

14 Stochastic Optimization Method for Analytic Continuation

Andrey S. Mishchenko

RIKEN Advanced Science Institute (ASI)

Cross-Correlated Materials Research Group (CMRG),

Wako 351-0198, Japan

Contents

1	Difficulties to solve ill-posed problems	4
1.1	Discrete form of integral equation	4
1.2	Sawtooth noise instability	4
2	Methods to solve ill-posed problems	6
2.1	Tikhonov-Phillips regularization method	7
2.2	Maximum entropy method	8
2.3	Stochastic sampling methods	8
2.4	Stochastic optimization method: relation to other stochastic sampling approaches	9
3	Stochastic optimization method: general description	10
3.1	Deviation measure	10
3.2	Parametrization of particular spectra	11
3.3	General overview: obtaining particular solution and its sum	13
3.4	General features of elementary updates	14
3.5	Global updates	15
3.6	Final solution and refinement	16
3.7	Elementary updates of class I	17
3.8	Elementary updates of class II	18
4	Practical aspects of the method	19
4.1	Choosing the number of global updates F	20
4.2	Choosing the number of particular solutions L	21
5	Tests of SOM	22
5.1	Test of SOM for imaginary time representation	22
5.2	Test of SOM for Matsubara representation	24

Obtaining dynamical properties from quantum Monte Carlo (QMC) simulations is a notoriously difficult problem because QMC provides a limited number of values of a dynamical correlation function

$$\{G(m), m = 1, M\} \quad (1)$$

either at Matsubara frequencies $i\omega_m$ or at imaginary time points τ_m , whereas dynamical information is associated with a spectral function $A(\omega)$ depending on the continuous energy variable ω . The procedure of obtaining the dynamical correlation function $A(\omega)$ from the known set of values $G(m)$ is called *analytic continuation*. One of the most complete overview of this problem for the case when the set of values $\{G(m), m = 1, M\}$ is obtained from numeric calculations can be found in Ref. [1]. Generally, the procedure requires solving the Fredholm integral equation of the first kind [2]

$$G(m) = \int_{-\infty}^{\infty} d\omega \mathcal{K}(m, \omega) A(\omega), \quad m = 1, \dots, M, \quad (2)$$

where $\mathcal{K}(m, \omega)$ is some known kernel which depends on what quantities are associated with $G(m)$ and $A(\omega)$.

One of numerous examples is when one wants to determine the Lehmann spectral function [3]. This function contains a lot of important information on quasiparticles. For example, the Lehmann function is proportional to the spectral response observed in experiments on angle resolved photoemission spectroscopy (ARPES) [4].

A typical quantity calculated in QMC is $G(m) = G(\tau_m)$ which is called imaginary time Green function (GF)

$$G(\tau_m) = \langle T_\tau c(\tau_m) c^\dagger(0) \rangle. \quad (3)$$

Here T_τ is the time ordering operator and c is the annihilation operator of a quasiparticle. The imaginary time GF satisfies the periodicity (anti-periodicity) relation

$$G(\tau + \beta) = \pm G(\tau) \quad (4)$$

with a period equal to the inverse temperature $\beta = 1/T$. Here upper (lower) sign is for boson (fermion) operators. Hence, there is an equivalent representation given by the values of the Fourier transform $G(m) = \mathcal{G}(i\omega_m)$ of the imaginary time GF

$$\mathcal{G}(i\omega_m) = \int_0^\beta d\tau e^{i\omega_m \tau} G(\tau) \quad (5)$$

at Matsubara frequencies $i\omega_m$ equal to $(2m + 1)i\pi/\beta$ [$2im\pi/\beta$] for fermion [boson] operators [3]. The quantity $\mathcal{G}(i\omega_m)$ is the GF in the Matsubara representation. Indeed, there is the inverse Fourier transform from the Matsubara representation to the imaginary time GF

$$G(\tau) = \frac{1}{\beta} \sum_{\omega_m} e^{-i\omega_m \tau} \mathcal{G}(i\omega_m). \quad (6)$$

It can be shown [3] that in the case when QMC data for the GF are obtained in the Matsubara representation (5), $G(m) = \mathcal{G}(i\omega_m)$, the kernel $\mathcal{K}(m, \omega) \equiv \mathcal{K}(i\omega_m, \omega)$ of the Eq. (2) is

$$\mathcal{K}(i\omega_m, \omega) = \pm \frac{1}{i\omega_m - \omega}, \quad (7)$$

where plus (minus) sign corresponds to boson (fermion) operators. On the other hand, if the QMC data are given in terms of the imaginary time GF, (3), the kernel $\mathcal{K}(m, \omega) \equiv \mathcal{K}(\tau_m, \omega)$ of the analytic continuation is

$$\mathcal{K}(\tau_m, \omega) = -\frac{\exp(-\tau_m\omega)}{\exp(-\beta\omega) \pm 1}, \quad (8)$$

where the positive (negative) sign is for fermion (boson) operators.

Another example is when the quantity of interest is the optical conductivity $\sigma(\omega)$ and the quantity supplied by QMC is the imaginary time current-current correlation function $G(m) = J(\tau_m)$. The kernel $\mathcal{K}(m, \omega) \equiv \mathcal{K}(\tau_m, \omega)$ in this case is

$$\mathcal{K}(\tau_m, \omega) = \frac{1}{\pi} \frac{\omega \exp(-\tau_m\omega)}{1 - \exp(-\beta\omega)}. \quad (9)$$

Indeed, the problem of solving the Fredholm equation of the first kind is encountered in many areas which are far from the particular problem of analytic continuation. For example, one has to solve an equation of the same type to restore the thermodynamic properties of the quantum systems from QMC [5] or to recover the variety of impurity traps in organic materials from the ESR spectra [6, 7]. Moreover, a similar equation has to be solved for medical X-ray and impedance tomography, image deblurring, and many other practical applications [8]. Indeed, because of the notorious practical importance of the problem there is a long history of the vast amount of attempts to develop methods giving solutions for this class of equations. The main difficulty with the type of equations considered above is the following: they belong to the class of *ill-posed* problems. The main characteristic feature of this class is that there is no unique solution in the mathematical sense. Hence, to solve such an equation, one has to introduce some additional information specifying what kind of solution is expected. Therefore, it is impossible to single out the best method for solving this class of equations because each specific problem requires its own approach.

In the following we give a historical, although incomplete, overview of the approaches invented to solve the Fredholm equations of the first kind and follow the development of the methods up to recent times. In Sec. 1, we introduce the most simple minded approach, the least-squares fit, and show why it is not suitable for *ill-posed* problems. We describe various approaches to *ill-posed* problems in Sec. 2. In particular, we discuss there Tikhonov-Phillips regularization method, the maximum entropy method, and several variants of the stochastic sampling method. The stochastic optimization method (SOM), which is the main topic of this chapter, is compared with other stochastic sampling methods in Sec. 2.4. We give a detailed description of the SOM and some practical recipes in Secs. 3 and 4, respectively. Some tests of SOM are presented in Sec. 5.

1 Difficulties to solve ill-posed problems

To outline the difficulties encountered in the problem of solving the Fredholm equation of the first kind it is convenient to transform the equation into its discrete analog. The discrete analog is a matrix equation which seems to be easily solvable by least-squares fit. However, this naive approach immediately fails due to the *ill-posed* nature of the problem and the solution shows a *sawtooth noise* instability. We introduce the discrete analog of the Fredholm equation in Sec. 1.1. The least-squares fit approach, the nature of the *ill-posed* problems, and the features of the *sawtooth noise* instability are described in Sec. 1.2.

1.1 Discrete form of integral equation

Approximating the spectral function by its values on a finite spectral mesh of N points

$$A(\omega) = \sum_{n=1}^N A(\omega_n) \delta(\omega - \omega_n), \quad (10)$$

the integral equation (2) can be rewritten in matrix form

$$G(m) = \sum_{n=1}^N \mathcal{K}(m, \omega_n) A(\omega_n), \quad m = 1, \dots, M, \quad (11)$$

or equivalently presented as

$$\vec{G} = \hat{\mathcal{K}} \vec{A}. \quad (12)$$

Here $\vec{G} \equiv (G(1), G(2), \dots, G(M))$ [$\vec{A} \equiv (A(\omega_1), A(\omega_2), \dots, A(\omega_N))$] is an M -dimensional [N -dimensional] vector and $\hat{\mathcal{K}}(m, n) \equiv \mathcal{K}(m, \omega_n)$ is an $M \times N$ matrix ($M \geq N$). The matrix $\hat{\mathcal{K}}(m, n)$ is known, depending on the kernel of the integral equation, \vec{A} is the vector to be determined, and \vec{G} is obtained by QMC with components known with some error-bars.

1.2 Sawtooth noise instability

In practice, the problem expressed by Eq. (12) is usually *ill-posed*, either because of non-existence or non-uniqueness of a solution \vec{A} [9]: Noise, which is always present in a given vector \vec{G} , leads to the situation when there is no solution \vec{A} that exactly satisfies Eq. (12). On the other hand, there is an infinite number of solutions which make the left hand side of Eq. (12) approximately equal to its right hand side. Therefore, one can not search for a unique vector \vec{A} but has to find some solution which is the best in some sense, or find a set of solutions which are good according to some sensible criterion. The above features are the fingerprints of *ill-posed* problems.

The most simple minded approach in such a case is to search for the least-squares minimum-norm solution \vec{A} which minimizes the *deviation measure* which is chosen in the form of the Euclidean residual norm

$$\| \hat{\mathcal{K}} \vec{A} - \vec{G} \|^2 = \sum_{m=1}^M \left| \sum_{n=1}^N \mathcal{K}(m, \omega_n) A(\omega_n) - G(m) \right|^2. \quad (13)$$

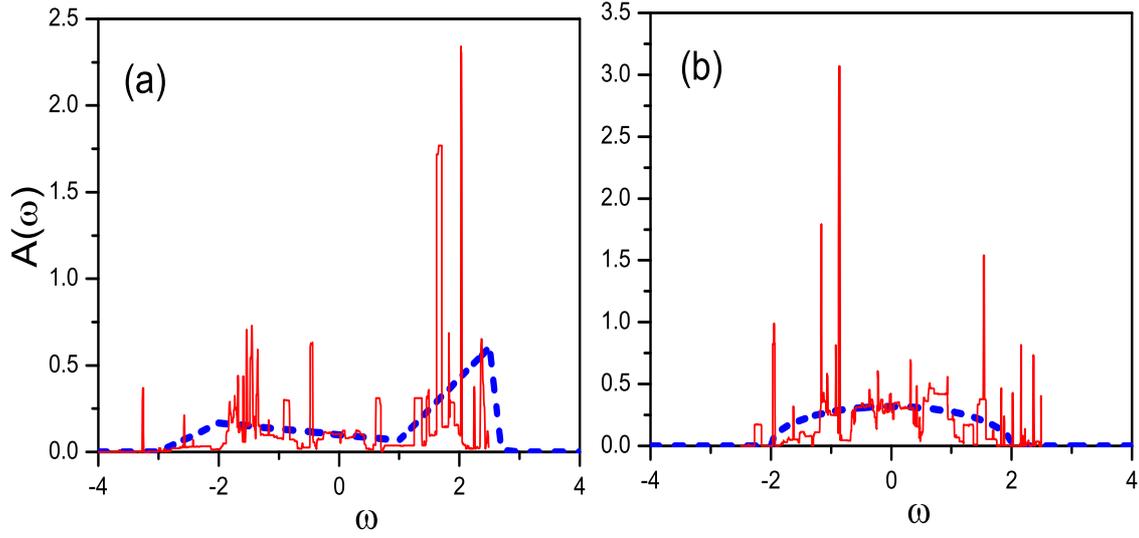


Fig. 1: Examples of sawtooth noise in a least-squares fit with restricted positive spectral function. The spikes in non-regularized solutions (red solid line in panels (a) and (b)) are much larger than the actual value of the spectra (blue dashed lines in (a) and (b)).

Indeed, one immediately surrenders to the *ill-posed* nature of the problem and tries to minimize, although the difference between the left- and right-hand sides of the Eq. (12) never reach zero. Choosing the Euclidean norm one admits the absence of a unique solution because the approximate equality in Eq. (12) can be defined in terms of an infinite number of another norms. For the Euclidean norm one can find the solution of the best least-squares fit in terms of the singular value decomposition of the matrix $\hat{\mathcal{K}}$ into a weighted sum of separable matrices [10]

$$\hat{\mathcal{K}} = \sum_{i=1}^r \sigma_i \vec{u}_i \otimes \vec{v}_i^\dagger, \quad (14)$$

where $\vec{u} \otimes \vec{v} \equiv \vec{u}(k)\vec{v}(l)$ is the matrix determined by the outer product of the left and right singular vectors of $\hat{\mathcal{K}}$, \vec{u} and \vec{v} , respectively. The real and nonnegative numbers $\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0$ are the singular values of $\hat{\mathcal{K}}$. The least-squares solution \vec{A} is given by the explicit expression

$$\vec{A} = \sum_{i=1}^r \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \vec{G}. \quad (15)$$

A problem arises from small singular values σ_i : Even very small statistical errors in \vec{G} induce large perturbations to the solution \vec{A} . These perturbations, called *sawtooth noise*, are typical for *ill-posed* problems and look like fast oscillations with amplitude much larger than the actual solution \vec{A} . The origin of the *sawtooth noise* is that the solution \vec{A} *over-fits* the statistical errors present in the input data \vec{G} . The *sawtooth noise* can even lead to large negative values of the otherwise positive actual solution \vec{A} . However, the least-squares fit under the condition of non-negativity of the spectral function \vec{A} results in sawtooth noise too (see Fig. 1).

2 Methods to solve ill-posed problems

It was shown in the previous section that the least-squares fit approach to the matrix form (12) of the integral equation (2) leads to results with *sawtooth noise* which can be very far from the actual solution. The noise arises from the small singular values σ_i of the integral kernel $\widehat{\mathcal{K}}$.

There are many methods developed to fight this noise. The simplest and most obvious one is the truncated singular value decomposition (TSVD), where the terms in Eq. (15), that correspond to several of the smallest singular values σ_i , are neglected. However, the above trick is just a particular example of a broad general class of approaches. Most, if not all, methods to circumvent the problem of spurious noise in the solution of the integral equation (2) can be united under the title *regularization methods*. As a particular example, the simplest regularization method for the matrix representation (12) of the integral equation (2) is based on the following trick. A regularization functional $\mathcal{F}(\vec{A})$, suppressing the oscillations of the solution \vec{A} , is added to the Euclidean norm (13)

$$\| \widehat{\mathcal{K}}\vec{A} - \vec{G} \|^2 + \gamma \mathcal{F}(\vec{A}) \quad (16)$$

and the *deviation measure* (16) is minimized instead. So, in general words, the regularized solution is sought as a minimizer of the deviation measure which is a weighted combination of the residual norm $\| \widehat{\mathcal{K}}\vec{A} - \vec{G} \|^2$ and a constraint $\gamma \mathcal{F}(\vec{A})$. Indeed, to construct the functional $\mathcal{F}(\vec{A})$ one needs some prior knowledge about solution the \vec{A} .

The functional (16) is historically the very first approach, named Tikhonov-Phillips regularization, that was developed to fight the sawtooth noise instability. However, to introduce a generic classification of the regularization approaches, it is convenient to use Bayesian statistical inference. According to the Bayes' theorem [11]

$$P[A|G] P[G] = P[G|A] P[A] , \quad (17)$$

where $P[A|G]$ is the posterior or conditional probability that the spectral function is A , provided the correlation function is G . Neglecting the normalization factor $P[G]$, which is independent of A , one gets

$$P[A|G] \sim P[G|A] P[A] , \quad (18)$$

where the ill-posed problem to find the most probable A given G is converted into the much easier problem of finding G given A , i.e., of maximizing the *likelihood function*, $P[G|A]$ taking into account simultaneously the *prior knowledge* about the spectrum $P[A]$. Note that any attempt to neglect the prior knowledge, i.e., to set $P[A] \equiv \text{const}$ and reduce the problem to the maximizing just the likelihood function, leads to the sawtooth noise instability of the solution. Notably, any method to regularize *ill-posed* problems can be presented in the form of a Bayesian approach and the distinctions between different approaches are restricted to the choice of the likelihood function $P[G|A]$ and the prior knowledge $P[A]$. Below, we introduce different possibilities of this choice. We describe the Tikhonov-Phillips regularization method in Sec 2.1, the maximum entropy method in Sec. 2.2, and several variants of the stochastic sampling method in Sec. 2.3. We finally consider the stochastic optimization method as an effective example of stochastic sampling methods in Sec. 2.4.

2.1 Tikhonov-Phillips regularization method

Historically, the approach called *Tikhonov-Phillips regularization method* (TPRM) has been invented independently in many different contexts and became the first approach to solve the above problems. The name comes from the first applications of the ideas to integral equations by A.N. Tikhonov [Tikhonoff] [12, 13] and D.L. Phillips [14] in the early 40ies of the last century. Independently, the regularization approach was applied in a different context to the discrete problem of matrix inversion [15–17] and is known in the statistical literature as ridge regression. However, leaving aside the differences in terminology and interpretations, the general idea is the following.

In the sense of Bayes' inference the TPRM is a choice where the likelihood function is

$$P[G|A] \sim \exp\{-\|\hat{\mathcal{K}}\vec{A} - \vec{G}\|^2\} \quad (19)$$

and the prior knowledge is

$$P[A] \sim \exp\{-\lambda^2 \|\hat{\Gamma}\vec{A}\|^2\}. \quad (20)$$

Thus, the *deviation measure* to minimize is the sum $\|\hat{\mathcal{K}}\vec{A} - \vec{G}\|^2 + \lambda^2 \|\hat{\Gamma}\vec{A}\|^2$. Here, the likelihood function requires the least-squares fit of \vec{G} while the constraint, where the solution \vec{A} is multiplied by a nonzero matrix $\hat{\Gamma}$, suppresses large absolute values of $A(\omega_k)$. Namely, the constraint removes spikes and, hence, large values of derivatives in the solution $[A(\omega_{k+1}) - A(\omega_k)]/[\omega_{k+1} - \omega_k]$.

The simplest modification of TPRM sets $\hat{\Gamma}$ as identity matrix $\hat{\Gamma} = \hat{I}$. In this case expression (15) for the solution \vec{A} takes the form

$$\vec{A} = \sum_{i=1}^r \left\{ \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \right\} \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \vec{G}. \quad (21)$$

It is clear that contributions, corresponding to small singular values $\sigma_i \ll \lambda$, are automatically filtered out by the factors in the curly brackets and large sawtooth spikes of the solution are suppressed. Thus *over-fitting* of the noise in the input data is avoided by restricting the possible solutions to the smooth ones. There are several approaches to find the optimal regularization parameter λ , *L-curve* [18, 19] and *U-curve* [20] methods in particular. These approaches consider relations between the Euclidean norm of the solution $\|\hat{\Gamma}\vec{A}\|^2$ and the residual $\|\hat{\mathcal{K}}\vec{A} - \vec{G}\|^2$.

An interesting modification of the TPRM is given in [21, 22]. The method expresses the solution \vec{A} in terms of an average over a correlation matrix $\langle \tilde{A}\tilde{A}^\dagger \rangle$ of possible solutions \tilde{A} . The knowledge of this correlation matrix provides a prior knowledge about the solution.

There are other methods which are based on the suppression of the large derivatives of the solution. These methods are based on the form of the functional (16), where the constraint $\gamma\mathcal{F}(\vec{A})$ is explicitly taken in a form which suppresses large derivatives of the solution [23, 24]. Many similar functionals can be found in earlier studies, see [9] for details on the rigorous mathematical treatment of *ill-posed* problems.

2.2 Maximum entropy method

One can criticize the first historical method to solve ill-posed problems as relying on unconditional smoothening of the solution. The constraint of the TPRM suppresses solutions with large derivatives. This can be a problem when the spectral function has sharp edges or narrow peaks. One of the recent approaches, the *Maximum Entropy Method* (MEM) [1], provides an attractive strategy to circumvent some problems of the TPRM.

MEM searches for the most probable “true” solution $A(\omega)$ among many possible particular solutions $\tilde{A}(\omega)$ assuming prior knowledge that the “true” solution $A(\omega)$ is close to a predefined function $D(\omega)$ called *default model*. The likelihood function of MEM is

$$P[G|\tilde{A}] = \exp\{-\chi^2[\tilde{A}]/2\}, \quad (22)$$

where

$$\chi^2[\tilde{A}] = \sum_{m=1}^M \mathcal{E}^{-1}(m)[G(m) - \tilde{G}(m)]^2, \quad (23)$$

and $\tilde{G}(m)$ is related to a particular solution $\tilde{A}(\omega)$ through $\tilde{G}(m) = \int_{-\infty}^{\infty} d\omega \mathcal{K}(m, \omega) \tilde{A}(\omega)$. The matrix $\mathcal{E}(m)$ is set by the noise in G and is related to the covariance matrix. The prior knowledge function is defined as

$$P[G|A] = \exp\{\alpha^{-1}S[\tilde{A}]\}, \quad (24)$$

where the entropy

$$S[\tilde{A}] = \int d\omega \tilde{A}(\omega) \ln[\tilde{A}(\omega)/D(\omega)] \quad (25)$$

characterizes the deviation of a particular solution $\tilde{A}(\omega)$ from the default model $D(\omega)$, a function that serves as the maximum entropy configuration. The regularization parameter α controls how much weight is given to the minimization of the *deviation measure* $\chi^2[\tilde{A}]$, i.e., to the resemblance of the solution $\tilde{A}(\omega)$ to the default model $D(\omega)$.

The MEM is superior to TPRM in cases where a lot of explicit information is known about $A(\omega)$. Moreover, one can avoid smoothening of large derivatives, typical for TPRM, given the knowledge about sharp parts of the “true” solution $A(\omega)$. The nonphysical smoothening can be avoided if sharp parts of the solution can be explicitly included into the default model. However, the method highly relies on the default model which can be a serious drawback if the most interesting features of the spectra are very sensitive to the form of the chosen default model [25].

2.3 Stochastic sampling methods

Any *stochastic sampling method* (SSM) uses a minimal prior knowledge about the solution, does not require any default model, and does not introduce any apparent smoothening of the solution. The characteristic feature of this class of methods is a change of the likelihood function

$P[\tilde{A}|G]$ into a likelihood functional (see, e.g., [26])

$$A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G], \quad (26)$$

where the “true” solution A is obtained as an average of particular solutions \tilde{A} , weighted by the likelihood function $P[\tilde{A}|G]$. An optimal likelihood function has to prefer solutions \tilde{A} with small *deviation measure* $\chi^2[\tilde{A}]$. Particular solutions \tilde{A} with too small $\chi^2[\tilde{A}]$ over-fit the data G and suffer from sawtooth noise. However, although it has not been proven formally, it is known that in practice the sawtooth noise can be self-averaging in a sum over a large enough number of solutions. One has to keep $\chi^2[\tilde{A}]$ not too restrictive because the sawtooth noise persists if most of the solutions in the functional (26) over-fit the input data. The requirement to take into account solutions \tilde{A} with large enough $\chi^2[\tilde{A}]$ sets up an *implicit regularization procedure*.

Starting from the very first practical design of a SSM by Sandvik [27], most of SSMs suggest the likelihood function in the form of a Boltzmann distribution

$$\mathcal{P}[A|G] = \exp\{-\chi^2[\tilde{A}]/\mathcal{T}\}, \quad (27)$$

where \mathcal{T} is treated as a fictitious temperature and χ^2 is defined by Eq. (23) and can be considered as a fictitious energy. Then, because of the above interpretation, one can use the Metropolis algorithm [28] to sample possible functions \tilde{A} . The prior knowledge function is usually defined by the condition that the spectral function $\tilde{A}(\omega)$ is positively definite and that the first few known frequency moments are conserved.

In principle, although introducing many useful details, the approaches suggested in Refs. [25, 29] belong to the same class as that introduced by Sandvik [27]. All three approaches do not use any default model for defining the prior knowledge function and, thus, are convenient in the problems where there is not much knowledge about how the resulting spectral function $A(\omega)$ has to look like.

A somewhat different approach is suggested in the statistical MEM (SMEM) by Beach [30] and Jarrel [31]. The method defines a dimensionless field $n(x)$ which is related to the default model $D(\omega)$. Then, averaging is performed over the dimensionless field using the likelihood Boltzmann distribution Eq. (27). The useful feature of the method is that, depending on its parameters, it can interpolate between two limiting cases when the spectrum is completely governed by the deviation measure (27) and when it is defined solely by the default model.

2.4 Stochastic optimization method: relation to other stochastic sampling approaches

The stochastic optimization method (SOM) [32], which is the main topic of this lecture, is a particular example of SSMs. SOM also does not use any default model, does not impose any apparent smoothening on the solution, and restricts prior knowledge to normalization and positivity of the solution.

A particular feature of SOM, which singles it out among SSMs, is that the sampling of solutions, which optimizes the deviation measure, is made without the artificial interpretation of the

likelihood function as a Boltzmann distribution [32,33]. A similar idea was also suggested later in the “generic inversion via falsification of theories” strategy [34]. Indeed, averaging over particular solutions (26), weighted by some likelihood function, has no relation to any real partition function. On the contrary, one simply has to average over a set of particular solutions, each of which fits the input data well enough. Therefore, the interpretation of $\chi^2[\tilde{A}]$ as an “energy” of some state in a system of a given temperature \mathcal{T} is a superfluous feature of the traditional SSMs. There is no real Hamiltonian and real temperature in the averaging procedure (26) and the net goal is to accumulate a large enough number of solutions which fit, but not over-fit, the data set $G(m)$.

Hence, it does not matter how the set of averaged “good enough” solutions is found. This is why the strategy to find particular solutions in SOM is completely different from other SSMs. On every step SOM starts from an arbitrary chosen initial particular solution and minimizes its deviation measure until a “good enough” fit is found. In this way SOM finds a large-enough number of “good” particular solutions and calculates an average

$$A(\omega) = \sum_{j=1}^L \xi_j \tilde{A}_j(\omega). \quad (28)$$

The simplest option is to set all coefficients equal to $\xi_j = 1/L$ for all L particular solutions whose deviation measure $\chi^2[\tilde{A}]$ is smaller than some selected value. A detailed description of the SOM, i.e., how to organize the process and how to choose its optimal parameters, is given in Secs. 3 and 4.

3 Stochastic optimization method: general description

In comparison to other SSMs, the SOM uses a slightly different measure $\chi^2[\tilde{A}]$, a considerably different way to parametrize a particular solution $\tilde{A}_j(\omega)$, and a completely different way to accumulate the particular spectral functions $\tilde{A}_j(\omega)$ for averaging. An important feature of SOM is that it treats the energy space continuously without imposing any finite ω -mesh. We describe the deviation measure and parametrization of the spectra in Secs. 3.1 and 3.2, respectively. Sections 3.3 and 3.4 discuss the way to obtain a particular solution and explain the general features of elementary updates which decrease the deviation measure of particular solutions. Global updates and the refinement of the solution are considered in Secs. 3.5, and 3.6, respectively. Finally, elementary updates of classes I and II are described in Secs. 3.7 and 3.8.

3.1 Deviation measure

The first step is to define the *deviation measure* determining which solution is a good approximation of the input data set G . The set G corresponds to some QMC data on imaginary times $G(m) = G(\tau_m)$ or at some Matsubara frequencies $G(m) = \mathcal{G}(i\omega_m)$, $m = 1, \dots, M$. Then the

deviation measure of SOM is given by expression

$$D[\tilde{A}] = \sum_{m=1}^M |\Delta(m)|. \quad (29)$$

Here $\Delta(m)$ is the *deviation function*

$$\Delta(m) = \frac{G(m) - \tilde{G}(m)}{\mathcal{S}(m)}, \quad (30)$$

which characterizes individual deviations of specific data points $G(m)$ from the values of the simulated function $\tilde{G}(m)$ defined by the particular spectral function \tilde{A} in terms of relation

$$\tilde{G}(m) = \int_{-\infty}^{\infty} d\omega \mathcal{K}(m, \omega) \tilde{A}(\omega). \quad (31)$$

The factors $\mathcal{S}(m)$ can be chosen as error-bars of the QMC data $G(m)$, if they are known.

However, there are plenty of sampling methods [32, 35–37], which can provide almost uniform, m -independent error-bars of the QMC data. These methods are usually used when $G(m)$ changes several orders of magnitude in the range $1 \leq m \leq M$. If the m -independent factor $\mathcal{S}(m) \equiv \mathcal{S}$ is put outside of the sum (29), the contribution of the data points with small $|G(m)|$ to the deviation measure $D[\tilde{A}]$ is evidently underestimated. In this case a reasonable choice for $\mathcal{S}(m)$ is to take $\mathcal{S}(m) = |G(m)|^d$, where $0 \leq d \leq 1$. Then, the contributions from the points with small and large values of $|G(m)|$ are equally represented in the sum when $d \rightarrow 1$.

3.2 Parametrization of particular spectra

We parameterize the spectral function \tilde{A} as a sum

$$\tilde{A}(\omega) = \sum_{t=1}^K \eta_{\{P_t\}}(\omega) \quad (32)$$

of rectangles $\{P_t\} = \{h_t, w_t, c_t\}$

$$\eta_{\{P_t\}}(\omega) = \begin{cases} h_t & , \quad \omega \in [c_t - w_t/2, c_t + w_t/2] , \\ 0 & , \quad \text{otherwise} , \end{cases} \quad (33)$$

determined by height $h_t > 0$, width $w_t > 0$, and center c_t .

A configuration

$$\mathcal{C} = \{\{P_t\}, t = 1, \dots, K\} \quad (34)$$

with the normalization constraint

$$\sum_{t=1}^K h_t w_t = I, \quad (35)$$

defines, according to Eqs. (31), (32), and (33), the function $\tilde{G}(m)$ at any point m . Figure 2 shows how the intersection of rectangles is understood in SOM.

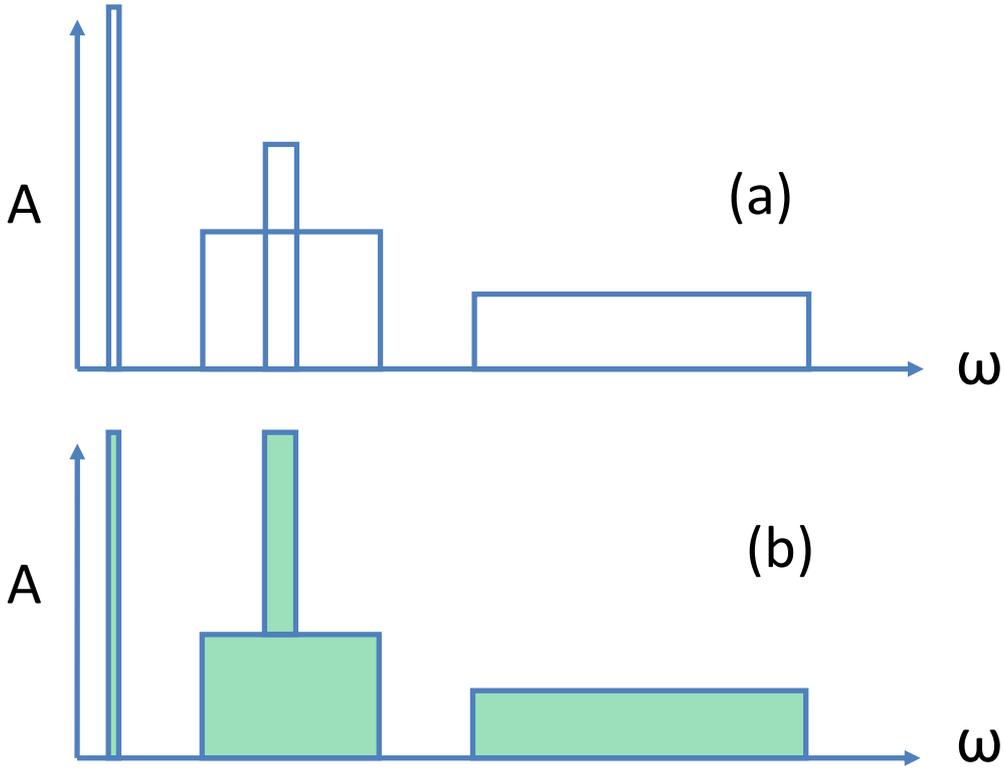


Fig. 2: An example of a configuration with $K = 4$. Panel (b) shows how the intersection of rectangles in panel (a) is treated.

Note that the specific type of the functions (33) is not crucial for the general features of the method although a simple form of the analytic expressions (31),(32), and (33) is of considerable importance for the performance of the method. If the analytic expression for $\tilde{G}(m)$ is not available for a given kernel \mathcal{K} , one tabulates the quantities

$$\Lambda(m, \Omega) = \int_{-\infty}^{\Omega} \mathcal{K}(m, x) dx, \quad m = 1, \dots, M \quad (36)$$

and finds the value of $\tilde{G}(m)$ using the following straightforward relation

$$\tilde{G}(m) = \sum_{t=1}^K h_t [\Lambda(m, c_t + w_t/2) - \Lambda(m, c_t - w_t/2)]. \quad (37)$$

In certain cases it is, however, straightforward to find analytic expressions. For example, let us consider the case when $G(m) = G(\tau_m)$ is a fermionic GF given by QMC at imaginary times (3) at zero temperature. In this case Eq. (8) reduces to $\mathcal{K}(\tau_\omega) = e^{-\tau_m \omega}$ and the spectral function is defined only at $\omega > 0$. It implies that in configuration (34) all $c_t - w_t/2 \geq 0$. Then, the explicit relation for the GF $\tilde{G}(\tau_m)$ in terms of the configuration \mathcal{C} is

$$\tilde{G}_{\mathcal{C}}(\tau_m) = \begin{cases} I & , \quad \tau_m = 0, \\ 2\tau_m^{-1} \sum_{t=1}^K h_t e^{-c_t \tau_m} \sinh(w_t \tau_m / 2) & , \quad \tau_m \neq 0. \end{cases} \quad (38)$$

Another example of an analytic expression is when one has a Matsubara GF (5) given via the kernel (7). In this case the analytic expression for $\tilde{\mathcal{G}}(i\omega_m)$ is

$$\tilde{\mathcal{G}}_C(i\omega_m) = \pm \sum_{t=1}^K h_t \ln \left[\frac{c_t - w_t/2 - i\omega_m}{c_t + w_t/2 - i\omega_m} \right], \quad (39)$$

where the plus (minus) sign is for boson (fermion) operators.

3.3 General overview: obtaining particular solution and its sum

Here we survey the whole procedure while the following sections add the necessary details.

First, $L \geq 10$ attempts to find particular solutions are performed. An attempt to obtain each particular solution \tilde{A}_j consists of two stages. The first is a random generation of initial configurations of rectangles and the second is a fixed number F of *global updates* decreasing the deviation measure D .

At first stage of the attempt j , some initial configuration $\mathcal{C}_j^{\text{init}}$, (34), is randomly generated. This means that a number of rectangles K is randomly chosen, with K in some range $1 < K < K_{\text{max}}$. The parameters $\{P_t\}$ of all rectangles are randomly generated under the constraint of normalization (35). Indeed one can impose further constraints, if some additional information is available.

Then F global updates are performed. The *global update* consists of a randomly chosen sequence of *elementary updates* which are described in next sections. A global update, which modifies the configuration $\mathcal{C}_j(r) \rightarrow \mathcal{C}_j(r+1)$, is accepted when $D[\tilde{A}_{r+1}] < D[\tilde{A}_r]$. For the particular solution \tilde{A}_j , obtained at each attempt after F global updates, one can control the quality of the fit of the input data using the deviation function $\Delta(m)$, (30). The number F of global updates is considered to be satisfactory, if the input data $G(m)$ are fit down to the noise level in a more than half of the L attempts. If not, the number of global updates F is increased and the procedure with $L \geq 10$ attempts is repeated.

Finally, when a satisfactory number of global updates F is found, an accumulation of $L \gg 10$ particular solutions \tilde{A}_j and their deviation measures $D[\tilde{A}_j]$ is performed. After L attempts there is a minimal deviation measure $\text{MIN}\{D[\tilde{A}_j]\}$, limited by the noise of the input data $G(m)$, and the rest of measures are larger. Tests show that to avoid over-fitting, i.e., to regularize the final solution, one has to include into the sum (28) all particular solutions whose deviation measures $D[\tilde{A}_j]$ are smaller than the double of the minimal deviation measure

$$D[\tilde{A}_j] \leq 2 \text{MIN}\{D[\tilde{A}_j]\}. \quad (40)$$

Such a choice of the regularization parameter is very similar to the strategy adopted in many other methods, e.g., SSM [29] or MEM [31]. Both in SOM and in many other methods the strategy is to keep differences between the fit $\tilde{G}(m)$ and data $G(m)$ of the order of the error-bars to avoid over-fitting. The inequality (40) is the way to introduce the regularization parameter in the most explicit manner.

3.4 General features of elementary updates

By elementary update we mean a random change of the configuration, which is either accepted or rejected in accordance with certain rules. There are two classes of elementary updates. The updates of class I do not alter the number of rectangles, K , changing only the values of the parameters from a randomly chosen set $\{P_t\}$. The updates of class II either add a new rectangle with randomly chosen parameters $\{h_{K+1}, w_{K+1}, c_{K+1}\}$, or remove a stochastically chosen rectangle t from the configuration. If a proposed change violates a constraint (38) (e.g., a change of w_t or h_t , or any update of the class II), then the necessary change of some other parameter set $\{P_{t'}\}$ is simultaneously proposed, to satisfy the requirement of the constraint.

The updates should keep the parameters of a new configuration within the domain of definition of the configuration \mathcal{C} . Formally, the domains of definition of a configuration (34) are $\Xi_{h_t} = [0, \infty]$, $\Xi_{c_t} = [-\infty, \infty]$, $\Xi_{w_t} = [0, \infty]$, and $\Xi_K \in [1, \infty]$. However, for the sake of faster convergence, one can reduce the domains of definition.

As there is no general *a priori* prescription for choosing reduced domains of definition, the rule of thumb is to start with maximal domains and then, after some rough solution is found, reduce the domains to reasonable values suggested by this solution. In particular, since the probability to propose a change of any parameter of a configuration is proportional to K^{-1} , it is natural to restrict the maximal number of rectangles $\Xi_K \in [1, K_{\max}]$ by some large number K_{\max} . To forbid rectangles with extremely small weight, which contribute to $\tilde{G}(\tau)$ less than the statistical errors of $G(\tau)$, one can impose the constraint $h_t w_t \in [S_{\min}, 1]$, with $S_{\min} \ll IK_{\max}^{-1}$. When there is some preliminary knowledge that an overwhelming majority of integral weight of the spectral function $\tilde{A}(\omega)$ is in a range $[\omega_{\min}, \omega_{\max}]$, one can restrict the domain of definition of the parameter c_t by $\Xi_{c_t} = [\omega_{\min}, \omega_{\max}]$. Then, to reduce the phase space one can choose $\Xi_{h_t} = [h_{\min}, \infty]$ and $\Xi_{w_t} = [w_{\min}, \min\{2(c_t - \omega_{\min}), 2(\omega_{\max} - c_t)\}]$.

While the initial configuration, the update type, and the parameter to be altered are chosen stochastically, the variation of the value of the parameters relevant to the update is optimized to maximize the decrease of D . Each elementary update of our optimization procedure (even that of the class II) is organized as a proposal to change some continuous parameter ξ by a randomly generated $\delta\xi$ in a way that the new value belongs to Ξ_ξ . Although proposals with smaller values of $\delta\xi$ are accepted with higher probability it is important, for the sake of better convergence, to propose sometimes changes $\delta\xi$ that probe the whole domain of definition Ξ_ξ . To probe all scales of $\delta\xi \in [\delta\xi_{\min}, \delta\xi_{\max}]$ we generate $\delta\xi$ with the probability density function $\mathcal{P} \sim (\max(|\delta\xi_{\min}|, |\delta\xi_{\max}|)/|\delta\xi|)^\gamma$, where $\gamma \gg 1$.

Calculating the deviation measures $D(\xi)$, $D(\xi + \delta\xi)$, $D(\xi + \delta\xi/2)$, and searching for the minimum of the parabolic interpolation, we find an optimal value of the parameter change

$$\delta\xi_{\text{opt}} = -b/2a, \quad (41)$$

where

$$a = 2(D(\xi + \delta\xi) - 2D(\xi + \delta\xi/2) + D(\xi))(\delta\xi)^{-2}, \quad (42)$$

and

$$b = (4D(\xi + \delta\xi/2) - D(\xi + \delta\xi) - 3D(\xi))\delta\xi. \quad (43)$$

In the case $a > 0$ and $\xi_{\text{opt}} \in \Xi_\xi$ we adopt as the update proposal $\tilde{\delta\xi}$ the value $\delta\xi$, $\delta\xi/2$, or $\delta\xi_{\text{opt}}$ for which the deviation measure $D(\xi + \tilde{\delta\xi})$ is the smallest. Otherwise, if the parabola minimum is outside Ξ_ξ , one has to compare only deviations for $\delta\xi$ and $\delta\xi/2$.

3.5 Global updates

The updating strategy has to provide for the efficient minimization of the deviation measure. It is highly inefficient to accept only those proposals that lead to the decrease of the deviation, since there is an enormous number of local minima with values $D_{\text{loc}}[\mathcal{C}]$ much larger than that obtained as minimal deviation measure $\text{MIN}\{D[\tilde{A}_j]\}$. As we observed in practice, these multiple minima drastically slow down (or even freeze) the process.

To optimize the escape from a local minimum, one has to provide a possibility of reaching a new local minimum with lower deviation measure through a sequence of less optimal configurations. It might seem that the most natural way of doing this would be to accept sometimes (with low enough probability) the updates leading to the increase of the deviation measure. However, this simple strategy turns out to be impractical. The reason is that the density of configurations per interval of deviation sharply increases with D . So that the acceptance probability for a deviation-increasing update should be fine-tuned to the value of D . Otherwise, the optimization process will be either non-convergent, or ineffective [if the acceptance probability is, correspondingly, either too large, or too small in some region of D].

A way out of the situation is to perform some sequence of T temporary elementary updates of a configuration $\mathcal{C}(0)$

$$\mathcal{C}(0) \rightarrow \mathcal{C}(1) \rightarrow \dots \rightarrow \mathcal{C}(r) \rightarrow \mathcal{C}(r+1) \rightarrow \dots \rightarrow \mathcal{C}(T), \quad (44)$$

where the proposal to update the configuration $\mathcal{C}(r) \rightarrow \mathcal{C}(r+1)$ is (temporary) accepted with probability

$$\mathcal{P}_{r \rightarrow r+1} = \begin{cases} 1 & , \quad D[\mathcal{C}(r+1)] < D[\mathcal{C}(r)], \\ Z(D[\mathcal{C}(r)]/D[\mathcal{C}(r+1)]) & , \quad D[\mathcal{C}(r+1)] > D[\mathcal{C}(r)]. \end{cases} \quad (45)$$

(Function Z satisfies the boundary conditions $Z(0) = 0$ and $Z(1) = 1$.) Then we choose out of the configurations $\{\mathcal{C}(r)\}$ (44) the one with minimal deviation measure and, if it is different from $\mathcal{C}(0)$, declare it to be the result of the global update, or, if this configuration turns out to be just $\mathcal{C}(0)$, reject the update.

We choose the function Z in the form

$$Z(x) = x^{1+d} \quad (d > 0), \quad (46)$$

which leads to comparatively high probabilities to accept small increases and hampers significant enlargements of the deviation measure. Empirically, we found out that the global update

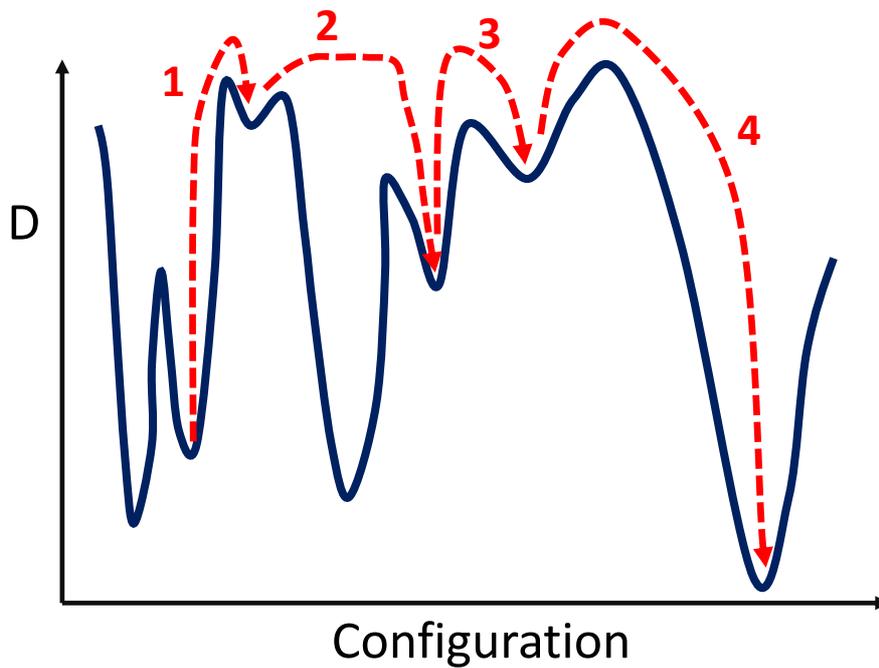


Fig. 3: Example of a global update with 4 elementary updates. The process transfers initial configuration by 4 elementary updates (dashed red arrows) from the initial minimum of the D -surface (solid blue line) to a lower minimum through the minima whose deviation measures are larger than that in the initial configuration.

procedure is most effective if one keeps the parameter $d = d_1 \ll 1$ at the first T_1 steps of sequence (44) (to leave local minimum) and then changes it to a value $d = d_2 \gg 1$ for the last $T - T_1$ elementary updates (to decrease the deviation measure). In our algorithm the values $T \in [1, T_{\max}]$, $T_1 \in [1, T]$, $d_1 \in [0, 1]$, and $d_2 \in [1, d_{\max}]$ were stochastically chosen for each global update run.

The two-step procedure for the global update is a method to reach the same goal as the tempering and annealing procedures used in SSM methods [27, 25, 29]. A temporary rise and consequent drop of the fictitious temperature \mathcal{T} is used in the standard SSMs. Similarly, temporary permission to grow up the deviation measure with the following directive to drop it down is introduced into SOM. An exchange between deep local minima of the deviation measures $D_{\text{loc}}[\mathcal{C}]$ has low probability. Therefore, the procedure which first rises and then drops the deviation measure arranges a path between some deep local minima through some shallow ones.

3.6 Final solution and refinement

After a set of L configurations

$$\{C_j^{\text{fin}}, j = 1, \dots, L\} \quad (47)$$

that satisfy criterion (40) is produced, the solution (28) can be obtained by summing up the rectangles, (33) and (47).

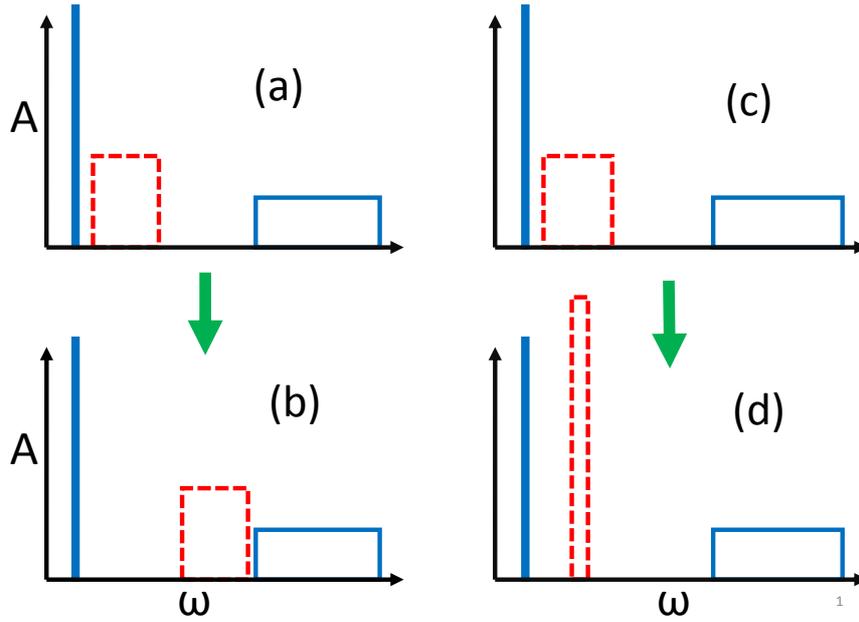


Fig. 4: Some of the elementary updates of class I: (a)→(b) is the shift of a rectangle (dashed red line); (c)→(d) is the change of the height of a rectangle (dashed red line) without changing its weight $h_t w_t$ and center c_t .

We employ, however, a more elaborated procedure, which we call refinement. Namely, we use the set (47) as a source of L_{ref} new independent starting configurations for further optimization. These starting configurations are generated as linear combinations of randomly chosen members of the set (47) with stochastic weight coefficients. Then, the refined final solution is represented as the average (28) of L_{ref} particular solutions resulting from the optimization procedure.

The main advantage of such a trick is that the initial configurations for the optimization procedure now satisfy the criterion (40) from the very beginning. Moreover, as any linear combination of a sufficiently large number R of randomly chosen parent configurations $\{C_{\eta}^{\text{fin}}, \eta = 1, \dots, R\}$ smoothes the sawtooth noise, the deviation of a refined configuration $C_{\text{ref}}^{\text{fin}}$ is normally lower than that of each additive one.

3.7 Elementary updates of class I

(A) *Shift of rectangle.* Change the center c_t of a randomly chosen rectangle t (Fig. 4a and 4b). The continuous parameter for optimization (41-43) is $\xi = c_t$ which is restricted to the domain of definition $\Xi_{c_t} = [\omega_{\min} + w_t/2, \omega_{\max} - w_t/2]$.

(B) *Change of width without change of weight.* Alter the width w_t of a randomly chosen rectangle t without changing of the rectangle weight $h_t w_t = \text{const}$ and center c_t (Fig. 4c and 4d). The continuous parameter for optimization is $\xi = w_t$ which is restricted by $\Xi_{w_t} = [w_{\min}, \min \{2(c_t - \omega_{\min}), 2(\omega_{\max} - c_t)\}]$.

(C) *Change of weight of two rectangles.* Change the heights of two rectangles t and t' (where t is a randomly chosen and t' is either randomly chosen or closest to t rectangle) without change of widths of both rectangles. The continuous parameter for optimization is the variation of

the height $\xi = h_t$ of rectangle t . To restrict the weights of the chosen rectangles to $[S_{\min}, 1]$ and preserve the total normalization (35) this update suggests to change $h_t \rightarrow h_t + \delta\xi$ and $h_{t'} \rightarrow h_{t'} - \delta\xi w_{t'}/w_t$ with $\delta\xi$ confined to the interval

$$S_{\min}/w_t - h_t < \delta\xi < (h_{t'} - S_{\min}/w_{t'})w_t/w_{t'} . \quad (48)$$

3.8 Elementary updates of class II

(D) *Adding a new rectangle.* To add a new rectangle one has to generate some new set $\{P_{\text{new}}\} = \{h_{\text{new}}, w_{\text{new}}, c_{\text{new}}\}$ and reduce the weight of some other rectangle t (either randomly chosen or closest) in order to keep the normalization condition (35). The reduction of the rectangle weight t is obtained by decreasing its height h_t .

The center of the new rectangle is selected at random according to

$$c_{\text{new}} = (\omega_{\min} + w_{\min}/2) + (\omega_{\max} - \omega_{\min} - w_{\min})r . \quad (49)$$

As soon as the value c_{new} is generated, the maximal possible width of a new rectangle is given by

$$w_{\text{new}}^{\max} = 2 \min(\omega_{\max} - c_{\text{new}}, c_{\text{new}} - \omega_{\min}) . \quad (50)$$

The continuous parameter for optimization $\delta\xi = h_{\text{new}}w_{\text{new}}$ is generated to keep the weights of both new rectangles and rectangle t larger than S_{\min}

$$\delta\xi = S_{\min} + r(h_t w_t - S_{\min}) . \quad (51)$$

Then, the value of the new rectangle height h_{new} for given $\delta\xi$ is generated to keep the width of new rectangles within the limits $[w_{\min}, w_{\text{new}}^{\max}]$

$$h_{\text{new}} = \delta\xi/w_{\text{new}}^{\max} + r(\delta\xi/w_{\min} - \delta\xi/w_{\text{new}}^{\max}) . \quad (52)$$

(E) *Removing a rectangle.* To remove some randomly chosen rectangle t , we enlarge the height $h_{t'}$ of some another (either randomly chosen or closest) rectangle t' according to the normalization condition (35). Since such a procedure does not involve a continuous parameter for optimization, we unite removing of rectangle t with the shift procedure (A) of the rectangle t' . Then, the proposal is the configuration with the smallest deviation measure.

(F) *Splitting a rectangle.* This update cuts some rectangle t into two rectangles with the same heights h_t and widths $w_{\text{new}_1} = w_{\min} + r(w_t - w_{\min})$ and $w_{\text{new}_2} = w_t - w_{\text{new}_1}$ (Fig. 5). Since removing a rectangle t and adding of two new glued rectangles does not change the spectral function we introduce the continuous parameter $\delta\xi$ which describes the shift of the center of the new rectangle with the smallest weight. The other rectangle is shifted in the opposite direction to keep the center of gravity of the two rectangles unaltered. The domain of definition Ξ_ξ obviously follows from the parameters of the new rectangles.

(G) *Glueing rectangles.* This update glues two (either randomly chosen or closest) rectangles t and t' into a single new rectangle with weight $h_{\text{new}}w_{\text{new}} = h_t w_t + w_{t'} h_{t'}$ and width $w_{\text{new}} = (w_t + w_{t'})/2$. The initial center of the new rectangle c_{new} corresponds to the center of gravity of rectangles t and t' . We introduce a continuous parameter by simultaneously shifting the new rectangle.

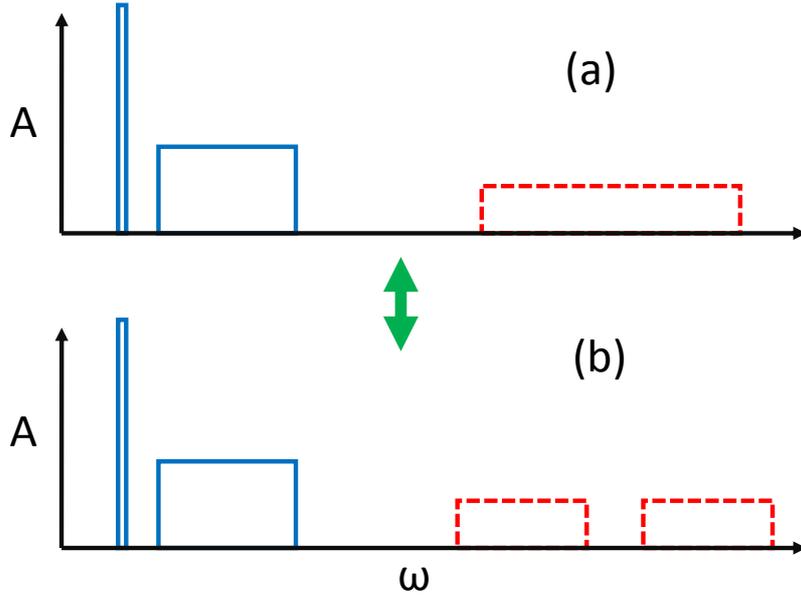


Fig. 5: Elementary updates of class II: (a)→(b) splitting of rectangle (dashed red line) and (b)→(a) gluing of rectangles without changing its total weight $h_t w_t$ and center of gravity c_t .

4 Practical aspects of the method

First, we summarize what is done by the SOM algorithm automatically and what are the numbers we need to determine in each particular case. The algorithm described in the previous section is able to search for as many particular solutions \tilde{A}_j as requested. For every attempt j to find a particular solution \tilde{A} it does the following steps:

- (i) Generate an initial configuration C_j^{init} with $K < K_{\max}$ rectangles. Each initial configuration C_j^{init} is statistically independent from the previous one C_{j-1}^{init} .
- (ii) Search for a particular solution \tilde{A}_j performing F global updates.
- (iii) Store the final configuration C_j^{fin} of solution \tilde{A}_j and its deviation measure $D[\tilde{A}_j]$.

After L attempts one obtains the final regularized solution as the sum

$$A(\omega) = \frac{1}{L_{\text{good}}} \sum_{j=1}^L \theta \left\{ 2\text{MIN}\{D[\tilde{A}_j]\} - D[\tilde{A}_j] \right\} \tilde{A}_j(\omega). \quad (53)$$

Here $\theta(x)$ is the θ -function: $\theta(x \geq 0)$ equals to unity and zero otherwise. L_{good} is the number of “good” fits

$$L_{\text{good}} = \sum_{j=1}^L \theta \left\{ 2\text{MIN}\{D[\tilde{A}_j]\} - D[\tilde{A}_j] \right\}, \quad (54)$$

restricted by the regularization condition that the deviation measure is less than twice of the minimal deviation measure $\text{MIN}\{D[\tilde{A}_j]\}$ found during L attempts.

To finalize the preparation of the method for solving different problems it is necessary to give recipes how to choose the numbers F (Sec. 4.1) and L (Sec. 4.2) in every particular case.

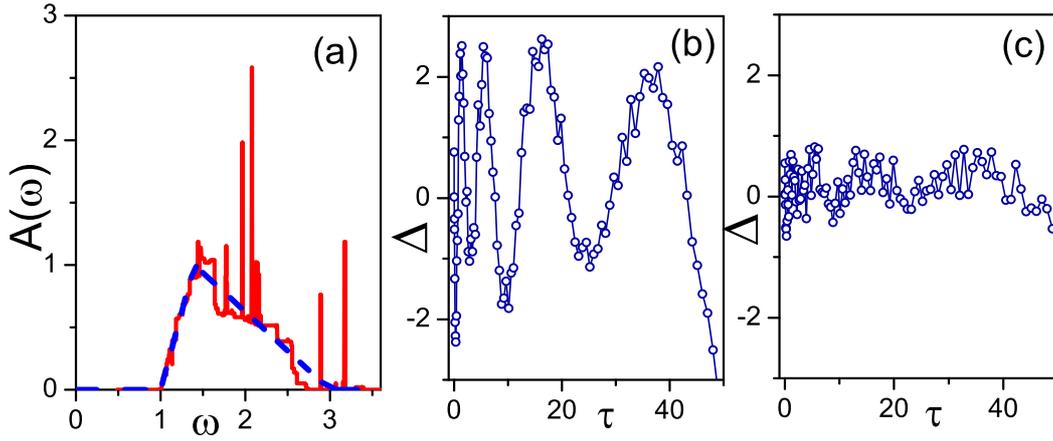


Fig. 6: (a) Typical spectrum $\tilde{A}_j(\omega)$ (red solid line), corresponding to a particular configuration C_j , compared to the actual spectrum (blue dashed line). Typical dependence of the deviation function $\Delta(m)$ (30) on imaginary times τ_m corresponding to a spectrum $\tilde{A}_j(\omega)$ which (b) under-fits and (c) over-fits the uncorrelated noise of imaginary time data.

4.1 Choosing the number of global updates F

To check whether a particular number F is large enough to reproduce the given data set (1) it is enough to perform about $L \approx 10$ attempts to find particular solutions \tilde{A}_j and consider the deviation functions $\Delta(m)$, (30), which correspond to each particular solution.

Note that a particular solution itself does not bear any important information on the quality of the fit of the data. Indeed, every particular solution contains sawtooth noise and typically looks like the solid red line in Fig. 6(a). One can claim that the comparison with the exact answer (dashed blue line) can give some insight into the quality, but in practice the exact answer is not known.

On the contrary, the deviation function $\Delta(m)$ gives direct information on the quality of the fit. Such a test of the quality of the fit requires uncorrelated noise in the QMC data, i.e., when the deviation from the exact solution $\delta G(m)$ for any m -point is independent from that in the neighboring m -point. Indeed, we take the uncorrelated nature of the noise for granted because the analytic continuation from *correlated* QMC data is a way to a wrong answer from the onset. For the sake of definiteness we consider here an example with imaginary time data. However, the case of the Matsubara representation is identical. Figure 6(b) shows an example when the input data are under-fitted by a particular solution. One can see that the typical period of oscillations of the deviation function $\Delta(m)$ around zero is much larger than the typical distance between the input data points τ_m . To the contrary, the fit shown in the Fig. 6(c) is noise-limited because the typical period of oscillations is comparable with a typical distance between data points. One can introduce a numeric criterion of the fit-quality

$$\kappa = \frac{1}{M-1} \sum_{m=2}^M \theta \{-\Delta(m)\Delta(m-1)\} \quad (55)$$

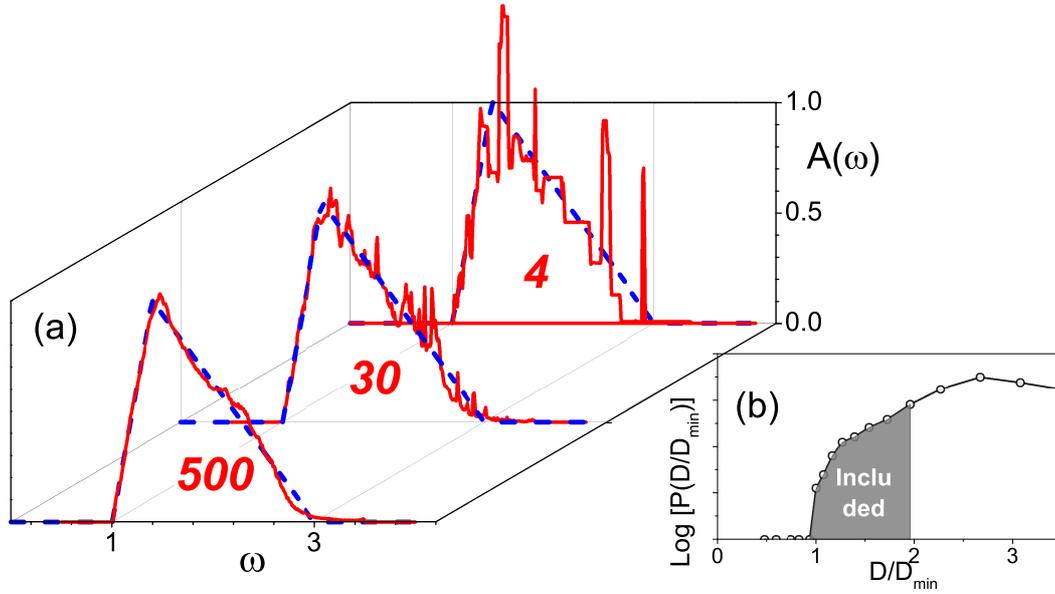


Fig. 7: (a) Self-averaging of the sawtooth noise after summation of 4, 30, and 500 solutions. (b) Typical probability distribution $P(D/D_{\min})$ of solutions with different deviation measures.

which is the ratio of number of intersections of zero by the function $\Delta(m)$ and number of the intervals between M input data points. Ideally, one would like to have $\kappa \rightarrow 1/2$ though it happens very rarely that $\kappa > 1/3$. Practically, a solution with fit-quality $\kappa > 1/4$ can be considered a good one.

Then, after $L \approx 10$ attempts to find particular solutions \tilde{A}_j , each by F global updates, the fit-qualities κ_j are considered. If $\kappa_j > 1/4$ for more than $L/2$ attempts, the number F is large enough. If not, it has to be increased.

4.2 Choosing the number of particular solutions L

SOM performs L attempts to find particular solutions \tilde{A}_j . The sum of particular solutions (28) becomes smoother as the number L of attempts increases. Figure 7(a) shows how the *sawtooth noise* self-averages when L increases. One can collect a distribution of the deviation measures and the typical picture is presented in Fig. 7(b). There is the minimal deviation measure D_{\min} which corresponds to the best fit of the noisy data $G(m)$ and there is a probability distribution which increases when D is slightly larger than the minimal deviation measure D_{\min} . Indeed, the distribution has some maximum and decreases for larger deviation measures D because there is the most probable deviation measure which is reached after F global updates. The shaded area in Fig. 7(b) shows which part of the distribution is included in the final solution (53) in order to regularize the sawtooth noise. We can not formulate any rigorous criterion when one can stop the accumulation of particular solutions. However, it looks reasonable to stop when there is no significant difference between the final spectra with L_{good} and $(1 - 1/3)L_{\text{good}}$ particular spectra included. We found that the above criterion is similar to that when one compares the sum (53) with $\theta \left\{ 2\text{MIN}\{D[\tilde{A}_j]\} - D[\tilde{A}_j] \right\}$ and with $\theta \left\{ 2(1 - 1/3)\text{MIN}\{D[\tilde{A}_j]\} - D[\tilde{A}_j] \right\}$.

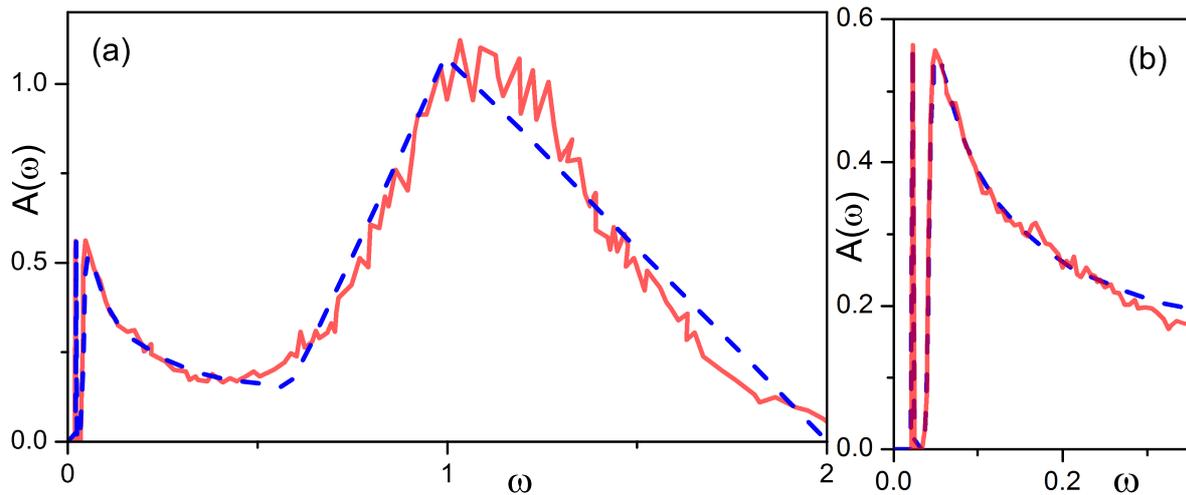


Fig. 8: The test spectrum (dashed blue line) and the spectrum obtained by SOM (solid red line). Panels (a) and (b) show the whole spectrum and its low energy part, respectively.

5 Tests of SOM

The procedure to check the SOM for different cases is the following. A spectral function $A(\omega)$, which is called *test spectrum*, is selected. Then, a set of input data with superimposed noise

$$\left\{ \tilde{G}(m) \left[1 + \frac{\mathcal{B}}{2} \mathcal{R} \right], m = 1, M \right\} \quad (56)$$

is generated. Finally, the SOM procedure is performed to restore the test spectrum.

The generation procedure of $\tilde{G}(m)$ uses a particular kernel \mathcal{K} , relation (31), and the test spectrum. Statistical noise is added with amplitude \mathcal{B} using a random number \mathcal{R} in the range

$$\mathcal{R} \in [-1, 1]. \quad (57)$$

We present tests for the imaginary time representation in Sec. 5.1. In particular, we test the case of a zero temperature GF in Sec. 5.1.1, finite temperature GF for fermions in Sec. 5.1.2, and finite temperature optical conductivity in Sec. 5.1.3. The test for GF in the Matsubara representation is presented in Sec. 5.2.

5.1 Test of SOM for imaginary time representation

5.1.1 Zero temperature Green function for a quasiparticle

For zero temperature the GF the kernel (8) for fermions reduces to $\mathcal{K}(\tau_\omega) = e^{-\tau_m \omega}$ and the spectral function $A(\omega)$ is defined only at $\omega > 0$. To check the accuracy of SOM, we tested it for a spectral density distribution that spreads over a large range of frequencies and simultaneously possesses fine structure in the low-frequency region [32]. The test spectrum was modeled as the sum of a delta-function with the energy $\varepsilon_\delta = 0.03$ and the weight $Z_\delta = 0.07$, and a continuous

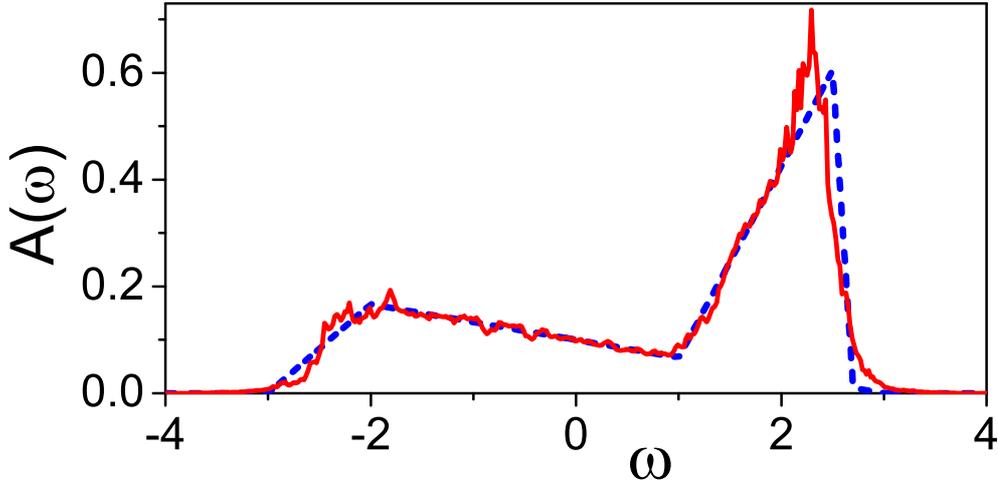


Fig. 9: The test spectrum (dashed blue line) and the spectrum obtained by SOM (solid red line) for the Lehmann spectral function of fermions at finite temperature.

high-frequency spectral density which starts at the threshold $\varepsilon_{\text{th}} = 0.04$. The continuous part of the spectral function A_{con} was modeled by the function

$$A_{\text{con}}(\omega) = \frac{Z_{\delta} \sqrt{\omega - \varepsilon_{\text{th}}}}{2\pi \sqrt{\varepsilon_{\text{gap}}} [(\omega - \varepsilon_{\text{th}}) + \varepsilon_{\text{gap}}]} \theta(\omega - \varepsilon_{\text{th}}) \theta(0.566 - \omega) \quad (58)$$

(here $\varepsilon_{\text{gap}} = \varepsilon_{\text{th}} - \varepsilon_{\delta}$ is a microgap) in the range $\omega \in [\varepsilon_{\text{th}}, 0.566]$ and by a triangle at higher frequencies (see the blue dashed line in the Fig. 8).

The GF $G(\tau)$ was calculated from the test spectrum in the $M = 300$ points $\tau_m = \tau_{\text{max}} m^2 / M^2$ in the time range from zero to $\tau_{\text{max}} = 1000$. The noise amplitude was chosen rather small $\mathcal{B} = 10^{-4}$. The restored spectral function reproduces both gross features of the high-frequency part (Fig. 8(a)) and the fine structure at small frequencies (Fig. 8(b)). The energy and the weight of the delta-function was restored with an accuracy of 10^{-4} . The final solution was obtained by averaging (53) $L_{\text{good}} = 1100$ particular solutions.

5.1.2 Finite temperature Green function for fermions

In this test the kernel is given by Eq. (8) for fermions and the spectral function $A(\omega)$ is defined in the whole range $-\infty < \omega < \infty$. The test spectrum was modeled by two triangles (blue dashed line in Fig. 9).

The GF $G(\tau)$ was calculated at finite temperature $\beta = 50$ on $M = 600$ uniformly spaced points in the range $[0, \beta]$. The noise amplitude was chosen rather small $\mathcal{B} = 10^{-4}$. The restored spectral function (red solid line in Fig. 9) reproduces the main features of the test spectrum. The final solution was obtained by averaging (53) $L_{\text{good}} = 150$ particular solutions.

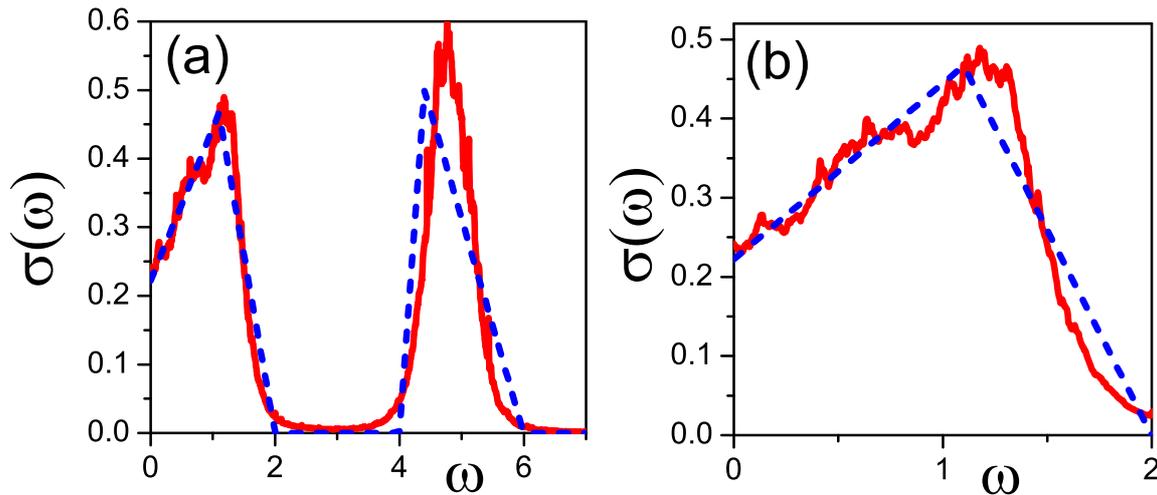


Fig. 10: The test spectrum (dashed blue line) and the spectrum obtained by SOM (solid red line) for optical conductivity at finite temperature. Panels (a) and (b) show the whole range and low energy part, respectively.

5.1.3 Finite temperature optical conductivity

In this test the kernel is given by Eq. (9) and the spectral function $\sigma(\omega)$ is symmetric $\sigma(\omega) = \sigma(-\omega)$. The test spectrum was modeled by two triangles (blue dashed line in Fig. 10). The current-current correlation function $J(\tau)$ was calculated at finite temperature $\beta = 20$ on $M = 200$ uniformly spaced points in the range $[0, \beta]$. The noise amplitude was chosen rather small $\mathcal{B} = 10^{-4}$. The restored optical conductivity (red solid line in Fig. 10) reproduces the main features of the test spectrum (Fig. 10(a)) and its low energy part (Fig. 10(b)). The final solution was obtained by averaging (53) $L_{\text{good}} = 200$ particular solutions.

5.2 Test of SOM for Matsubara representation

In this test the kernel is given by Eq. (7) and the spectral function $A(\omega)$ is defined in the whole ω -range. The test spectrum was modeled by two triangles (blue dashed line in Fig. 11(c)).

The GF $\mathcal{G}(i\omega_n)$ was calculated at finite temperature $\beta = 30$ for the first $M = 200$ positive Matsubara frequencies $i\omega_n$ and the analytic continuation was done directly from the set of GFs in Matsubara representation. The noise amplitude was $\mathcal{B} = 10^{-4}$. In Fig. 11 one can see rather good overall agreement between the test and the restored spectra. The final solution was obtained by averaging (53) $L_{\text{good}} = 200$ particular solutions.

The real and imaginary parts of the GF in Matsubara representation are shown in Fig. 11(a). the imaginary time GF, calculated from the test spectrum by Eq. (31), is shown in Fig. 11(b) by a solid black line. The imaginary time GF, calculated from the first 200 Matsubara components in the inverse Fourier transform (6), is shown in Fig. 11(b) by green circles. It is seen that the first 200 components of the GF in the Matsubara representation are not enough to describe the imaginary time GF at small values of τ . This discrepancy is a direct indication that the

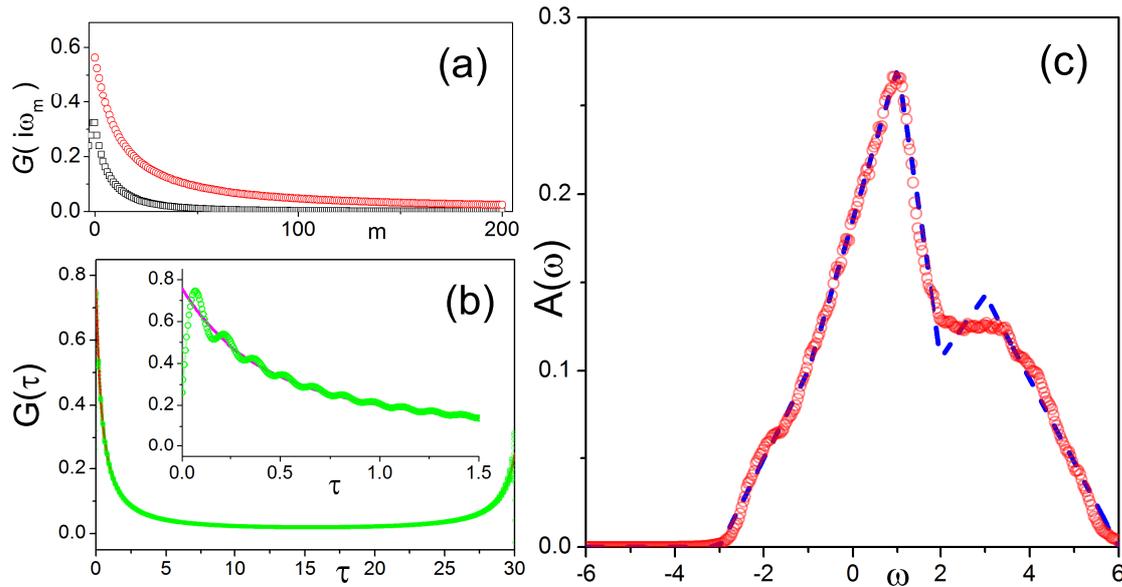


Fig. 11: (a) First 200 Fourier components of the real (red circles) and imaginary (black squares) part of the GF in Matsubara representation obtained from the GF in imaginary time. (b) Imaginary time GF (solid line) and imaginary time GF obtained from first the $M = 200$ GFs in Matsubara representation. The inset shows low imaginary times. (c) Actual spectrum (dashed blue line) and that restored from 200 Matsubara components (red solid line).

transformation of the QMC data from one representation into the other is a step which can lose information. Namely, it is dangerous to transform the Matsubara representation into the imaginary time representation because even a large number of Matsubara points still can lead to spurious oscillations of the imaginary time GF (Fig. 11(b)). Hence, it is preferable to make the analytic continuation from the same representation as that in which the QMC data are obtained.

Conclusion

We presented the stochastic optimization method for analytic continuation. The method was considered in relation with numerous other methods handling ill-posed problems. It was concluded that the method is the best for problems when one has to avoid any artificial smoothing of the spectral function and when there is no *a priori* knowledge about the expected solution. The method was successfully applied to many problems. The exponential kernel $\mathcal{K}(m, \omega) = \exp[-\tau_m \omega]$ for zero temperature was considered in Refs. [32, 38–57] and various kernels, ranging from Fermi distribution to the Matsubara frequency representation, are considered in Refs. [58–62]. The method was also used for Gaussian kernels in Refs. [6, 7]. Indeed, the broad area of the solved problems and the successful tests for problems which were not considered before give confidence that the method has considerable potential when application to problems where initial *a priori* knowledge is not available.

The author acknowledges support of RFBR 10-02-00047a and fruitful discussions with B.V. Svistunov and N.V. Prokof'ev.

References

- [1] M. Jarrell and J.E. Gubernatis, *Phys. Rep.* **269**, 133 (1996)
- [2] R. Kress, *Linear Integral Equations* (Springer, New York, 1999)
- [3] G.D. Mahan, *Many Particle Physics* (Plenum Press, New York, 1990)
- [4] A Damascelli, Z. Hussain, and Z.X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003)
- [5] C. Huscroft, R. Gass, and M. Jarrell, *Phys. Rev. B* **61**, 9300 (2000)
- [6] H. Matsui, A.S. Mishchenko, and T. Hasegawa, *Phys. Rev. Lett.* **104**, 056602 (2010)
- [7] A.S. Mishchenko, H. Matsui, and T. Hasegawa, *Phys. Rev. B* **85**, 085211 (2012)
- [8] J. Kaipio and S. Erkki, *Statistical and Computational Inverse Problems*, (Applied Mathematical Sciences Vol. 160) (Springer, Berlin, 2005)
- [9] A.N. Tikhonoff and V.Y. Arsenin, *Solutions of Ill-Posed Problems* (Winston & Sons, Washington, 1977)
- [10] G.H. Golub and C. Reinsch, *Numerische Mathematik* **14**, 403 (1970)
- [11] R.T. Cox, *The Algebra of Probable Inference* (Johns Hopkins University Press, 1961);
A. Papoulis, *Probability and Statistics* (Prentice Hall, New York, 1990)
- [12] A.N. Tikhonoff, *Doklady Akademii Nauk SSSR* **39**, 195 (1943)
- [13] A.N. Tikhonoff, *Doklady Akademii Nauk SSSR* **151**, 501 (1963)
(*Soviet Mathematics* **4**, 1035 (1963))
- [14] D.L. Phillips, *J. ACM* **9**, 84 (1962)
- [15] A.E. Hoerl, *Chemical Engineering Progress* **58**, 54 (1962)
- [16] A.E. Hoerl and R.W. Kennard, *Technometrics* **12**, 55 (1970)
- [17] M. Foster, *J. Soc. Industr. Appl. Math.* **9**, 387 (1961)
- [18] P.C. Hansen, *Soc. Industr. Appl. Math. Rev.* **34**, 561 (1992)
- [19] P.C. Hansen and D.P. O’Leary, *Soc. Industr. Appl. Math. J. Sci. Comput.* **14**, 487 (1993)
- [20] D. Krawchuk-Stando and M. Rudnicki, *Int. J. Appl. Math. Comput. Sci.* **17**, 157 (2007)
- [21] I.S. Krivenko and A.N. Rubtsov, arXiv:cond-mat/0612233
- [22] I.S. Krivenko and A.N. Rubtsov, *JETP Lett.* **94**, 768 (2012)

-
- [23] M. Jarrell and O. Biham, Phys. Rev. Lett. **63**, 2504 (1989)
- [24] S.R. White, D.J. Scalapino, R.L. Sugar, and N.E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989)
- [25] K. Vafayi and O. Gunnarsson, Phys. Rev B **76**, 035115 (2007)
- [26] J. Skilling, J. Microsc. **190**, 28 (1998)
- [27] A.W. Sandvik, Phys. Rev B **57**, 10287 (1998)
- [28] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953)
- [29] O.F. Syljuasen, Phys. Rev B **78**, 174429 (2008)
- [30] K.S.D. Beach, arXiv:cond-mat/0403055
- [31] S. Fuchs, T. Pruschke and M. Jarrell, Phys. Rev E **81**, 056701 (2010)
- [32] A.S. Mishchenko, N.V. Prokof'ev, A. Sakamoto and B.V. Svistunov, Phys. Rev. B **62**, 6317 (2000)
- [33] S.R. White, *Computer Simulation Studies of Condensed Matter Physics III* (Springer, Heidelberg, 1991), p. 145
- [34] E. Vitali, M. Rossi, L. Reatto and D.E. Galli, Phys. Rev. B **82**, 174510 (2010)
- [35] D.M. Ceperley, J. Comput. Phys. **51**, 404 (1983)
- [36] D.M. Ceperley and B.J. Alder, J. Chem. Phys. **81**, 5833 (1984)
- [37] N.V. Prokof'ev, B.V. Svistunov, and I.S. Tupitsyn, Phys. Rev. Lett. **82**, 5092 (1999)
- [38] A.S. Mishchenko, N.V. Prokof'ev, and B.V. Svistunov, Phys. Rev. B **64**, 033101 (2001)
- [39] A.S. Mishchenko, N.V. Prokof'ev, A. Sakamoto, and B.V. Svistunov, Int. J. Mod. Phys. B **15**, 3940 (2001)
- [40] A.S. Mishchenko, N. Nagaosa, N.V. Prokof'ev, A. Sakamoto, and B.V. Svistunov, Phys. Rev. B **66**, 020301(R) (2002)
- [41] A.S. Mishchenko, N. Nagaosa, N.V. Prokof'ev, A. Sakamoto, and B.V. Svistunov, Phys. Rev. Lett. **91**, 236401 (2003)
- [42] A.S. Mishchenko and N. Nagaosa, Phys. Rev. Lett. **93**, 036402 (2004)
- [43] A.S. Mishchenko and N. Nagaosa, Phys. Rev. B **73**, 092502 (2006)
- [44] A.S. Mishchenko and N. Nagaosa, J. Phys. Chem. Solids **67**, 259 (2006)

- [45] G. De Filippis, V. Cataudella, A.S. Mishchenko, C.A. Perroni, and J.T. Devreese, *Phys. Rev. Lett.* **96**, 136405 (2006)
- [46] A.S. Mishchenko, Proceedings of the International School of Physics “Enrico Fermi”, Course CLXI, 177-206 (2006)
- [47] A.S. Mishchenko and N. Nagaosa, *Polarons in Complex Matter*, Springer Series in Material Science, Springer, ed. by A.S. Alexandrov, 503-544 (2007)
- [48] V. Cataudella, G. De Filippis, A.S. Mishchenko, and N. Nagaosa, *Phys. Rev. Lett.* **99**, 226402 (2007)
- [49] A.S. Mishchenko, in “Computational Many-Particle Physics”, ed. by H. Fehske, R. Scheider and A. Weisse, *Lect. Notes Phys.* 739, pp. 367-395 (Springer, Berlin Heidelberg 2008)
- [50] A.S. Mishchenko, N. Nagaosa, Z.-X. Shen, G. De Filippis, V. Cataudella, T.P. Devereaux, C. Bernhard, K.W. Kim, and J. Zaanen, *Phys. Rev. Lett.* **100**, 166401 (2008)
- [51] V. Cataudella, G. De Filippis, A.S. Mishchenko, and N. Nagaosa, *J. Supercond. Nov. Magn.*, **22**, 17 (2009)
- [52] A.S. Mishchenko, N. Nagaosa, A. Alvermann, H. Fehske, G. De Filippis, V. Cataudella, and O.P. Sushkov, *Phys. Rev. B*, **79**, 180301(R) (2009)
- [53] A.S. Mishchenko, *Usp. Phys. Nauk* **179**, 1259 (2009) [*Phys. Usp.* **52**, 1193 (2009)]
- [54] A.S. Mishchenko, *Advances in Condensed Matter Physics* **2010**, 306106 (2010)
- [55] G.L. Goodvin, A.S. Mishchenko, and M. Berciu, *Phys. Rev. Lett.* **107**, 076403 (2011)
- [56] A.S. Mishchenko, N. Nagaosa, K.M. Shen, Z.-X. Shen, X.J. Zhou, T.P. Devereaux, *Europhys. Lett.* **95**, 57007 (2011)
- [57] G. De Filippis, V. Cataudella, A.S. Mishchenko and N. Nagaosa, *Phys. Rev. B*, **85**, 094302 (2012)
- [58] S.S. Aplesnin, *Zh. Eksp. Teor. Fiz.* **124** 1080 (2003) [*JETP* **97** 969 (2003)]
- [59] H. Hafermann, S. Brener, A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, *J. Phys.: Condens. Matter* **21**, 064248 (2009)
- [60] H. Hafermann, M.I. Katsnelson, and A.I. Lichtenstein, *Europhys. Lett.*, **85**, 37006 (2009)
- [61] E. Gorelov, M. Karolak, T.O. Wehling, F. Lechermann, A.I. Lichtenstein, and E. Pavarini, *Phys. Rev. Lett.* **104**, 226401 (2010)
- [62] E. Gorelov, J. Kolorenč, T. Wehling, H. Hafermann, A.B. Shick, A.N. Rubtsov, A. Landa, A.K. McMahan, V.I. Anisimov, M.I. Katsnelson, and A.I. Lichtenstein, *Phys. Rev. B* **82**, 085117 (2010)