1 Introduction

2 Theory

2.1 The Maximum Entropy Method (MEM)

Given a N dimensional noisy data set $\vec{G}$ and a model characterized by the M dimensional parameter vector $\vec{A}$. The model is assumed to represent a valid relation between G_i and $\vec{A}$. This means that for an exact value G_i and perhaps some other known parameters, one could in principle determine the correct vector of unknown parameters $\vec{A}$. However since $\vec{G}$ is noisy the uncertainty about the true value of G_i induces a uncertainty on the true value of $\vec{A}$. Therefore it makes sense to determine a probability distribution $p(\vec{A}|\vec{G})$ instead of one single solution for $\vec{A}$.
Using Bayes' theorem the following relation for the wanted prob. distribution $p(\vec{A}|\vec{G})$ can be found

$$p(\vec{A}|\vec{G}) \propto p(\vec{G}|\vec{A})p(\vec{A})$$

This is the well known $posterior \propto likelihood * prior$ relation from Bayesian statistics. Hence if it is possible to find the likelihood and posterior distributions one has quantitative information about the probability for $\vec{A}$ to be true given $\vec{G}$.

2.1.1 The likelihood function

Following Bryan [reference] for the maximum entropy method the likelihood function is restricted to the functional form

$$p(\vec{G}|\vec{A}) \propto exp(-L(\vec{F},\vec{G}))$$

where $\vec{F}$ is leneary related to $\vec{A}$, $\vec{F} = \mathbf{K}\vec{A}$ by the matrix $\mathbf{K}$ which represents a valid relation between $\vec{A}$ and $\vec{G}$.

2.1.2 The prior distribution

In contrast to the likelihood function which can be found by data manipulation and reasoning in most cases very exact, the prior is quiet hard to find in a correct way. This is the typical drawback of Bayesian inference methods. The core of the maximum entropy method is now the used prior distribution. Following again Brayn [reference] the prior for an unnormalized positive additive density $\vec{A}$ is given by

$$p(\vec{A}|\alpha,\vec{m}) \propto exp(\alpha S(\vec{m},\vec{A}))$$

where $\alpha \in \mathbb{R}^+$ is a unknown parameter and S is the entropy of $\vec{A}$ relative to a default model $\vec{m}$

$$S = \sum_{m=1}^{M} A_m - m_m - A_m log(A_m/m_m)$$

It is noteworthy that on the one hand this restricts the parameter vector $\vec{A}$ to be possible to be interpreted as *positive additive density*. On the other hand the choice for α and $\vec{m}$ are of big influece and have to handled with care. This subject will be addressed in more detail in a later chapter.

2.1.3 The posterior distribution

Combining now both general forms of the likelihood and prior distribution within the maximum entropy framework the posterior distribution for $\vec{A}$ is up to a normalization constant given by

$$p(\vec{A}|\vec{m}, \alpha, \vec{G}) \propto exp(\alpha S(\vec{m}, \vec{A}) - L(\vec{F}, \vec{G})) = exp(Q) \tag{1}$$

Hence for given α, expression (1) reaches its maximum probability for $\hat{\vec{A}}$ which maximizes $Q = \alpha S - L$. This means we have to find $\hat{\vec{A}}$ such that

$$\vec{\nabla} Q(\hat{\vec{A}}) = \alpha \vec{\nabla} S(\hat{\vec{A}}) - \vec{\nabla} L(\hat{\vec{A}}) = 0 \tag{2}$$

The numercial solution of equation (2) is the central issue of the MEM algorithm.

2.2 Analytic Continuation of QMC Data using MEM

Most quantum Monte Carlo (QMC) simulations produce Green's functions $G(\tau)$ of Matsubara imaginary time $\tau = it$. However the real time/frequency results $G(t)/G(\omega)$ are crucial since most experiments probe quantities related to the real time/frequency Green's functions. Fortunately the relation between $G(\tau)$ and the imaginary part of $G(\omega)$, is linear and given by

$$G(\tau) = \int d\omega A(\omega) K(\tau, \omega) \tag{3}$$

where the so caled Lehmann spectral function is given by $A(\omega) = -\frac{1}{\pi} \Im G(\omega)$ and $K(\tau, \omega)$ is a kernel, different for fermionic, bosonic or anomalous case. Therefore if it is possible to reconstruct $A(\omega)$ from given $G(\tau)$ one has the information about the real frequency Green's function $G(\omega)$. Why????
In this report we will restrict our self to the fermoinic case. For fermions the Lehmann spectral function is positve definite, $G(\tau)$ is periodic with inverse temperatur $\beta = 1/k_B T$ and the Kernel is given by

$$K(\tau, \omega) = \frac{exp(-\tau\omega)}{1 + exp(-\beta\omega)} \tag{4}$$

2.2.1 Discretized version of the problem

Because of the methodically given uncertainty of Quantum Monte Carlo simulations, doing QMC for N different imaginary times τ_n will produce a N dimensional noisy data set $\vec{G}$ where G_n is the mean of all QMC steps.
The idea is now to find a way to extract $A(\omega)$ form the noisy data set $\vec{G}$ using the maximum entropy method.
First of all we note that for the MEM formalism a valid model wich predicts G_n for a given M dimensional parameter vector $\vec{A}$ is necessary. This can be achieved if expression (3) is approximated as Rieman sum

$$G_n = G(\tau_n) = \int_a^b d\omega A(\omega) K(\tau_n, \omega) \approx \sum_{m=1}^{M} A_m K(\tau_n, \omega_m) \tag{5}$$

where $A_m = \Delta\omega A(\omega_m)$, $\omega_m = \Delta\omega m$ and $\Delta\omega = (b-a)/M$ (a,b have to be choosen in a sensible way). After this discretizaton we have a parameter vector $\vec{A} = (A_1, ..., A_M)^T$ and a true linear model $\vec{G} = \mathbf{K}\vec{A}$ where $\mathbf{K} \in \mathbb{R}^{N \times M}$ and $K_{nm} = K(\tau_n, \omega_m)$.

2.2.2 The likelihood function

As already mentioned to apply the Maximum Entropy Method the likelihood function hast to have the functional form $p(\vec{G}|\vec{A}) \propto exp(-L(\vec{F}, \vec{G}))$ where $\vec{F}$ is leneary related to $\vec{A}$ by $\vec{F} = \mathbf{K}\vec{A}$ what is already fulfilled by (5). For QMC data it is possible to achieve a multivariate gaussian shape of the likelihood function, such that $L = 1/2((\vec{G}-\vec{F})^T diag\{1/\sigma_n^2\}(\vec{G}-\vec{F}))$.

Since for the purpose of this educational project work no real QMC data was available we will only give a short overview of the main aspects how QMC data has to be manipulated in principle to reach the desired form of the likelihood function.
For each of the N imaginary times τ_n, $G(\tau_n) = G_n$ is calculated a plenty of times in N_{QMC} QMC steps, with results G_n^i and each with a different error. Hence one can interpret the relative freqency as probability distribution

$$p_{QMC}(G_n) = n(G_n^i = G_n)/N_{QMC}$$

The resulting distribution $p_{QMC}(G_n)$ is not gaussian and also correlated between different QMC steps. To get rid of this problem one perfoms a rebinning of the data. This means one considers the average of n_b succeding measurement as new datapoint

$$G_n^b = \sum_{(b-1)n_b+1}^{bn_b} \frac{G_n^i}{n_b}$$

So instead of N_{QMC} datapoints for each τ_n we have now $N_b = N_{QMC}/n_b$ datapoints. This rebinning has now two desired effects wich we can understand if the procedure is considered as logical 2 step rebinning.

1. In a first rebinning we get rid of the correlations between the succeding QMC steps. (As long as the bin size n_b is choosen big enough compared to the correlation lenght.)

2. Since correlations are removed the rebinned data represents a set of independent and identical drawn random variables. Hence for the second rebinning step we can argue using the central limit theorem that the resulting random variable should be gaussian distributed.

Remark: To find the optimal binsize n_b the current method is to compare higher moments (skewness, kurtosis) of the rebinned data to a data set of equal length N_b, drawn by perfect gaussian. The optimal size for n_b is reached if the moments for both data samples converge.

This gaussian distribution can be approximated by

$$p(G_n^b) = \frac{1}{\sqrt{2\pi}\sigma_n} exp(-\frac{(G_n^b - \overline{G}_n)^2}{2\sigma_n^2})$$

where $\overline{G}_n = \sum_b G_n^b / N_b$ and $\sigma_n^2 = \sum_b (G_n^b - \overline{G}_n)^2/(N_b - 1)$ are calculated in the usual way from the data. It is helpfull to keep in mind, that this distributions is only an approximation and not the true one due to the errors in $\overline{G}_n$ and σ_n. But with this information we can argue using again the CLT that the true distribution of the mean for each time step τ_n is again gaussian

$$p(\overline{G}_n) = \frac{1}{\sqrt{2\pi}\sigma_n^{real}} exp(-\frac{(\overline{G}_n - \mu_n)^2}{2\sigma_n^{real}})$$

where μ_n and σ_n^{real} represent unknown true values. We now assume that for a true $G(\tau_n)$ the observed QMC results are distributed in a way that $G(\tau_n)$ is given by the mean of the observed QMC results. Since the mean is conserved under all rebinning and averaging steps we have $G(\tau_n) = \mu_n$. Using (5) we can argue that for a given spectral function $A(\omega)$ the observed distribution for $\overline{G}_n$ is given by

$$p(\overline{G}_n|\vec{A}) = \frac{1}{\sqrt{2\pi}\sigma_n^{real}} exp(-\frac{(\overline{G}_n - \sum_m K_{nm} A_m)^2}{2\sigma_n^{real}})$$

In a last step we approximate $\sigma_n^{real} \approx \sigma_n/\sqrt{N_b}$ motivated by the CLT and reformulate the whole problem in a compact multivariate form. This gives a likelihood function

$$p(\vec{\overline{G}}|\vec{A}) \propto exp\left(-\frac{1}{2}(\vec{\overline{G}} - \mathbf{K}\vec{A})^T diag\{\frac{N_b}{\sigma_n^2}\}(\vec{\overline{G}} - \mathbf{K}\vec{A})\right) \quad (6)$$

where we have assumed statistical independence between the different $\overline{G}_n$. For simplicity of notation we will set $\vec{\overline{G}} \to \vec{G}$ and $\sigma_n/\sqrt{N_b} \to \sigma_n$ in the upcomming parts of this report. Such that the final form of the likelihood function given by

$$
\begin{aligned}
p(\vec{G}|\vec{A}) &\propto exp\left(-\frac{1}{2}(\vec{G}-\mathbf{K}\vec{A})^T diag\{\frac{1}{\sigma_n^2}\}(\vec{G}-\mathbf{K}\vec{A})\right) \\
&= exp\left(-\frac{1}{2}(\vec{G}-\vec{F})^T diag\{\frac{1}{\sigma_n^2}\}(\vec{G}-\vec{F})\right) \\
&= exp(-L(\vec{F},\vec{G}))
\end{aligned} \tag{7}
$$

2.2.3 The prior distribution

Since $A(\omega)$ is positive definite $\vec{A}$ has all properties neccesary to apply the form of the MEM prior.

$$p(\vec{A}|\alpha,\vec{m}) \propto exp(\alpha S(\vec{A},\vec{m}) \tag{8}$$

2.2.4 The posterior distribution

Putting all things together we found an expression for the probability of $\vec{A}$ to be the true spectral function in case observing the data $\vec{G}$ given α and $\vec{m}$.

$$p(\vec{A}|\alpha,\vec{m},\vec{G}) \propto exp(\alpha S - L(\vec{F},\vec{G})) = exp(Q(\vec{A})) \tag{9}$$

Therefore for given α and $\vec{m}$ we can calculate the most probable $\hat{\vec{A}}$. However this tells us nothing about the plausibility of the values for α and $\vec{m}$ which have significat influence of the obtained results.

2.2.5 Approaches to treat α

We will present 2 common ways how to deal with the uncertainty about how to choose α. The so called *Classic* and *Bryan's method*.
If we introduce $p(\alpha)$ the prior probability distribution for α (which we assume to be independen on $\vec{m}$ we can find a posterior distribution for α using (9) as

$$p(\alpha|\vec{G},\vec{m}) = \int d\vec{A} p(\alpha,\vec{A}|\vec{G},\vec{m}) = \int d\vec{A} p(\vec{A}|\alpha,\vec{m},\vec{G})p(\alpha) \tag{10}$$

Making a gaussian approximation $exp(Q(\vec{A})) \approx exp(Q(\hat{\vec{A}}) + \frac{1}{2}\delta\vec{A}^T\nabla\nabla Q(\hat{\vec{A}})\delta\vec{A})$ of (9) we can approximate the intragral by

$$p(\alpha|\vec{G},\vec{m}) \propto \prod_m \left(\frac{\alpha}{\alpha+\lambda_m}\right)^{1/2} exp(Q(\hat{\vec{A}}(\alpha))p(\alpha) \tag{11}$$

where λ_m are the eigenvalues of $diag\{\vec{A}^{1/2}\}\nabla\nabla L(\vec{A})diag\{\vec{A}^{1/2}\}$ evaluated at $\hat{\vec{A}}(\alpha)$.
The *Classic method* is now to choose $\hat{\alpha}$ which maximizes (11). Seting $\partial_\alpha p(\alpha|\vec{G},\vec{m}) =$

0 and assuming $\partial_\alpha \lambda_m \approx 0$ and that the prior $p(\alpha)$ will be owerwhelmed by the data, this leads to

$$-2\hat{\alpha}S \approx \sum_m \frac{\lambda_m}{\lambda_m + \hat{\alpha}} \tag{12}$$

Brayn's method in contrast tries to use the wohle information contained in $p(\alpha|\vec{G}, \vec{m})$. Instead of choosing one single value for α one takes the expection value of $\vec{\hat{A}}(\alpha)$.

$$\vec{\overline{A}} = \int d\alpha \vec{\hat{A}}(\alpha) p(\alpha|\vec{G}, \vec{m}) \tag{13}$$

3 Numerical algorithm

In this section we describe the numerical algorithm to perform the analytic continuation of quantum Monte Carlo data by the maximum entropy method. As described in Sec. 2 the quantity $Q = \alpha S - L$ has to be maximized with respect to $\vec{A}$ in order to find the most probable $\vec{A}$ given the noisy Greens function $\vec{G}$. To maximize Q we calculate the gradient of Q with respect to $\vec{A}$ and set it to zero:

$$\vec{\nabla} Q = \alpha \vec{\nabla} S - \vec{\nabla} L = 0 \tag{14}$$

Then Equ. 14 leads to:

$$-\alpha \log\left(\frac{A_i}{m_i}\right) = \sum K_{ji} \frac{\partial L}{\partial \vec{F}} \tag{15}$$

where:

$$\vec{F} = \mathbf{K}\vec{A} \text{ and } \vec{\nabla} L = \frac{\partial \vec{F}}{\partial \vec{A}} \frac{\partial L}{\partial \vec{F}} = \mathbf{K}^T \frac{\partial L}{\partial \vec{F}} \tag{16}$$

The solution $\vec{A}$ to the problem can then be represented in terms of a vector $\vec{v}$

$$\log\left(\frac{\vec{A}}{\vec{m}}\right) = \mathbf{K}^T \vec{v} \tag{17}$$

where Equ. 17 has to be read component wise. Now, a singular value decomposition of $\mathbf{K}$ is performed with $\mathbf{K} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T$. Please note that both $\mathbf{V}$ and $\mathbf{U}$ are orthonormal matrices and therefor $\mathbf{V}^{-1} = \mathbf{V}^T$ and $\mathbf{U}^{-1} = \mathbf{U}^T$. Here $\mathbf{\Sigma}$ has only nonzero components on its diagonal which are called the singular values of $\mathbf{K}$. Because $\mathbf{K}^T$ and $\mathbf{U}$ share the same vector space the solution $\vec{A}$ can also be represented by a new vector $\vec{u}$

$$A_i = m_i \exp\left(\sum U_{in} u_n\right) \tag{18}$$

Now Bryan argues that unless $\mathbf{K}$ is of full rank the components of $\vec{u}$ will not be independent. Because of the limited precision of the computer and the singular value decomposition some of the singular values of $\mathbf{K}$ will effectively be zero. The search for the optimal $\vec{u}$ can therefor be reduced to the nonzero singular values of $\mathbf{K}$. Let s be the number of nonzero singular values the search can then be limited to the s-dimensional space which Bryan calls the singular space. Bryan's method therefor first reduces all relevant matrices to the singular space. The vector $\vec{u}$ is now of length s, the number of columns of $\mathbf{V}$ and $\mathbf{U}$ are reduced to s and $\mathbf{\Sigma}$ is now a $s \times s$ square matrix. Making use of Equ. 18 and $\mathbf{K} = \mathbf{V\Sigma U}^T$ Equ. 15 can be rewritten as:

$$-\alpha \mathbf{U}\vec{u} = \mathbf{U\Sigma V}^T \frac{\partial L}{\partial \vec{F}} \tag{19}$$

Multiplying Equ. 19 by $\mathbf{U}^T$ on both sides it reduces to

$$-\alpha \vec{u} = \mathbf{\Sigma V}^T \frac{\partial L}{\partial \vec{F}} \equiv g \tag{20}$$

or

$$-\alpha \vec{u} - g = 0 \tag{21}$$

Equ. 21 can be solved by a multidimensional Newton search iteratively

$$\mathbf{J}\vec{\delta u} = -\alpha \vec{u} - g \tag{22}$$

where $\mathbf{J} = \alpha \mathbf{I} + \frac{\partial \vec{g}}{\partial \vec{u}}$ is the Jacobian and $\mathbf{I}$ the identity matrix. With $\mathbf{W} = \frac{\partial^2 L}{\partial^2 \vec{F}}$, $\mathbf{M} = \mathbf{\Sigma V}^T\mathbf{WV\Sigma}$ and $\mathbf{T} = \mathbf{U}^T\vec{A}\mathbf{U}$ Equ. 22 reads

$$((\alpha + \mu)\mathbf{I} + \mathbf{MT})\vec{\delta u} = -\alpha \vec{u} - g \tag{23}$$

At each iteration step of the Newton search the step length $\vec{\delta u}$ muss be restricted for the stability of the algorithm. Therefor, a Levenberg-Marquardt parameter μ is added in Equ.23 to ensure stability. Bryan proposes $\vec{\delta u}^T T \vec{\delta u} \leq \sum m_i$ as a maximum step length for the algorithm. However, we found that sometimes even if this criterion is fulfilled the algorithm can be instable due to numerical overflow in Equ. 5. Therefor, we use the criterion $\parallel \vec{A} \parallel^2 \leq \sum_i m_i$. The Newton search can be made more efficient by diagonalizing Equ. 23. First we diagonalize $\mathbf{T}$:

$$\begin{aligned} \mathbf{TP} &= \mathbf{P\Gamma}, \\ \mathbf{\Gamma} &= \mathbf{diag}\{\gamma_i\} \end{aligned} \tag{24}$$

Then we define

$$\mathbf{B} = \mathbf{diag}\{\gamma_i^{\frac{1}{2}}\}\mathbf{P}T\mathbf{MP}\mathbf{diag}\{\gamma_i^{\frac{1}{2}}\} \tag{25}$$

and again solve the eigenvalue equation

$$\begin{aligned} \mathbf{BR} &= \mathbf{R\Lambda}, \\ \mathbf{\Lambda} &= \mathbf{diag}\{\lambda_i\} \end{aligned} \tag{26}$$

Please note that $\mathbf{P}$ and $\mathbf{R}$ are orthogonal matrices and γ_i and λ_i the eigenvalues of $\mathbf{T}$ and $\mathbf{B}$. Then to finally diagonalize Equ. 23 we define

$$\mathbf{Y} = \mathbf{P}\mathbf{diag}\{\gamma_i^{\frac{1}{2}}\}\mathbf{R} \tag{27}$$

With $\mathbf{Y}^{-T}\mathbf{Y}^{-1} = \mathbf{T}$ and $\mathbf{Y}^{-1}\mathbf{M}\mathbf{Y}^{-T} = \mathbf{\Lambda}$ Equ. 23 can be rewritten as

$$[(\alpha+\mu)\mathbf{I} + \mathbf{\Lambda}]\mathbf{Y}^{-1}\vec{\delta u} = \mathbf{Y}^{-1}[-\alpha\vec{u} - \vec{g}] \tag{28}$$

which leads to s independent equations for $\mathbf{Y}^{-1}\vec{\delta u}$. Now Equ. 23 can be rewritten to

$$(\alpha+\mu)\vec{\delta u} = -\alpha\vec{u} - g - \mathbf{M}\mathbf{Y}^{-T}\mathbf{Y}^{-1}\vec{\delta u} \tag{29}$$

So to finally we first solve Equ. 28 for $\mathbf{Y}^{-1}\vec{\delta u}$, use it in Equ. 29 to solve for $\vec{\delta u}$ and calculate the new value for $\vec{u}_{n+1} = \vec{u}_n + \vec{\delta u}$. The iteration is terminated if $\sum_i |\vec{u}_{n+1} - \vec{u}_n| \leq 10^{-10}$.

4 Results

In this section we investigate the performance of the Maximum entropy method by applying it to synthetic Greens function in order to recover the spectrum $A(\omega)$. We generate the Greens function data by first calculating the spectrum $A(\omega)$ and then using Equ. to calculate the Greens function given our spectrum and the kernel $K(\tau,\omega)$. After that we corrupt the Greens function by Gaussian noise as the "real" Greens function obtained by Quantum Monte Carlo methods always suffer from noise. In this evaluation we solely use the spectrum of the BCS superconductor for our investigation which can be calculated by

$$A(\omega) = \begin{cases} \frac{1}{W}\frac{|\omega|}{\sqrt{\omega^2-\Delta^2}} & \text{, if } \Delta < |\omega| < \frac{W}{2} \\ 0 & \text{, else} \end{cases} \tag{30}$$

where W denotes the bandwidth and 2Δ the gap magnitude. In Fig. we show an example of the spectrum and the resulting Greens function for $W = 0.9$ and $\Delta = 10$.

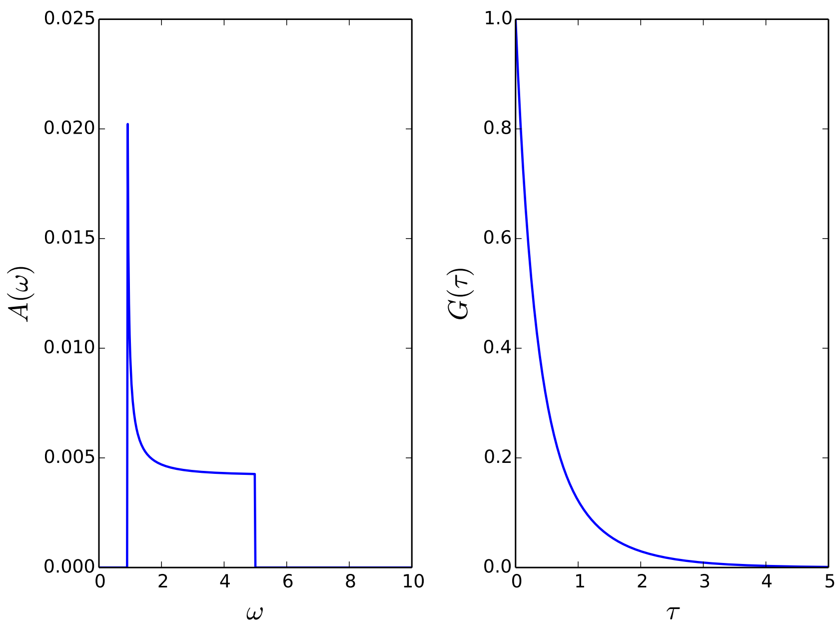

Figure 1: Example BCS spectrum $A(\omega)$ (left) and resulting Greens function $G(\tau)$ (right). The spectrum is calculated according to Equ. 30 with $\Delta = 0.9$ and $W = 10$.

The results obtained by Maximum entropy highly depend on the regularization parameter α. We demonstrate the influence of α by estimating the spectrum shown in Fig. 1 for three different values for $\alpha = 0.5, 2.5, 10$. The results are shown in Fig. 2

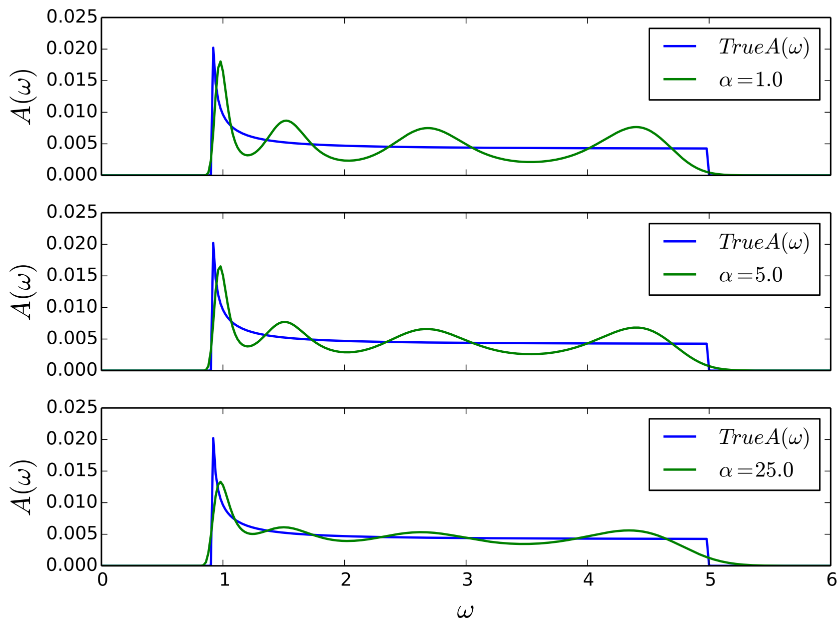

Figure 2: Influence of the regularization parameter α on the performance of the Maximum Entropy method. The spectrum is calculated according to Equ. 30 with $\Delta = 0.9$ and $W = 10$.

Another important parameter is the choice of the minimum singular value θ which determines the dimension of the singular space. We investigate the impact of θ on the performance of the Maximum Entropy method in Fig. 3

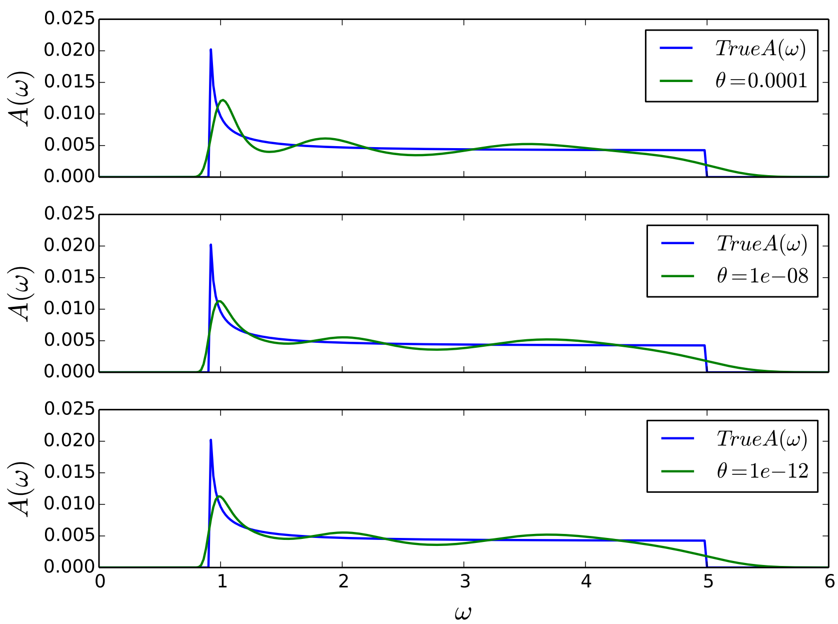

Figure 3: Influence of the minimum singular value θ on the performance of the Maximum Entropy method. The spectrum is calculated according to Equ. 30 with $\Delta = 0.9$ and $W = 10$.

Finally we show the difference between classic MEM and the Bryan method in choosing α. As discussed in Sec. !!Reference!! the classic MEM uses the maximum value of the probability of α given A and G while Bryan calculates the final $\hat{A}$ by $\hat{A} = \int A(\alpha) P_\alpha d\alpha$. In Fig. we show the probability p_α and the resulting spectra calculated by classic and Bryan's method

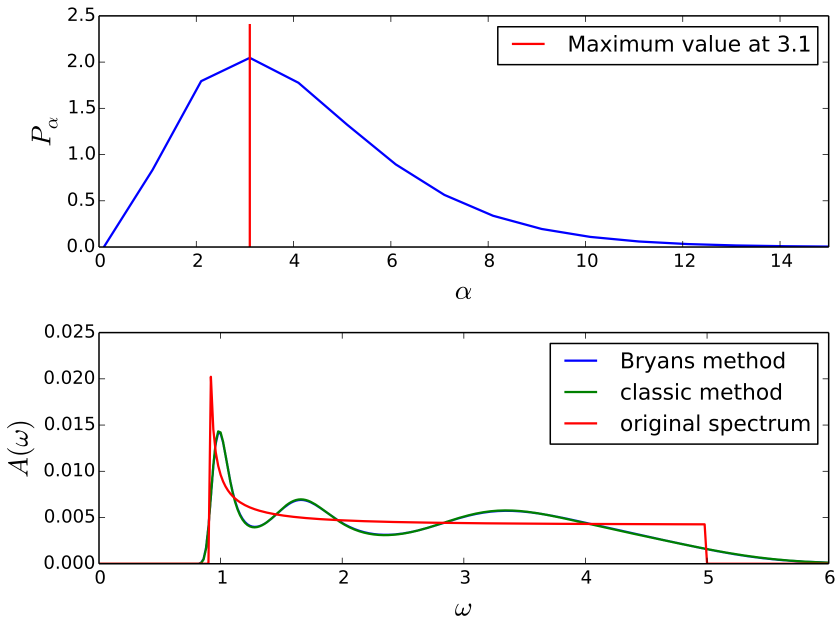

Figure 4: caption

5 Conclusions