Maximum Entropy Method

Chris Allton

Swansea University

Chicheley Hall
January 2015

Outline

Background

The Problem

A Solution

Tests

BACKGROUND

The Task

Given data D

Find fit F by maximising $P(F|D)$

Bayes Theorem

Need to maximise $P(F|D)$

Bayes Theorem:

$$P(F|D)P(D) = P(D|F)P(F)$$

i.e. $$P(F|D) = \frac{P(D|F)P(F)}{P(D)}$$

But $P(D|F) \sim e^{-\chi^2}$ $\longrightarrow$ minimising χ^2 $\neq$ maximising $P(F|D)$

$\longrightarrow$ *Maximum Likelihood Method* wrong??

No! Since for simple $F(t) = Ze^{-Mt}$, $P(F) = P(Z, M) \sim$ const

Bayes Theorem

Need to maximise $P(F|D)$

Bayes Theorem:

$$P(F|D)P(D) = P(D|F)P(F)$$

i.e. $$P(F|D) = \frac{P(D|F)P(F)}{P(D)}$$

But $P(D|F) \sim e^{-\chi^2}$ $\longrightarrow$ minimising χ^2 $\neq$ maximising $P(F|D)$

$\longrightarrow$ *Maximum Likelihood Method* wrong??

No! Since for simple $F(t) = Ze^{-Mt}$, $P(F) = P(Z, M) \sim \text{const}$

Priors

Actually $P(F = \text{elephant}) \equiv 0$

$\longrightarrow$ "priors" which encode any additional information

(a.k.a. predisposition, prejudices, impartialities, biases, prediliction, subjectivity, . . .)

E.g. in L.G.T. $P(M < 0) \equiv 0$

Maximum Likelihood Method applies this prior implicitly

Can encode prior information with "entropy" $= S$ (dis-information)

Define $\mathcal{I}(F) =$ "Information content" of F

"Bland" F has $\mathcal{I}(F) \sim 0$ and $S \gg 0$

"Spiky" F has $\mathcal{I}(F) \gg 0$ and $S \equiv 0$

Priors

Actually $P(F = \text{elephant}) \equiv 0$

$\longrightarrow$ "priors" which encode any additional information

(a.k.a. predisposition, prejudices, impartialities, biases, prediliction, subjectivity, . . .)

E.g. in L.G.T. $P(M < 0) \equiv 0$

Maximum Likelihood Method applies this prior implicitly

Can encode prior information with "entropy"$= S$ (dis-information)

Define $\mathcal{I}(F) =$ "Information content" of F

"Bland" F has $\mathcal{I}(F) \sim 0$ and $S \gg 0$

"Spiky" F has $\mathcal{I}(F) \gg 0$ and $S \equiv 0$

Entropy

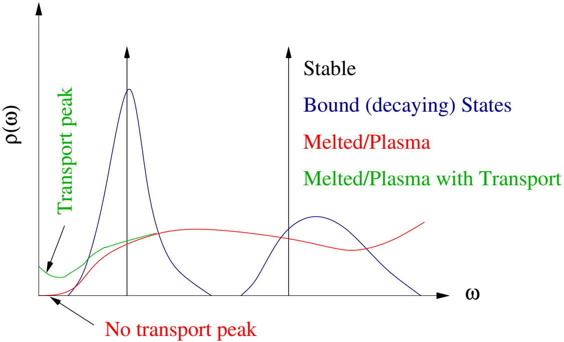
	No Data	Data
No Prior	$\mathcal{I}(F) \equiv 0$	F from min χ^2
Prior	$F \equiv$ prior	F from max $P(F\|D)$

$$P(F) = e^{-S}$$

THE PROBLEM

Spectral Functions

$$
\begin{aligned}
C(t) &= \int \rho(\omega) K(\omega, t) d\omega \\
\text{where} \quad K(\omega, t) &= \text{kernel} \sim e^{-\omega t} \\
\rho(\omega) &= \text{spectral function} \\
&= \sum_i Z_i \, \delta(\omega - M_i) \quad \text{(confined phase)}
\end{aligned}
$$



The problem

$$\rho \equiv F \quad \text{and} \quad C \equiv Data$$

Need to maximise $P(\rho|C) = \dfrac{P(C|\rho)P(\rho)}{P(C)} \sim e^{-\chi^2+S}$

The problem:

$C(t)$ known at $\mathcal{O}(10)$ t-points

but $\rho(\omega)$ should be known at $\mathcal{O}(10^3)$ points

Naively: $\mathcal{I}(\rho) \gg \mathcal{I}(C)$

$\mathcal{I}(\text{output}) \gg \mathcal{I}(\text{input})$!!!

Without S system is underconstrained

$\longrightarrow$ many solutions with $\chi^2 = 0$

A SOLUTION

A Solution: Maximum Entropy Method

Can define Entropy

$$S = \int \left[\rho(\omega) - m(\omega) - \rho(\omega) log \left(\frac{\rho(\omega)}{m(\omega)} \right) \right] d\omega$$

$m(\omega)$ is the default model which encodes prior information
Typical $m(\omega)$ is free theory result

Discretise $K(\omega, t) \longrightarrow K_{w,n}$

Express $\rho_w = m_w \exp \sum_{n=1}^{N} V_{w,n} u_n$

where $V_{w,m}$ is from a SVD of $K_{w,n}$.

Note $N \leq N_t$ by construction!

$\longrightarrow \mathcal{I}(\text{output}) \leq \mathcal{I}(\text{input})$

TESTS

MEM versus Burnier-Rothkopf

NRQCD lattice studies for p-waves:

$$\rho^{\text{MEM}} \neq \rho^{\text{BR}}$$

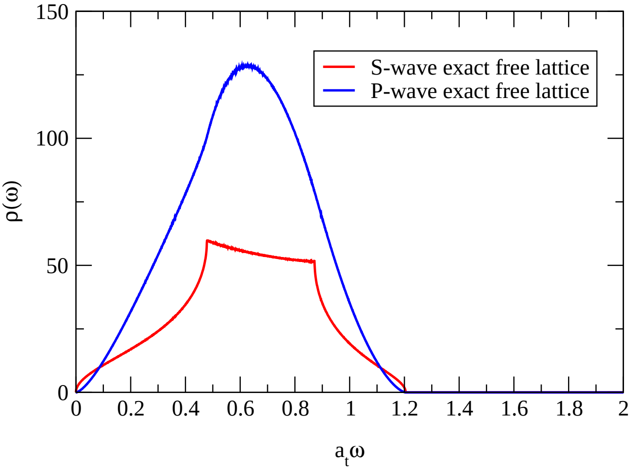
MEM says P-waves melt at $\sim T_C$
BR says P-waves don't melt at $\sim T_C$

Use correlator data which is known and unbound (melted):

free fermions

Theoretical Free Lattice Spectral Function (NRQCD)

$$\rho(\omega)^{\text{theory}} = \sum_{k_x,k_y,k_z} \delta(\omega - \hat{E}(k_x,k_y,k_z))$$



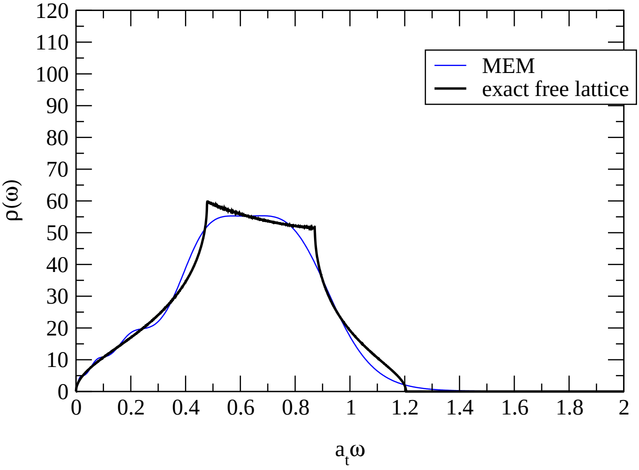
Free Lattice Correlation Functions (NRQCD)

- Use FASTSUM Collaboration's "2nd generation" parameters
- Set links $U_\mu(x) = e^{igaA_\mu(x)} \equiv 1$ i.e. *free*
- Use computer code to generate $C(t)$
- Use same stat errors and t-correlations as interacting case

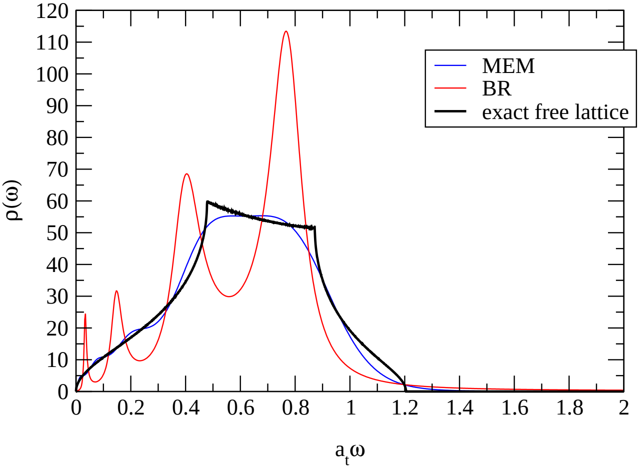
Normalisation (Sum Rule) for NRQCD:

$$C(t) = \int \rho(\omega) e^{-\omega t} d\omega \quad \longrightarrow \quad C(t=0) = \int \rho(\omega) d\omega$$

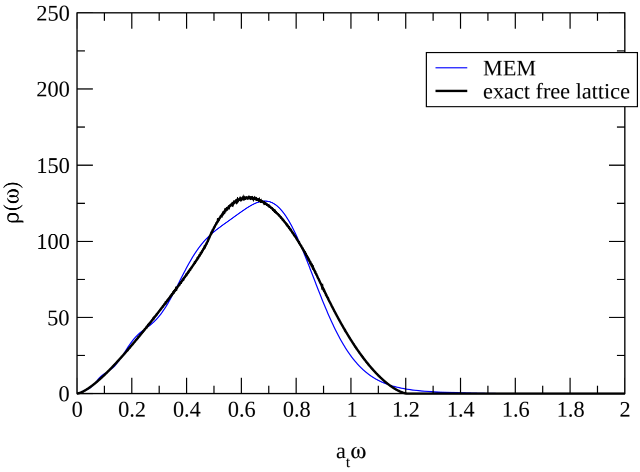
MEM Result - S-wave



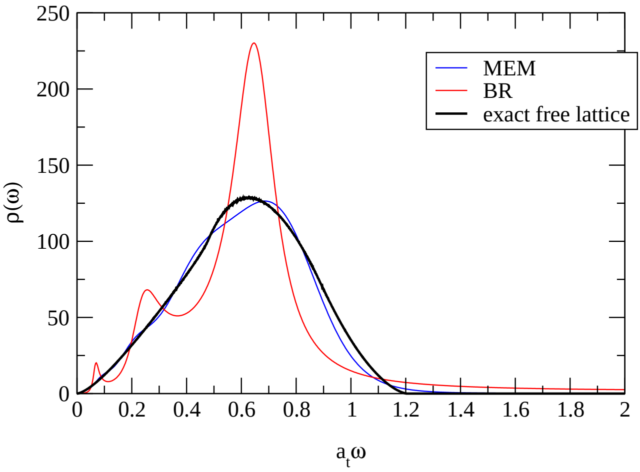
MEM & BR Result - S-wave



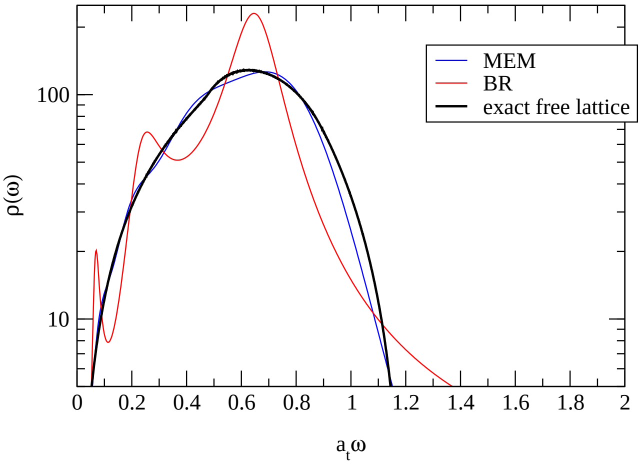
MEM Result - P-wave



MEM & BR Result - P-wave



MEM & BR Result - P-wave (log scale)



MEM systematics

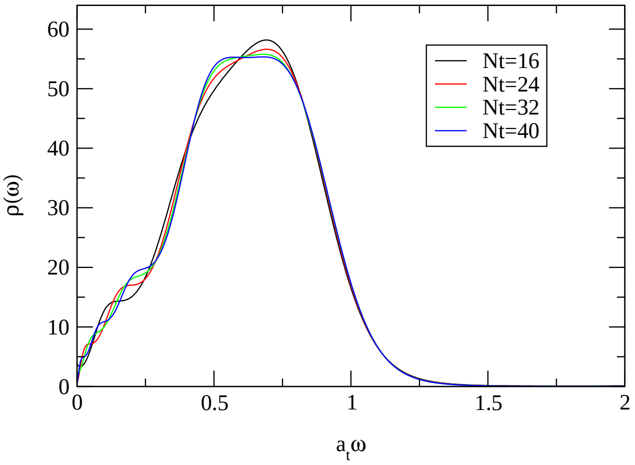
- default model
- time range
- energy discretisation: $\omega = \{\omega_{\text{min}}, \omega_{\text{min}} + \Delta\omega \; \ldots \omega_{\text{max}}\}$
- number of configs
- numerical precision

(All true also for BR)

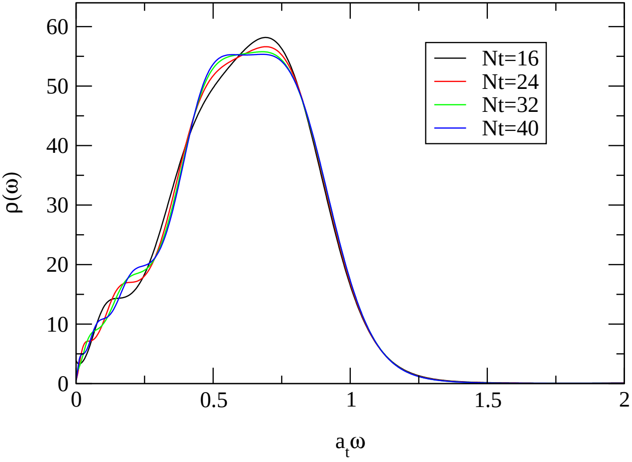
Recall $\mathcal{I}(\rho) \leq N_t$ for MEM

Can vary this in free case by varying N_t

N_t systematics - S-wave

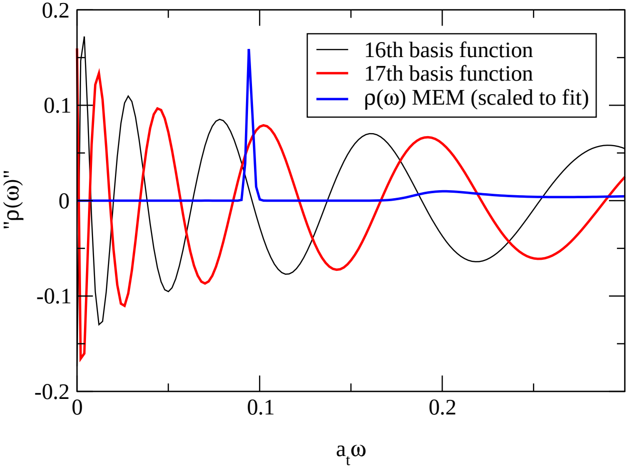


N_t systematics - P-wave



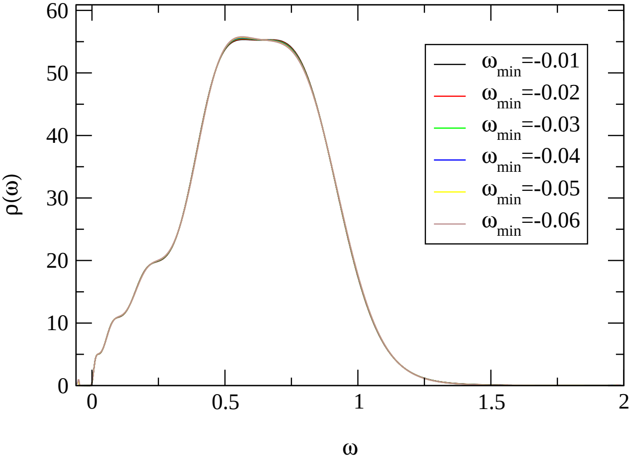
Feature Resolution

MEM can reproduce features smaller than the characteristic size of its basis functions:

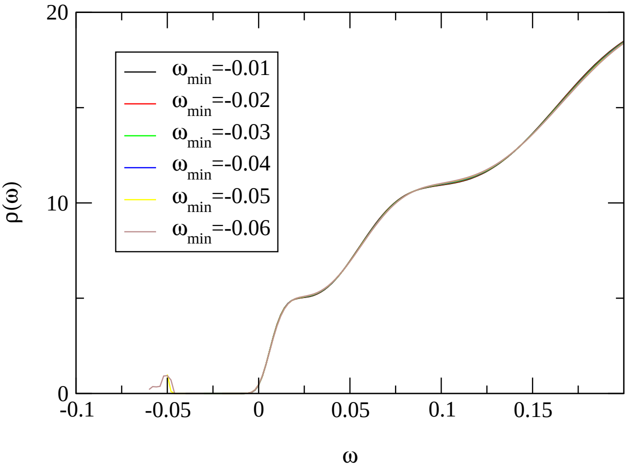


Negative ω_{min}

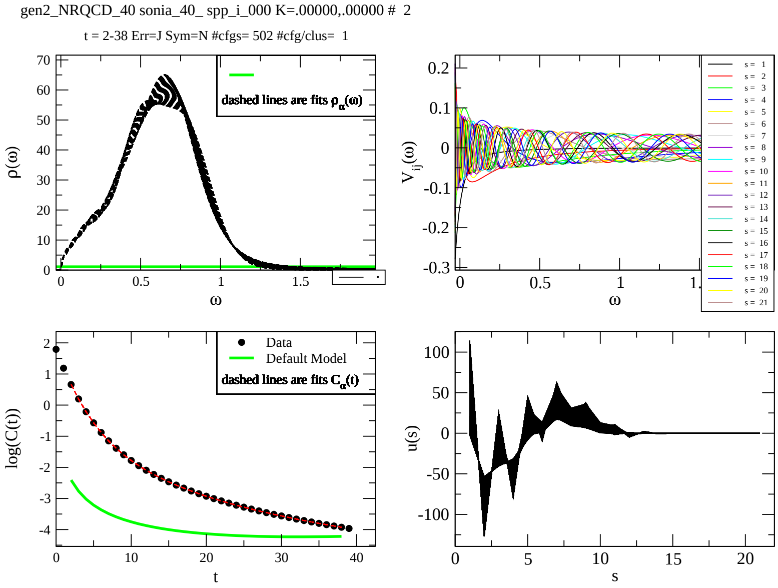
For $\omega_{min} < 0.06$ MEM does not converge



Negative ω_{min}



MEM: more than you ever wanted to know



"Information Theorem"

Recall: $\mathcal{I}(\rho(\omega)) \;\le\; \mathcal{I}(C(t))$

Due to *t*-correlations: $\mathcal{I}(C(t)) < N_t$

Each spectral feature requires 3 real numbers to describe:

position, height, width

$$\longrightarrow \mathcal{I}(\rho(\omega)) \sim 3 \times N_{\text{features}}$$

By construction, MEM works in a space of dim $\le \mathcal{I}(C(t))$
$\longrightarrow$ "Information Theorem" automatically satisfied

Other P-wave studies

2 studies of NRQCD P-waves using $C(t)$ only and *not* spectral functions:

G. Aarts et al, Phys. Rev. Lett. 106 (2011) 061602
S. Kim, P. Petreczky and A. Rothkopf, PoS LATTICE 2013 (2014) 169

Both found a difference in P-wave $C(t)$ as T increased past T_C

SLIDES TO HELP ME ANSWER TRICKY QUESTIONS

SLIDES TO HELP ME ANSWER DUMB QUESTIONS