Sparse modeling for Quantum Monte-Carlo simulation

Masayuki Ohzeki
Graduate School of Information Sciences, Tohoku University, Sendai, 980-8579, Japan
E-mail: mohzeki@tohoku.ac.jp

Abstract. We show a new kind of applications of the sparse modeling to a traditional problem in the condensed-matter physics. In the quantum Monte-Carlo simulation, we observe a huge amount of data for investigation of the details of the low-energy behavior for interacting many-body systems. Although the real-time behavior is actually under investigation, the quantum Monte-Carlo simulation is performed on the imaginary time for restriction of the method. Thus we need a technique for analytical continuation connecting between the real and imaginary-time functions. However the analytical continuation can be problematic because the problem consists of solving the ill-conditioned equation. In the present study, by employing an adequate regularization, we solve efficiently the ill-conditioned equation in the analytical continuation. As a result, we have a novel way to perform the analytical continuation and find an intermediate representation between imaginary-time and real-frequency domains.

1. Introduction
In the present manuscript, we provide an elementary guide for our recent papers from many-body physics to fascinating applications of the sparse modeling [1, 2]. In the analysis on the condensed matter, it is harmful to directly deal with the theoretical model for investigating some nontrivial aspects. We usually employ various approximate method and numerical calculations. In the quantum many-body systems, most of the analyses are performed with recourse to the imaginary-time framework for computing the real-time behavior. Then one needs to utilize the analytical continuation to transform the estimated quantities in imaginary-time regions to real-frequency regions, which can be compared to the experimental results. The representative method to deal with the quantum many-body systems is the elaborate diagrammatic approach widely used for investigating both of static and dynamics responses of the systems [3, 4]. Variants of the quantum Monte-Carlo simulations (QMC) also take full advantage of imaginary-time descriptions [5, 6]. We consider the case of the QMC in the following.

The analytical continuation can be formulated as the inverse problem. Given the relationship between imaginary-time description $G$ and the real-frequency behavior $\rho$ as

$$G = K \rho. \quad (1)$$

we find a reasonable solution of $\rho$ from $G$ obtained by the QMC. However the inverse problem can be ill-conditioned one since the singular values of $K$ decay very fast. Then most of the independent components in $\rho$ lead to no contributions to $G$. Furthermore any noise in $G$ significantly affect $\rho$. In order to obtain a reasonable solution, one often employs the maximum
entropy method (MaxEnt or MEM) [7, 8, 9]. By regarding the non-negative quantity $\rho$ as the probability distribution, one maximize the entropy of the distribution function, while minimizing the consistency with Equation (1). The inclusion of an entropy term to find a reasonable solution leads to a bias toward spectra with large entropy. In some sense, the obtained solution can be most probable one given the QMC data. However, one do not know a priori how to induce the effect of the entropy in finding the solution. Despite extensive efforts from various aspects, the method remains arbitrariness how to solve the inverse problems.

Recent development of solving the inverse problem is the stochastic way [10, 11, 12]. In the stochastic analytical continuation, the desired quantity $\rho$ is sampled according to the squared error between $G$ and $K\rho$ and the “temperature” controlling the variance of the sampling as the simulated annealing [13].

On the other hand, to solve the inverse problem on the QMC, we propose the application of the sparse modeling. In particular, we employ the procedure of the least-absolute value shrinkage and selection operators (LASSO) by considering the least square estimation with regularization of the $L_1$ norm. The $L_1$ is a core technology in the sparse modeling. The $L_1$ norm promotes the sparse solution compared to the case of the simple least square estimation. For instance, the relevant interaction from many snapshots of a vast number of components can be chosen by employing the $L_1$ norm as well as the standard cost function to analyze the data [14, 15, 16].

Here we assume that the relevant information, which can be obtained by the QMC data, should be sparse in the analytical continuation. The straightforward approach by use of the LASSO implies that some sparsity on the desired quantity can be expected. We do not stick this idea for the QMC data. Instead, we aim at efficiently cut down the effects from the noisy QMC data to the desired quantity by considering the feature of the Kernel matrix $K$.

The remaining of the paper consists of the following five sections. The second section shows the formulation of the Matsubara Green function for investigating the quantum many body systems. In the third section, we formulate the inverse problem to be solved and review the MEM in short. The fourth section explains the application of the sparse modeling to the QMC data following the literature [1]. We show how to solve the inverse problem in the fifth section. In the last section, we summarize our study.

2. Matsubara Green function

We first review the Matsubara Green function, which is a central quantity to be investigated in quantum many-body physics. In quantum many-body physics, one of the target is to estimate the density of states $\rho(\omega)$. From the density of states, we can compute the interested quantities as the internal energy, specific heat, thermal conductivity etc. of the system. In order to obtain the density of states, we compute the Matsubara Green function defined by the following thermal expectation:

$$G(\tau) = \left\langle T\hat{a}(\tau)\hat{a}^\dagger(0) \right\rangle, \quad (2)$$

where $\hat{a}(t)$ and $\hat{a}^\dagger(t)$ are the annihilation and creation operators of the particle in the system and the bracket denotes the thermal expectation defined as

$$\langle \hat{A} \rangle = \text{Tr} \left\{ \hat{\rho} \hat{A} \right\}. \quad (3)$$

The density operator $\hat{\rho}$ is

$$\hat{\rho} = \frac{1}{Z} \exp \left( -\beta \hat{H} \right), \quad (4)$$

where $\beta = 1/T$ is the inverse temperature and the symbol T denotes the time-ordered product. The quantum Monte-Carlo simulation is one of the standard tools to compute the Matsubara
Green function. The problem is how to reach the density of states from the observations in the quantum Monte-Carlo simulation.

Let us consider the Fourier transformation of the Matsubara Green function \( G(\tau) \). Then we impose the aperiodic condition for the fermionic case and the periodic one for the bosonic one, respectively.

\[
G(\tau) = \frac{1}{\beta} \sum_l \tilde{G}(i\omega_l) \exp(-i\omega_l\tau),
\]

where \( \omega_l \) is the Matsubara frequency \( i(2l+1)\pi/\beta \) for the fermion case and \( i2l\pi/\beta \). The density of states can be given by the imaginary components of the Fourier transformed Matsubara Green function as

\[
\rho(\omega) = -\frac{1}{\pi} \text{Im} \tilde{G}(\omega)
\]

The residue theorem can reduce the above equality to

\[
G(\tau) = \frac{1}{2\pi i} \oint_C dz \exp(-\tau z) \frac{\rho(z)}{1 \pm \exp(-\beta z)} \tilde{G}(z).
\]

Here we perform the analytical continuation for \( \tilde{G}(i\omega_l) \rightarrow \tilde{G}(z) \) and +/- denotes the fermionic/bosonic case. In addition, following the definition of the Matsubara Green function, we find the relationship with the density of states as, by use of the Cauchy principal value,

\[
\tilde{G}(z) = \int_{-\infty}^{\infty} d\omega \frac{\rho(\omega)}{z-\omega}.
\]

When we derived the above relationship between the (Fourier-transformed) Matsubara Green function and the density of states, we imposed the causality of the dynamics of the particles created at 0 until annihilation at \( \tau \). Therefore

\[
G(\tau) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega \int_{C} dz \frac{\exp(-\tau z)}{1 \pm \exp(-\beta z)} \frac{\rho(\omega)}{z-\omega}.
\]

The integrand has the singular values at \( z = \omega \) and \( \exp(-\beta z) = \mp 1 \), which leads to \( z = i\pi(2l+1)/\beta \) for the fermionic case and \( z = i2l\pi/\beta \) for the bosonic case, respectively. We take the singular values along the real axis and obtain

\[
G(\tau) = \int_{-\infty}^{\infty} d\omega \frac{\exp(-\tau \omega)}{1 + \exp(-\beta \omega)} \rho(\omega).
\]

for the fermionic case. On the other hand,

\[
G(\tau) = \int_{-\infty}^{\infty} d\omega \frac{\exp(-\tau \omega)}{1 - \exp(-\beta \omega)} \rho(\omega).
\]

for the bosonic case. As a consequence, we relate the density of states through the integral transformation to the Matsubara Green function.

3. Standard method: Maximum Entropy Method

The inverse problem to be solved in our problem is as follows:

\[
G(\tau) = \int_{-\infty}^{\infty} d\omega K_\pm(\tau,\omega) \rho(\omega),
\]
where $0 \leq \tau \leq \beta$ and $K(\omega)$ is the kernel, which is defined as

$$K_{\pm}(\tau, \omega) = \frac{\exp(-\tau \omega)}{1 \pm \exp(-\beta \omega)}. \quad (13)$$

Here, the $\pm$ stands for the fermionic case by $+$ and bosonic one by $-$. The analytical continuation problem is the inverse problem to obtain $\rho(\omega)$ by solving the above equation from the observations $G(\tau)$, which can be obtained by the quantum Monte-Carlo simulation. However the observation $G(\tau)$ suffers from the presence of the noise in the statistical error in the quantum Monte-Carlo simulation. In addition, the kernel decays exponentially in a region of the large absolute value of $\omega$. Thus the quantity on the right-hand side of Equation (12) can be insensitive to variations of $\rho(\omega)$. As a consequence, we have a plenty of number of plausible solutions satisfying Equation (12) within a given accuracy.

The maximum entropy method is a standard tool for solving the above ill-conditioned problem. In the following, let us deal with the discrete version of Equation (12) as in Equation (1), where $G = \{G(\tau_i)\}$, $\rho = \{\rho(\omega_i)\}$ and $(K)_{kl} = K(\tau_k, \omega_l)$. In addition, we define the squared error as

$$\chi^2(G, \rho) = \frac{1}{\sigma^2} \|G - K\rho\|_2^2 \quad (14)$$

where $\sigma^2$ is the variance of the noise in the observation, which is related to the statistical error in the quantum Monte-Carlo simulation. Hereafter we consider the case $\sigma^2 = 1$ for simplicity. The straightforward manipulation of the minimization of $\chi^2$ yields a consistent results satisfying Equation (1). The resultant solution is not expected to be a real answer because we have enormous candidates of the solution if we do not employ any regularization term. To find a reasonable solution, we employ the Bayesian inference. In the Bayesian inference, we define a certain probabilistic model and select a prior distribution to estimate the desired quantity. We here construct the following probabilistic model. The likelihood function for the density of states given the QMC data is defined by the Gaussian distribution as

$$P(G|\rho) \propto \exp \left( -\frac{1}{2} \chi^2(G, \rho) \right) \quad (15)$$

Second we set the prior distribution as

$$P(\rho) \propto \exp(-\alpha S(\rho)) \quad (16)$$

where $S(\rho)$ is the “entropy” defined as

$$S(\rho) = \sum_k \left( \rho(\omega_k) - m(\omega_k) - \rho(\omega_k) \log \left( \frac{\rho(\omega_k)}{m(\omega_k)} \right) \right) \quad (17)$$

Here $m(\omega)$ is called the “default” model, which represents some expected structure in $\rho(\omega)$ a priori. In addition, we often impose the following constraints on $\rho$ as

$$1 = \sum_k \rho(\omega_k) \quad \rho(\omega_k) \geq 0. \quad (18)$$

We define the set satisfying the above constraints as $P$. The parameter $\alpha$ controls the strength of the prior distribution in the posterior distribution. By using the Bayesian theorem, we reach the posterior distribution.

$$P(\rho|G) \propto P(G|\rho)P(\rho) \quad (19)$$
Therefore the maximum a posteriori estimation leads to the following optimization problem

$$\max_{\rho \in \mathcal{P}} \left\{ \alpha S - \frac{1}{2} \chi^2(\mathbf{G}, \rho) \right\}. \tag{20}$$

This is the MaxEnt or MEM [7, 8, 9].

The MEM is used in many fields rather than quantum many-body physics. For example, one of the applications of MEM is for deconvolution in radio astronomy [17].

The criticism of the MEM can be raised naturally. How can we determine the default model in the prior distribution? There is no definite answer.

4. Sparse modeling for analytical continuation

Instead of the MEM, we propose an alternative solver stemming from sparse modeling. We invent a new algorithm to solve the inverse problem (1) inspired by the compressed sensing (CS) in the framework of the sparse modeling. In the CS, we consider the $L_1$-norm minimization of the $N$-dimensional unknown vector $\mathbf{x}$ under an equality constraint $\mathbf{y} = A\mathbf{x}$, where $\mathbf{y}$ is the $M$-dimensional observed vector and $A$ is the sensing matrix. The penalty method to solve the constrained minimization problem leads to an well-known $L_1$ and $L_2$-norm minimization problems, LASSO, defined as

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \| \mathbf{y} - A\mathbf{x} \|^2_2 + \lambda \| \mathbf{x} \|^1 \right\}. \tag{21}$$

In the CS, the unknown vector is assumed to be sparse. The minimization of the $L_1$ norm selects a relatively sparse solution from a plenty of the solutions satisfying $\mathbf{y} = A\mathbf{x}$ when $M < N$.

The situation in the inverse problem (1) is very similar to the case of the CS. The exponentially decaying components in the kernel matrix decrease the essentially relevant elements in the observation $\mathbf{G}$. To avoid the irrelevant elements of the observation $\mathbf{G}$, we perform the singular value decomposition (SVD) on the kernel matrix as $\mathbf{K} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ is the unitary matrix and $\Lambda$ is the diagonal matrix containing the singular values. In addition, we define the transformed vectors as $\mathbf{\rho}' = \mathbf{V}^T\mathbf{\rho}$ and $\mathbf{G}' = \mathbf{U}^T\mathbf{G}$. Then the squared error is unchanged under the SVD as

$$\chi^2(\mathbf{G}, \mathbf{\rho}) = \| \mathbf{G}' - \Lambda \mathbf{\rho}' \|^2_2 \tag{22}$$

In this expression, $\Lambda$ plays a roll of some filters. The components of $\mathbf{\rho}$ multiplied by small values in $\Lambda$ do not affect significantly the observed vectors $\mathbf{G}'$. We then avoid several components of $\mathbf{\rho}$ automatically selected by using the $L_1$ norm. For this purpose, let us consider the following optimization problem

$$\min_{\rho \in \mathcal{P}} \left\{ \frac{1}{2} \chi^2(\mathbf{G}, \mathbf{\rho}) + \lambda \| \mathbf{\rho}' \|^1 \right\}. \tag{23}$$

By $L_1$ norm, decreasing the number of nonzero elements encourages the data fit in the significant contributions with the large values of $\Lambda$ and avoids irrelevant contributions with the small values of $\Lambda$. Unless the case of the CS, we do not necessarily the sparsity of the vector representing the density of states. In order to avoid the irrelevant contributions for reconstruction of $\mathbf{G}$, we employ the $L_1$ norm. Although the part of the squared error takes a form satisfying the separability, the norm can be written as the simple summation as $\sum_l (G'_l - \Lambda l \rho'_l)^2$, the constraints on the density of states makes the optimization problem nontrivial. To solve the optimization problem, the alternating direction of the multiplier method (ADMM) is available and efficient.

Furthermore, we can generalize the case of the noise intensity dependent on the observed frequency $\tau_l$ and correlated with each other as

$$\chi^2(\mathbf{G}, \mathbf{\rho}) = (\mathbf{G} - K\mathbf{\rho})^T \Sigma^{-1} (\mathbf{G} - K\mathbf{\rho}) \tag{24}$$
where $\Sigma^{-1}$ is the covariance matrix of the noise. We similarly perform the SVD and transformation of $G'$ and $\rho'$ as

$$
\chi^2(G, \rho) = (G' - \Lambda \rho')^T U^T \Sigma^{-1} U (G' - \Lambda \rho')
$$

(25)

The property of the kernel matrix is reflected on the covariance matrix as $U^T \Sigma^{-1} U$.

5. ADMM procedure

We provide the ADMM procedure for readers below. The ADMM is to efficiently solve the optimization problem with the non-differential cost function and several constraints. In our problem, we utilize the augmented Lagrangian method for the constraints $\rho'$ equals artificial variables $z$ and $z'$ for simply solving the original optimization problem and the standard Lagrangian multiplier method for the constraint $\sum_l \rho(\omega_l) = 1$, that is $1^T V \rho' = 1$ as

$$
\min_{\rho} \left\{ \frac{1}{2} \chi^2(G, \rho) + \lambda \| z \|_1 + \frac{\mu}{2} \| \rho' - z + u \|_2^2 - \nu (1^T V \rho' - 1) + \lim_{\gamma_i \to \infty} \gamma_i \Theta(z'_i) + \frac{\mu'}{2} \| V \rho' - z' + u' \|_2^2 \right\}.
$$

(26)

Taking derivative with respect to $\rho'$, we find the optimal solution on $\rho'$ for the fixed values $\mu$, $\nu$ and $z$.

$$
\rho = (\Lambda U^T \Sigma^{-1} U - \mu + \mu')^{-1} (\Lambda U^T \Sigma^{-1} U G' + \mu (z - u) + \mu V^T (z' - u') + \nu V^T 1)
$$

(27)

$$
\equiv \xi + \nu \eta.
$$

(28)

The Lagrange multiplier $\nu$ changes as satisfying the constraint $\sum_l \rho(\omega_l) = 1$. Thus

$$
\nu = \frac{1 - \sum_l \xi_l}{\sum_l \eta_l}.
$$

(29)

Two artificial variables are solved simply as

$$
z = S_{1/\mu} (\rho' + u)
$$

(30)

$$
z' = P_+ (V \rho' + u'),
$$

(31)

where $S_\lambda(x)$ is the Soft-Thresholding function defined as

$$
S_\lambda(x) = \begin{cases} 
  x - \lambda & (x > \lambda) \\
  0 & (-\lambda \leq x \leq \lambda) \\
  x + \lambda & (x < -\lambda)
\end{cases}
$$

(32)

and $P_+$ is a projection operator onto non-negative quadrant as $P_+(z) = \max \{ z, 0 \}$. Following the auxiliary Lagrangian method, the Lagrangian multipliers are updated as

$$
u = \nu + (\rho' - z)
$$

(33)

$$
u' = \nu' + (V \rho' - z').
$$

(34)

When $\Sigma^{-1}$ is diagonal, the computation of the inverse matrix can be simplified into the inverse of the diagonal elements.
6. Summary
In the present manuscript, we deal with the analytical continuation problem as an application of the sparse modeling. The analytical continuation includes a typical inverse problem and demands some regularization to be solved efficiently. The standard method was the MEM, which utilizes the prior distribution to infer the density of states via the MAP estimation. However it is difficult to choose an appropriate prior distribution in advance without any knowledge on the problem. Instead of the MEM, we employ the CS-like algorithm from the sparse modeling to elucidate the relevant contributions from observations. Because the sensing matrix has elements exponentially decaying, we perform the SVD in advance to filter out several irrelevant components. The components multiplied by small values of the singular values are irrelevant for reconstruction of the observations. We employ the $L_1$ norm to reduce the number of components and solve efficiently the inverse problem. As in the literature [1], the method can efficiently find reasonably the density of states without any prior knowledge. In addition, the proposed method elucidate the sparse representation for the QMC data [2, 18]. The sparse representation is useful for compression of the QMC data and finding the physical relevant feature from it. The resultant representation seems to consist of the highly nontrivial polynomial satisfying the integral equation stemming from the kernel. The future study aims at solving the nontrivial structure of the sparse representation of the kernel appearing in the analytical continuation in the quantum many-body physics. In addition, various techniques find inherent function from data as neural networks are available [19]. As other directions, in the stochastic analytical continuation, we may utilize the accelerated stochastic method [20, 21, 22, 23, 24]. In addition, the QMC data may be divided into several batch in order to efficiently optimize the cost function and avoid the overfit the noisy data. Then the Bayesian learning can be one of the useful techniques because it can predict the certainty of the estimation [25].

Acknowledgments
The present study was financially supported by MEXT KAKENHI No. 25120008, CREST, Japan Science and Technology Agency, JSPS KAKENHI No. 16H04382 and ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

References
[1] Otsuki J, Ohzeki M, Shinaoka H and Yoshimi K 2017 Phys. Rev. E 95(6) 061302
[2] Shinaoka H, Otsuki J, Ohzeki M and Yoshimi K 2017 Phys. Rev. B 96(3) 035147
[3] Abrikosov A, Gorkov L, Dzyaloshinski I and Daizyoshinski I 1975 Methods of Quantum Field Theory in Statistical Physics Dover Books on Physics Series (Dover Publications) ISBN 9780486632285
[4] Baym G and Kadanoff L P 1961 Phys. Rev. 124(2) 287-299
[5] Gubernatis J, Kawashima N and Werner P 2016 Quantum Monte Carlo Methods: Algorithms for Lattice Models (Cambridge University Press) ISBN 9781316483121
[6] Gull E, Millis A J, Lichtenstein A I, Rubtsov A N, Troyer M and Werner P 2011 Rev. Mod. Phys. 83(2) 349-404
[7] Jarrell M and Gubernatis J 1996 Physics Reports 269 133 – 195 ISSN 0370-1573
[8] Gunnarsson O, Haverkort M W and Sangiovanni V 2010 Phys. Rev. B 81(15) 155107
[9] Bergeron D and Tremblay A M S 2016 Phys. Rev. E 94(2) 023303
[10] Sandvik A W 1998 Phys. Rev. B 57(17) 10287–10290
[11] Fuchs S, Pruschke T and Jarrell M 2010 Phys. Rev. E 81(5) 056701
[12] Sandvik A W 2016 Phys. Rev. E 94(6) 063308
[13] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Science 220 671–680
[14] Yamanaka S, Ohzeki M and Decelle A 2015 Journal of the Physical Society of Japan 84 024801
[15] Ohzeki M 2015 Journal of the Physic Society of Japan 84 054801
[16] Ezaki T, Watanabe T, Ohzeki M and Masuda N 2017 Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 375
[17] Starck J L, Pantin E and Murtagh F 2002 Publications of the Astronomical Society of the Pacific 114 1051
[18] Shinaoka H, Otsuki J, Haule K, Wallerberger M, Gull E, Yoshimi K and Ohzeki M 2018 Phys. Rev. B 97(20) 205111 URL https://link.aps.org/doi/10.1103/PhysRevB.97.205111

[19] Arai S, Ohzeki M and Tanaka K 2018 Journal of the Physical Society of Japan 87 033001

[20] Ichiki A and Ohzeki M 2013 Phys. Rev. E 88(2) 020101

[21] Ichiki A and Ohzeki M 2015 Phys. Rev. E 91(6) 062105

[22] Ohzeki M and Ichiki A 2015 Phys. Rev. E 92(1) 012105

[23] Takahashi K and Ohzeki M 2016 Phys. Rev. E 93(1) 012129

[24] Ohzeki M and Ichiki A 2015 Journal of Physics: Conference Series 638 012003

[25] Ohzeki M 2016 Journal of Physics: Conference Series 699 012019