Stochastic optimization method for analytic continuation

A. S. Mishchenko

RIKEN (Institute of Physical and Chemical Research), Japan
RRC “Kurchatov Institute”, Russia
Stochastic optimization method for analytic continuation

A. S. Mishchenko

RIKEN (Institute of Physical and Chemical Research), Japan
RRC “Kurchatov Institute”, Russia

1. Quantities one can get by QMC – correlation functions.
2. Examples of useful correlation functions.
3. Extracting physical information: necessity of analytic continuation.
4. General problem to solve: Fredholm integral equation of kind I.
5. Where similar problems are encountered?
6. Why the naïve methods fail?
8. More sophisticated methods: MaxEnt and Stochastic sampling.
9. Stochastic optimization method (SOM) as the utmost accomplishment of stochastic methods principles.
Diagrammatic Monte Carlo and new method of analytic continuation

A. S. Mishchenko

RIKEN (Institute of Physical and Chemical Research), Japan
Diagrammatic Monte Carlo and new method of analytic continuation

A. S. Mishchenko

RIKEN (Institute of Physical and Chemical Research), Japan
Examples of problems where one can get an important correlation function

Simple but not the simplest example: polaron

\[ \hat{H}_0 = \sum_k \epsilon(k) a_k^\dagger a_k + \sum_q \omega_q b_q^\dagger b_q \]

\[ \hat{H}_{\text{int}} = \sum_{k,q} V(k,q) (b_q^\dagger - b_{-q}) a_{k-q}^\dagger a_k + h.c. \]
Examples of problems where one can get an important correlation function

Simple but not the simplest example: polaron

\[
\hat{H}_0 = \sum_k \epsilon(k) a_k^\dagger a_k + \sum_q \omega_q b_q^\dagger b_q
\]

\[
\hat{H}_{\text{int}} = \sum_{k,q} V(k,q) (b_q^\dagger - b_{-q}) a_{k-q}^\dagger a_k + \text{h.c.}
\]

Green function by QMC

\[
G_k(\tau) = \langle \text{vac} | a_k(\tau) a_k^\dagger | \text{vac} \rangle
\]
Examples of problems where one can get an important correlation function

Simple but not the simplest example: polaron

\[ \hat{H}_0 = \sum_k \epsilon(k) a_k^\dagger a_k + \sum_q \omega_q b_q^\dagger b_q \]

\[ \hat{H}_{\text{int}} = \sum_{k,q} V(k, q)(b_q^\dagger - b_{-q}) a_{k-q}^\dagger a_k + h.c. \]

Green function by QMC

\[ G_k(\tau) = \langle \text{vac} | a_k(\tau) a_k^\dagger | \text{vac} \rangle \]

No simple connection to measurable properties
**Physical properties under interest: Lehman function**

Lehmann spectral function (LSF)

\[ L_k(\omega) = \sum_\nu \delta(\omega - E_\nu(k)) |\langle \nu | a_k^\dagger | \text{vac} \rangle|^2 \]

LSF has poles (sharp peaks) at the energies of stable (metastable) states. It is a measurable (in ARPES) quantity.

Noninteracting one is simple:

\[ L_{k, \text{NONINT}}^k(\omega) = \delta(\omega - \epsilon(k)) \]
Physical properties under interest: Lehmann function.

Lehmann spectral function (LSF)

\[ L_k(\omega) = \sum_\nu \delta(\omega - E_\nu(k)) |\langle \nu |a_k^\dagger|\text{vac}\rangle|^2 \]

LSF has poles (sharp peaks) at the energies of stable (metastable) states. It is a measurable (in ARPES) quantity.

LSF of one particle at T=0 can be determined from equation:

\[ G_k(\tau) = \int_0^\infty d\omega L_k(\omega) e^{-\omega\tau} \]

Fredholm first kind.
Physical properties under interest: Z-factor and energy

Lehmann spectral function (LSF)

\[
L_k(\omega) = \sum_\nu \delta(\omega - E_\nu(k)) |\langle \nu | a_k^\dagger | \text{vac} \rangle|^2
\]

\[
G_k(\tau) = \int_0^\infty d\omega \ L_k(\omega) \ e^{-\omega \tau}
\]

If the state with the lowest energy in the sector of given momentum is stable

\[
L_k(\omega) = Z^{(k)} \delta(\omega - E(k)) + \ldots
\]

The asymptotic behavior is

\[
G_k(\tau \gg \max \left[ \omega_{q,\kappa}^{-1} \right]) \rightarrow Z^{(k)} \exp[-E_{g.s.}(k)\tau]
\]
Physical properties under interest: Z-factor and energy

The asymptotic behavior is

$$G_k(\tau \gg \max \left[ \omega_{q,\kappa}^{-1} \right]) \rightarrow Z^{(k)} \exp[-E_{g.s.}(k)\tau]$$

Fig. 12.1. Typical behavior of the GF of a polaron and determination of $Z^{(k)}$-factor and energy of the ground state from the fit of the linear asymptotics.
Physical properties under interest: Lehmann function.

Lehmann spectral function (LSF)

\[ L_k(\omega) = \sum_\nu \delta(\omega - E_\nu(k)) \left| \langle \nu | a_k^\dagger | \text{vac} \rangle \right|^2 \]

LSF of one particle at T=0 can be determined from equation:

\[ G_k(\tau) = \int_0^\infty d\omega \, L_k(\omega) \, e^{-\omega \tau} \]

Solving of this equation is a notoriously difficult problem

\[ L_k(\omega) = \mathcal{F}_\omega^{-1} \left[ G_k(\tau) \right] \]
**Physical properties under interest: Lehmann function.**

Lehmann spectral function (LSF)

\[ L_k(\omega) = \sum_{\nu} \delta(\omega - E_\nu(k)) \left| \langle \nu | a_k^\dagger | \text{vac} \rangle \right|^2 \]

LSF of one particle at T=0 can be determined from equation:

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m,\omega) \, A(\omega) \]

Solving of this equation is a notoriously difficult problem

\[ L_k(\omega) = \mathcal{F}_\omega^{-1} [ G_k(\tau) ] \]
Solution of integral equation

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega) \]

\[ L_{\kappa}(\omega) = \hat{F}_{\omega}^{-1} \left[ G_{\kappa}(\tau) \right] \]
Exciton

\[ \hat{H}_0^{\text{par}} = \sum_k \varepsilon_a(k) a_k^\dagger a_k + \sum_k \varepsilon_h(k) h_k h_k^\dagger \]

\[ \hat{H}_{a-h} = -N^{-1} \sum_{p,k,k'} U(p,k,k') a_{p+k}^\dagger h_{p-k}^\dagger h_{p-k'} a_{p+k'} \]

Examples of problems where one can get an important correlation function

Coulomb attraction
Examples of problems where one can get an important correlation function

\[ \hat{H}_{\text{par}} = \sum_k \varepsilon_a(k) a_k^\dagger a_k + \sum_k \varepsilon_h(k) h_k^\dagger h_k \]

\[ \hat{H}_{\text{a-h}} = -N^{-1} \sum_{pkk'} U(p, k, k') a_{p-k}^\dagger h_{p+k}^\dagger h_{p-k'} a_{p+k'} \]

\[ \hat{H}_{\text{par-bos}} = i \sum_{\kappa=1}^Q \sum_{k, q} (b_{q, \kappa}^\dagger - b_{-q, \kappa}) \left[ \gamma_{aa, \kappa}(k, q) a_{k-q}^\dagger a_k + \gamma_{hh, \kappa}(k, q) h_{k-q}^\dagger h_k + \gamma_{ah, \kappa}(k, q) h_{k-q}^\dagger a_k \right] + \text{h.c.} \]

\[ \hat{H}_{\text{bos}} = \sum_{\kappa=1}^Q \sum_q \omega_{q, \kappa} b_{q, \kappa}^\dagger b_{q, \kappa} \]
Examples of problems where one can get an important correlation function

\[ \hat{H}_0^{\text{par}} = \sum_{k} \epsilon_a(k) a_k^\dagger a_k + \sum_{k} \epsilon_h(k) h_k h_k^\dagger \]

\[ \hat{H}_{\text{a-h}} = -N^{-1} \sum_{pkk'} U(p, k, k') a_{p+k}^\dagger h_{p-k}^\dagger h_{p-k'} a_{p+k'} \]

\[ + \hat{H}_{\text{el-ph}} \]

Coulomb attraction

Phonon exchange
Examples of problems where one can get an important correlation function

Exciton-polaron: two-particle Green function

\[ G_{k}^{pp'}(\tau) = \langle \text{vac} | a_{k+p'}(\tau) h_{k-p'}(\tau) h_{k-p}^{\dagger} a_{k+p}^{\dagger} | \text{vac} \rangle \]

\[ I(\omega) = \tilde{F}_{\omega}^{-1} \left[ \sum_{pp'} G_{k=0}^{pp'}(\tau) \right] \]

Optical absorption

Coulomb attraction

Phonon exchange
Examples of problems where one can get an important correlation function

\[ \mathcal{I}(\omega) = \mathcal{F}^{-1}_\omega \left[ \sum_{pp'} G_{k=0}^{pp'}(\tau) \right] \]

Optical absorption

Also Fredholm integral equation of the first kind

Coulomb attraction

Phonon exchange
Exact solution for optical spectra of exciton-polaron
A. S. Mishchenko and N. Nagaosa, CMRG, RIKEN ASI

Diagrammatic Monte Carlo

FIG. 1. A typical diagram for $\Gamma^{\text{MP}0}(\tau)$. Solid (dashed) lines represent $G(V)$ propagators, solid circles (squares) designate Coulomb (QP-phonon) interactions, and dotted lines are the phonon propagators. Imaginary time runs from left to right.

Coulomb attraction
No particle-phonon coupling

Exciton-polaron

$H = \sum_k e_k \epsilon^+_k \epsilon_k + \sum_k \epsilon_\sigma(k) h_k^\dagger h_k^\sigma + \sum_q \omega_q b_q^\dagger b_q$

$- \sum_{kq} \frac{\delta_n(q)}{\sqrt{N}} e_k^\dagger e_k \epsilon_k + \frac{\delta_n(q)}{\sqrt{N}} h_k^\dagger h_k^\sigma [b_q^\dagger + b_q]

- \sum_{pkk'} \frac{U(p,k,k')}{N} e_k^\dagger h_{p-k}^\dagger h_{p-k}^\sigma e_k^\dagger e_k$
There are a lot of problems where one has to solve Fredholm integral equation of the first kind.

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m,\omega) \, A(\omega) \]
Many-particle Fermi/Boson system in imaginary times representation

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ K(m, \omega) \ A(\omega) \]

\[ K(\tau_m, \omega) = -\frac{\exp(-\tau_m \omega)}{\exp(-\beta \omega) \pm 1} \]
Many-particle Fermi/Boson system in Matsubara representation

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ K(m, \omega) \ A(\omega) \]

\[ G(\omega_m) = \int_{0}^{\beta} d\tau \ e^{i\omega_m \tau} \ G(\tau) \]

\[ G(\tau) = \frac{1}{\beta} \sum_{\omega_m} e^{-i\omega_m \tau} \ G(i\omega_m) \]

\[ K(i\omega_m, \omega) = \pm \frac{1}{i\omega_m - \omega} \]
Optical conductivity at finite $T$ in imaginary times representation

\[ G(m) = \int_{-\infty}^{\infty} d\omega \kappa(m, \omega) A(\omega) \]

\[ G(i\omega_m) = \int_0^\beta d\tau e^{i\omega_m \tau} G(\tau) \]

\[ G(\tau) = \frac{1}{\beta} \sum_{\omega_m} e^{-i\omega_m \tau} G(i\omega_m) \]

\[ \kappa(\tau_m, \omega) = \frac{1}{\pi} \frac{\omega \exp(-\tau_m \omega)}{1 - \exp(-\beta \omega)} \]
Image deblurring with e.g. known 2D noise $K(m, \omega)$

$G(m) = \int_{-\infty}^{\infty} d\omega \ K(m, \omega) \ A(\omega)$

$m$ and $\omega$ are 2D vectors

$K(m, \omega)$ is a 2D x 2D noise distribution function

Original | one $\lambda$
---|---
Blurred & noisy | three $\lambda$'s
Tomography image reconstruction (CT scan)

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ K(m,\omega) \ A(\omega) \]

\( m \) and \( \omega \) are 2D vectors

\( K(m,\omega) \) is a 2D x 2D distribution function
What is dramatic in the problem?

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega) \]
ILL-POSED!

What is dramatic in the problem?

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ K(m,\omega) \ A(\omega) \]

Aircraft stability
Nuclear reactor operation
Image deblurring
A lot of other…
We cannot obtain an exact solution not because of some approximations of our approaches.

Instead, we have to admit that the exact solution does not exist at all!
ILL-POSED!

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m,\omega) \, A(\omega) \]

1. No unique solution in mathematical sense
ILL-POSED!

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m,\omega) \, A(\omega) \]

1. No unique solution in mathematical sense

2. Some additional information is required which specifies which kind of solution is expected
1. No unique solution in mathematical sense
   No function A to satisfy the equation

2. Some additional information is required which specifies which kind of solution is expected. In order to choose among many approximate solutions.
ILL-POSED!

Physics department: Max Ent.

$G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega)$

How to solve?
ILL-POSED!

Physics department: Max Ent.

Engineering department: Tikhonov Regularization

How to solve?

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m, \omega) \, A(\omega) \]
ILL-POSED!

How to solve?

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m, \omega) \, A(\omega) \]

Physics department: Max Ent.

Statistical department: ridge regression

Engineering department: Tikhonov Regularization
ILL-POSED!

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m,\omega) \, A(\omega) \]

Next player: stochastic methods

Physics department: Max Ent.

Engineering department: Tikhonov Regularization

Statistical department: ridge regression
Not settled!

- Still highly competitive field
- Many approaches developed, some specific ones are better for some specific cases
- Different approaches speak different languages, need some unified analysis
- Comparison of different methods, not just self-advertising, is needed
ILL-POSED!

Historically first: 1943: Tikhonov Regularization

Physics department: Max Ent. (Mark Jarrel)

Next player: stochastic Methods Since 1998

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega) \]
The easiest way to explain the problem is to turn to a discrete form of the Fredholm equation

\[
G(m) = \int_{-\infty}^{\infty} d\omega \, \kappa(m,\omega) \, A(\omega)
\]

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

\[
A(\omega) = \sum_{n=1}^{N} A(\omega_n) \delta(\omega - \omega_n)
\]

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

\[
G(m) = \sum_{n=1}^{N} \kappa(m,\omega_n) A(\omega_n), \quad m = 1, \ldots, M
\]

or equivalently presented as

\[
\tilde{G} = \tilde{K} \tilde{A}
\]
The easiest way to explain the problem is to turn to a discrete form of the Fredholm equation:

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

Because of noise present in the input data $G(m)$ there is no unique $A(\omega_n) = A(n)$ which exactly satisfies the equation.
Because of noise present in the input data $G(m)$ there is no unique $A(\omega_n) = A(n)$ which exactly satisfies the equation.

Hence, one can search for the least-square fitted solution $A(n)$ which minimizes:

$$
\| \tilde{K}A - \tilde{G} \|^2 = \sum_{m=1}^{M} \left| \sum_{n=1}^{N} \kappa(m, \omega_n) A(\omega_n) - G(m) \right|^2
$$
Choosing the Euclidean norm one admits the absence of unique solution because there is an infinite number of deviation norms.
Unique solution for the least-square fit through singular values decomposition of the matrix $K$
Unique solution for the least-square fit through singular values decomposition of the matrix $K$

Explicit expression:

$$\hat{A} = \sum_{i=1}^{r} \frac{\hat{u}_i^\dagger \otimes \hat{v}_i}{\sigma_i} \hat{G}$$

Typical singular values:

$$\sigma(i)$$
Saw tooth noise instability due to small singular values.

Explicit expression:
Saw tooth noise instability due to small singular values.

Truncating small singular values (from 1 to 11)
Tikhonov regularization to fight with the saw tooth noise instability.
Tikhonov regularization to fight with the saw tooth noise instability.
Tikhonov functional to minimize (\( \Gamma \) is unitary matrix):
General formulation of methods to solve ill-posed problems in terms of Bayesian statistical inference.

\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega) \]
Bayes theorem:


\[ G(m) = \int_{-\infty}^{\infty} d\omega \ \kappa(m,\omega) \ A(\omega) \]

\( P[A|G] \) – conditional probability that the spectral function is A provided the correlation function is G
Bayes theorem:


- conditional probability that the spectral function is A provided the correlation function is G

To find it is just the analytic continuation
\[ G(m) = \int_{-\infty}^{\infty} d\omega \, K(m, \omega) \, A(\omega) \]


\( P[G|A] \) is *easier* problem of finding \( G \) given \( A \): likelihood function

\( P[A] \) is prior knowledge about \( A \):

**Analytic continuation**

P[G|A] is easier problem of finding G given A: likelihood function

P[A] is prior knowledge about A:

All methods to solve the above problem can be formulated in terms of this relation
Historically first method to solve the problem of Fredholm kind I integral equation.

Tikhonov regularization method (1943)

A.N. Tikhonov, Dokladyu Akademii Nauk SSSR (1943)

A.N. Tikhonov, Dokladyu Akademii Nauk SSSR (1963) (Soviet mathematics)

Historically first method to solve the problem of Fredholm kind I integral equation.

A.N. Tikhonov, Dokladyu Akademii Nauk SSSR (1943)

The regularization method was developed not by …… in 1977, it was rediscovered…..
Historically first method to solve the problem of Fredholm kind I integral equation.

**Tikhonov regularization method (1943)**

\[
G(m) = \int_{-\infty}^{\infty} d\omega \, \mathcal{K}(m, \omega) \, A(\omega)
\]

\[
P[G|A] \sim \exp\{-\| \hat{\mathcal{K}}A - \bar{G} \|^2\}
\]

\[
P[A] \sim \exp\{-\lambda^2 \| \hat{\Gamma}A \|^2\}
\]

\[
\| \hat{\mathcal{K}}A - \bar{G} \|^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} \mathcal{K}(m, \omega_n)A(\omega_n) - G(m)\]

Tikhonov regularization method (1943)

If $\Gamma$ is unit matrix:

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

$$P[G|A] \sim \exp\{-\| \hat{\kappa} \vec{A} - \vec{G} \|^{2}\}$$

$$P[A] \sim \exp\{-\lambda^2 \| \hat{\Gamma} \vec{A} \|^{2}\}$$

$$\| \hat{\kappa} \vec{A} - \vec{G} \|^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} \kappa(m, \omega_n)A(\omega_n) - G(m)$$
\[ \| \hat{K}A - \bar{G} \|^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} \kappa(m, \omega_n)A(\omega_n) - G(m) \|^2 \]

\[ \| \hat{K}A - \bar{G} \|^2 + \lambda^2 \| \hat{\Gamma}A \|^2 \]

\[ P[G|A] \sim \exp\{-\| \hat{K}A - \bar{G} \|^2\} \]

\[ P[A] \sim \exp\{-\lambda^2 \| \hat{\Gamma}A \|^2\} \]

\[ A = \sum_{i=1}^{r} \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \bar{G} \]

\[ A = \sum_{i=1}^{r} \left\{ \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \right\} \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \bar{G} \]

Filter factors
Tikhonov regularization to fight with the saw tooth noise instability.
Tikhonov regularization to fight with the saw tooth noise instability.

\[
\tilde{A} = \sum_{i=1}^{r} \frac{\vec{u}_i \otimes \vec{v}_i}{\sigma_i} \vec{G}
\]

\[
\tilde{A} = \sum_{i=1}^{r} \left( \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \right) \frac{\vec{u}_i \otimes \vec{v}_i}{\sigma_i} \vec{G}
\]
Maximum entropy method


\[ P[G|\tilde{A}] = \exp\{-\chi^2[\tilde{A}]/2\}, \]

\[ \chi^2[\tilde{A}] = \sum_{m=1}^{M} \epsilon^{-1}(m) [G(m) - \tilde{G}(m)]^2 \]

\[ P[\tilde{A}] = \exp\{\alpha^{-1}S[\tilde{A}]\}, \]

\[ S[\tilde{A}] = \int d\omega \tilde{A}(\omega) \ln[\tilde{A}(\omega)/D(\omega)] \]
Maximum entropy method


\( D(\omega) \) is default model

\[
P[A] = \exp\{\alpha^{-1} S[\tilde{A}]\},
\]

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

Prior knowledge function
Maximum entropy method


1. One has escaped extra smoothening.

2. But one has got default model as an extra price.

\[ P[A] = \exp\{\alpha^{-1} S[\tilde{A}]\}, \]

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

Prior knowledge function
Maximum entropy method

\[ P[A|G] \sim P[G|A] \cdot P[A] \]

1. We want to avoid extra smoothening.

2. We want to avoid default model as an extra price.

\[
P[A] = \exp\{\alpha^{-1} S[\tilde{A}]\},
\]

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

Prior knowledge function
Maximum entropy method

\[ P[A|G] \sim P[G|A] \ P[A] \]

Both items (extra smoothening and arbitrary default model) can be somehow circumvented by the group of stochastic methods.

\[
P[A] = \exp\{\alpha^{-1} S[\tilde{A}]\},
\]

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

Prior knowledge function
Stochastic methods

\[ P[A|G] \sim P[G|A] \ P[A] \]

Both items (extra smoothening and arbitrary default model) can be somehow circumvented by the group of stochastic methods.
Stochastic methods

\[ P[A|G] \sim P[G|A] \cdot P[A] \]

The main idea of the stochastic methods is:

1. Restrict the prior knowledge to the minimal possible level (positive, normalized, etc…).

2. Change the likelihood function to the likelihood functional.
Stochastic methods

\[ P[A|G] \sim P[G|A] \ P[A] \]

The main idea of the stochastic methods is:

1. Restrict the prior knowledge to the minimal possible level (positive, normalized, etc…). **Avoids default model.**

2. Change the likelihood function to the likelihood functional. **Avoids saw-tooth noise.**
Stochastic methods

\[ P[A|G] \sim P[G|A] \ P[A] \]

Change the likelihood function to the likelihood functional. Avoids sawtooth noise.

\[ A = \int d\tilde{A} \tilde{A} \ P[\tilde{A}|G] \]

\[ P[A|G] = \exp\left\{-\chi^2[\tilde{A}]/\mathcal{T}\right\} \]

\[ \chi^2[\tilde{A}] = \sum_{m=1}^{M} \mathcal{E}^{-1}(m)[G(m) - \bar{G}(m)]^2 \]
Stochastic methods

Likelihood functional. **Avoids sawtooth noise.**

\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G] \]

\[ P[A|G] = \exp\{-\chi^2[\tilde{A}]/\mathcal{T}\} \]

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

Sandvik, Phys. Rev. B 1998, is the first practical attempt to think stochastically.
Stochastic methods


\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G] \]

\[ P[A|G] = \exp\{-\chi^2[\tilde{A}]/\mathcal{T}\} \]

\[ \chi^2[\tilde{A}] = \sum_{m=1}^{M} \varepsilon^{-1}(m)[G(m) - \tilde{G}(m)]^2 \]

Sandvik, Phys. Rev. B 1998, is the first practical attempt to think stochastically.

SOM was suggested in 2000. Mishchenko et al, Appendix B in Phys. Rev. B.
Some applications of SOM:

Stochastic methods

Likelihood functional. **Avoids sawtooth noise.**

\[
A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G]
\]

\[
P[\tilde{A}|G] = \exp\{-\chi^2[\tilde{A}]/\mathcal{T}\}
\]

\[
\chi^2[\tilde{A}] = \sum_{m=1}^{M} \xi^{-1}(m)[G(m) - \tilde{G}(m)]^2
\]

Sandvik, Phys. Rev. B 1998, is the first practical attempt to think stochastically.

SOM was suggested in 2000. Mishchenko et al, Appendix B in Phys. Rev. B.
Stochastic methods


\[ A = \int d\tilde{\mathcal{A}} \tilde{\mathcal{A}} \, P[\tilde{\mathcal{A}} | G] \]

\[
P[A | G] = \exp\left\{-\chi^2[\tilde{\mathcal{A}}] / \mathcal{I}\right\}
\]

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

Sandvik, Phys. Rev. B 1998, is the first practical attempt to think stochastically.

SOM was suggested in 2000. Mishchenko et al, Appendix B in Phys. Rev. B.

Beach, 2004, cond-mat
Stochastic methods

Likelihood functional. **Avoids sawtooth noise.**

\[
A = \int d\tilde{A} \tilde{A} P[\tilde{A} | G],
\]

\[
P[A | G] = \exp\{-\chi^2[\tilde{A}] / T\}
\]

\[
\chi^2[\tilde{A}] = \sum_{m=1}^{M} \varepsilon^{-1}(m) [G(m) - \tilde{G}(m)]^2
\]

**Sandvik, Phys. Rev. B 1998, is the first practical attempt to think stochastically.**

**Beach, 2004, cond-mat**

**SOM was suggested in 2000. Mishchenko et al, Appendix B in Phys. Rev. B.**

**Other variants after 2004**
Stochastic methods

Likelihood functional. **Avoids sawtooth noise.**

\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G], \]

\[ P[A|G] = \exp\left\{-\chi^2[\tilde{A}]/\mathcal{T}\right\}, \]

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

**What is the special need for the stochastic sampling methods?**
Stochastic methods

1. Avoid saw-tooth noise.
2. Avoid over-smoothing of the $\delta$-function

Typical spectrum of QP at $T=0$.

What is the special need for the stochastic sampling methods?
Stochastic methods

1. Avoid saw-tooth noise.
2. Avoid over-smoothing of the $\delta$-function.

Tikhonow regularization over-smoothes the $\delta$-function.

MaxEnt – default model has to fix $\delta$-function in advance.

What is the special need for the stochastic sampling methods?
Stochastic methods

1. Avoid saw-tooth noise.
2. Avoid over-smoothing of the $\delta$-function

Typical spectrum of QP at $T=0$.

Stochastic methods is a way to circumvent these problems.

What is the special need for the stochastic sampling methods?
Stochastic methods

\[ A = \int d\tilde{A} \, \tilde{A} \, P[\tilde{A} | G] \]

\[ P[A | G] = \exp\{-\chi^2[\tilde{A}] / \mathcal{T}\} \]

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

Back to Sandvik approach
Stochastic methods

One does not search for a single solution but samples through difference "configurations" (spectral functions $A$). Using the likelihood function $P$ which is characterized by fictitious "temperature" $T$ and fictitious "energy" $\chi^2$.

$$A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G]$$

$$P[A|G] = \exp\left\{-\chi^2[\tilde{A}]/T\right\}$$

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

\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G]. \]

\[ P[A|G] = \exp\left\{-\chi^2[\tilde{A}]/T\right\} \]

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

One does not search for a single solution but samples through different “configurations” (spectral functions A). Using the likelihood function P which is characterized by fictitious “temperature” T and fictitious “energy” \( \chi^2 \).

One interprets \( \chi^2 \) as an “energy” of fictitious Hamiltonian and T as a fictitious “temperature”. Hence, one involves the Metropolis algorithm for Monte Carlo to sample through configurations A.
Stochastic methods

\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A} | G] \]

\[ P[A | G] = \exp\{-\chi^2[\tilde{A}] / T\} \]

\[ \chi^2[\tilde{A}] = \sum_{m=1}^{M} \varepsilon^{-1}(m)[G(m) - \tilde{G}(m)]^2 \]

One interprets \( \chi^2 \) as an “energy” of fictitious Hamiltonian and \( T \) as a fictitious “temperature”. Hence, one involves the Metropolis algorithm for Monte Carlo to sample through configurations.

1. \( T \) is not too high. Otherwise \( A \) is far from spectra which fit well the correlation function \( G \).
2. \( T \) is not too small otherwise we are back again to the sawtooth noise problem. Over-fitting of the noise.
Stochastic methods

One interprets $\chi^2$ as an “energy” of fictitious Hamiltonian and $T$ as a fictitious “temperature”. Hence, one involves the Metropolis algorithm for Monte Carlo to sample through configurations.

1. $T$ is not too high. Otherwise $A$ is far from spectra which fit well the correlation function $G$.
2. $T$ is not too small otherwise we are back again to the sawtooth noise problem. Ower-fitting of the noise.

Simple rule $T = M$
One interprets $\chi^2$ as an “energy” of fictitious Hamiltonian and $T$ as a fictitious “temperature”. Hence, one involves the Metropolis algorithm for Monte Carlo to sample through configurations.

Which features of Sandvik method are artificial?

1. There is no real Hamiltonian and $T$ and, hence, one has no requirement to sample through Metropolis

2. Algorithm is not effective at low $T$ and use the tempering procedures with sampling at different $T$s.
Stochastic optimization method.

1. One has to sample through solutions $A(\omega)$ which fit the correlation function $G$ well.

2. One has to make some weighted sum of these well solutions $A(\omega)$.

$$A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G].$$

$$A(\omega) = \sum_{j=1}^{L} \xi_j \tilde{A}_j(\omega).$$
SOM is very similar to Sandvik method but circumvent its artificial features and, as a result, turns out more effective.
Stochastic optimization method.

One collects and averages large amount of “well” solutions and take an average.

1. What is the likelihood function (deviation measure of fit quality?)
2. How the spectrum is parameterized
3. How to find one “well” solution?
4. When the number of solutions is enough?
5. Tests.
Stochastic optimization method.

1. What is the likelihood function (deviation measure of fit quality)?

The deviation measure of SOM is given by expression

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

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

\[ \Delta(m) = \frac{G(m) - \tilde{G}(m)}{S(m)}, \] (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

\[ \tilde{G}(m) = \int_{-\infty}^{\infty} d\omega \kappa(m,\omega) \tilde{A}(\omega). \] (31)
We parameterize the spectral function $\tilde{A}$ as a sum

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

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

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

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

A configuration

$$\mathcal{C} = \{\{P_t\}, t = 1, ..., K\}$$

with the normalization constraint

$$\sum_{t=1}^{K} h_t w_t = I ,$$

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

Parameterization of the particular solution:

No predefined mesh for the energy ($\omega$) space.

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

Contribution of rectangle to

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

\[
\tilde{A}(\omega) = \sum_{t=1}^{K} \eta_{\{P_t\}}(\omega)
\]

of rectangles \( \{P_t\} = \{h_t, w_t, c_t\} \)

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

determined by height \( h_t > 0 \), width \( w_t > 0 \), and center \( c_t \).

A configuration

\[
\mathcal{C} = \{\{P_t\}, t = 1, \ldots, K\}
\]

with the normalization constraint

\[
\sum_{t=1}^{K} h_t w_t = I,
\]

If no analytic expression.

One tabulates:

\[
\Lambda(m, \Omega) = \int_{-\infty}^{\Omega} K(m, x) \, dx, \quad m = 1, \ldots, M
\]

Contribution:

\[
\tilde{G}(m) = \sum_{i=1}^{K} h_t \left[ \Lambda(m, c_t + w_t/2) - \Lambda(m, c_t - w_t/2) \right]
\]
Stochastic optimization method.

Contribution of rectangle to

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

\[
\tilde{A}(\omega) = \sum_{t=1}^{K} \eta_{\{P_t\}}(\omega)
\]

of rectangles \( \{P_t\} = \{h_t, w_t, c_t\} \)

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

determined by height \( h_t > 0 \), width \( w_t > 0 \), and center \( c_t \).

A configuration

\[ C = \{\{P_t\}, t = 1, ..., K\} \]

with the normalization constraint

\[
\sum_{t=1}^{K} h_t w_t = I,
\]

Particular cases. Imaginary time, \( T=0 \):

Kernel is

\[
K(m, \omega) = \exp(-i\tau_m \omega)
\]

Contribution:

\[
\tilde{G}(m) = \int_{-\infty}^{\infty} d\omega \ K(m, \omega) \ \tilde{A}(\omega)
\]

\[
\tilde{G}_C(\tau_m) = \begin{cases} 
  I, & \tau_m = 0, \\
  2\tau_m^{-1} \sum_{t=1}^{K} h_t e^{-c_t \tau_m} \sinh(w_t \tau_m/2), & \tau_m \neq 0.
\end{cases}
\]
Stochastic optimization method.

Contribution of rectangle to

Particular cases. Matsubara, any T:

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

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

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

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

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

A configuration

$$\mathcal{C} = \{\{P_t\}, t = 1, \ldots, K\}$$

with the normalization constraint

$$\sum_{t=1}^{K} h_t w_t = I,$$

Kernel is

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

Contribution:

$$\tilde{G}_C(i\omega_m) = \pm \sum_{t=1}^{K} h_t \ln \left[ \frac{c_t - w_t/2 - i\omega_m}{c_t + w_t/2 - i\omega_m} \right]$$
Particular solution $L^{(i)}(\omega)$ for LSF is presented as a sum of a number $K$ of rectangles with some width, height and center.
How to find one of solutions?
Stochastic Optimization method.

- Particular solution $L^{(i)}(\omega)$ for LSF is presented as a sum of a number $K$ of rectangles with some width, height and center.
- Initial configuration of rectangles is created by random number generator (i.e. number $K$ and all parameters of rectangles are randomly generated).
Stochastic Optimization method.

- Particular solution $L^{(i)}(\omega)$ for LSF is presented as a sum of a number $K$ of rectangles with some width, height and center.
- Initial configuration of rectangles is created by random number generator (i.e. number $N$ and all parameters of of rectangles are randomly generated).
- Each particular solution $L^{(i)}(\omega)$ is obtained by a naïve method without regularization (though, varying number $K$).

**Deviation measure for configuration:**

The deviation measure of SOM is given by expression

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

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

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

and

$$D[\tilde{L}_k(\omega)] = \int_0^{\tau_{\text{max}}} \left| \tilde{G}_k(\tau) - G_k(\tau) \right| G_k^{-1}(\tau) d\tau .$$
How to minimize the deviation?

Which updates?
Stochastic Optimization method: update procedures.

Shift of rectangular.
Stochastic Optimization method: update procedures.

Change of height of rectangular with fixed Z-factor.
Stochastic Optimization method: update procedures.

Split of rectangular.
Stochastic Optimization method: update procedures.

Glue two rectangles.
Parameters for changing are obtained by optimizing some continuous parameter making quadratic (intra)extrapolation.

For example: Measure of deviation for the shift of rectangle is calculated for distances $x$, $2x$, and $3x$ and then 3 points $D(x)$ is reproduced by parabola.

Variable $x$ can be any other continuous parameter of the update.
Stochastic Optimization method: update procedures.

Deviation measure.

CONFIGURATON

Accept only updates which decrease the deviation measure

WRONG STRATEGY
Stochastic Optimization method: update procedures.

Always accept with some probability some updates which decrease the deviation measure.

WRONG STRATEGY: Sandvik 1998, Beach 2004
Stochastic Optimization method: update procedures.

**CONFIGURATION**

*Shake-off two-step strategy:*

Step 1: Increase of deviation measure is allowed during $M$ steps with high probability.

Step 2: Only decrease of deviation measure is allowed during last $K$ steps.
Shake-off two-step strategy:
Step 1: Increase of deviation measure is allowed during $M$ steps with high probability
Step 2: Only decrease of deviation measure is allowed during last $K$ steps.

The $K+M$ chain is rejected if the final deviation $D$ is larger than the initial deviation.
Stochastic Optimization method: update procedures.

Shake-off two-step strategy:
Step 1: Increase of deviation measure is allowed during $M$ steps with high probability
Step 2: Only decrease of deviation measure is allowed during last $K$ steps.

$K+M$ chain is accepted if final $D$ is smaller than initial
How to judge that one of solutions is “GOOD”
How to judge that one of solutions is “GOOD”

\[ \kappa > \frac{1}{4} \text{ (Ideal limit } \kappa = \frac{1}{2} \text{)} \]

**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 \( \tau_m \) corresponding to a spectrum \( \tilde{A}_j(\omega) \) which (b) under-fits and (c) over-fits the uncorrelated noise of imaginary time data.

**deviation measure of SOM is given by expression**

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

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

\[
\Delta(m) = \frac{G(m) - \tilde{G}(m)}{S(m)}.
\]
Stochastic Optimization method.

- Particular solution $L^{(i)}(\omega)$ for LSF is presented as a sum of a number $K$ of rectangles with some width, height and center.
- Initial configuration of rectangles is created by random number generator (i.e. number $K$ and all parameters of rectangles are randomly generated).
- Each particular solution $L^{(i)}(\omega)$ is obtained by a naïve method without regularization (though, varying number $K$).
- Final solution is obtained after $M$ steps of such procedure

$$L(\omega) = M^{-1} \sum_i L^{(i)}(\omega)$$

- Each particular solution has saw tooth noise
- Final averaged solution $L(\omega)$ has no saw tooth noise though not regularized with sharp peaks/edges!!!
We can find many particular solutions each of which fits the input data reasonably.
We can find many particular solutions each of which fits the input data reasonably.

Which particular solutions one has to take into account?
Self-averaging of the saw-tooth noise.

Fig. 7. Comparison of the actual spectral function (dashed line) with the results of spectral analysis after averaging over (a) $M = 4$, (b) $M = 28$, and (c) $M = 500$ particular solutions.
Self-averaging of the saw-tooth noise.

Fig. 7. Comparison of the actual spectral function (dashed line) with the results of spectral analysis after averaging over (a) \( M = 4 \), (b) \( M = 28 \), and (c) \( M = 500 \) particular solutions.
Self-averaging of the saw-tooth noise.

Fig. 7. Comparison of the actual spectral function (dashed line) with the results of spectral analysis after averaging over (a) $M = 4$, (b) $M = 28$, and (c) $M = 500$ particular solutions.
Fig. 7: (a) Self-averaging of the sawtooth noise after summation of 4, 30, and 500 solutions. (b) Typical probability distribution $P(D/D_{\text{min}})$ of solutions with different deviation measures.
Which particular solutions one has to take into account?

One has to include solution with deviation measure \( D[A] \) which is less that twice of minimal \( \text{MIN}\{D[A]\} \)

\[
D[A] < 2 \text{MIN}\{D[A]\}
\]

*deviation measure* of SOM is given by expression

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

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

\[
\Delta(m) = \frac{G(m) - \tilde{G}(m)}{S(m)} ,
\]
Sandvik method

\[ A = \int d\tilde{A} \tilde{A} P[\tilde{A}|G] \]

\[ P[A|G] = \exp\left\{-\chi^2[\tilde{A}]/T\right\} \]

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

1. T is not too high. Otherwise A is far from spectra which fit well the correlation function G.
2. T is not too small otherwise we are back again to the sawtooth noise problem. Over-fitting of the noise.

Simple rule rule \( T = M \)
Tikhonov functional:
similar strategy for choice of $\lambda$

\[
\| \hat{K} \hat{A} - \hat{G} \|^2 = \sum_{m=1}^{M} \left| \sum_{n=1}^{N} K(m, \omega_n) A(\omega_n) - G(m) \right|^2
\]

\[
\| \hat{K} \hat{A} - \hat{G} \|^2 + \lambda^2 \| \hat{\Gamma} \hat{A} \|^2
\]

\[
\vec{A} = \sum_{i=1}^{r} \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \vec{G}
\]

\[
\vec{A} = \sum_{i=1}^{r} \left\{ \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \right\} \frac{\vec{u}_i^\dagger \otimes \vec{v}_i}{\sigma_i} \vec{G}
\]
Max Ent

Similar strategy everywhere: equate noise contribution with regularization contribution

Avoid over-fitting
Similar strategy everywhere: equate noise contribution with regularization contribution


Arsenin (1986): the art of finding solution for ill posed problem lies in an intuition which tells us when to stop improve the deviation before the noise of input data overruns the information contained in the input data.
Which particular solutions one has to take into account?

One has to include solution with deviation measure \( D[A] \) which is less that twice of minimal \( \text{MIN}\{D[A]\} \)

\[
D[A] < 2 \text{ MIN}\{D[A]\}
\]

**Deviation measure of SOM is given by expression**

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

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

\[
\Delta(m) = \frac{G(m) - \tilde{G}(m)}{S(m)}.
\]
Some tests
Some tests

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ k(m,\omega) \ A(\omega) \]

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

\[ R \in [-1, 1] \]

\[ B = 10^{-4} \]
Some tests

Particular cases.

Imaginary time, $T=0$:

Kernel is

$$K(m,\omega) = \exp(-\tau m \omega)$$

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.
Some tests

Particular cases.

Imaginary time, finite $T$, fermions

Kernel is

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

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.
Some tests

Particular cases.

Imaginary time, finite T, optical conductivity

Kernel is

\[ G(m) = \int_{-\infty}^{\infty} d\omega \ K(m,\omega) \ A(\omega) \]

\[ K(\tau_m,\omega) = \frac{1}{\pi} \frac{\omega \exp(-\tau_m \omega)}{1 - \exp(-\beta \omega)} \]

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.
Some tests

Particular cases.

Matsubara frequencies, finite $T$, fermions

Kernel is

$$G(m) = \int_{-\infty}^{\infty} d\omega \, K(m,\omega) \, A(\omega)$$

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).
Particular cases.

Imaginary time, finite $T$, optical conductivity

Kernel is

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

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.
Analytical continuation of spectral data from imaginary time axis to real frequency axis using statistical sampling

K. Vafayi and O. Gunnarsson
Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany
(Received 27 February 2007; revised manuscript received 2 May 2007; published 19 July 2007)

We present a method for performing analytical continuation of spectral data from imaginary time to real frequencies based on a statistical sampling method. Compared with the maximum entropy method (MEM), an advantage is that no default model needs to be introduced. For the problems studied here, the statistical sampling method gives comparable or slightly better results than MEM using quite accurate default models.

DOI: 10.1103/PhysRevB.76.035115

PACS number(s): 72.15.Eb, 02.70.Ss

ity in Eq. (7) as a weight function. Comparing with the maximum entropy method (MEM), an advantage is that there is no need to provide a default model, which influences the MEM results if the method is close to its limit of applicability. For the problems considered here, the statistical sampling method gives comparable or slightly better results than MEM using default models close to the exact result. The
Stochastic Optimization method.

- Particular solution $L^{(i)}(\omega)$ for LSF is presented as a sum of a number $K$ of rectangles with some width, height and center.
- Initial configuration of rectangles is created by random number generator (i.e. number $K$ and all parameters of rectangles are randomly generated).
- Each particular solution $L^{(i)}(\omega)$ is obtained by a naïve method without regularization (though, varying number $K$).
- **Final solution is obtained after $M$ steps of such procedure**

\[ L(\omega) = M^{-1} \sum_i L^{(i)}(\omega) \]

- Each particular solution has saw tooth noise
- **Final averaged solution $L(\omega)$ has no saw tooth noise though not regularized with sharp peaks/edges!!!!**
Conclusions:

1. Analytic continuation is ill posed problem.
2. Similar Fredholm I integral equation problem in many applications.
4. All methods bear similar strategy of regularization: not to over-fit the noise.
5. Each method is the best in each particular case. There is no universal method which is “the best” for all cases.
6. We are still on the way to improve the analytic continuation.
7. Combinations of methods might help.
Questions?
New Method for Low Temperature analysis of the ESR spectra

Andrey Mishchenko
CMRG, RIKEN

Collaborations:

Tatsuo Hasegawa (AIST)
Hiroyuki Matsui (AIST)

New Method for Low Temperature analysis of the ESR spectra

1. Nature of the inhomogeneous ESR lineshape and line narrowing
3. Analysis of the lineshape of an electron trapped by an impurity
4. Analysis of the fine structure of the ESR line can give a complete information on the distribution of the traps versus localization parameters
Basics of ESR

Transition between Zeeman split levels under the influence of the electromagnetic field.

For example, the frequency is fixed and magnetic field $B$ is varied. Then, the intensity of signal $I(B)$ is

$$I(B) \sim \delta(B - B_0)$$
Basics of ESR

Transition between Zeeman split levels under the influence of the electromagnetic field.

For example, the frequency is fixed and magnetic field $B$ is varied. Then, the intensity of signal $I(B)$ is

$$I(B) \sim \delta(B-B_0)$$

Hyperfine splitting
Basics of ESR

Transition between Zeeman split levels under the influence of the electromagnetic field.

For example, the frequency is fixed and magnetic field $B$ is varied. Then, the intensity of signal $I(B)$ is

$$I(B) \sim \delta(B-B_0)$$

Hyperfine splitting

In complex system at low temperatures the lineshape is set by the sum of random contributions coming from hyperfine and superhyperfine interactions.
In complex system at low temperatures the lineshape is set by the sum of random contributions coming from hyperfine and superhyperfine interactions.

The distribution of the sum of random variable is Gaussian:

\[
G(B) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left[ -\frac{(B - B_0)^2}{2\sigma_0^2} \right]
\]

\[
S(B) = \frac{dG(B)}{dB}
\]
If the electron is spread over N molecules, one has a distribution over sum of random variables. Then, according to the **Central Limit Theorem** the distribution is Gaussian with more narrow dispersion \( \sigma \):

\[
G(B) = \frac{1}{\sqrt{2\pi \sigma^2_0}} \exp \left[ \frac{-(B - B_0)^2}{2\sigma^2_0} \right]
\]

\[
\langle B \rangle = \frac{1}{N} (B_1 + B_2 + \ldots + B_N)
\]

\[
\sigma = \frac{\sigma_0}{\sqrt{N}}
\]
Basics of ESR

The ESR signal from single molecule in this case is Gaussian. The standard expression describing the hyperfine structure of one molecule reads\textsuperscript{10}

\[
f(B) = \frac{\sum_{m_1=-n_1I_1}^{n_1I_1} \cdots \sum_{m_k=-n_kI_k}^{n_kI_k} P(m_1, \ldots, m_k) \times \frac{1}{\pi \left( B - \sum_{i=1}^{k} A_i m_i \right)^2 + \Gamma^2}}{\Gamma}. \tag{1}
\]

Here \(k\) is the number of the groups of equivalent nuclei, \(n_i\) is the number of the equivalent nuclei in the \(i\)th group, \(I_i\) is nuclear spin in the \(i\)th group, \(\Gamma\) is the linewidth of each peak, \(P\) is the intensity of each peak and \(B\) is magnetic field. If protons (\(I = 1/2\)) are the only paramagnetic nuclei, as it is, e.g., in the case for pentacene molecule, \(P\) is given as

\[
P(m_1, \ldots, m_k) = \prod_{i=1}^{k} \frac{C_{2n_iI_i}^{m_i+n_iI_i}}{(2I_i+1)-n_i}, \tag{2}
\]

where \(C_{2n_iI_i}^{m_i+n_iI_i}\) are binomial coefficients.

\[N\] molecules, \(A_i = A_i/N\), \(n_i = n_i/N\)
If the electron is spread over N molecules, one has a distribution over sum of random variables. Then, according to the **Central Limit Theorem** the distribution is Gaussian with more narrow dispersion $\sigma$:

\[
\langle B \rangle = \frac{1}{N} (B_1 + B_2 + \ldots + B_N)
\]

\[
\sigma = \frac{\sigma_0}{\sqrt{N}}
\]

Above knowledge is from the theory of inhomogeneous lineshape in molecules. In solids? When an electron is localized on the trap, there is a charge distribution $f(i)$ and one needs to look at the distribution of different variable $\langle B \rangle$:
If the electron is spread over $N$ molecules, one has a distribution over sum of random variables. Then, according to the Central Limit Theorem the distribution is Gaussian with more narrow dispersion $\sigma$:

$$<B> = \frac{1}{N} (B_1 + B_2 + \ldots + B_N)$$

$$\sigma = \frac{\sigma_0}{\sqrt{N}}$$

Above knowledge is from the theory of inhomogeneous lineshape in molecules. In solids? When an electron is localized on the trap, there is a charge distribution $f(i)$ and one needs to look at the distribution of different variable $<B>$:

Experiment reveals non-Gaussian signal. Maybe this is the reason.
If the electron is spread over N molecules, one has a distribution over sum of random variables. Then, according to the Central Limit Theorem the distribution is Gaussian with more narrow dispersion $\sigma$:

$$<B> = \frac{1}{N} (B_1 + B_2 + \ldots + B_N)$$

$$\sigma = \frac{\sigma_0}{\sqrt{N}}$$

Above knowledge is from the theory of inhomogeneous lineshape in molecules. In solids? When an electron is localized on the trap, there is a charge distribution $f(i)$ and one needs to look at the distribution of variables $<B>$:

$$\sum_i f(i) = 1$$

$$<B> = \sum_i f(i) B_i$$
If the electron is spread over N molecules, one has a distribution over sum of random variables. Then, according to the **Central Limit Theorem** the distribution is Gaussian with more narrow dispersion $\sigma$:

$$<B> = \frac{1}{N} (B_1 + B_2 + \ldots + B_N)$$

$$\sigma = \frac{\sigma_0}{\sqrt{N}}$$

Above knowledge is from the theory of inhomogeneous lineshape in molecules. In solids? When an electron is localized on the trap, there is a charge distribution $f(i)$ and one needs to look at the distribution of variables $<B>$:

Numerical simulations show that (although CLT does not work in this case) the distribution is still Gaussian.

$$\sum_i f(i) = 1$$

$$<B> = \sum_i f(i) B_i$$
CLT for non-uniformly distributed variables: nontrivial!

The uniform distribution is the best case for narrowing

In extreme limit of localized case $N_{\text{eff}} \to 1$. 
CLT for non-uniformly distributed variables: nontrivial!

\[
N_{\text{eff}} = \left[ \sum_{i} p(i)^2 \right]^{-1/2}
\]
However, experimental signal is not Gaussian which means that there is no traps with some given value of $N_{\text{eff}}$ which dominate.

For $T<50K$ the wave saturation experiment shows that all carriers are localized and no broadening except nonhomogeneous one is expected!
However, experimental signal is not Gaussian which means that there is no traps with some given value of $N_{\text{eff}}$ which dominate.

Two kinds of traps:

$$I(B) = \alpha G(N_{\text{eff}}^1, B-B_0) + \beta G(N_{\text{eff}}^2, B-B_0)$$
However, experimental signal is not Gaussian which means that there is no traps with some given value of $N_{\text{eff}}$ which dominate.

### Experimental ESR in pentacene

![Graph showing ESR peak](image)

**Two kinds of traps:**

$$I(B) = \alpha G(N_{\text{eff}}^1, B-B_0) + \beta G(N_{\text{eff}}^2, B-B_0)$$

**Three kinds of traps:**

$$I(B) = \alpha G(N_{\text{eff}}^1, B-B_0) + \beta G(N_{\text{eff}}^2, B-B_0) + \gamma G(N_{\text{eff}}^3, B-B_0)$$
Experimental ESR in pentacene

Broader view: distribution of traps

\[
G_{N_{\text{eff}}}(B) = \frac{1}{\sqrt{2\pi[\sigma_0/\sqrt{N_{\text{eff}}}]^2}} \exp \left[\frac{(B - B_0)^2}{2[\sigma_0/\sqrt{N_{\text{eff}}}]^2}\right]
\]

\[
I(B) = \int_{1}^{+\infty} D(N_{\text{eff}}) G_{N_{\text{eff}}}(B - B_0) dN_{\text{eff}}
\]

\[
S(B) = \int_{1}^{+\infty} D(N_{\text{eff}}) \left[\frac{d}{dB} G_{N_{\text{eff}}}(B - B_0)\right] dN_{\text{eff}}
\]

\[
G(\tau) = \int_{0}^{+\infty} \rho(\omega) K[\tau, \omega]
\]
Broader view:
distribution of traps

Note, such interpretation requires that all molecules of molecular crystal are equally oriented with respect to surface.

\[ G_{N_{eff}}(B) = \frac{1}{\sqrt{2\pi \sigma_0 / \sqrt{N_{eff}}}} \exp \left[ \frac{(B - B_0)^2}{2\sigma_0 / \sqrt{N_{eff}}^2} \right] \]

\[ I(B) = \int_{1}^{+\infty} D(N_{eff}) G_{N_{eff}}(B - B_0) dN_{eff} \]
\[ S(B) = \int_{1}^{+\infty} D(N_{eff}) \left[ \frac{d}{dB} G_{N_{eff}}(B - B_0) \right] dN_{eff} \]

\[ G(\tau) = \int_{0}^{+\infty} \rho(\omega) K[\tau, \omega] \]
Fredholm integral equation of the 1-st kind: so called ill posed problem

\[ G_{N_{\text{eff}}} (B) = \frac{1}{\sqrt{2\pi[\sigma_0/\sqrt{N_{\text{eff}}}]^2}} \exp \left[ \frac{(B - B_0)^2}{2[\sigma_0/\sqrt{N_{\text{eff}}}]^2} \right] \]

\[ I(B) = \int_{1}^{+\infty} D(N_{\text{eff}}) G_{N_{\text{eff}}} (B - B_0) dN_{\text{eff}} \] 
\[ S(B) = \int_{1}^{+\infty} D(N_{\text{eff}}) \left[ \frac{d}{dB} G_{N_{\text{eff}}} (B - B_0) \right] dN_{\text{eff}} \]

\[ G(\tau) = \int_{0}^{+\infty} \rho(\omega) K[\tau, \omega] \]

Previously (A.S. Mishchenko et al, PRB v. 62, 6317 (2000)) a method more flexible and less capricious than MEM was developed for solving the analytic continuation problem -

Stochastic Optimization method
\[ G_{N_{\text{eff}}}(B) = \frac{1}{\sqrt{2\pi[\sigma_0/\sqrt{N_{\text{eff}}}]^2}} \exp \left[ \frac{(B - B_0)^2}{2(\sigma_0/\sqrt{N_{\text{eff}}})^2} \right] \]

\[ I(B) = \int_1^{+\infty} D(N_{\text{eff}}) G_{N_{\text{eff}}}(B - B_0) dN_{\text{eff}} \]

\[ S(B) = \int_1^{+\infty} D(N_{\text{eff}}) \left[ \frac{d}{dB} G_{N_{\text{eff}}}(B - B_0) \right] dN_{\text{eff}} \]

\[ G(\tau) = \int_0^{+\infty} \rho(\omega) K[\tau, \omega] \]

![Graph showing density of N_{\text{eff}}]
\[ G_{N_{eff}}(B) = \frac{1}{\sqrt{2\pi [\sigma_0/\sqrt{N_{eff}}]^2}} \exp \left\{ \frac{(B - B_0)^2}{2[\sigma_0/\sqrt{N_{eff}}]^2} \right\} \]

\[ I(B) = \int_{1}^{+\infty} D(N_{eff}) G_{N_{eff}}(B-B_0) dN_{eff} \quad S(B) = \int_{1}^{+\infty} D(N_{eff}) \left[ \frac{d}{dB} G_{N_{eff}}(B-B_0) \right] dN_{eff} \]

\[ G(\tau) = \int_{0}^{+\infty} \rho(\omega) K[\tau, \omega] \]
There are 100-s methods to fit the signal.

We say for the 1-st time that we do not need any fit!!!
ESR spectrum of organic FET

Reliability of result: