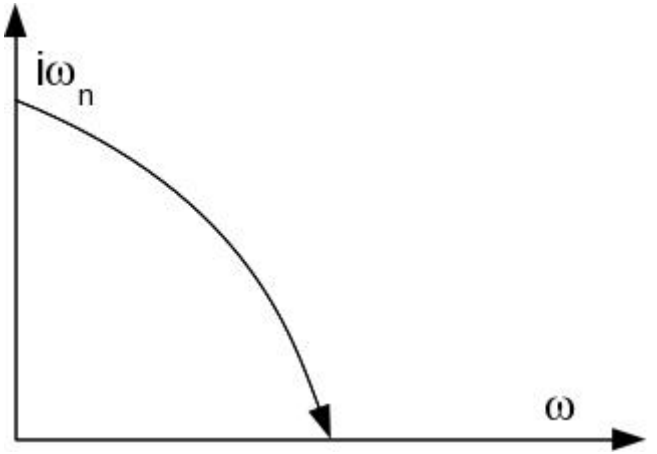
The Maximum Entropy Method

For Wick rotation



Mark Jarrell

Outline

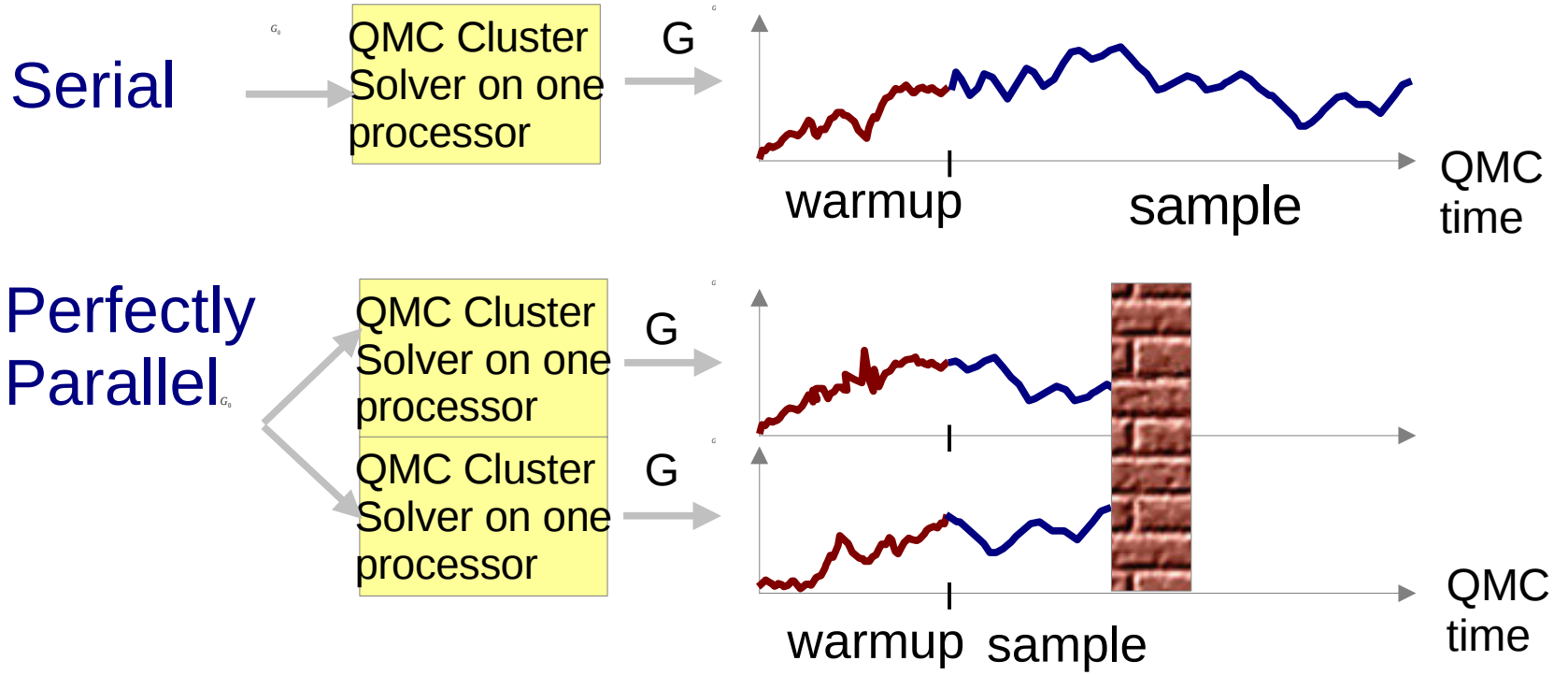
- The Problem
- Bayes Theorem and the MEM
- An Algorithm
  - How things fail (cautionary notes)
- Some examples
  - Model selection
  - Spectra normalization
  - The annealing method (use it!)

Outline

- The Problem
- Bayes Theorem and the MEM
- An Algorithm
  - How things fail (cautionary notes)
- Some examples
  - Model selection
  - Spectra normalization
  - The annealing method

Assumtion

- You have an ergodic QMC code that gives you $G(i?_n)$, $G(\tau)$, $?(\tau)$...

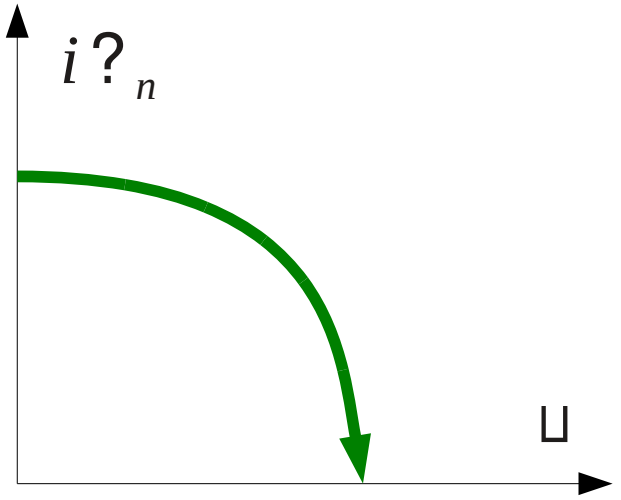


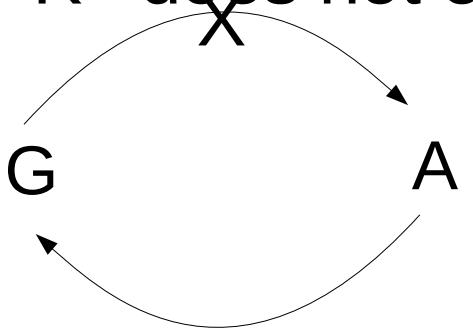
- It generates data that can be made statistically independent (rebinning, covariance...)

The Problem

$$G(\tau)=\int \frac{A(\omega)e^{(-\tau\omega)}}{1+e^{(-\beta\omega)}}$$

$$G(i?_n)=\int \frac{A(\square)}{i\square_n-\square}$$



- $A(\square)=-1/?\Im G_R(\square)$
- If we write G=KA
  - Det(K) =0
  - K^{-1} does not exist
- 

-
-
-
- Many A give the same G (within the error)

The kernel

- The kernel K in G = KA is
  - For Bosons $A(\omega) = \frac{-1}{\pi\omega}\mathrm{Im}G(\omega) \geq 0$

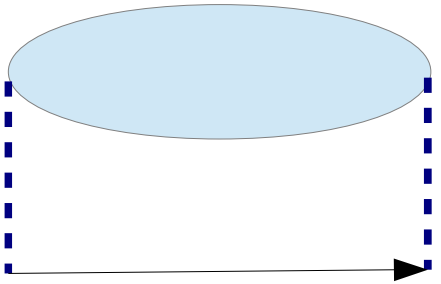
$$K(\tau,\omega) = \frac{\omega[e^{-\tau\omega} + e^{-(\beta-\tau)\omega}]}{1 - e^{-\beta\omega}} \text{ or } K(i\omega_n,\omega) = \frac{2\omega^2}{\omega^2 + \omega_n^2}$$

  - For Fermions $A(\omega) = \frac{-1}{\pi}\mathrm{Im}G(\omega) \geq 0$

$$K(\tau,\omega) = \frac{\exp(-\tau\omega)}{1 + \exp(-\beta\omega)} \text{ or } K(i\omega_n,\omega) = \frac{1}{i\omega_n - \omega}$$

Desired Properties of the Solution

- Want A with no spurious features
  - A should have only the information required to reproduce G and satisfy *a priori* constraints
- Want to be able to incorporate *a priori* information
  - Exact results
  - Sum rules



$$\sim \frac{u^2 n(1-n)}{i?_n}$$

$$\int d? A(\square) = 1$$

References and Codes

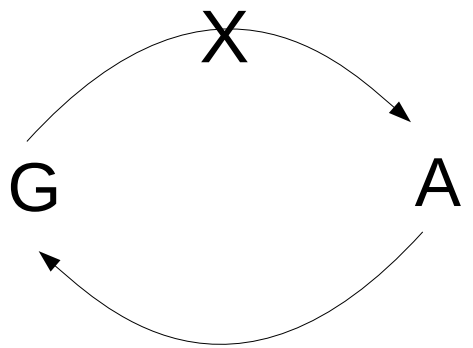
- Codes (1987)
  - http://www.phys.lsu.edu/~jarrell/Research/MEM/MEM.html
- References
  - H.-B. Schuttler et al, PRL 55, 1204 (1985); PRB 34,4744 (1986).
  - S.R. White, PRL. 63, 1523 (1989); M. Jarrell, PRL 63, 2504 (1989).
  - M. Jarrell, and J.E. Gubernatis, Phys. Rep. 269, 3, p133-195, (1996).
  - A.W. Sandvik, PRB 57, 1028710290 (1998); K.S.D. Beach, preprint, arXiv:cond-mat/0403055 (2004).
  - S. Fuchs, PRB, 83.235113.
  - R.K. Bryan, Eur. Biophys. J. 18, 165 (1990).
  - **More: in workshop proceedings and previous talk**
- Primer on MEM (neutrons) by D.S. Sivia (right)
  - http://www.fas.org/sgp/othergov/doe/lanl/pubs/00326658.pdf
  - Books on Amazon

Outline

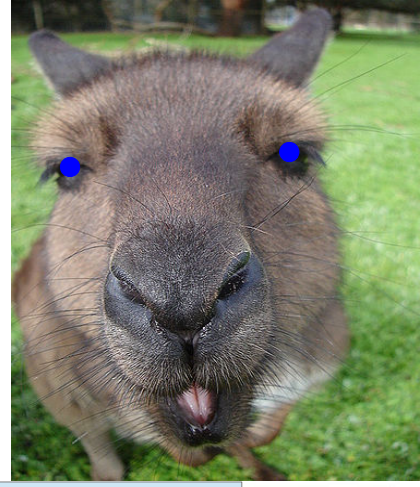
- The Problem
- Bayes Theorem and the MEM
- An Algorithm
  - How things fail (cautionary notes)
- Some examples
  - Model selection
  - Spectra normalization
  - The annealing method

Bayes Theorem

- We cannot answer the question "*given G, what is A*"?
- Instead we ask, "*given G, what is the most probable A*"?
  - Optimize P(A|G)
- Use Bayes Theorem
  - P(A,G)=P(A|G)P(G) = P(G|A) P(A)
  - P(G) is a constant associated with QMC
  - P(A|G) $\propto$ P(G|A) P(A)
    - What are P(G|A) and P(A)?



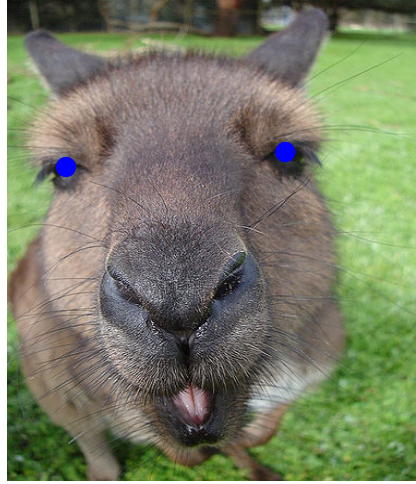
The Prior Probability P(A)



- Suppose I tell you that
  - 1/3 of kangaroos are left handed
  - 1/3 of kangaroos are blue eyed
- What fraction, p_1 , of kangaroos are both left handed and blue eyed?

A

Correlations and Kangaroos



Positive correlation

blue-eyed \ Left-handed	T	F
T	$p_1=x$, $0<X<1/3$	$p_2=1/3-x$
F	$p_3=1/3-x$	$p_4=1/3+x$

blue-eyed \ Left-handed	T	F
T	$p_1=1/3$	$p_2=0$
F	$p_3=0$	$p_4=2/3$

$p_1+p_2=1/3$

$p_1+p_3=1/3$

$p_1+p_2+p_3+p_4=1$

Not enough Information!

Negative correlation

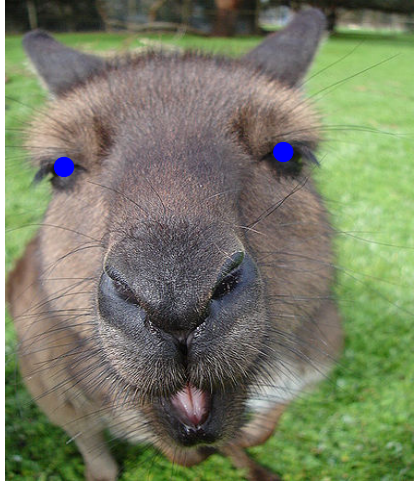
blue-eyed \ Left-handed	T	F
T	$p_1=0$	$p_2=1/3$
F	$p_3=1/3$	$p_4=1/3$

No correlation

blue-eyed \ Left-handed	T	F
T	$p_1=1/9$	$p_2=2/9$
F	$p_3=2/9$	$p_4=4/9$

A

Shannon Entropy



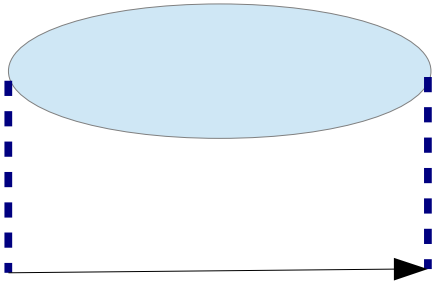
- Homework: show that p_1=1/9 is the result obtained by maximizing the Shannon entropy subject to the know constraints
  - $-\sum_i p_i \ln(p_i) - \sum_j ?_j (constraint_j)$
  - i.e., $p_1 + p_2 =$ /3
- By maximizing the Shannon entropy subject to known constraints, we ensure that the answer we obtain has
  - the least information-->data+*a priori*
  - No spurious features

Regularizing function	p_1	Correlation
$-\sum p_i \log p_i$	0.11111	None
$-\sum p_i^2$	0.08333	Negative
$\sum \log p_i$	0.13013	Positive
$\sum \sqrt{p_i}$	0.12176	Positive

S is defined relative to a model

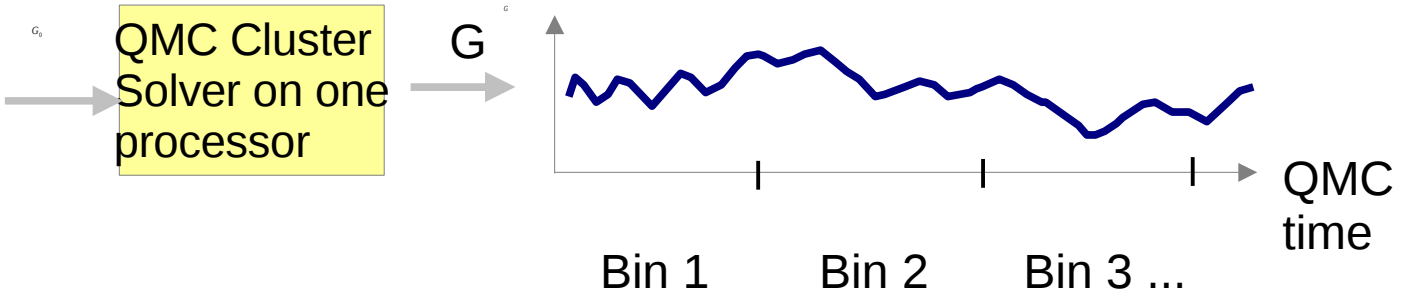
$$S = -\sum_i A_i \ln A_i / m_i - A_i + m_i$$

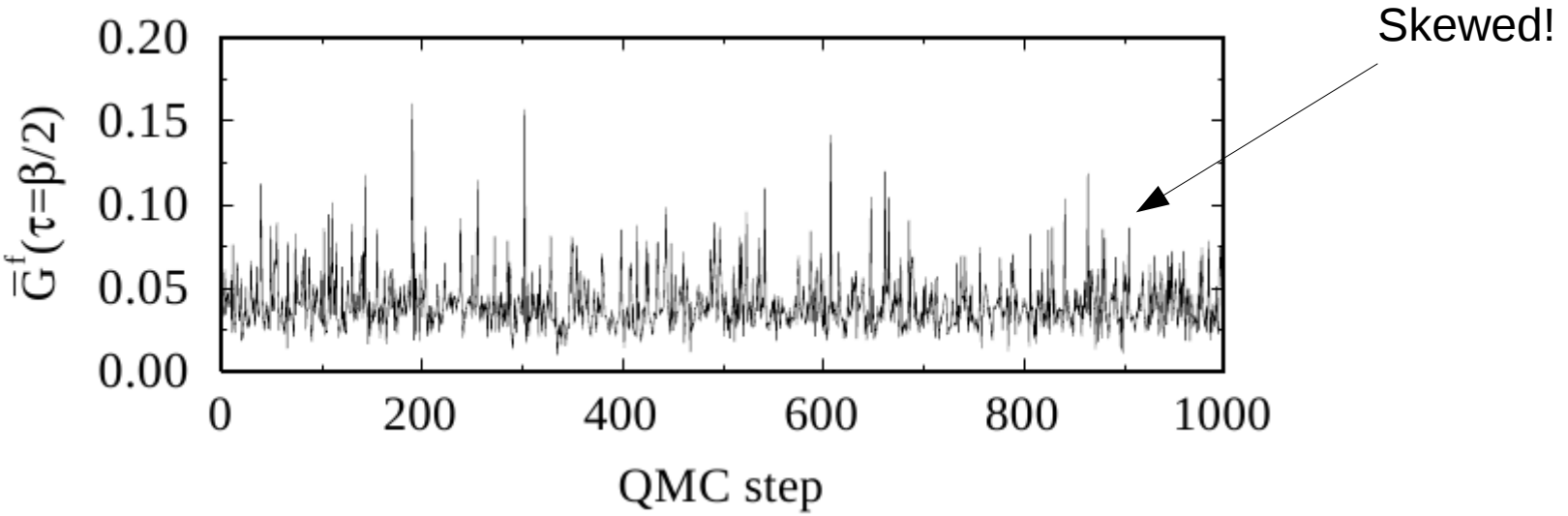
- The model allows us to introduce *a priori* information
  - Exact result at high T
  - Exact result at high frequency



- The model should not be overly informative unless you are certain the information introduced is exact
- It is important that the model not have wrong information

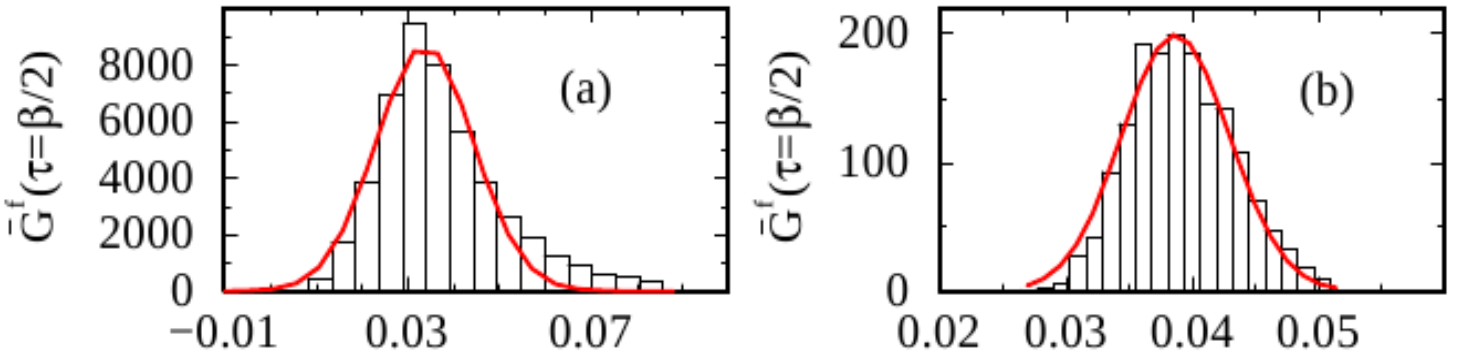
The Likelihood function P(G|A)



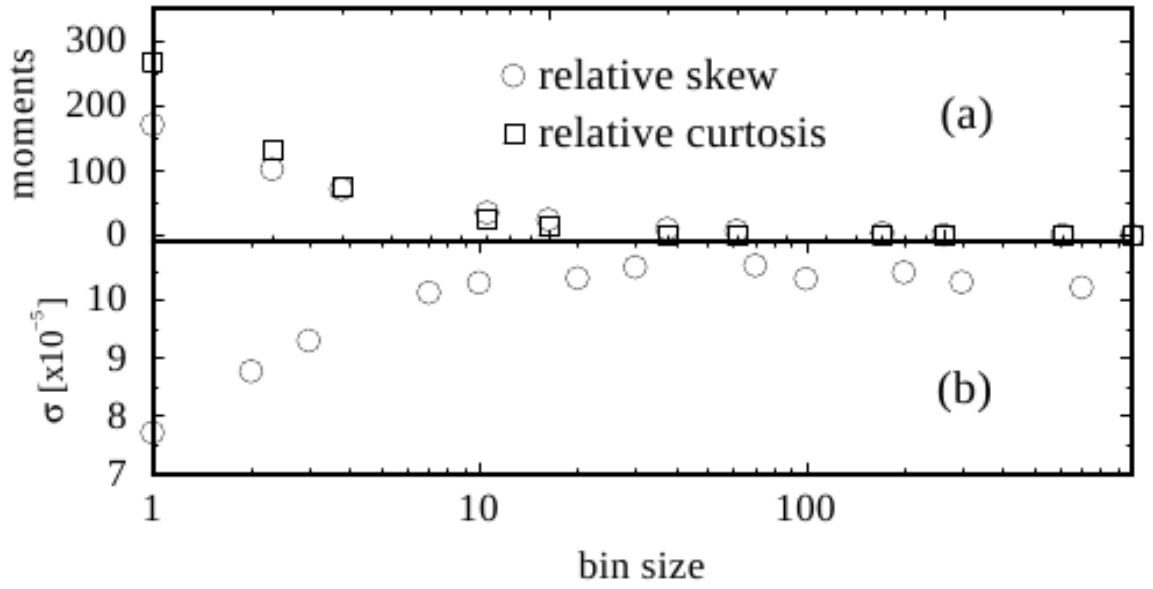


Symmetric PAM f-electron local Green's function Gf (τ = β/2) plotted as a function of the QMC step for U = 2, V = 0.6, and β = 20.

The Likelihood function



Distribution of the data shown in the last slide (a) and after rebinning (b). The solid line is a Gaussian fit. In (b) the data was processed by packing it sequentially into **bins of 30** measurements each. Below, moments of the data versus bin size.

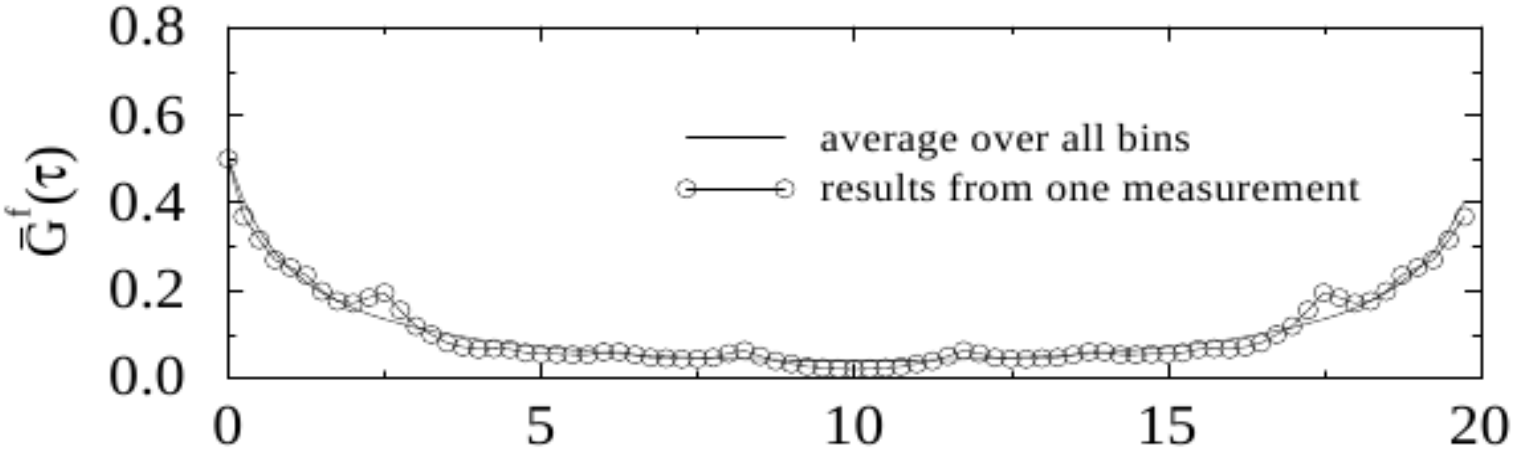


$$\chi^2 = \sum_{l=1}^{L} \left(\frac{\bar{G}_l - \sum_j K_{l,j} A_j}{\sigma_l} \right)^2$$

P(G|A)= exp -? ^{2}/2

Right?

The likelihood function: correlations in Masubara time or frequency



- Calc. the covariance C
- Diagonalize C
- Rotate G=KA
- Calculate P(G|A)= exp -?^{2}/2

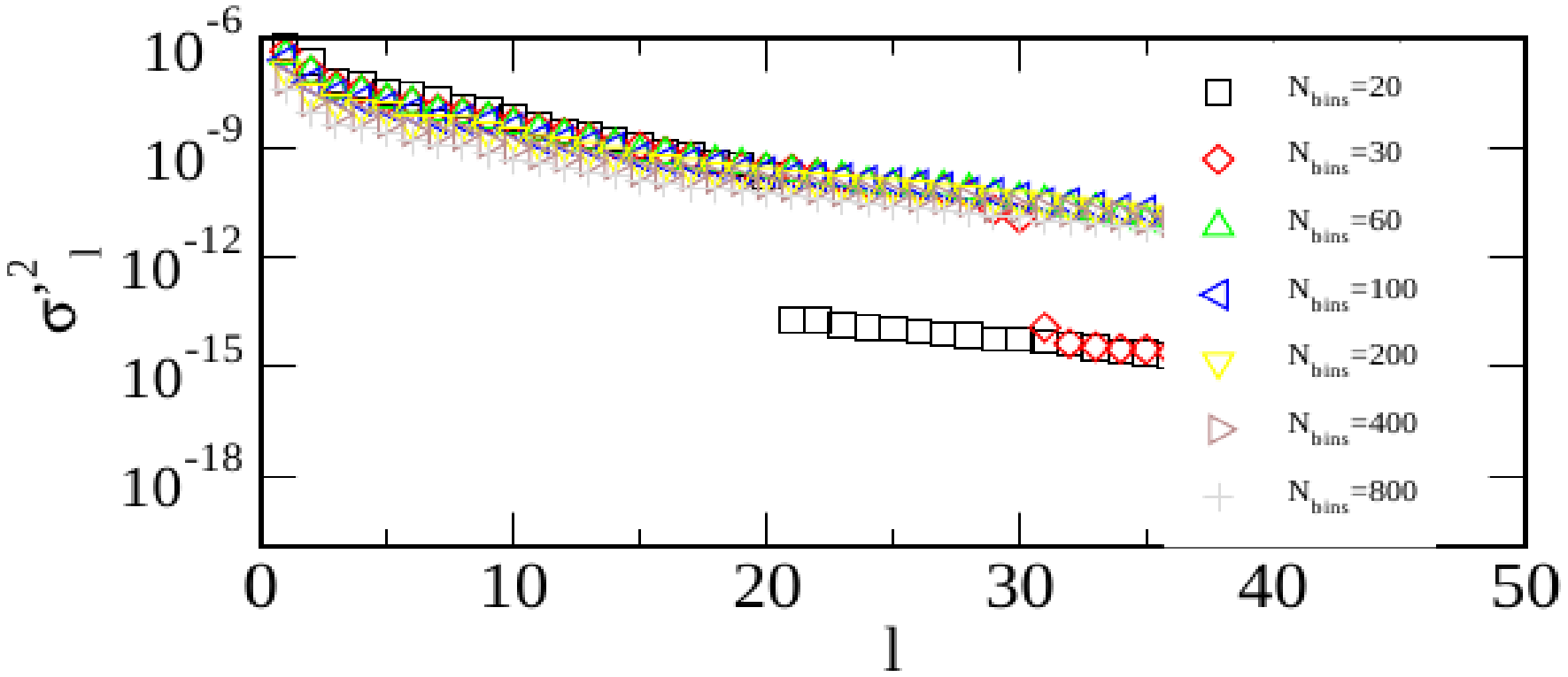
$$C_{lk} = \frac{1}{N_{bins}(N_{bins}-1)} \sum_{j=1}^{N_{bins}} (\langle \bar{G}_l \rangle - \bar{G}_l^j)(\langle \bar{G}_k \rangle - \bar{G}_k^j) .$$

$$\mathbf{U}^{-1}\mathbf{C}\mathbf{U} = \sigma_i'^2 \delta_{ij} .$$

$$\mathbf{K}' = \mathbf{U}^{-1}\mathbf{K} \qquad \bar{\mathbf{G}}' = \mathbf{U}^{-1}\bar{\mathbf{G}} .$$

$$\chi^2 = \sum_l \left(\frac{\bar{G}_l' - \sum_j K_{l,j}' A_j}{\sigma_l'} \right)^2$$

Cautionary note: Calculate C

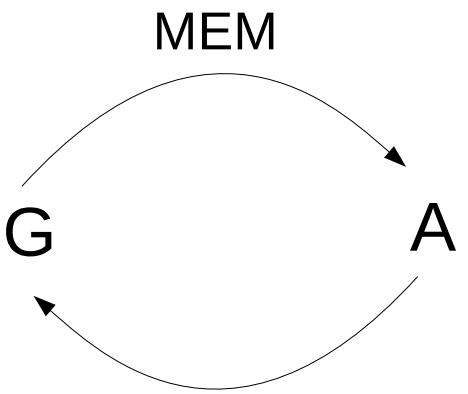


Eigenvalue spectra of the covariance of the PAM f G for different numbers of bins of data. Each bin contains 100 measurements and L = 41. When Nbins < 2L, the eigenvalue spectrum develops a sharp break. Generally we need

$$N_{bins} \geq 2L$$

The Maximum Entropy Method

- The question *Given G what is the most probable A?*
  - Maximize $P(A|G) \propto P(G|A)\,P(A)$ (Bayes theorem).
  - $P(A) \propto \exp(\alpha S)$, $P(G|A) \propto \exp(-\chi^2/2)$
  - $S = -\sum A_i \ln(A_i/m_i) - A_i + m_i$
- What about α?
  - $P(A,\alpha|G) = P(A|G,\alpha)\,P(\alpha)$
  - $P(\alpha) = 1/\alpha$, Jeffrey prior
  - $P(\alpha|G) = \int dA\, P(A,\alpha|G)$
  - If you can perform $\int dA$, then you are done!
    - But..The space of A is very large!



Outline

- The Problem
- Bayes Theorem and the MEM
- An Algorithm
  - How things fail (cautionary notes)
- Some examples
  - Model selection
  - Spectra normalization
  - The annealing method

A Maximum Entropy Algorithm

- To perform $\int dA$ we use Gaussian approximation $d^N A / \prod_i \sqrt{A_i}$
  - Use a measure
    - S is flat

$$P(A|\alpha, m) = \frac{1}{Z_S} \exp\left\{\alpha\left(-\sum A_i \ln A_i/m_i - A_i + m_i\right)\right\}$$

$$Z_S = \int \frac{d^N A}{\prod_i \sqrt{A_i}} \exp\left\{\alpha\left(-\sum A_i \ln A_i/m_i - A_i + m_i\right)\right\}$$

$$S \approx \frac{1}{2}\delta A^T \nabla\nabla S|_{A=m} \delta A = -\frac{1}{2}\delta A^T \{1/m\} \delta A$$

$$Z_S \approx \int \frac{d^N A}{\prod_i \sqrt{A_i}} \exp\left\{\alpha\left(-\frac{1}{2}\delta A^T \{1/m\} \delta A\right)\right\} = (2\pi/\alpha)^{N/2}$$

http://bayes.wustl.edu/etj/etj.html

- The likelihood function is normalized
  - $P(\bar{G}|A) = e^{-\chi^2/2}/Z_L \qquad \chi^2 = \sum \frac{\left(\bar{G}'_l - \sum_i K'_{li} A_i\right)^2}{\sigma'^2_l}$
  - $Z_L = \int d^L G \exp\left\{\frac{1}{2}\sum_{l=1}^{L} \frac{\left(\bar{G}'_l - G_l\right)^2}{\sigma'^2_l}\right\} = (2\pi)^{L/2} \prod_l \sigma'_l$
- In the Gaussian approximation, MEM amounts to optimizing

$$P(A|G,m,\alpha) \propto P(G|A,m,\alpha)P(A|m,\alpha) = \frac{\exp(\overbrace{\alpha S - \chi^2/2}^{Q})}{Z_S Z_L}$$

- This may be done with, e.g., a Newton Search
  - See the Workshop Proceedings

A

Three flavors of MEM: selection of α

- *Historic* MEM
  - Adjust α so that $\chi^2 = L$
- *Classic* MEM
  - Find α that maximizes P(α|G,m)
- *Bryan's method*
  - Marginalize A over P(α|G,m)
  - Find the optimal A for each α, A(α)
  - find $\bar{A} = \int d\alpha \hat{A}(\alpha) P(\alpha|\bar{G}, m)$.

For Classic and Bryan's Method we need P(?|G,m)

$$P(\alpha|\bar{G},m) = \int \frac{d^N A}{\prod_i \sqrt{A_i}} P(A,\alpha|\bar{G},m)\,.$$

$$P(A,\alpha|\bar{G},m) = P(A|\bar{G},m,\alpha)P(\alpha) \propto \frac{\exp(\alpha S - \chi^2/2)}{Z_S Z_L} P(\alpha)$$

P(?) ∝1/?

$$P(\alpha|\bar{G},m) = \int \frac{d^N A}{\prod_i \sqrt{A_i}} \frac{\exp(\alpha S - \chi^2/2)}{Z_S Z_L \alpha} = \frac{Z_Q}{Z_S Z_L \alpha}$$

Z_Q is calculated in a similar fashion to Z_S. We expand about the maximum of Q at $A = \hat{A}$ so that $\exp\{\alpha S - \chi^2/2\} \approx \exp\{Q(\hat{A}) + \frac{1}{2}\delta A^T \nabla\nabla Q|_{\hat{A}}\, \delta A\} = \exp\{Q(\hat{A}) + \frac{1}{2}\delta A^T \{\frac{1}{2} \nabla\nabla \chi^2|_{\hat{A}} - \{\alpha/\hat{A}\}\}\delta A\}$. We again make a Gaussian approximation to the integral, and if λ_i are the eigenvalues of $\frac{1}{2}\{A^{1/2}\} \nabla\nabla\chi^2|_{\hat{A}} \{A^{1/2}\}$, then

$$P(\alpha|\bar{G},m) = \frac{1}{W_\alpha}\prod_i \left(\frac{\alpha}{\alpha+\lambda_i}\right)^{1/2} \frac{e^{Q(\hat{A})}}{\alpha}$$

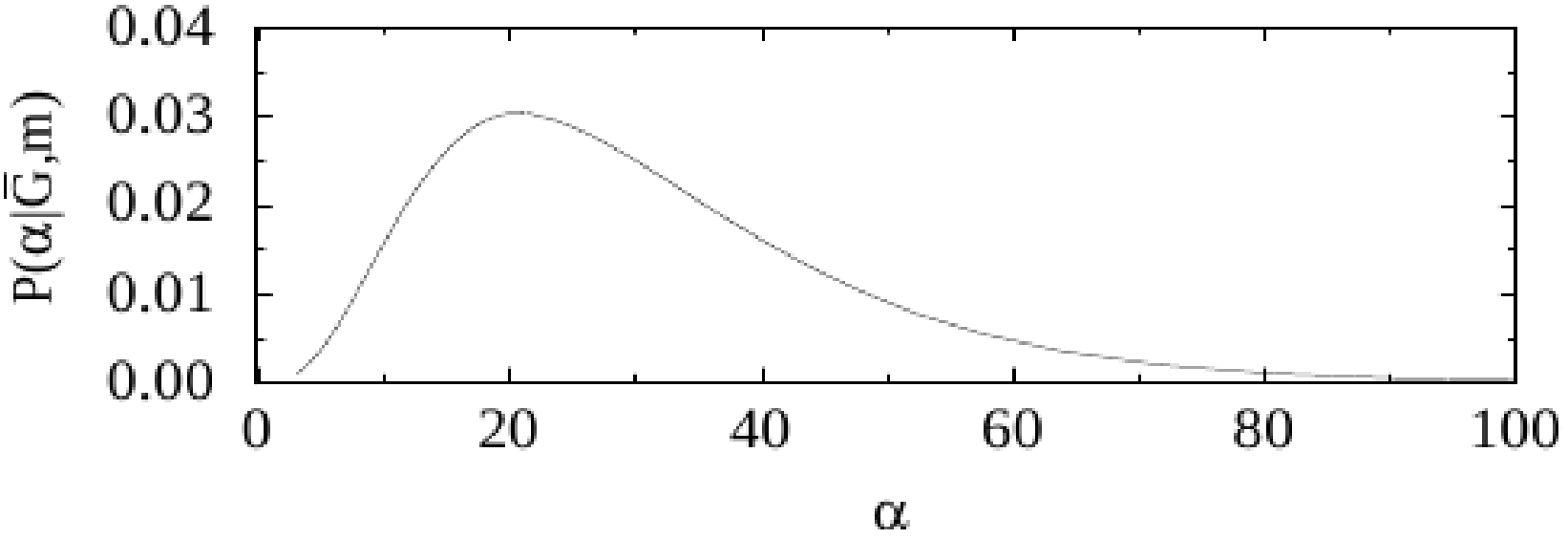
$$W_\alpha = \int \frac{d\alpha}{\alpha}\prod_i \left(\frac{\alpha}{\alpha+\lambda}\right)^{1/2} e^{Q(\hat{A})}\,.$$

- For strong data, P(α|G,m) is dominated by the product and the exponential, so that the Classic optimal α, $\hat{\alpha}$ is given by

$$-2\hat{\alpha}S \approx \sum_i \frac{\lambda_i}{\hat{\alpha}+\lambda_i}\,.$$

Bryan vs. Classic vs. Historic MEM

- Historic should not be used
- For strong data, the Bryan and Classic MEM are the same
- For weaker data (or m), P(α|G,m) is heavily skewed and the Brvan MEM should be used



Cautionary Note!

- For very weak data or model m, α becomes small so that Q=αS-? ^{2}/2 becomes flat in some directions A (for example high frequencies)

$$G(\tau)=\int \frac{A(\omega)e^{(-\tau\omega)}}{1+e^{(-\beta\omega)}}$$

- In this case, the Gaussian approximations to the integrals ∫dA breaks down, and the code runs away to α––>0
  - Most common problem encountered with MEM
  - Fix with better data or better m (annealing method)

Error Estimates

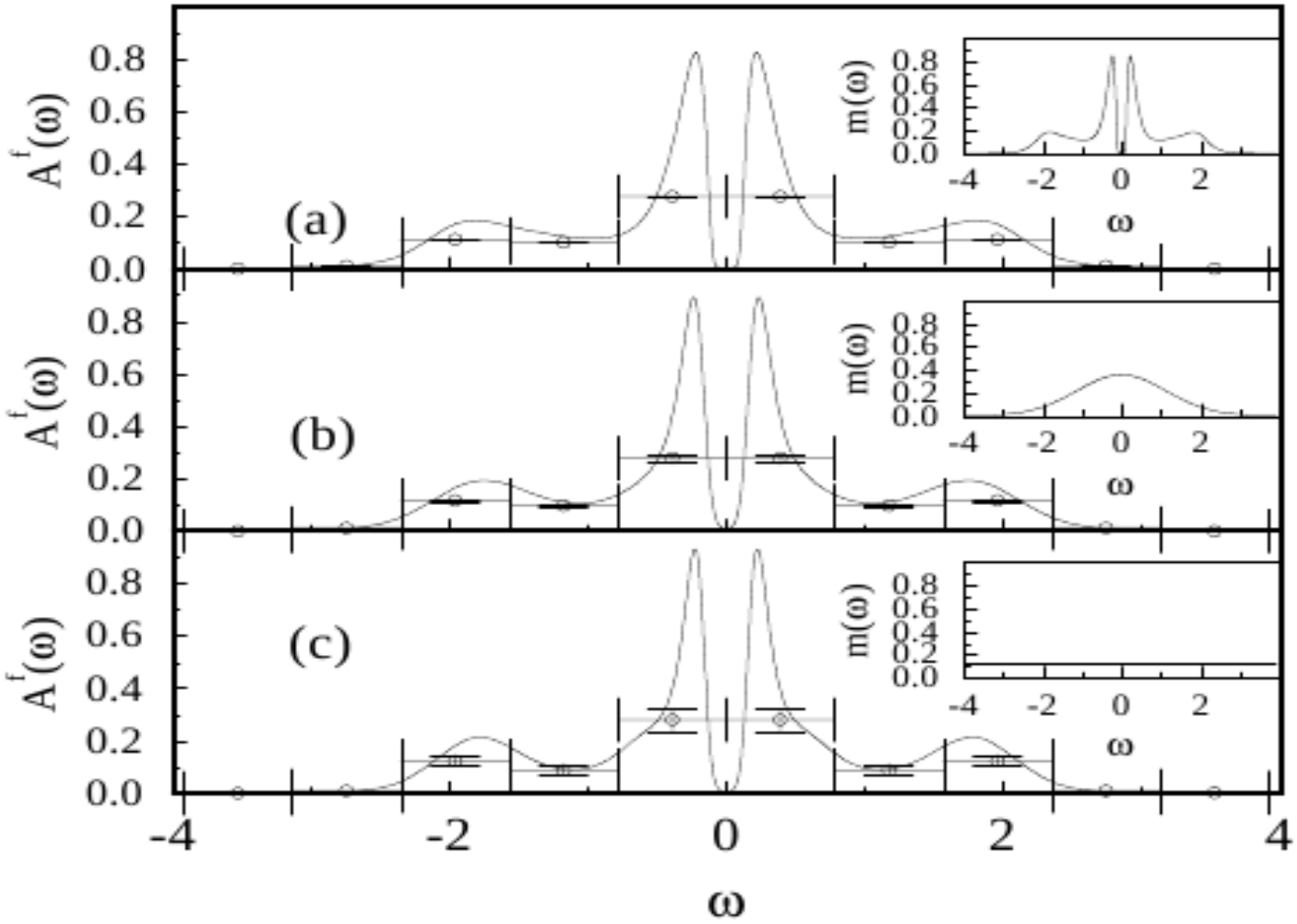
- The distribution of A is given by
  - $P(A|\bar{G}, m, \alpha) \propto e^{-\frac{1}{2}\delta A^T \cdot \nabla\nabla Q|_{\hat{A}} \cdot \delta A}$
- Thus the covariance of A is
  - $\langle \delta A(\omega) \delta A(\omega') \rangle = -(\nabla\nabla Q|_{\hat{A}})^{-1}$
- We may estimate the error of integrals of A
  - $H = \int d\omega A(\omega) h(\omega)$
  - $\left\langle (\delta H)^2 \right\rangle = \int\int d\omega d\omega' \, h(\omega) h(\omega') \, \langle \delta A(\omega) \delta A(\omega') \rangle$
- We know $\nabla\nabla Q|_{\hat{A}}$ since it is used in the Newton search to find A(α) (see workshop proc.)

Outline

- The Problem
- Bayes Theorem and the MEM
- An Algorithm
  - How things fail (cautionary notes)
- Some examples
  - Model selection
  - Spectra normalization
  - The annealing method

Error Propagation

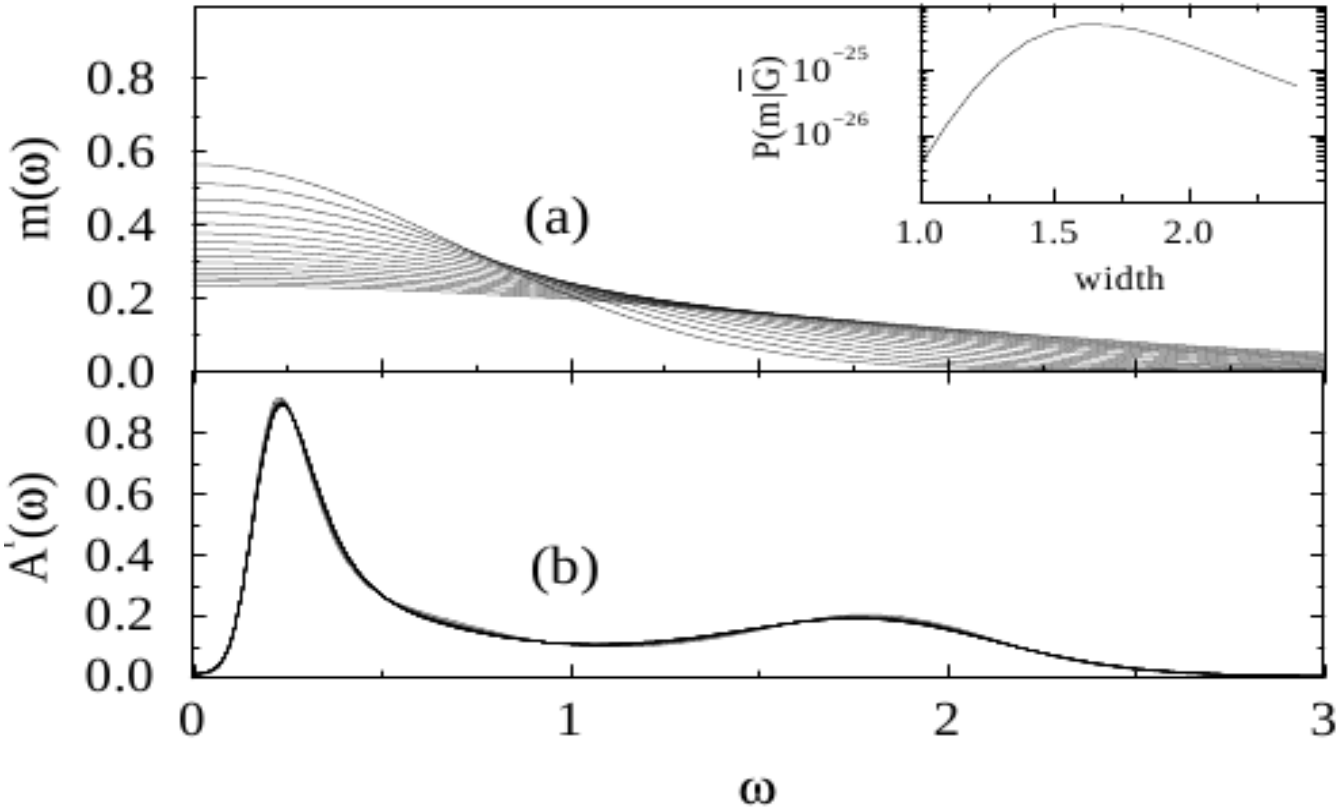
- Most often, we let h be a box function, so that we calculate errors on regions of A



The PAM f-electron density of states Af (ω) generated using (a) a perturbation theory, (b) a Gaussian, and (c) a flat default model.

Model Selection: the symmetric PAM

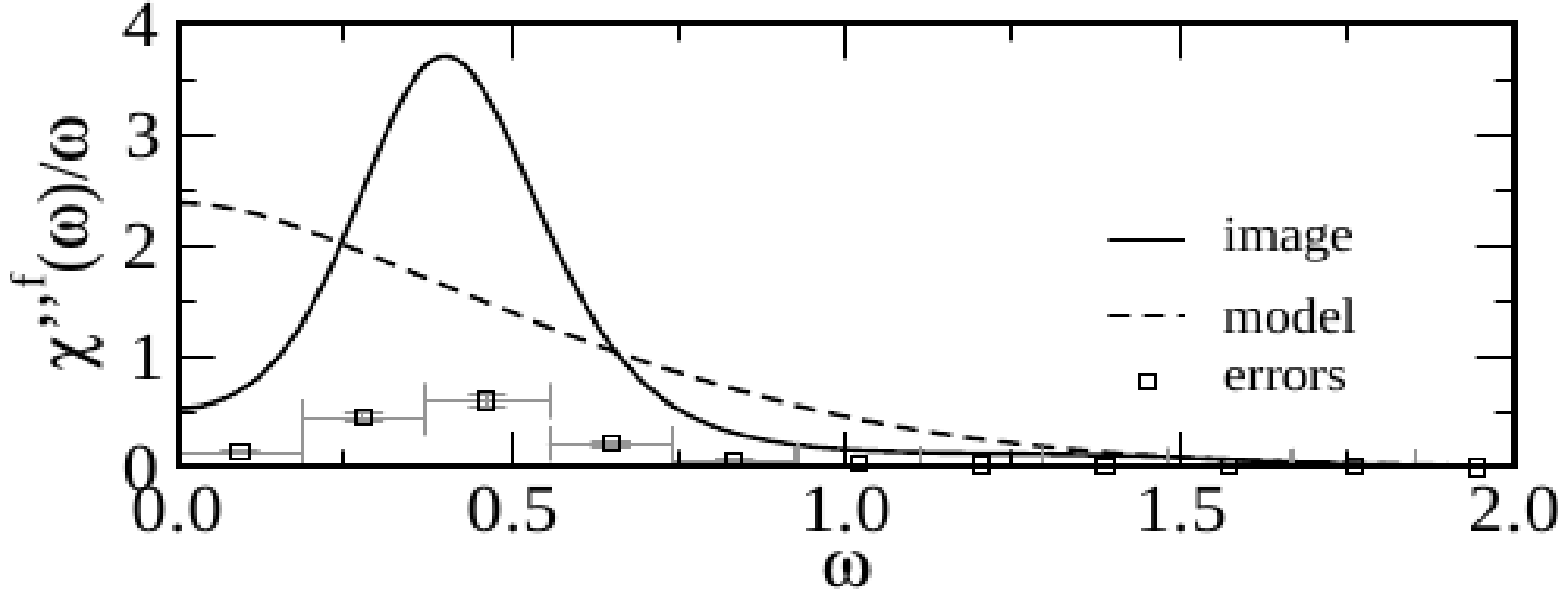
- I previously mentioned
  - $P(m|\bar{G}) = \int d\alpha P(\alpha|m,\bar{G})P(m)$
- I see no reason to believe P(m) is not flat, so
  - $P(m|\bar{G}) \propto W_\alpha$
- E.g.,



Two Particle Spectra

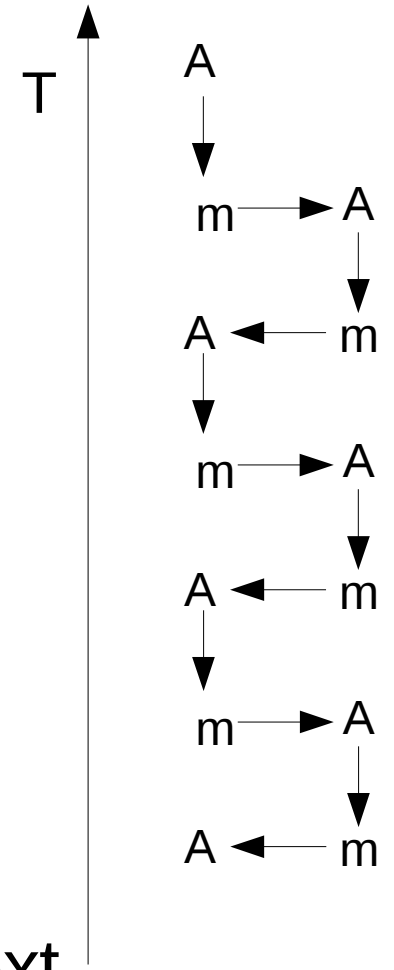
- Here, we let A=χ''(ω)/ω, which is positive definite

$$\chi(\tau) = \int_0^\infty d\omega \frac{\omega[e^{-\tau\omega} + e^{-(\beta-\tau)\omega}](\chi''(\omega)/\omega)}{1 - e^{-\beta\omega}}$$

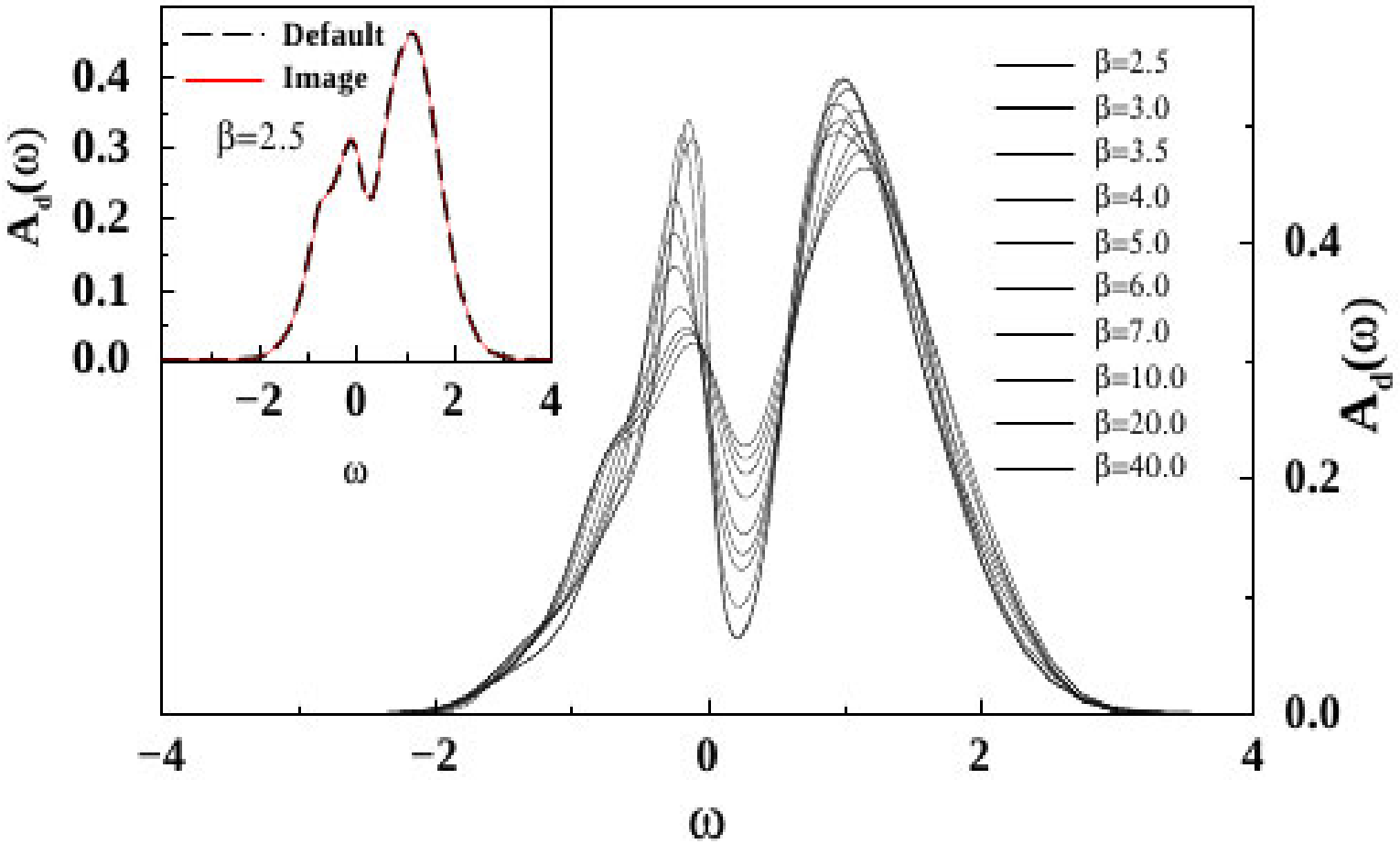


The Annealing Method

- The Gaussian approximations used for integrals $\int dA$ break down when
  - The data is weak
  - The model is weak
  - Annealing addresses these problems.
- Need a series of data in temperature
  - Start at high T where
    - Data is always strong
    - We can often calculate A exactly
  - Use the A at high T as the m or the next lower T, etc.



Annealing method: PAM d-DOS



- T steps chosen so that A changes slowly
- Model is nearly exact at high T
- Not Bayesian (purists close your eyes)!
- Inexpensive! High T runs cheap!
- Best results.

The evolution of the d-electron density of states of the asymmetric PAM when U = 1.5, V = 0.6 nd = 0.6, and nf = 1.0. At high temperatures, as shown in the inset, the spectra

Matsubara Freq. Self energy

- The Self energy also has a Hilbert transform

$$?(k,?)-\square_H=\int dx\frac{-1/?\square''(k,x)}{\square-x}$$

- It is convenient to normalize the spectrum, using

$$?(k,?)-\square_H\sim U^2n(1-n)/\square$$

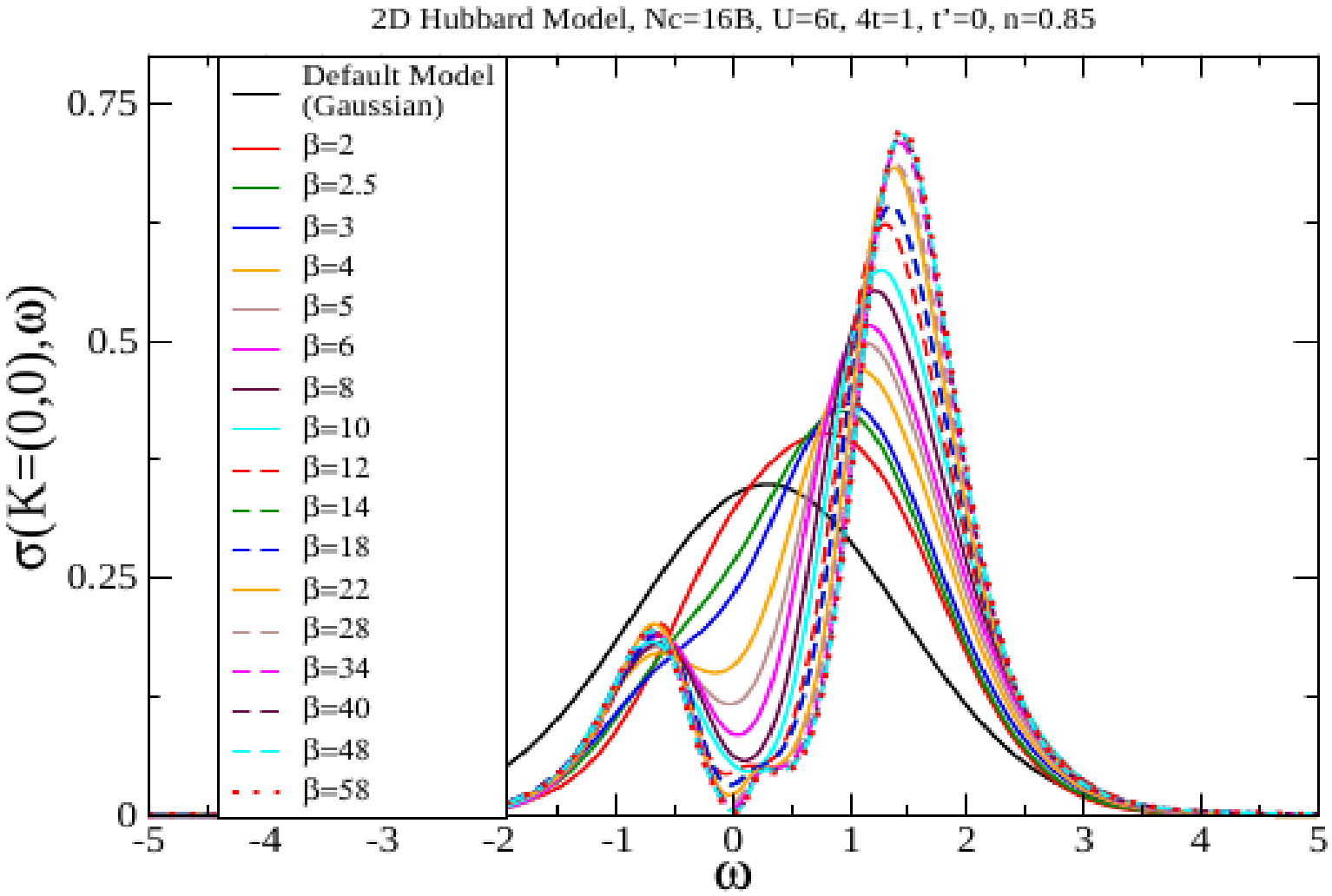
$$\int dx-1/?\square''(k,x)=U^2n(1-n)$$

- We continue the spectrum so that

$$:\sigma(\mathbf{k},\omega)=-\frac{1}{\pi}\Sigma''(\mathbf{k},\omega)/U^2\chi_{\sigma,\sigma},\int d\omega\sigma(\mathbf{k},\omega)=1.$$

$$\frac{\Sigma(\mathbf{k},i\omega_n)-\Sigma_H}{U^2\chi_{\sigma,\sigma}}=\int d\omega\frac{\sigma(\mathbf{k},\omega)}{i\omega_n-\omega};$$

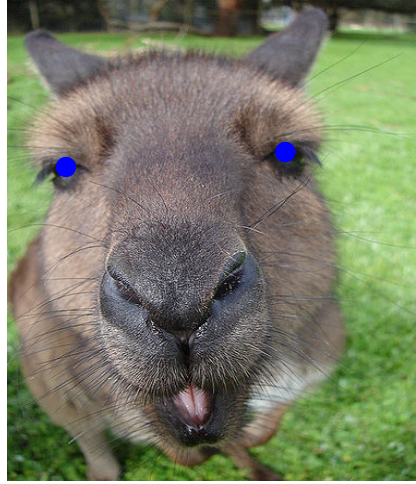
Rubtsov CTQMC/DCA of Hubbard Model



Normalized self energy spectra σ(k, ω) calculated by annealing for the Hubbard model with the DCA with U = 6t (4t = 1), t′ = 0, k = (0, 0) cluster size Nc = 16 and filling n = 0.85 with an optimized Gaussian default model.

Calculations of H. Chen

Steps to ensure a good MEM



1. Rebin your data to remove correlations in QMC time.
2. Generate sufficient bins of data so that **Nbins > 2L** where L is the number of Matsubara time or frequencies used.
3. If a self consistent method, such as DCA, is used to generate the data, be sure that the **error in G from the previous iteration is negligible**.
4. When possible, **normalize your data** so that the spectrum integrates to one.
5. Calculate the covariance of the data making sure that: (1) the eigenvalue spectrum is continuous (if not, increase Nbins), and (2) that the covariance matrix is well enough conditioned to allow it to be diagonalized (if not, the data is oversampled in Matsubara time).
6. Diagonalize the covariance, rotate the data and kernel into the diagonal frame.
7. Choose a good default model, hopefully you can **use the annealing technique. U**se a non-informative model unless the information in the model is exact.
8. When possible, **use Bryan's MEM** for marginalizing over α.
9. Systematically improve your data until the calculated spectrum converges.
10. When the annealing method is used, if the temperature step appears large (i.e. the spectrum changes abruptly) introduce data at additional intermediate temperatures.
11. If your model is not exact, **try different non-informative default models**. A reliable result is independent of the model. You may also want to use the model with the highest posterior probability (Bryan's method).

Computational Materials Science in Louisiana

- International Computational Materials Science Seminar Series
- Shared graduate courses in Computation and Materials
- Google LA-SiGMA (or visit http://lasigma.loni.org/)

For more information, visit:
http://lasigma.loni.org

LA-SiGMA is supported by the National Science Foundation award number #EPS-1003897.

Mardi Gras Workshop