Maximum Entropy Method of Analytic Continuation

Applications and Improvements

by
Ryan Levy

A thesis submitted in partial fulfillment
of the requirements for the honors degree of
Bachelors of Science in Physics
in the University of Michigan
2016

Advisor: Emanuel Gull

Abstract

In many condensed matter problems, calculations are done on the imaginary axis using a wick rotation. Moving from this basis back onto the real axis, known as analytic continuation, is an ill-posed problem. As several techniques utilize this method, such as Keldysh formalism and matsubara frequency calculations, moving to the real axis becomes ever more important. Using the principle of maximum entropy, an algorithm and code (`Maxent`) is developed to take steps toward a reliable continuation procedure. We present the background, pitfalls, and improvements made to this technique.

Acknowledgments

I first would like to thank Prof. Emanuel Gull, who provided me this opportunity to work on an interesting research project, in addition to his overall guidance. I also need to thank the Gull group members: James LeBlanc, Andrey Antipov, Alex Gaenko, Joe Paki, Xi Chen, Qiaoyuan Dong, Sergei Iskakov, and Hanna Terletska, for all of their support and knowledge. I express my extreme gratitude to James, who has answered numerous questions in the past two years, and without whom this work would have never been possible.

I would also like to thank Drs. Rachel Goldman, Simon Huang, and Markus Eisenbach who helped set me on the path toward this thesis. Dr. Goldman for providing me with my first opportunity to contribute to science and Dr. Huang for his patience and many hours introducing me to materials science. And finally Dr. Eisenbach who gave me an opportunity to perform research in both computational physics and high performance computing.

Contents

1 Introduction

Recently, a focus of condensed matter research has been the emergent properties of many-body systems. Interesting physics can arise from the many interactions within these systems, giving rise properties that the individual components (e.g. electrons) did not have. This includes both the strongly-correlated electron regime as well as non-equilibrium dynamics. To combat the complexity of interesting systems, much of the mathematical work is done in the complex plane where powerful tools are available. One tool, the use of the pole's of the fermi/bose distribution function, allows thermodynamic quantities to be calculated by extending energies onto the complex plane to be evaluated via contour integration or other complex space techniques. These poles are known as Matsubara frequencies, and have a robust formalism to perform integrals/sums[1].

A Wick rotation can also extend real time to imaginary time for a similar result. The time evolution operator, $\sim e^{iHt}$ when Wick rotated ($t \to i\tau$) becomes $\sim e^{-H\tau}$which is a decaying exponential rather than an oscillatory function. Imaginary time is then related to the Matsubara frequencies by Fourier transform, allowing the choice of basis to be one of mathematical convenience. Rarely, however, do these techniques yield well-behaved solutions on the real axis, which is the root of our struggle.

The goal of analytic continuation is to calculate the real valued function given its imaginary axis values. Analytically, this can be accomplished by taking a complex input ix (where x is a real number) and replacing it with $x \pm i\epsilon$ and taking the limit as $\epsilon \to 0$. This cannot be applied to discrete numerical data, such as data from a quantum Monte Carlo calculation. The purpose of this thesis is to apply analytic continuation to discrete data such as this, where there are no direct analytic methods.

In the the remainder of this chapter, we introduce the formalism and background to Green's functions and the problem of analytic continuation. Chapters 2 and 3 are devoted to the development of the maximum entropy (maxent) formalism and code (`Maxent`). Chapter 4 is an illustration of the physics that `Maxent` can probe, and in chapters 5, 6, and 7 are attempts at improvements through the default model, Green's function representation, and error bar calculation respectively.

1.1 Green's functions and Analytic Continuation

We begin by outlining the Green's function formalism in the context of analytic continuation. Given a Green's function G at inverse temperature $\beta = 1/T$ with momentum $\mathbf{k}$, given as a function of imaginary time τ, we define the Fourier transform between τ and Matsubara frequency $i\omega_n$ as

$$G(\mathbf{k}, i\omega_n) = \int_0^\beta e^{i\omega_n \tau} G(\mathbf{k}, \tau), \tag{1.1}$$

and the inverse transform

$$G(\mathbf{k}, \tau) = \frac{1}{\beta} \sum_{i\omega_n} e^{-i\omega_n \tau} G(\mathbf{k}, i\omega_n). \tag{1.2}$$

The Matsubara frequencies are $(2n+1)\pi/\beta$ and $2n\pi/\beta$ for fermionic and bosonic frequencies respectively. These definitions require that $G(\mathbf{k},\tau) < 0$ for all τ, as well as being wholly defined within $[0,\beta]$. All Green's functions are assumed to be normalized, which implies that their high frequency expansion goes as

$$G(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n} + \frac{c_2}{(i\omega_n)^2} + \dots \tag{1.3}$$

where the coefficient of $1/i\omega_n$ is fixed at 1. The equivalent definition for Green's functions in the time domain is that

$$-G(0) - G(\beta) = 1. \tag{1.4}$$

The momentum dependent spectral function $A(\mathbf{k},\omega)$ is defined by

$$A(\mathbf{k},\omega) = -\frac{1}{\pi}\mathrm{Im}\left[G(\mathbf{k},\omega)\right], \tag{1.5}$$

where the energy-dependent local density of states $D(\omega)$ can be found by summing over all momenta

$$D(\omega) = \sum_{\mathbf{k}} A(\mathbf{k},\omega). \tag{1.6}$$

Throughout the remainder of this document I will refer to the spectral function as $A(\omega)$, with an implicit $\mathbf{k}$.

To solve for the spectral function given G, we define a kernel such that

$$G(i\omega_n) = \int d\omega\, A(\omega)K(i\omega_n,\omega); \quad G(\tau) = \int d\omega\, A(\omega)K(\tau,\omega). \tag{1.7}$$

By formulating the problem of analytic continuation in this form, the algorithm is easily generalized for an arbitrary kernel. A list of several popular kernels appears in Tab. 1.1.

(a) Fermion Kernels

Kernel Name	Kernel
Frequency	$\frac{1}{i\omega_n - \omega}$
Time	$-\frac{e^{-\tau\omega}}{1+e^{-\omega\beta}}$

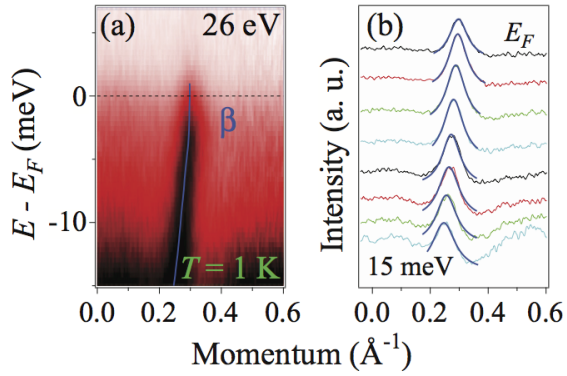
(b) Bosonic Kernels

Kernel Name	Kernel
Frequency	$\frac{1}{i\omega_n + \omega}$
Time	$\frac{1}{2}\omega\frac{\left[e^{-\omega\tau} + e^{-\omega(\beta-\tau)}\right]}{1-e^{-\omega\beta}}$

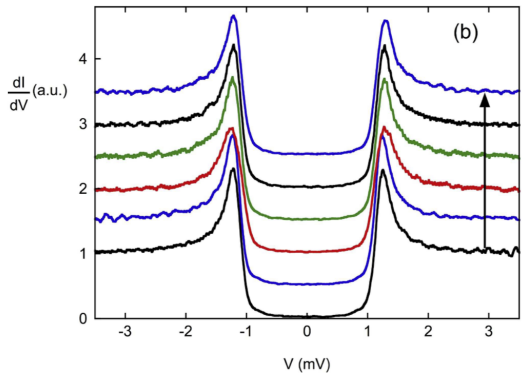
Table 1.1: Examples of Kernels

1.2 Measuring the Spectral Function

The spectral function can be measured directly and indirectly through experimental techniques. Using Angle-resolved photoemission spectroscopy (ARPES), the momentum dependent spectral function is measured directly[2]. An example of this is shown in Fig. 1.1(a). Indirectly, the density of states can be found using a scanning tunneling microscope (STM)[3]. The local density of states (which is the momentum averaged



(a) ARPES measurement for SmB_6[2]. On the left is raw data near the fermi energy E_F, while the right figure shows Lorentzian fits for different cuts.



(b) Density of states from a STM measurement of a Nb wire at 100 mK. The arrow indicates direction along the wire[3].

Figure 1.1: Experimental measurements of the spectral function

spectral function) can be found by observing dI/dV, the change in current per change in voltage, such as in Fig. 1.1(b). Both techniques provide a useful method for comparing theory to experiment.

1.3 Problems with Analytic Continuation

The formulation given in eq (1.7) is ill-posed, that is infinitely many solutions of $A(\omega)$, with major feature variations, will produce the proper $G(i\omega_n)$ within some error. This is illustrated in Fig. 1.2, where three distinct looking $A(\omega)$ and their $G(i\omega_n)$ are shown.

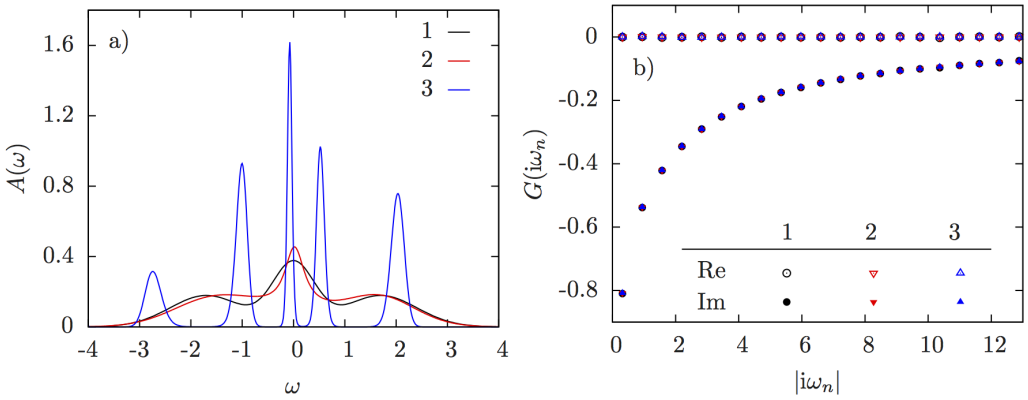


Figure 1.2: Three distinct spectra functions labeled 1,2,3 and their counterparts in imaginary frequency using the fermionic kernel at $\beta = 10$ (from [4])

Generally only a 1000 Matsubara frequency points are needed to encapsulate all the information of the Green's function; this gives a high density of information useful for dynamical calculations but prone to trouble when solving the problem of analytic continuation. For example, if we convert the integral of eq (1.7)

into a sum

$$\int d\omega\, K(i\omega_n, \omega) \Rightarrow \sum_{\omega} K(i\omega_n, \omega)\Delta\omega, \tag{1.8}$$

forming the matrix $\mathbf{K}$ of elements $K_{ni} = K(i\omega_n, \omega_i)\Delta\omega_i$ and vector $\mathbf{A}$ and $\mathbf{G}$ of elements $A(\omega_i)$ and $G(\omega_i)$ respectively to rewrite the problem as

$$\mathbf{G} = \mathbf{K}\mathbf{A}. \tag{1.9}$$

Matrix inversion on eq (1.9) will produce noisy, senseless data, an example of which is illustrated in Fig. 1.3.

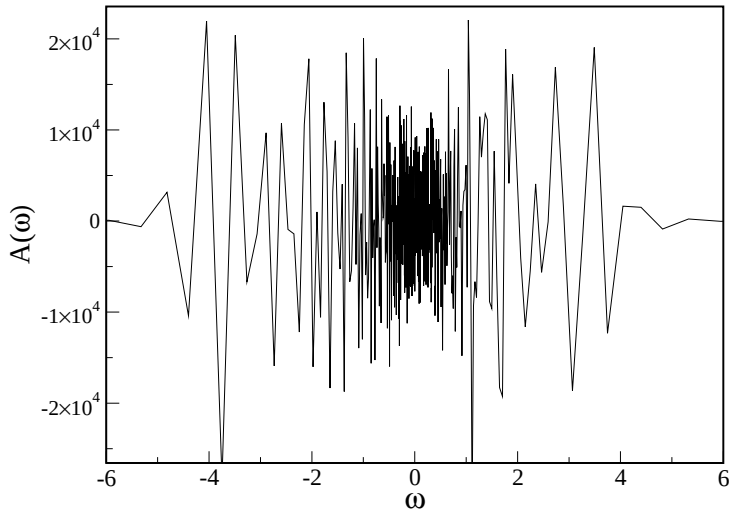


Figure 1.3: Example of matrix inversion of eq (1.9)

1.4 Quantifying a Good Solution

Because of the ill-posed nature of our problem, illustrated heavily above in sec. 1.3, a spectral function on the real axis is not enough to validate the solution. Only on the original imaginary time or frequency axis can a solution be properly vetted, or rather the spectral function is considered to be a possible solution. This process of moving a spectral function back to the imaginary axis is known as back-continuation. Once such a back-continued solution is found, one possible way to quantify the goodness of fit is χ^2, which is written as

$$\chi^2 = \sum_{n=0}^{N} \frac{(G(i\omega_n) - \int d\omega\, K(i\omega_n, \omega)A(\omega))^2}{\sigma_n^2}, \tag{1.10}$$

for N imaginary axis points; for a spectral function to back-continue within error bars one would expect to find $\chi^2 \simeq N$. In many cases, the systematic error from the continuation procedure is too large to obtain a good χ^2 value and thus only the overall features of the spectral function is relevant, not absolute measurements.

2 Maximum Entropy and Bayesian Inference

First, we motivate the maximum entropy method (maxent) mathematically, and then proceed to further motivate it by providing a physical picture by analogy to the free energy.

2.1 Bayesian Statistics

At this point, the spectral function has not been assumed to be anything more than a solution to eq (1.7), but to utilize Bayesian inference the following rules are exploited,

$$\begin{aligned} A(\omega) \geq 0 \qquad & \text{fermions}, \\ \text{sign}(\omega)A(\omega) \geq 0 \qquad & \text{bosons}, \\ \int_{-\infty}^{\infty} d\omega\, A(\omega) < \infty \qquad & \text{both}, \end{aligned} \tag{2.1}$$

which come about through physical properties of the system [5]. The last rule gives the freedom to renormalize $A(\omega)$ to be 1, and restore the true normalization after the maxent procedure. From these rules, we can take $A(\omega)$ (or $\text{sign}(\omega)A(\omega)$ for the bosonic case) as a probability density function.

Bayes's theorem says given two events a and b**,** the joint probability of a and b $\Pr[a, b]$ is given by

$$\Pr[a, b] = \Pr[a|b]\Pr[b] = \Pr[b|a]\Pr[a] \tag{2.2}$$

where $\Pr[a]$ is the probability of event a and $\Pr[a|b]$ is the conditional probability of a given b. These probabilities must be normalized such that

$$\begin{aligned} \int da \Pr[a] &= 1, \\ \int da \Pr[a|b] &= 1. \end{aligned} \tag{2.3}$$

The goal is to now use Bayes's theorem on the imaginary axis data. In general, our input G (taken to be in a frequency representation) is measured through some procedure (e.g. Monte Carlo) and represents an approximation of the true value within an error bar. Thus the input can be viewed as an event which is the measurement of some exact $G(i\omega_n)$, where the measurement (our input) is denoted as $\tilde{G}(i\omega_n)$. Using this viewpoint, the following definitions can be applied in Bayesian language,

$$\begin{aligned} \Pr[A|\tilde{G}] \qquad & \text{posterior probability}, \\ \Pr[\tilde{G}|A] \qquad & \text{likelihood function}, \\ \Pr[A] \qquad & \text{prior probability}, \\ \Pr[\tilde{G}] \qquad & \text{the evidence}. \end{aligned}$$

The question now becomes, given the evidence $\Pr[\tilde{G}]$, what is the posterior probability of the spectral function; this produces a relationship of

$$\Pr[A|\tilde{G}] = \Pr[\tilde{G}|A]\Pr[A]/\Pr[\tilde{G}]. \tag{2.4}$$

Note that $\Pr[\tilde{G}]$ becomes an overall scaling factor, which I will ignore[5].

2.1.1 Likelihood Function

The logarithm of the likelihood function L should behave such that $\Pr[\tilde{G}|A] \propto e^{-L}$ so that as the spectrum A becomes more likely, L is minimized and our probability is quickly maximized. Under a rigorous analysis[6], the central limit theorem gives that the asymptotic behavior of the likelihood function is

$$e^{-L} = e^{-\chi^2/2}, \tag{2.5}$$

where χ^2 is the familiar description of goodness of fit, explicitly

$$\begin{aligned} \chi^2 &= \sum_{n,m}^{N} (\bar{G}_n - G(i\omega_n))^* C_{nm}^{-1} (\bar{G}_m - G(i\omega_m)) \\ &= \sum_{m}^{N} \frac{\left(\int d\omega\, K(i\omega_m, \omega) A(\omega) - G(i\omega_m)\right)^2}{\sigma_m^2}, \end{aligned}$$

where $\bar{G}_m = \int d\omega\, K(i\omega_m, \omega) A(\omega)$ and the simplification to the second line is from a diagonal covariance matrix C.

While χ^2 is a useful object to work with, overfitting ($\chi^2 \simeq 0$) is a problem that plagues the most simplistic least-squares fitting routine. In order to overcome this, Bayesian updating, and prior probability methods are employed.

2.1.2 Entropy and Prior Probability

For a problem where an infinite number of solutions exist, one may try to regularize the solution using a specific function. By finding the extrema of the regularizing function (generally a maximum), a specific solution can be obtained systematically. While there are numerous sensible regularizing functions, one particularly useful one that provides few unjustified correlations between elements of A is Shannon entropy $S = \int d\omega\, p(\omega) \ln[p(\omega)]$ for probability distribution p. As argued in sec. 2.1, $A(\omega)$ can be thought of as a probability distribution and suitable to replace p. In order to provide a fixed reference point to the entropy, a default model $d(\omega)$ is introduced such that

$$\begin{aligned} S &= \int d\omega \left[A(\omega) - d(\omega) - A(\omega) \ln \left[\frac{A(\omega)}{d(\omega)} \right] \right] \\ &= -\int d\omega\, A(\omega) \ln \left[\frac{A(\omega)}{d(\omega)} \right] \end{aligned}$$

where the constant terms vanish if A and d have the same normalization. The default model, useful in that the specifics of which ω values to choose for A has been eliminated, can be another degree of freedom to select when performing calculations. This subject is extensively covered later in this thesis.

Finally, the central limit theorem can be used once again to show

$$\Pr[A] \propto e^{\alpha S}. \tag{2.6}$$

The addition of some constant α is used as a regulator for the scale of the entropy, compared to the χ^2 term. Returning to 2.4, it can be written as

$$\begin{aligned} \Pr[A|\tilde{G}] &= \frac{e^{-\chi^2/2}e^{\alpha S}}{\Pr[\tilde{G}]} \\ &= \frac{e^{-Q}}{\Pr[\tilde{G}]}. \end{aligned}$$

We define the objective

$$Q = \frac{1}{2}\chi^2 - \alpha S[A], \tag{2.7}$$

which is minimized for the true spectral function.

Now with inclusion of the α parameter, the Bayesian updating statement can be rewritten to explicitly show its dependence. First, eq (2.4) is rewritten as

$$\begin{aligned} \Pr[A, \alpha|\tilde{G}] &= \Pr[\tilde{G}|A, \alpha]\Pr[A, \alpha]/\Pr[\tilde{G}] \\ &= \Pr[\tilde{G}|A]\Pr[A|\alpha]\Pr[\alpha]/\Pr[\tilde{G}], \end{aligned}$$

which using normalization and other techniques can be written as

$$\Pr[A, \alpha|\tilde{G}] = \Pr[\alpha]\frac{e^{-Q}}{\Pr[\tilde{G}]}, \tag{2.8}$$

where $\Pr[\alpha]$ can either be taken as $1/\alpha$ (Jeffreys prior) or a constant.

2.2 Physical Motivation

The motivation for the objective Q in eq (2.7) can be supplemented with physical arguments. In statistical mechanics, at equilibrium the free energy F will be minimized. At equilibrium, the internal energy U is minimized or the entropy S is maximized. Since $F = U - TS$ there is interplay between U and S[7]. Q can therefore be thought of as a quasi-free-energy where the parameter α, much like T, contributes to the interplay between U and S, where U is represented by χ^2.

2.3 Historic, Classic, and Bryan's Method of Maxent

With eq (2.7) determined, it can be used to find $A(\omega)$ for a given α value, denoted A_α, which minimizes the objective. Unfortunately α provides another degree of freedom and method to determine a particular solution, and there are 3 approaches to maxent which attempt to do this. The first, *historic maxent*, completely disregards the Bayesian statistics and simply seeks a solution that provides a good χ^2. Due to the ill-posed nature of analytic continuation, this does not produce good results simply because there are far too many spectral functions that produce good χ^2 values.

An improvement to historic maxent is then to find the most probable α value, known as *classic maxent*. A range of α values are tested, then using eq (2.8) the most probable spectral function is selected. In a paper by Bryan [8], the classic maxent technique was improved. Bryan uses the probability associated with each $A_\alpha$to create a weighted average spectral function $\tilde{A}$

$$\tilde{A}(\omega) \equiv \int d\alpha \, A_\alpha(\omega) \mathrm{Pr}[\alpha|\tilde{G}]. \tag{2.9}$$

The probability $\mathrm{Pr}[\alpha|\tilde{G}]$ can be found using,

$$\mathrm{Pr}[\alpha|\tilde{G}] = \mathrm{Pr}[\alpha] \int \mathcal{D}A \, \frac{e^{-Q}}{\mathrm{Pr}[\tilde{G}]}, \tag{2.10}$$

which is an integral over A. If a fitting routine finds features that are similar for two close Q values, the features can be averaged into the final spectral function in a systematic way, while suppressing problematic features for a given α.

Generally, both classic maxent and Bryan's method produce nearly identical results. On occasion, a known feature may be averaged out in Bryan's method, or classic maxent produces too sharp a spectral function, in which case selecting one method over the other is easier.

Another important remark is that the normalization of these probabilities is generally not known. Because we only know the measurement of G (denoted $\tilde{G}$), the exact values of the probability have no context. They are still useful in comparison only to the same measurement input $\tilde{G}$ and are a helpful tool in determining the success of parameter choices, e.g. α values.

3 Implementation Details

3.1 Development of the `Maxent` code

The primary result of this thesis is the development and testing of a code, `Maxent`, designed to integrate into the ALPSCore library[9]. It was originally partially developed[4] using ALPS[10], then rewritten to use ALPSCore. Along with the basic rewrite, there was a focus on user friendly design, (user) error checking, unit tests, and additional functionality. The added functionality is the topic of chapters 5, 6 and 7. `Maxent` is written in C++ and in addition to ALSPCore use the boost library and Eigen3 [11].

3.2 Implementation

Shown in Algorithm 3.1 is a breakdown of the algorithm employed. The second step, a decomposition of the kernel to singular space, was originally formulated by Bryan[8]. The Levenberg-Marquardt routine is a powerful non-linear fit routine which moves between a gradient descent and a Gauss-Newton method[12].

Algorithm 3.1 `Maxent` Analytic Continuation Procedure

1. Read in G data and set up default model, kernel, and ω grid
2. Decompose kernel using SVD
   a) Throw out eigenvalues smaller than precision (known as singular space)
3. Given a range of α, use a Levenberg-Marquardt fitting routine to find A_α
   a) As $\alpha \to 0$ there is a least-squares fit, while $\alpha \to \infty$ produces the default model
4. Compute the most-probable A_α (classic) and perform a weighed-average over all A_α (Bryan)

The new version of `Maxent` is now hosted on a Github repository which includes issue tracking and an installation guide. Examples and how to use and interpret the program are also now online, the data for this partially comprises chapter 4. Other examples include how to analytically continue the self-energy (results shown in Fig. 5.1) and bosonic Green's functions.

In addition to the `Maxent` codebase, a set of utilities are incorporated. Included is a Kramers-Kronig relation to determine the Green's function real part from the imaginary part, and vice versa and a program to convert Green's function in the time domain to the Legendre basis.

4 Application - DMFT

Of particular interest are materials that at high temperature act as metals but when cooled become insulators[13]. This transition from metal to Mott insulator is known as the Mott transition. Mott insulators are materials that under traditional band theory calculations should be metals, but in actuality are insulators. Mott understood that this behavior is due to neglecting electron-electron interactions, as well as treating electrons as Bloch waves, delocalized waves in momentum space[14].

`Maxent` is a tool that, at its most basic level should be robust enough to capture a Mott transition. The combination of generating datasets with DMFT and analytically continuing them with `Maxent` provides a useful toolkit to physicists.

4.1 The Hubbard Model

A simple and popular lattice model to study is the Hubbard model; Fig. 4.1 shows a 2D illustration. This model describes a lattice of sites; each site contains a single orbital "atom" which can have electron eigenstates $|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle$, corresponding to energies $0, \epsilon_0$, and $U+2\epsilon_0$. Here I've used ϵ_0 as the single band energy and U the interaction energy between electrons[1]. The Hamiltonian that describe this is given as

$$H = -\sum_{\langle ij\rangle,\sigma} t_{ij}\left(c^{\dagger}_{i\sigma}c_{j\sigma} + h.c.\right) + U\sum_{i} n_{i\uparrow}n_{i\downarrow}, \tag{4.1}$$

where t_{ij} is the hopping term, describing electrons of spin σ between sites i and j. Note that $n_{i\sigma}$ is the standard density operator $n_{i\sigma} = c^{\dagger}_{i\sigma}c_{i\sigma}$.

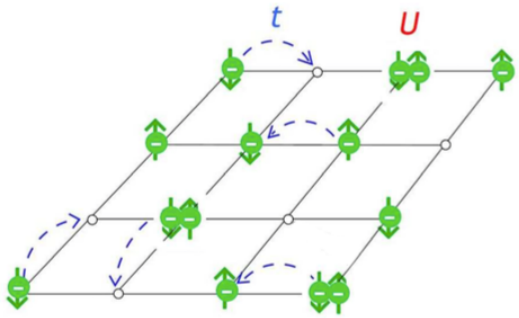


Figure 4.1: Illustration of the Hubbard Model in 2D[15]

The Hubbard model has the known property that for different values of U at half-filling, the density of states (DOS) will exhibit a Mott insulator transition[14]. This transition would show the finite DOS when in the metallic phase and above a threshold U, an insulating DOS with zero states at the fermi level. This will be illustrated in sec. 4.3.

4.2 DMFT

Dynamical Mean Field Theory is one method for calculating a local Green's function for the Hubbard model. It extends the usual mean field theory (MFT) by including a dynamical field rather than an effective static

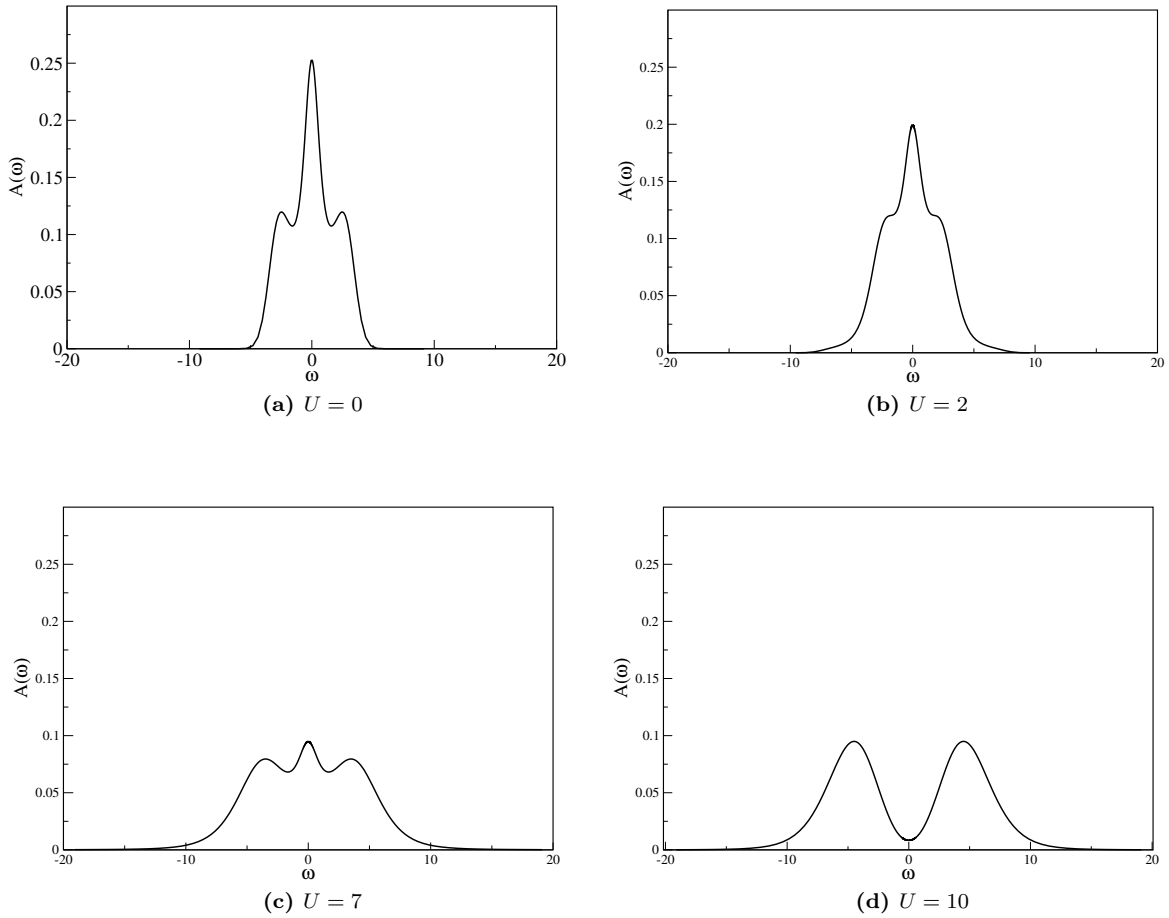
[1] In the interest of simplicity we will use only single orbital atoms, but multi-orbital extensions are possible.

one. In the non-interacting and atomic limit, DMFT is also found to give the exact solution, as well as for a the problem in infinite dimensions [14].

The implementation of DMFT is used with permission from E. Gull, et al[16].

4.3 Maxent and the Mott Transition

By looking at the Hubbard model in 2D with a single site, there is no momentum dependence on the spectral function, and thus the spectral function is identically the DOS. Using a $\beta = 2$, $t_{ij} = 1$ Hubbard model, Fig. 4.2(a→d) show the spectral function $A(\omega)$ for various U obtained from `Maxent`.



(a) $U = 0$ **(b)** $U = 2$

(c) $U = 7$ **(d)** $U = 10$

Figure 4.2: Spectral functions from `Maxent` (Bryan's method) using DMFT data at $\beta/t = 2$

5 Improvements: Choice of Default Model

In general, the choice of default model should not affect the results of the maximum entropy method. A good result from this procedure, is characterized by no change in the results given different sensible default models. Sensible models are continuous functions where there is a lack of un-physical features, e.g. data at half-filling is symmetric about the fermi level.

This idea is unfortunately tied to the quality of the non-linear fitting routine and the ability to find the absolute minima. For a large singular space or small errors, the fitting routine may routinely hit local minima instead, the results of such an issue shown in Fig. 5.1. `Maxent`'s fitting procedure is heavily biased on a "good" initial guess for the spectral function, which is taken to be the default model. Thus the default model becomes a way to assist the fitting routine into moving past these local minima.

5.1 Perturbation Theory

Consider the interacting Hubbard model at half-filling

$$H = -t \sum_{<ij>,\sigma} \left(c^{\dagger}_{i\sigma} c_{j\sigma} + h.c. \right) + U \sum_{i} n_{i\uparrow} n_{j\downarrow}. \tag{5.1}$$

The Green's function solution to this Hamiltonian is

$$G(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n - \epsilon_k - \Sigma(\mathbf{k}, i\omega_n)}, \tag{5.2}$$

where ϵ_k is the tight-binding dispersion and Σ the self-energy. Analytically this is continued to the real axis by taking $i\omega_n \to \omega + i\Gamma$ with $\Gamma \ll 1$.

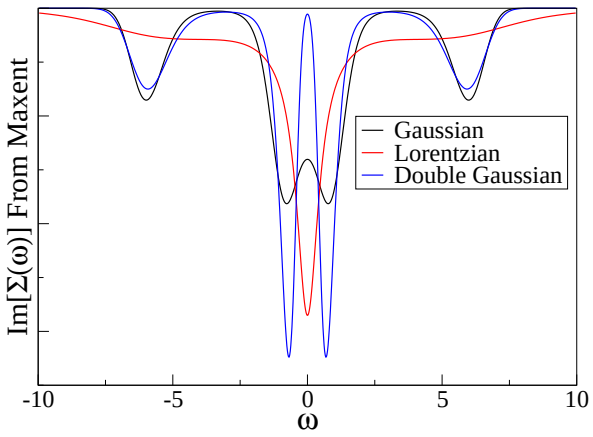


Figure 5.1: A false local minima for a Lorentzian default model, in the continuation of the $\beta = 2$, $U = 10$ DMFT-produced self-energy

Using second order perturbation theory, we can expand the self-energy using the Schwinger-Dyson equation of motion[17]

$$\Sigma(\mathbf{k}, \nu) = -\frac{U^2}{\beta^2} \sum_{\mathbf{k'q}} \sum_{i\nu' i\omega} G(\nu', \mathbf{k'}) G(\nu' + \omega, \mathbf{k'} + \mathbf{q}) G(\nu + \omega, \mathbf{k} + \mathbf{q}), \tag{5.3}$$

where ω is a bosonic frequency, ν, ν' are fermionic frequencies, $\mathbf{k'}, \mathbf{k}, \mathbf{q}$ are momenta values within the first Brillouin zone, and the Hartree term has been neglected. The $i\nu'$ and $i\omega$ frequency sum can be completed

analytically to produce the form

$$\Sigma(\mathbf{k},\nu) = U^2 \sum_{\mathbf{k'q}} \frac{n_F(\epsilon_{\mathbf{k'+q}})\left[1 - n_F(\epsilon_{\mathbf{k'}})\right] + n_F(\epsilon_{\mathbf{k+q}})\left[n_F(\epsilon_{\mathbf{k'}}) - n_F(\epsilon_{\mathbf{k'+q}})\right]}{i\nu + \epsilon_{\mathbf{k'+q}} - \epsilon_{\mathbf{k'}} - \epsilon_{\mathbf{k+q}}}, \tag{5.4}$$

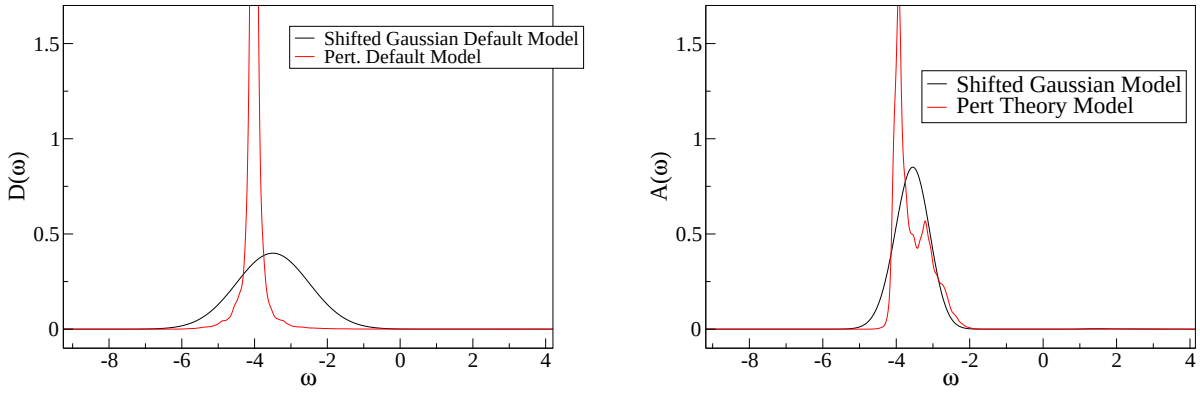
where $n_F(\epsilon_{\mathbf{k}}) = \dfrac{1}{\exp\left[\epsilon_{\mathbf{k}}\beta\right] + 1}$ is the fermi function at inverse temperature β. The integral over k' and q was completed numerically with a Monte-Carlo integration method[18].

5.2 $U = 1$ Test Case

Using the results of eq (5.4) the solution can be written as

$$G(k, \omega + i\Gamma) = \frac{1}{\omega + i\Gamma - \epsilon_k - \Sigma(k,\omega)}. \tag{5.5}$$

Using continuous time DMFT[16] at the $\mathbf{k} = (0,0)$ point for $U = 1, \beta = 2$ on a Betts 2D lattice[19] with 16 sites, the Matsubara frequency Green's function can be determined. The DMFT results were then continued using `Maxent`, with a shifted Gaussian default model (width of $\sigma = 1$) and a separate output generated using a perturbation theory default model at $\Gamma = 0.01$ shown in Fig. 5.2 (a) and (b). While the shifted Gaussian model produces a nice smooth peaked function, the perturbation theory model output is simply the input, with added noise.



(a) Input Default Models

(b) `Maxent` results (Bryan's method)

Figure 5.2: Test case figures

Despite the significant differences in shape, both examples appear to be similar on the Matsubara axis in Fig. 5.3. In fact, the perturbation default model was as good as the gaussian, with only 9% of the Matsubara points producing a smaller error than the Gaussian model. The sharp features of the perturbation theory makes this method a less appealing approach due to the risk of significant noise.

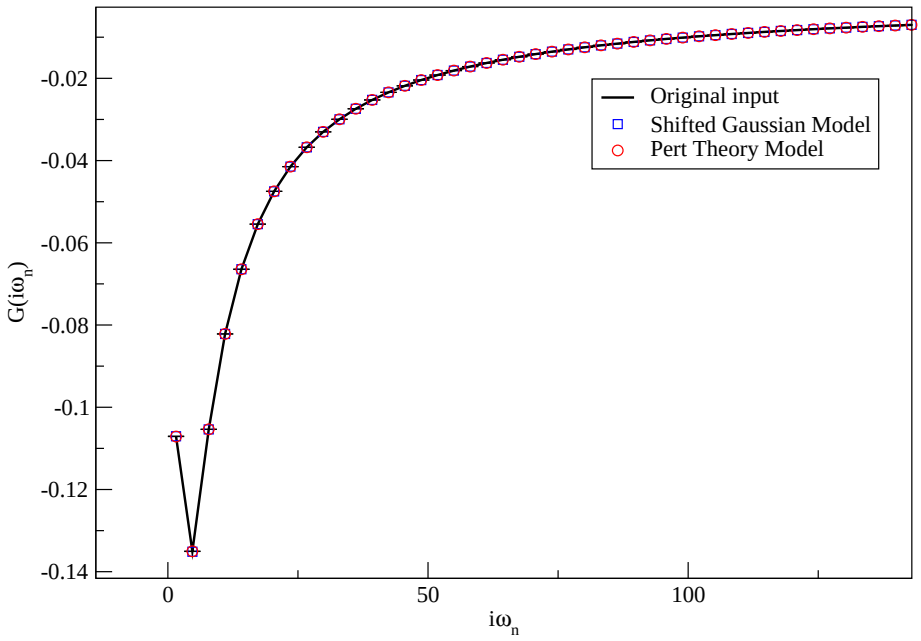


Figure 5.3: Results from Figure Fig. 5.2 back-continued to the Matsubara axis

6 Legendre Representation

6.1 Summary

Representing a Green's function in a different basis, specifically the Legendre basis, has two major benefits over the time or frequency representation. Because of the density of information, the storage size for a Green's function is significantly reduced. This is advantageous when storing large or multiple Green's function objects. The other benefit is that the kernel is at most on the order of about $40 \times N_\omega$. For a small amount of ω grid points N_ω, `Maxent` is considerably faster compared to other data representations. For each momentum point that G has, one must perform a separate run of `Maxent`. To increase resolution in ω and **k** while doing `Maxent` calculations in reasonable time, the Legendre representation becomes preferable.

6.2 Mathematical Background

Recall the generalized expression for analytic continuation

$$\mathbf{G} = \int \mathbf{K}(\omega)\mathbf{A}(\omega)d\omega \tag{6.1}$$

which for the case of imaginary-time representation

$$G(\tau) = \int -\frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}} A(\omega)d\omega \tag{6.2}$$

where $\pm$ corresponds to fermions/bosons.

As formulated by Boehnke, et al. [20], a representation of the same Green's function in the basis of Legendre polynomials given by the transforms

$$G(\tau) = \sum_\ell \frac{\sqrt{2\ell+1}}{\beta} P_\ell\left[x(\tau)\right] G_\ell \tag{6.3}$$

$$G_\ell = \sqrt{2\ell+1} \int_0^\beta d\tau\, P_\ell\left[x(\tau)\right] G(\tau) \tag{6.4}$$

where P_ℓ are the Legendre polynomials and $x(\tau) = 2\tau/\beta - 1 \in [-1, 1]$. Using the orthogonality of the Legendre polynomials and eq (6.4), eq (6.2) can become

$$\begin{aligned} G_\ell &= -\int_{-\infty}^{\infty} d\omega A(\omega) \int_0^\beta d\tau \sqrt{2\ell+1} \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}} P_\ell(x(\tau)) \\ &= -\int_{-\infty}^{\infty} d\omega A(\omega) K(\ell, \omega) \end{aligned}$$

with the *Legendre Kernel* defined as

$$K(\ell,\omega) \equiv \sqrt{2\ell+1}\int_{-1}^{1} dx\, \frac{e^{-(1+x)\beta\omega/2}}{1 \pm e^{-\beta\omega}} P_\ell(x). \tag{6.5}$$

The Legendre Green's functions have an even higher density of information than the standard Matsubara formalism, where Green's functions sampled with 1000s of points can be represented by approximately 40 G_ℓ values.

Currently this technique is restricted to τ space, due to the infinite nature of the Matsubara frequencies. However, it is relatively easy to move $\tau \to \ell \to i\omega_n$ or $i\omega_n \to \tau \to \ell$ if enough information of the system (high frequency tails) is known.

6.3 Conversion from τ to Legendre

When converting $G(\tau)$ to G_ℓ, one selects a maximum ℓ value, denoted ℓ_{max}. As ℓ_{max} passes some threshold, the Legendre polynomial are highly oscillatory so that the added information is only noise. This introduces systematic error into the data, which is carefully accounted for by choice of ℓ_{max}. This is illustrated in Fig. 6.1. Here, selecting $\ell_{max} \in [10, 40]$ is acceptable to contain all the information needed.

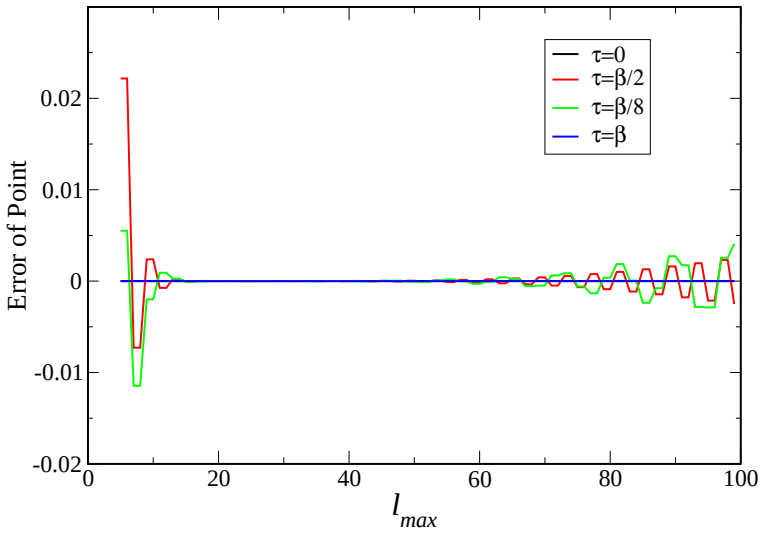
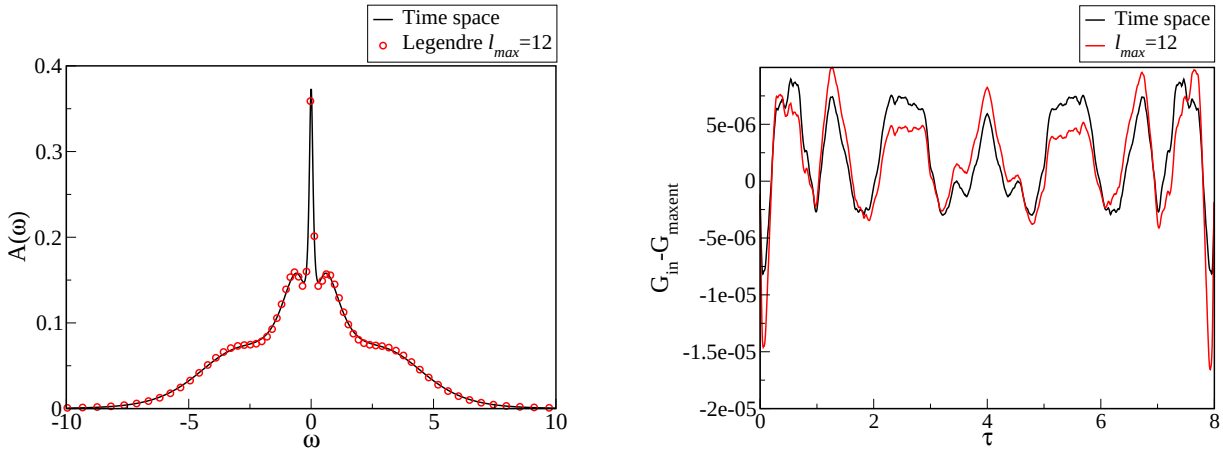


Figure 6.1: Convergence of ℓ_{max} cutoff for an interacting ($U = 4$) system at $\beta = 8$

Using a bootstrapping technique, the error bars from both the Legendre conversation process ($\tau \to \ell$) and back ($\ell \to \tau$) can be determined. As long as the majority of the back continued points are within error bars, the conversation is considered successful. High frequency information can be embedded if known beforehand[20], to assist in containing the endpoints of $G(\tau)$.

6.4 Maxent Output

As a test case, the DMFT results from a 2D Hubbard model, with $U = 4$, $\beta = 8, t = 1$, is used. The expected maxent result is easily obtained, and as shown in Fig. 6.2(a). To compute the comparison of the error for each of the different techniques, both spectral functions were continued back to the imaginary axis, with the difference of the input data shown in Fig. 6.2(b). Both have the same error on the order of 10^{-6}, however the Legendre data has a less accurate representation of the high frequency information (which for $G(\tau)$ are contained in the endpoints, see eq (1.4)). This data was calculated at a significantly faster runtime compared to the time space Green's function, shown in Fig. 6.3.



(a) `Maxent` Comparison **(b)** Error of Analytic Continuation

Figure 6.2: Legendre Representation

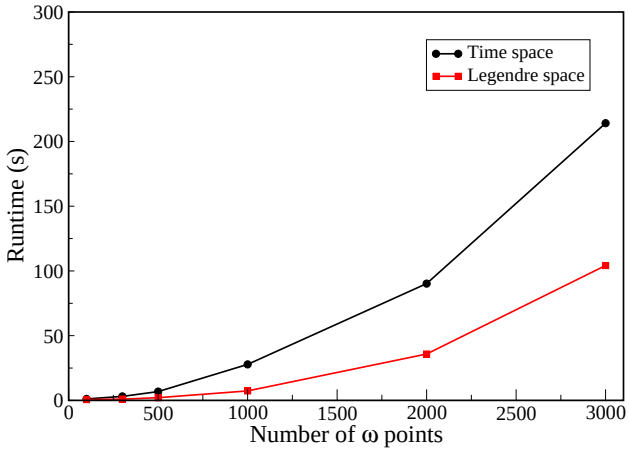


Figure 6.3: Runtime of `Maxent` varying the number of ω points

7 Producing Error Bars from Maxent

7.1 Bootstrap method

To calculate the error of a non-linear process $f(A)$, one can employ a bootstrap routine. This routine takes an input data set A and generates gaussian noise about each point with width of the point's error. Repeating this N times generates N curves A' which are then put through the process $f(A')$; the standard deviation of the N output sets are then calculated, which represents the error of the process f.

7.2 Bootstrap and Maxent

Jarrell and Gubernatis in their analysis of the maximum entropy method[5], defined for a given most probable α and corresponding A_α, the positive-definite matrix $\Gamma_{ij} = \alpha\delta_{ij} + \Lambda_{ij}$ where

$$\begin{aligned} \Lambda_{ij} &= \sqrt{A_i}\frac{\partial^2 L}{\partial A_i \partial A_j}\sqrt{A_j}\Big|_{A=A_\alpha} \\ &= \sqrt{A_i}[K^T C^{-1} K]_{ij}\sqrt{A_j}\Big|_{A=A_\alpha} \end{aligned}$$

where C is the covariance matrix on the input data and K the kernel. After moving into a convenient basis where the curvature of the entropy is constant and moving back, one can write that the covariance X of the spectral function output is given as

$$X_{ij} = \langle \delta A_i \delta A_j \rangle \approx \sqrt{A_i}\,[\Gamma^{-1}]_{ij}\sqrt{A_j}. \tag{7.1}$$

Note that this is only a second order approximation to the covariance matrix, and thus will fail for certain datasets.

In order to use the bootstrap routine to determine the error from the covariance matrix, the covariance must be diagonal. Thus we can determine a basis where the covariance is diagonal, bootstrap in this basis, and transform back. The transformation matrices are given by solutions to the eigenvalue problem,

$$\begin{aligned} \Gamma &= U^{-1}DU = U^T DU, \\ \implies \Gamma^{-1} &= UD^{-1}U^T, \end{aligned}$$

where D is a diagonal matrix of eigenvalues and U is a matrix whose columns are eigenvectors of Γ. Because Γ is positive-definite, we can use the fact that U is orthogonal. Thus we have a diagonal covariance X_D which can be written

$$[X_D]_{ij} = \left(\sqrt{A_i}U_{ij}\right) D_{ij}^{-1} \left(U_{ij}^T\sqrt{A_j}\right). \tag{7.2}$$

We can perform a bootstrap in the diagonal basis, then transform back using U. Note that this is using the

metric of $\sqrt{A}$, so after bootstrapping we must square our vector to get out the original A. This overcomes the fact that $\sqrt{A}$ during bootstrap may become negative.

7.3 Calculation of Error Bars

Algorithm 7.1 Bootstrap Routine

1. Generate U, D^{-1}
2. Transform $\sqrt{A}$ to $\left\{\sqrt{A}\right\}_u = U\sqrt{A}$
3. Add Gaussian noise of width $\sigma_i = D_i^{-1}$ (bootstrap)

   $\left\{\sqrt{A}\right\}_u \rightarrow \left\{\sqrt{A_{noise}}\right\}_u = \left\{\sqrt{A}\right\}_u + (\text{gaussian noise of width } \sigma)$
4. Move $\sqrt{A_u}$ to non-diagonal basis

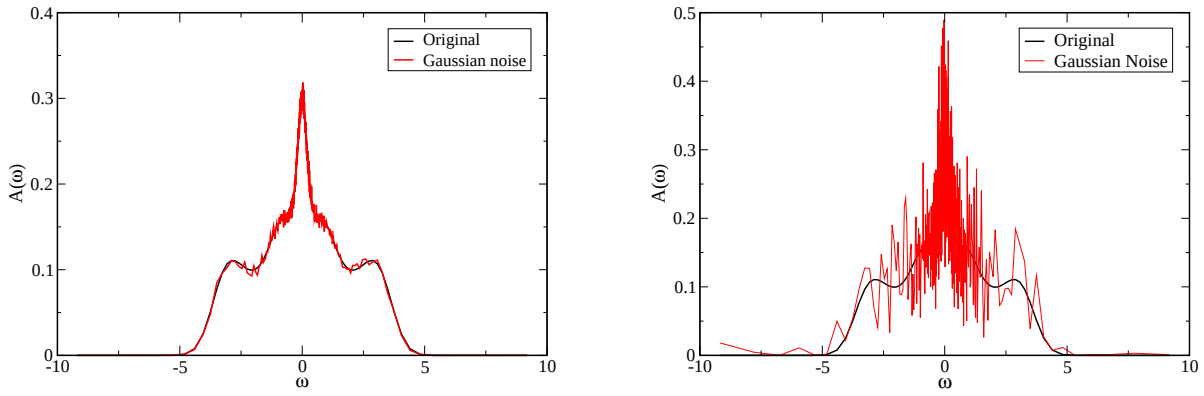
   $\left\{\sqrt{A_{noise}}\right\}_u \rightarrow \sqrt{A_{noise}} = U^T \left\{\sqrt{A_{noise}}\right\}_u$
5. Repeat 3,4 N times
6. Using N different $\sqrt{A_{noise}}$, determine mean and stddev
7. Plot square of mean value with error bar of x

   $x_i = 2A_i(\text{stddev})_i$

The algorithm for the calculation of the error bars is explicitly shown in Algorithm 7.1.

Using the $U = 0$, $\beta = 8$, the Green's function $G(\tau)$ with constant errors of 10^{-5} as example data, the Γ matrix can be calculated. First, the contribution of the off diagonal elements of the covariance matrix must be understood. By taking Gaussian noise of width σ, the eigenvalues of Γ^{-1} can be understood.

Shown in Fig. 7.1(a) and (b) are the results of of Algorithm 7.1 with a constant gaussian width σ. As σ approaches the order of the spectral function, significant noise will be apparent. For that case, the bootstrap routine will require a large N in order to capture the propagation of error through the transformation.



(a) $\sigma = 0.01$ **(b)** $\sigma = 0.1$

Figure 7.1: Gaussian width of constant σ transformed (bootstrap with $N = 1$)

A full bootstrap with real D^{-1} values is shown in Fig. 7.2. The plot shows the original spectrum with noise around important features near $\omega = 0$, which will contribute to higher error bars in that region.

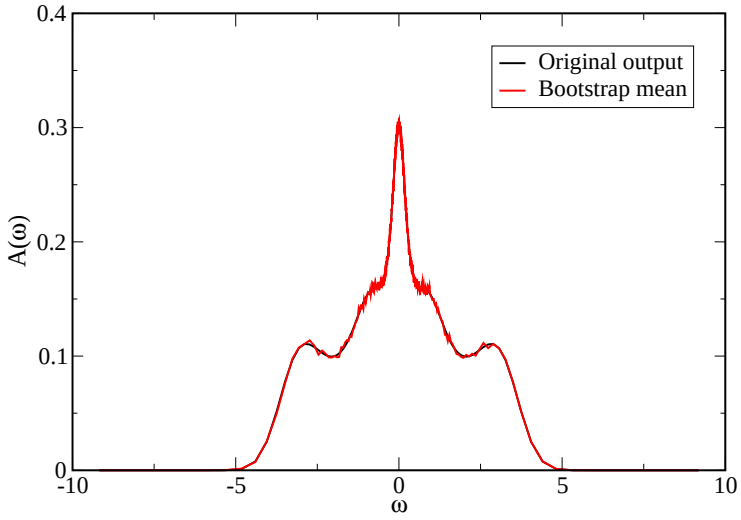


Figure 7.2: Mean value after a bootstrap using $N = 10000$ and $\sigma_i = D_i^{-1}$

If we take the standard deviation of the bootstrap routine as an error bar we obtain a plot that has almost constant error bars of ~ 0.35, shown in Fig. 7.3. This plot is what we expected. The high frequency limit of the spectral function, something that is well known, has less uncertainty than the features closer to $\omega = 0$. In other words, the precise height of the peak is unknown but is a sensible solution (i.e. a solution that obeys the sum rules). Still, the χ^2 value is about 9.8, so that this spectral function is still one solution to the analytic continuation of the data.

7.4 Default Model Variance

Another source of error comes from the fitting routine itself and the dependence on the default model. This should give some semblance to the systematic error in the fitting routine, however it most likely underestimates the error. Another concern is that the `Maxent` routine appears to be $O(n^2)$ where n is the number of real frequencies, and so when one wants fine grids, running multiple default models can take a significant amount of time.

An example calculation of the default model dependence is shown in Fig. 7.4(a→c). Default models, shown in Fig. 7.4(a) were input to `Maxent`, which produced the Bayesian averaged spectral functions in Fig. 7.4(b). There is only subtle variation between the three outputs, unlike the covariance approximation. Error bars in Fig. 7.4(c) represent the variation between these default models.

These calculations provide an idea as to the precision of the analytic continuation procedure. The nature of this problem unfortunately shows that the maximum entropy method is not precise enough in general to make exact measurements of the spectral function, rather an overall shape can be determined.

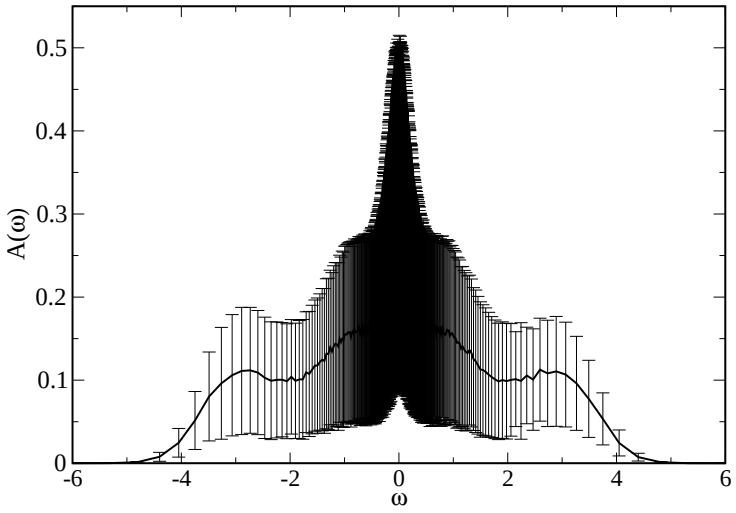
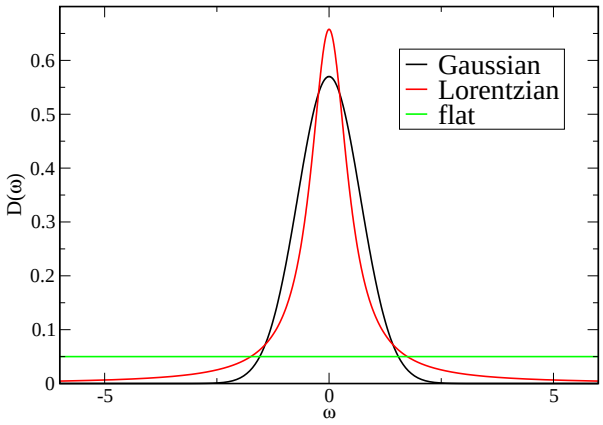
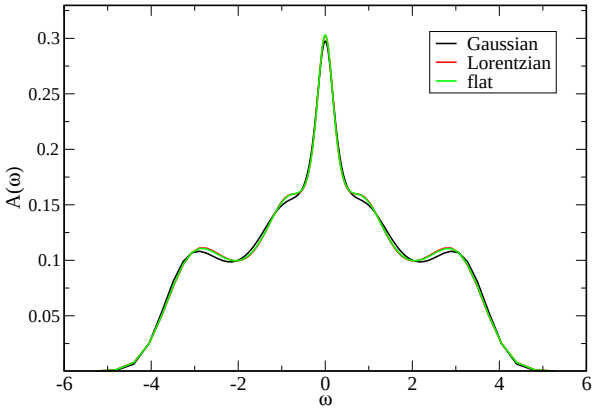


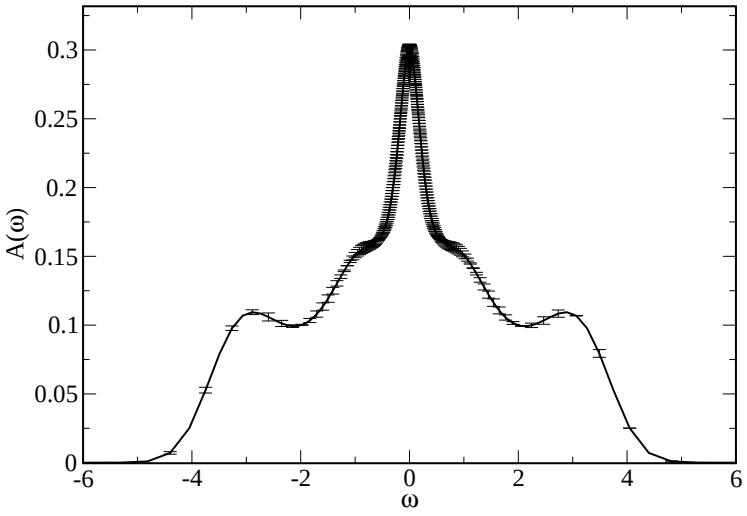
Figure 7.3: Error bars using Standard Error of the bootstrap routine $N = 10000$



(a) Default Models



(b) Bayesian averaged output for the three default models used



(c) Mean value with standard deviation as the error bar

Figure 7.4: Default Model Comparison from `Maxent`'s avspec

Conclusion

The maximum entropy method is a procedure for solving the ill-posed question of analytic continuation in the context of condensed matter physics. It has strong motivation from both mathematics and physics, with improvements built upon Bayesian updating. This can yield both the most probable spectral function (for a given default model) as well as a Bayesian averaged spectral function.

In this thesis we present a C++ code `Maxent` to perform the maximum entropy procedure. We then characterize and expand its features. At the most basic, `Maxent` is reliable enough to recover the Mott insulator transition from DMFT data. Using knowledge from perturbation theory may provide the ability to create an analytic expression for a spectral function, but it does not assist in the `Maxent` calculation. The perturbation theory default models reduced the smoothness of our curve, which is disadvantages for practical calculations. On the other hand, representing the Green's function in the Legendre basis was successful and encapsulates the Green's function in fewer data points. Finally, two methods were presented for obtaining an error bar from `Maxent`. The error bars obtained suggest that the `Maxent` routine could find many solutions within error bars and may not be suitable for precise measurements in a general case.

There are other techniques that also perform analytic continuation that were not covered here. Each has pitfalls similar to maxent and none can reliably recover the exact spectrum for all cases. Having knowledge of what does not work yields stronger tools and intuition to tackle the ill-posed problem. As the imaginary time techniques continue to improve, the importance of a reliable and precise analytic continuation procedure increase; and only through the work of methods like the maximum entropy method can important information be extracted.

The results of this work and the `Maxent` code will be made available as part of an upcoming publication[21].

Bibliography

[1] G. Mahan, *Many-Particle Physics*, Physics of Solids and Liquids (Springer US, 2000).

[2] N. Xu, C. E. Matt, E. Pomjakushina, X. Shi, R. S. Dhaka, N. C. Plumb, M. Radović, P. K. Biswas, D. Evtushinsky, V. Zabolotnyy, J. H. Dil, K. Conder, J. Mesot, H. Ding, and M. Shi, Phys. Rev. B **90**, 085148 (2014).

[3] J. Senzier, P. S. Luo, and H. Courtois, Applied Physics Letters **90** (2007), http://dx.doi.org/10.1063/1.2436651.

[4] S. Fuchs, *Thermodynamic and spectral properties of quantum many-particle systems*, Ph.D. thesis, University of Göttingen (2011).

[5] M. Jarrell and J. E. Gubernatis, Physics Reports **269**, 133 (1996).

[6] *1994 Proceedings of the Statistical Computation of the American Statistical Society* (1994).

[7] C. Kittel, *Introduction to Solid State Physics* (Wiley, 2004).

[8] R. K. Bryan, Eur Biophys J **18**, 165 (1990).

[9] A. Gaenko, E. Gull, A. E. Antipov, L. Gamper, and G. Carcassi, (2015), 10.5281/zenodo.17398.

[10] B. Bauer *et al.*, Journal of Statistical Mechanics: Theory and Experiment **2011**, P05001 (2011).

[11] G. Guennebaud, B. Jacob, *et al.*, "Eigen v3," http://eigen.tuxfamily.org (2015).

[12] D. W. Marquardt, Journal of the society for Industrial and Applied Mathematics **11**, 431 (1963).

[13] G. Kotliar and D. Vollhardt, Physics Today **57**, 53 (2004).

[14] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).

[15] D. Vollhardt, AIP Conference Proceedings **1297**, 339 (2010).

[16] E. Gull, P. Werner, S. Fuchs, B. Surer, T. Pruschke, and M. Troyer, Computer Physics Communications **182**, 1078 (2011).

[17] G. Rohringer, A. Valli, and A. Toschi, Phys. Rev. B **86**, 125114 (2012).

[18] T. Hahn, Computer Physics Communications **168**, 78 (2005).

[19] D. D. Betts, H. Q. Lin, and J. S. Flynn, Canadian Journal of Physics **77**, 353 (1999).

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

[21] R. Levy, J. P. F. LeBlanc, and E. Gull, "Implementation of the maximum entropy method for analytic continuation," *unpublished*.