2(b). This spectrum is transformed into $G^{\mathrm{exact}}(\tau)$ by performing the integral in Eq. (1) with $\beta = 100$. Introducing Gaussian noise $\boldsymbol{\eta}$ with the standard deviation $\sigma = 10^{-3}$, we obtained input data $G_i^{\mathrm{input}} = G^{\mathrm{exact}}(\tau_i) + \eta_i$ [Fig. 2(a)]. The discretization parameters are $M = 4001$ for τ and $N = 1001$ for ω, and cutoff is $\omega_{\mathrm{max}} = 4$.

Figure 2(b) shows the spectrum $\rho^{\mathrm{SpM}}(\omega)$ reconstructed by SpM. Results for three different values of λ are plotted: an optimal choice $\lambda = 10^{-1.8} \equiv \lambda_{\mathrm{opt}}$ in (b2), and larger and smaller values in (b1) and (b3), respectively. How to estimate λ_{opt} will
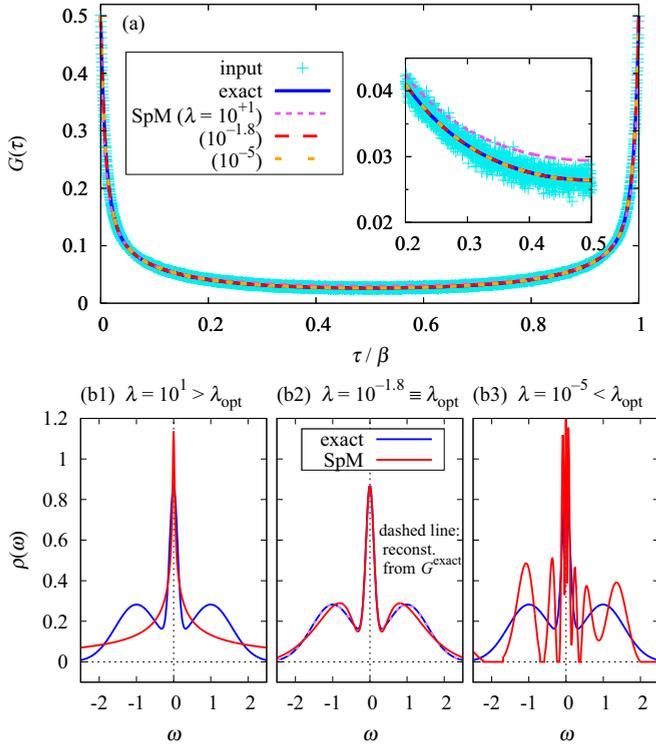
FIG. 2. (a) Imaginary-time data: the exact G^{exact} without noise, the input G^{input} with noise, and the result G^{SpM} recovered after analytical continuations. (b) Real-frequency spectra: the exact ρ^{exact} and the result ρ^{SpM} reconstructed from G^{input}. Here, ρ^{exact} consists of three Gaussians with parameters (position, width, weight) = $(0, 0.15, 0.2), (\pm 1, 0.8, 0.4)$. The dashed line in (b2) (almost overlapping with ρ^{exact}) shows a spectrum reconstructed from G^{exact} for comparison ($\lambda = 10^{-12}$).
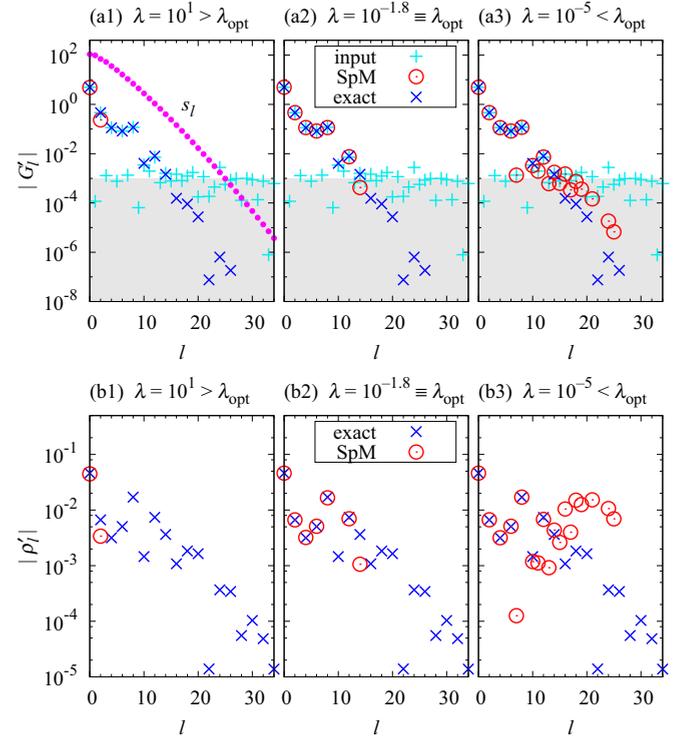

FIG. 3. (a) Imaginary-time and (b) real-frequency data represented in the SV basis. The shaded area in (a) indicates regions below $\sigma = 10^{-3}$. The closed circles in (a1) show the singular values s_l.

be discussed later. In the optimal case, a reasonable agreement is seen around $\omega = 0$. The deviation around $\omega = \pm 1$ is due to the noise, since the whole spectrum can be reconstructed in the absence of noise [dashed line in Fig. 2(b2)]. This deviation indicates the limitation in reconstructing the real-frequency spectrum from noisy imaginary-time data, which will be discussed later. The spectrum becomes featureless for strong regularization ($\lambda > \lambda_{\text{opt}}$), while artificial spikes appear for weak regularization ($\lambda < \lambda_{\text{opt}}$). The latter is typical overfitting behavior.

Here, we go back to the imaginary-time data in Fig. 2(a) and check agreement between the input $\boldsymbol{G}^{\text{input}}$ and the SpM result $K\boldsymbol{\rho}^{\text{SpM}} \equiv \boldsymbol{G}^{\text{SpM}}$. In the optimal case, $\boldsymbol{G}^{\text{SpM}}$ shows a perfect agreement with the exact data without noise rather than $\boldsymbol{G}^{\text{input}}$, meaning that the noise has been removed. Note that an equally good agreement is also seen in $\lambda < \lambda_{\text{opt}}$, indicating that the two apparently different spectra, Figs. 2(b2) and 2(b3), are equally "good" solutions in terms of $\chi^2(\boldsymbol{\rho})$. We will see below that those spectra are clearly distinguished by taking the L_1 regularization into account.

In Fig. 3, the data in Fig. 2 are represented in the basis defined with SVD in Eq. (7) [termed as singular value (SV) basis hereafter]. We first remark that the imaginary-time data, G'_l, decay exponentially as shown in Fig. 3(a). It follows that the input data have only about six elements above the

magnitude of the noise, $\sigma = 10^{-3}$, and other information is lost. The optimal solution turns out to select those elements properly, while fewer (more) elements are selected in the results for $\lambda > \lambda_{\text{opt}}$ ($\lambda < \lambda_{\text{opt}}$). Correspondingly, the spectra are represented with two elements for $\lambda > \lambda_{\text{opt}}$, seven elements for $\lambda = \lambda_{\text{opt}}$, and many elements including incorrect values for $\lambda < \lambda_{\text{opt}}$ as shown in Fig. 3(b).

The data in Fig. 3(a) are highly suggestive: Only a few elements of $\boldsymbol{G}'$ possess relevant information unaffected by noise. This is an intrinsic feature of imaginary-time quantities as explained below. In the SV basis, Eq. (3) may be written as $G'_l = s_l \rho'_l$. Hence, the fast decay of G'_l originates in s_l rather than ρ'_l, meaning that it does *not* depend on particular models. It follows that large-l components of G'_l are inevitably buried in noise [30]. Since large-l bases correspond to highly oscillatory functions, the fine structure of (unknown) exact $\rho(\omega)$ are essentially lost in imaginary-time QMC data. The present regularization scheme extracts the full information of the input $G(\tau)$, which, however, gives only limited information of the real-frequency counterpart.

Finally, we discuss how to find an optimal value of λ. Figure 4(a) shows λ dependence of the square error $\chi^2(\boldsymbol{\rho}')$ in Eq. (8). As λ decreases from the strong regularization regime, $\chi^2(\boldsymbol{\rho}')$ first drops rapidly and then becomes more or less saturated below $\lambda \sim 10^{-2}$, which is an indication of overfitting. We obtained reasonable spectra in a wide region around this kink, namely, $\lambda \simeq 10^{-2.6}$–$10^{-1.4}$ (colored area in Fig. 4). Hence, an optimal value may be determined as follows. We first define a function $f(\lambda) = a\lambda^b$ which connects the left and right endpoints of $\chi^2(\boldsymbol{\rho}')$ [dashed line in Fig. 4(a)]. Then,
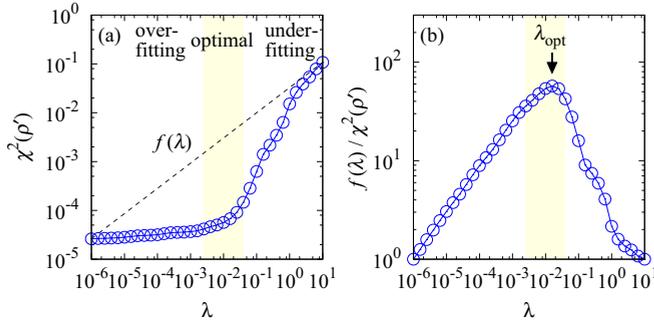
FIG. 4. (a) The square error $\chi(\rho')$ as a function of λ. (b) Reduction of $\chi(\rho')$ relative to $f(\lambda)$ shown in (a).

the peak in the ratio $f(\lambda)/\chi^2(\rho')$ corresponds to the position of the kink in $\chi^2(\rho')$ [Fig. 4(b)]. In this way, we obtained $\lambda_{opt} = 10^{-1.8} \approx 1.6 \times 10^{-2}$. A similar method was adopted in the literature [31,32].

IV. REQUIRED ACCURACY OF QMC DATA

In this section, we discuss a relation between accuracy of QMC data and capability of reproducing spectrum. Let us consider a situation where the overall structure of $\rho(\omega)$ is established in the literature, but its finer structure is controversial. Practical problems of interest are, e.g., a peaklike structure at the edge of a Mott gap [33], and spin excitations in the square-lattice Heisenberg model [34,35]. For investigating such issues with QMC, it is highly convenient if one can know accuracy of $G(\tau)$ necessary to achieve reliable $\rho(\omega)$ for a specific problem. As an example to illustrate our idea, we consider a two-peak spectrum $\rho^{expect}(\omega)$ shown in Fig. 5(a), and suppose that the existence of the sharper peak located at $\omega = 2.2$ is a matter of issue. Our principal interest here is
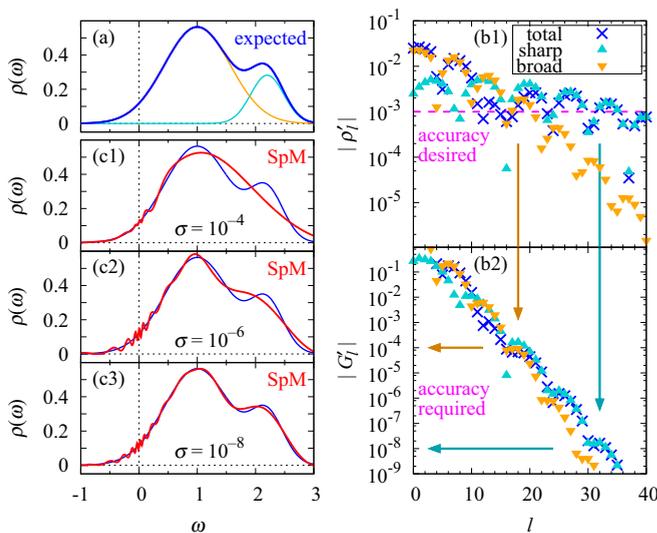
how much accuracy is required for QMC data to verify the existence or absence of the two-peak structure.

We first transform $\rho^{expect}(\omega)$ into ρ'_l in the SV basis. Figure 5(b1) shows ρ'_l, where the contribution of each peak is plotted separately. The data for the sharper peak decays slower than the broader peak. In general, a narrower peak at a higher energy yields slower decay. We consider that each peak can be reconstructed with sufficient accuracy from the components satisfying, e.g., $\rho'_l \gtrsim 10^{-3}$, which correspond to $l \lesssim 18$ and 32 for the broader and sharper peaks, respectively. Those boundaries are converted into those in $G(\tau)$: As illustrated in Figs. 5(b1) and 5(b2), we finally obtain the required accuracies $10^{-4} \equiv \sigma_{broad}$ and $10^{-8} \equiv \sigma_{sharp}$ to reproduce the broader and shaper peaks, respectively.

Now we test these estimations by SpM calculations. We prepared three sets of input $G(\tau)$ with different noise levels, $\sigma = 10^{-4}$, 10^{-6}, and 10^{-8}. The computed spectra (with optimal value of λ_{opt}) are shown in Figs. 5(c1)–5(c3). The result for $\sigma = 10^{-4}$ exhibits no separable two-peak structure. As σ decreases, the higher-energy peak grows, and finally at $\sigma = 10^{-8}$, the distinct two-peak structure is observed as expected. The above estimation of the required accuracy thus turns out to be reasonable. If QMC data with sufficient accuracy (σ_{sharp} in the above example) does not yield the expected structure, then we can conclude its absence.

V. SUMMARY

The essence of our algorithm is twofold. First, the SVD enables an efficient representation of imaginary-time input data and real-frequency spectra. Second, the L_1 regularization selects out bases having relevant information, removing noise automatically. We can thus perform stable analytical continuations without any tuning parameters. Instead of using prior knowledge as in other methods, our scheme makes full use of the ill-conditioned nature of the kernel, which has been the source of the problem in analytical continuations, but it now brings a significant advantage in reducing redundant bases. With this advantage, it is also possible to estimate QMC accuracy that is required to resolve the essential features of a given spectral function. It will stimulate future investigations of verifying controversial features of spectral functions.

Remarkably, our results indicated that imaginary-time data contain much less information than its size in the presence of noise. The resultant compact representation (G'_l) may be used not only for analytical continuations but also for imaginary-time-based calculations such as diagrammatic expansions and QMC measurements. This possibility will be pursued in a separate paper [36].

FIG. 5. (a) A model spectrum $\rho^{expect}(\omega)$ consisting of two Gaussian peaks: (position, width, weight) = (1, 0.8, 0.8), (2.2, 0.4, 0.2). (b) Corresponding ρ'_l and G'_l. (c) Spectra computed with the SpM scheme from three sets of $G(\tau)$ with different noise levels, $\sigma = 10^{-4}$, 10^{-6}, and 10^{-8}.

APPENDIX: ADMM

In the present scheme, we need to find $\boldsymbol{\rho}'$ that minimizes $F(\boldsymbol{\rho}')$ in Eq. (9) subject to the constraints in Eq. (5). How to solve it is presented in this Appendix. Following the conventional notation, we change the variables as $\boldsymbol{G} \to \boldsymbol{y}$ and $\boldsymbol{\rho} \to \boldsymbol{x}$. Then, the cost function reads

$$F(\boldsymbol{x}') = \tfrac{1}{2}\|\boldsymbol{y}' - S\boldsymbol{x}'\|_2^2 + \lambda\|\boldsymbol{x}'\|_1, \tag{A1}$$

and the constraint is represented as

$$x_j \geqslant 0, \quad \langle\boldsymbol{x}\rangle \equiv \sum_j x_j = 1, \tag{A2}$$

where $\boldsymbol{x} = V\boldsymbol{x}'$. Here, notational convention is that vectors without a prime denote quantities represented in the original $\tau - \omega$ basis (dimension M or N) and those with a prime in the SV basis [dimension $L = \min(M,N)$]. We can reduce L without affecting the result, which will be discussed later.

We apply the ADMM algorithm developed by Boyd *et al.* [29]. Following the ADMM procedure, we introduce auxiliary vectors $\boldsymbol{z}$ and $\boldsymbol{z}'$, and consider minimization of the function

$$\widetilde{F}(\boldsymbol{x}',\boldsymbol{z}',\boldsymbol{z}) = \frac{1}{2\lambda}\|\boldsymbol{y}' - S\boldsymbol{x}'\|_2^2 - \nu(\langle V\boldsymbol{x}'\rangle - 1)$$
$$+ \|\boldsymbol{z}'\|_1 + \lim_{\gamma\to\infty}\gamma\sum_j\Theta(-z_j), \tag{A3}$$

subject to

$$\boldsymbol{z}' = \boldsymbol{x}', \quad \boldsymbol{z} = V\boldsymbol{x}'. \tag{A4}$$

The sum rule is imposed by the Lagrange multiplier ν, and non-negativity is represented by an infinite potential γ that acts on negative elements (Θ is the Heaviside step function). The essence of this method is that a minimization is performed separately for each vector $\boldsymbol{x}'$, $\boldsymbol{z}'$, and $\boldsymbol{z}$, and their consistency is imposed afterwards gradually. The ADMM algorithm thus achieves flexibility of handling plural constraints.

The constraints in Eq. (A4) are treated by the augmented Lagrange multiplier method. A brief description of the treatment of the first constraint, $\boldsymbol{z}' = \boldsymbol{x}'$, is given below. Two kinds

of coefficients play a cooperative role: (normalized) Lagrange multipliers $\boldsymbol{u}'$ which couple with $(\boldsymbol{z}' - \boldsymbol{x}')$ and a coefficient μ' of a penalty term $\|\boldsymbol{z}' - \boldsymbol{x}'\|_2^2$. The parameter $\mu' (= 1$–$100)$ controls speed of convergence, while $\boldsymbol{u}'$ is iteratively updated together with its conjugate variable $\boldsymbol{z}'$. Similarly, we introduce μ and $\boldsymbol{u}$ for the second constraint, $\boldsymbol{z} = V\boldsymbol{x}'$.

Omitting detailed derivations (we refer readers to Ref. [29]), we present below update formulas used in actual computations:

$$\boldsymbol{x}' \leftarrow \left(\frac{1}{\lambda}S^tS + (\mu' + \mu)\mathbf{1}\right)^{-1}$$
$$\times \left(\frac{1}{\lambda}S^t\boldsymbol{y}' + \mu'(\boldsymbol{z}' - \boldsymbol{u}') + \mu V^t(\boldsymbol{z} - \boldsymbol{u}) + \nu V^t\boldsymbol{e}\right)$$
$$\equiv \boldsymbol{\xi}_1 + \nu\boldsymbol{\xi}_2, \tag{A5a}$$
$$\boldsymbol{z}' \leftarrow \mathcal{S}_{1/\mu'}(\boldsymbol{x}' + \boldsymbol{u}'), \tag{A5b}$$
$$\boldsymbol{u}' \leftarrow \boldsymbol{u}' + \boldsymbol{x}' - \boldsymbol{z}', \tag{A5c}$$
$$\boldsymbol{z} \leftarrow \mathcal{P}_+(V\boldsymbol{x}' + \boldsymbol{u}), \tag{A5d}$$
$$\boldsymbol{u} \leftarrow \boldsymbol{u} + V\boldsymbol{x}' - \boldsymbol{z}, \tag{A5e}$$

where $e_i = 1$ and

$$\nu = \frac{1 - \langle V\boldsymbol{\xi}_1\rangle}{\langle V\boldsymbol{\xi}_2\rangle}. \tag{A6}$$

$\mathcal{P}_+$ denotes a projection onto a non-negative quadrant, i.e., $\mathcal{P}_+z_j = \max(z_j, 0)$ for each element. $\mathcal{S}_\alpha(\boldsymbol{x})$ is the elementwise soft thresholding function defined by

$$\mathcal{S}_\alpha(x) = \begin{cases} x - \alpha & (x > \alpha) \\ 0 & (-\alpha \leqslant x \leqslant \alpha) \\ x + \alpha & (x < -\alpha). \end{cases} \tag{A7}$$

Starting with the initial conditions, $\boldsymbol{x}' = \boldsymbol{z}' = \boldsymbol{u}' = \mathbf{0}$ and $\boldsymbol{z} = \boldsymbol{u} = \mathbf{0}$, the updates in Eqs. (A5) are repeated until convergence is reached. Since updates require only matrix-vector products, computational cost is quite cheap.

The most time-consuming part is the SVD of the kernel K. Once it is done, we convert the input $\boldsymbol{y}$ into the SV basis, $\boldsymbol{y}' = U^t\boldsymbol{y}$, and perform the rest of the calculations in this representation. At this stage, we can safely drop bases having small s_l of the order of rounding errors, e.g., $s_l < 10^{-10}$. This does not affect the result, since components of those bases finally become zero by the L_1 regularization. For the calculations in this Rapid Communication, the number of bases is reduced from $(M, N) = (4001, 1001)$ to $L = 50$. Therefore, considerable speedup of the iteration can be achieved.

[1] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Satistical Physics* (Dover Publications, New York, 1963).

[2] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011).

[3] J. Gubernatis, N. Kawashima, and P. Werner, *Numerical Approaches to Spatial Correlations in Strongly Interacting Fermion Systems* (Cambridge University Press, Cambridge, UK, 2016).

[4] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).

[5] M. Jarrell and J. Gubernatis, Phys. Rep. **269**, 133 (1996).

[6] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni, Phys. Rev. B **81**, 155107 (2010).

[7] D. Bergeron and A.-M. S. Tremblay, Phys. Rev. E **94**, 023303 (2016).

[8] A. W. Sandvik, Phys. Rev. B **57**, 10287 (1998).

[9] A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).

[10] S. Fuchs, T. Pruschke, and M. Jarrell, Phys. Rev. E **81**, 056701 (2010).

[11] K. S. D. Beach, arXiv:cond-mat/0403055.

[12] A. W. Sandvik, Phys. Rev. E **94**, 063308 (2016).

[13] K. S. D. Beach, R. J. Gooding, and F. Marsiglio, Phys. Rev. B **61**, 5147 (2000).

[14] A. Östlin, L. Chioncel, and L. Vitos, Phys. Rev. B **86**, 235107 (2012).

[15] A. Dirks, M. Eckstein, T. Pruschke, and P. Werner, Phys. Rev. E **87**, 023305 (2013).

[16] F. Bao, Y. Tang, M. Summers, G. Zhang, C. Webster, V. Scarola, and T. A. Maier, Phys. Rev. B **94**, 125149 (2016).

[17] J. Schött, I. L. M. Locht, E. Lundin, O. Grånäs, O. Eriksson, and I. Di Marco, Phys. Rev. B **93**, 075104 (2016).

[18] R. Levy, J. P. F. LeBlanc, and E. Gull, Comput. Phys. Commun. **215**, 149 (2017).

[19] O. Goulko, A. S. Mishchenko, L. Pollet, N. Prokof'ev, and B. Svistunov, Phys. Rev. B **95**, 014102 (2017).

[20] G. Bertaina, D. E. Galli, and E. Vitali, Adv. Phys. **2**, 302 (2017).

[21] L.-F. Arsenault, R. Neuberg, L. A. Hannah, and A. J. Millis, arXiv:1612.04895.

[22] Note the sign of $G(\tau)$: In our definition, $G(\tau)$ is positive definite in $0 \leqslant \tau \leqslant \beta$.

[23] In bosonic cases, K and ρ are defined by $K_{ij} \equiv \omega K_-(\tau_i, \omega_j)$ and $\rho_j \equiv [\rho(\omega_j)/\omega]\Delta\omega$. Then, all descriptions below are applicable.

[24] We can determine the spectral sum c by $c = G(\tau = 0) + G(\beta)$ or by the high-frequency tail $-c/(i\omega_n)$.

[25] For non-negative spectra in MaxEnt, see Ref. [37].

[26] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, UK, 1988).

[27] R. Tibshirani, J. R. Stat. Soc. B **58**, 267 (1996).

[28] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Found. Trends Mach. Learn. **3**, 1 (2011).

[30] A similar feature was observed when expanding $G(\tau)$ in terms of the Legendre polynomials [38].

[31] A. Decelle and F. Ricci-Tersenghi, Phys. Rev. Lett. **112**, 070603 (2014).

[32] S. Yamanaka, M. Ohzeki, and A. Decelle, J. Phys. Soc. Jpn. **84**, 024801 (2015).

[33] S. Nishimoto, F. Gebhard, and E. Jeckelmann, J. Phys.: Condens. Matter **16**, 7063 (2004).

[34] B. Dalla Piazza, M. Mourigal, N. B. Christensen, G. J. Nilsen, P. Tregenna-Piggott, T. G. Perring, M. Enderle, D. F. McMorrow, D. A. Ivanov, and H. M. Rønnow, Nat. Phys. **11**, 62 (2015).

[35] A. Masaki-Kato, A. S. Mishchenko, T. Shirakawa, N. Nagaosa, and S. Yunoki (unpublished).

[36] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, arXiv:1702.03054.

[37] A. Reymbaut, D. Bergeron, and A.-M. S. Tremblay, Phys. Rev. B **92**, 060509 (2015).

[38] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, Phys. Rev. B **84**, 075145 (2011).