Maximum Entropy Method Implementation Using MATLAB

Jooyoung Park
School of Electrical and Computer Engineering, Cornell University, Ithaca, NY

I. Introduction

Calculating spectral properties of large matrices is an extremely demanding in regards to computational power of CPUs. *Kernel polynomial method* (KPM) is an efficient linear Chebyshev approximation to this spectra using Chebyshev moment data [1]. However, KPM suddenly truncates Chevyshev series data, resulting in Gibbs phenomenon. To reduce the discrepancies between an actual information and the approximation, we use the *maximum entropy method* (MEM) approach. MEM solves a convex non-linear optimization problem where the solution maximizes the information theoretic relative entropy of the spectrum subject to data constraints [2].

A. Contributions

An implementation of MEM described in [2] is presented here. The MEM program is based on a KPM function implemented in MATLAB. Chebyshev moments data acquired by the KPM is fed as an initial moments information into the MEM. First, at section Sec. II ,we go over mathematics and algorithms of MEM in [2]. Then, the MATLAB implementation of the MEM is briefly described in Sec. III. Consequently, we compare the numerical outcomes of the KPM, MEM, and the original information.

B. Notation

Lowercase and uppercase boldface letters designate vectors and matrices, respectively. For a matrix $\mathbf{H}$, we define its inverse to be $\mathbf{H}^{-1}$. Entry of m-th row and m'-th column of a matrix $\mathbf{H}$ is denoted as $H_{mm'}$. The m-th entry of a vector $\vec{\boldsymbol{\lambda}}$ is λ_m.

II. The Maximum Entropy Method Overview

In this section, we succinctly review the maximum entropy method described in [2]. Consider a function of KPM to a density of states,

$$D(x) = \frac{1}{\pi\sqrt{1-x^2}}\left[1 + \sum_{m=1}^{M} \hat{\mu}_m g_m^M T_m(x)\right] \tag{1}$$

Let $\phi = \cos^{-1}(x)$. Then $T_m(x) = \cos(m\phi)$. Define $D(\phi) = \sin(\phi)D(x)$ then density of states in ϕ domain is given by,

$$D(\phi) = \frac{1}{\pi}\left[1 + \sum_{m=1}^{M} \hat{\mu}_m g_m^M \cos(m\phi)\right] \tag{2}$$

We can evaluate the Chebyshev moments by L point fast Fourier transforms (FFT) on $D(\phi)$,

$$\mu_m \overset{(a)}{\approx} \sum_{l=0}^{L} \cos(m\phi_l) D(\phi_l) \Delta\phi \tag{3}$$

where $\cos((L+1)\phi_l) = 0$ is satisfied for $0 \le l \le L$. We choose $L = M \times K \times I$ so as to match $M \times K$ degree moments of kernel polynomials $\hat{\mu}_m g_m^{M\times K}$ and I for accuracy improvement. Nyquist-Shannon sampling theorem says (a) is exact only when $D(\phi)$ is band-limited. So naturally, the goal here

is to minimize numerical errors on the high frequency components of MEM solutions. The quality of moments data is given by χ^2 statistics,

$$\chi^2 = \sum_{m=0}^{M} \left(\frac{\hat{\mu}_m - \mu_m}{\sigma_m} \right)^2 \tag{4}$$

where $\hat{\mu}_m$ and μ_m are m-th Chebyshev moments data from KPM and its counterpart data from MEM, consecutively. σ_m is standard deviation of m-th Chebyshev moment samples. The optimization problem of minimizing numerical errors is to maximize the relative entropy given by,

$$S = \int_0^\pi \left(D(\phi) - D_\circ(\phi) - D(\phi) \ln \left(\frac{D(\phi)}{D_\circ(\phi)} \right) \right) d\phi \tag{5}$$

subject to data constraints. $D_\circ(\phi)$ is the density of states from KPM without the knowledge of initial data. This becomes the problem of primal optimization problem given by maximizing,

$$Q_p = S - \frac{\chi^2}{2\alpha} \tag{6}$$

where parameter α sets a balance between the fit, χ^2, and the information gap between $D(\phi)$ and $D_\circ(\phi)$, $-S$. This is a convex optimization problem, however, difficult to solve in practice. A dual optimization problem solves the same problem in a relatively simpler way for implementation. A $\vec{\boldsymbol{\lambda}}$ parameter is introduced here,

$$\lambda_m = \frac{\mu_m - \hat{\mu}_m}{\alpha \sigma_m^2} \tag{7}$$

Then the maximum entropy $D(\phi)$ satisfying primal optimization problem is,

$$D(\phi) = D_\circ(\phi) \left(- \sum_{m=0}^{M} \lambda_m \cos(m\phi) \right) \tag{8}$$

The dual problem is to maximize,

$$Q_d = \left(\int_0^\pi D(\phi) d\phi \right) + \sum_{m=0}^{M} \left(\hat{\mu}_m \lambda_m + \frac{\alpha \sigma_m^2 \lambda_m^2}{2} \right) \tag{9}$$

Define,

$$\begin{aligned} \xi_m &= \frac{\partial Q_d}{\partial \lambda_m} \\ &= \hat{\mu}_m - \mu_m + \alpha \sigma_m^2 \lambda_m^2 \end{aligned} \tag{10}$$

Here, (7) is satisfied when $\xi_m = 0$. Then, (7) can be solved by Newton-Rahpson iteration,

$$\mathbf{H}[n](\vec{\boldsymbol{\lambda}}[n+1] - \vec{\boldsymbol{\lambda}}[n]) = \vec{\boldsymbol{\xi}}[n] \tag{11}$$

beginning from step $n = 0$. $\mathbf{H}[n]$ is the Hessian of the dual problem (9),

$$\begin{aligned} H_{mm'} &= \frac{\partial^2 Q_d}{\partial \lambda_m \partial \lambda'_m} \\ &= \frac{\mu_{m+m'} + \mu_{|m-m'|}}{2} + \alpha \sigma_m^2 \delta_{mm'} \end{aligned} \tag{12}$$

Then,

$$Q_d = Q_p + \sum_{m=0}^{M} \frac{\xi_m^2}{2\alpha\sigma_m^2} \tag{13}$$

Here, $Q_d > Q_p$ and the typical stopping criterion is 2% difference between the two.

III. Maximum Entropy Method Algorithm Implementation

The algorithm for maximum entropy method is described here.

Algorithm 1: Maximum Entropy Method Algorithm

Data: $\hat{\boldsymbol{\mu}}$, g^M, and $\hat{\boldsymbol{\sigma}}$ from KPM
Result: $\vec{\boldsymbol{\mu}}$ that matches extrapolated $\hat{\boldsymbol{\mu}}\mathbf{g}^{M\times K}$ by maximizing relative entropy under constraints

```
%initialization%
μ⃗ ← concat(μ̂g^M, {0}^{M×(K−1)});
Calculate initial χ², τ ← 1/2, α ← χ², λ⃗, ξ⃗;
D_∘[0 : L] ← Initial data using equation (2) sampled at L points;
while χ² > M do
    Backup to α_old, χ²_old, ξ⃗_old, λ⃗_old, μ⃗_old;
    while (Q_d − Q_p)/Q_d > 0.02 do
        Calculate ξ⃗, H;
        λ⃗ ← (λ⃗ − H⁻¹_{M×M} ξ⃗);
        D[0 : L] ← Update L point sampled D by equation (8) and D_∘;
        μ⃗ ← Update μ⃗ using equation (3) and D from above;
        Calculate S, χ², Q_p, Q_d in order;
    end
    if χ² > χ²_old then
        τ ← τ/2;
        α ← α_old(1 − τ);
        Restore χ²_old, ξ⃗_old, λ⃗_old, μ⃗_old;
    else
        α ← α(1 − τ);
    end
end
Output D[0 : L];
```

The outer loop in Algorithm 1 starts at $\alpha \leftarrow \chi^2$ where χ^2 is the fit of KPM model. The initial step down size is $\frac{1}{2}$ while the size is reduced by half whenever a instability is detected by a significant increase in χ^2. We also restore the previous conditions as we reduce the step down size. Typical stopping criterion is $\chi^2 < M$.

The inner loop in Algorithm 1 iteratively solves the optimization problems of maximizing (6) and (9) under constraints. On each Newton-Rahpson iteration, $\vec{\boldsymbol{\lambda}}$ is solved, $D(\phi)$ and $\vec{\boldsymbol{\mu}}$ are updated. Use this updated values to calculate Q_p and Q_d and determine if the stopping criterion is met. If not, move on to the next iteration and reuse the previous solutions as the starting point.

IV. Numerical Simulations

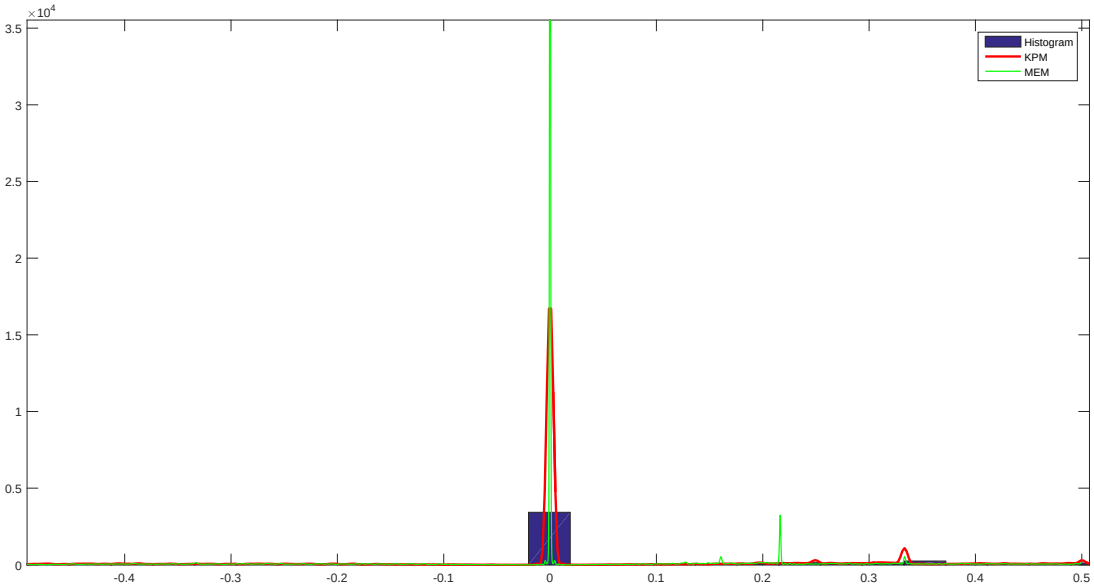

Fig. 1. as19991115 - K = 4, I = 4

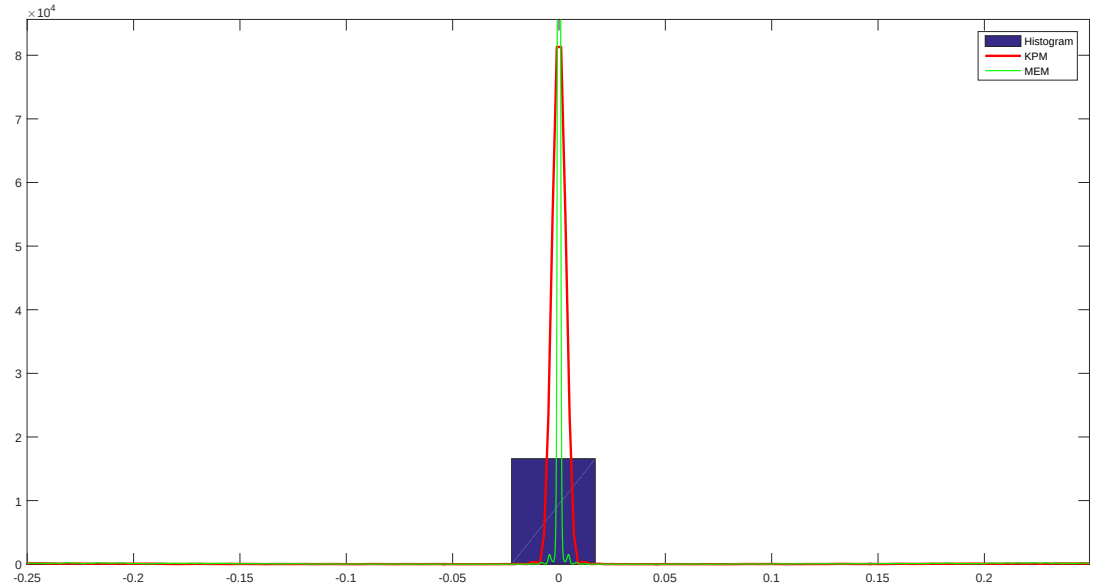

Fig. 2. as-caida20060911 - K = 4, I = 4

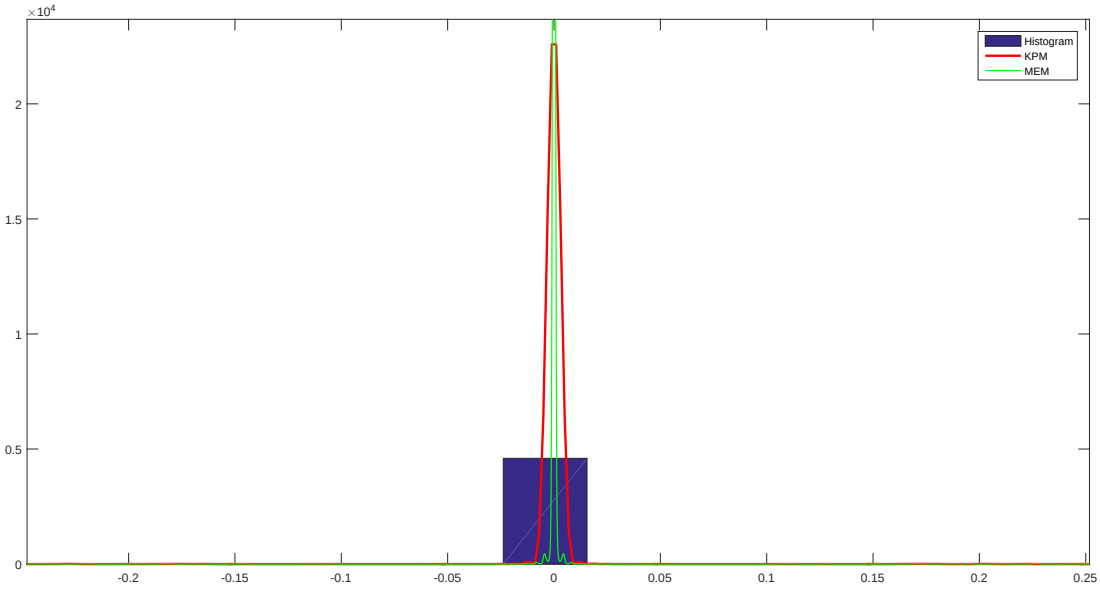

Fig. 3. Erdos02-cc - K = 4, I = 4

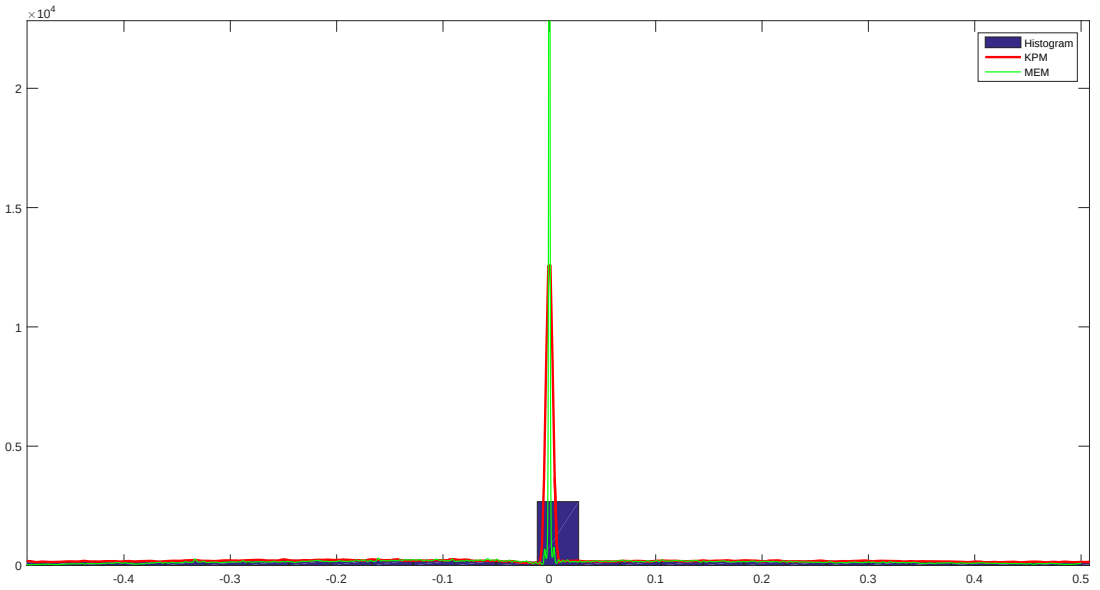

Fig. 4. homo-cc - K = 4, I = 4

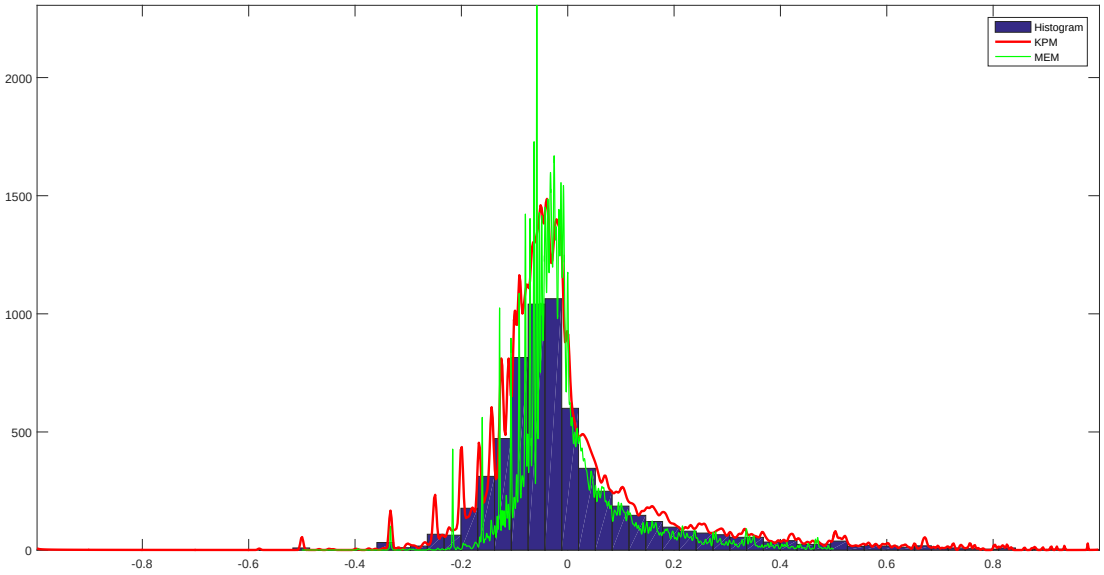

Fig. 5. marvel-chars-cc - K = 4, I = 4

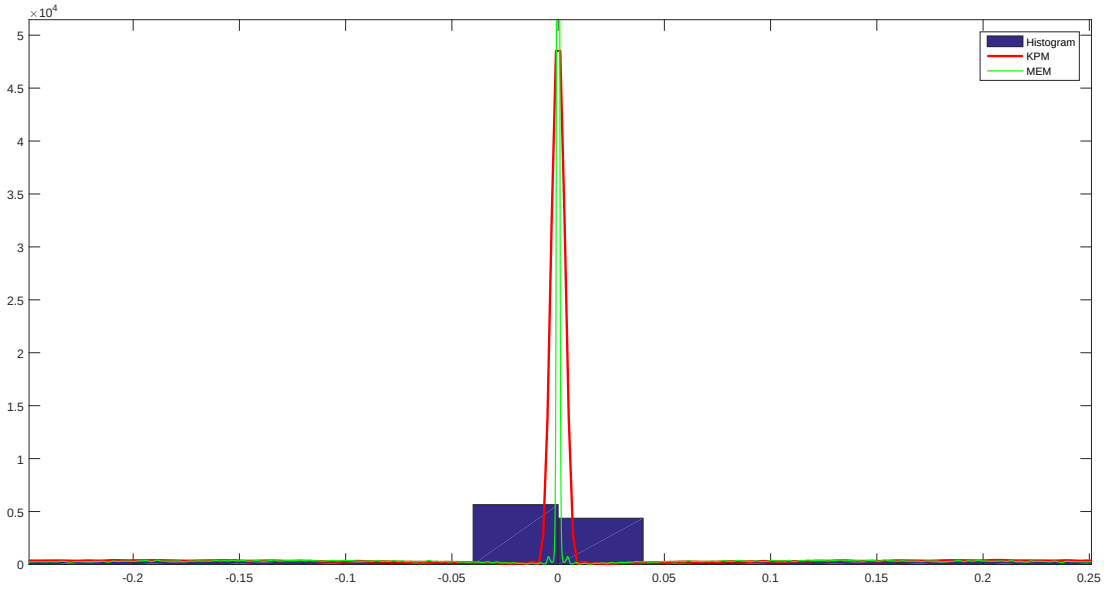

Fig. 6. marvel-comics-cc - K = 4, I = 4

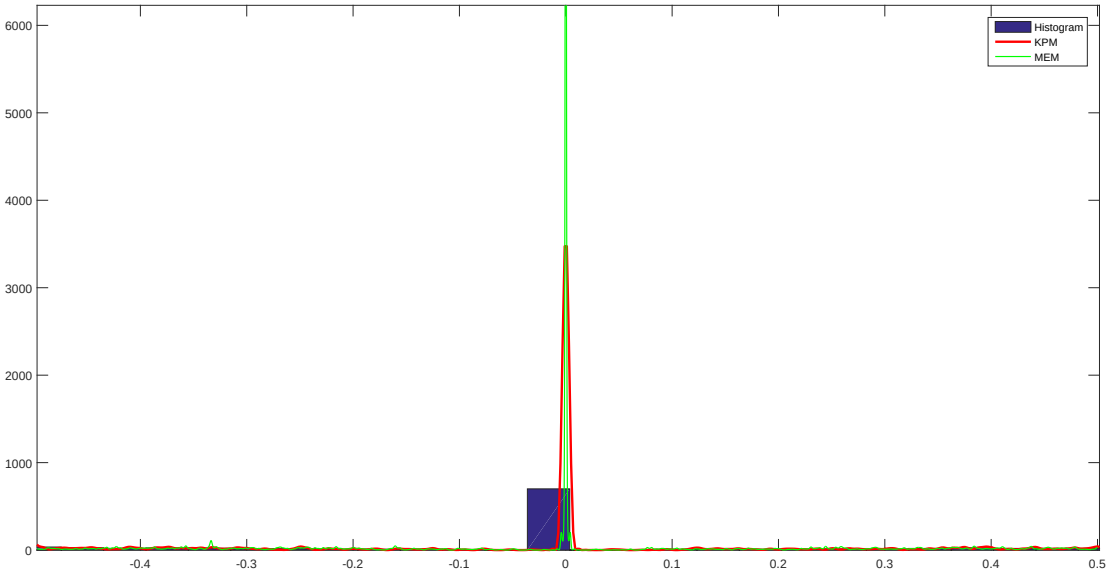

Fig. 7. musm-cc - K = 4, I = 4

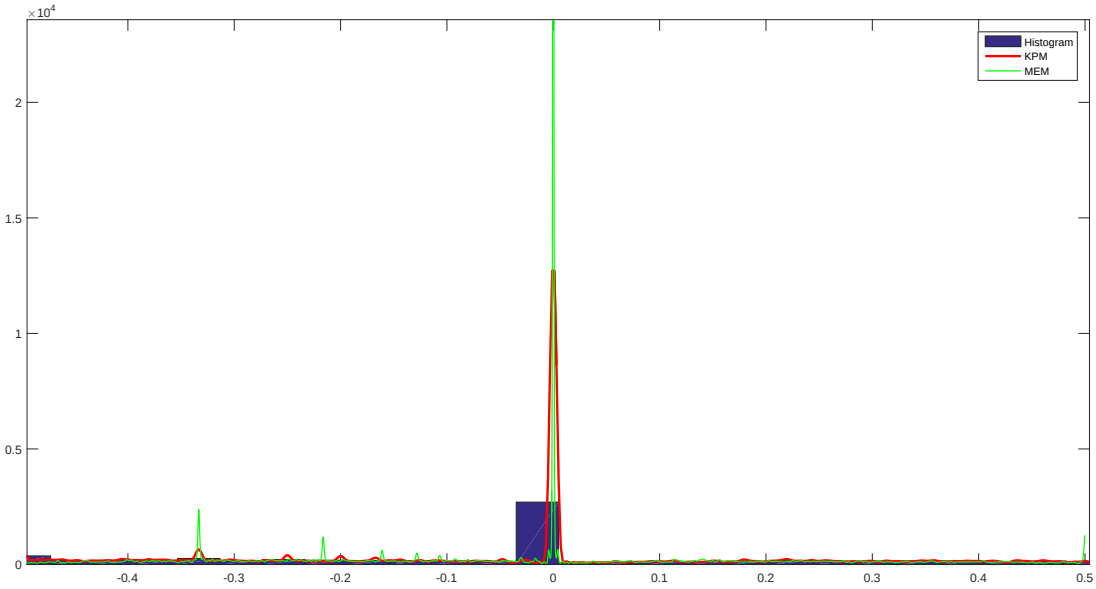

Fig. 8. pgp-cc - K = 4, I = 4

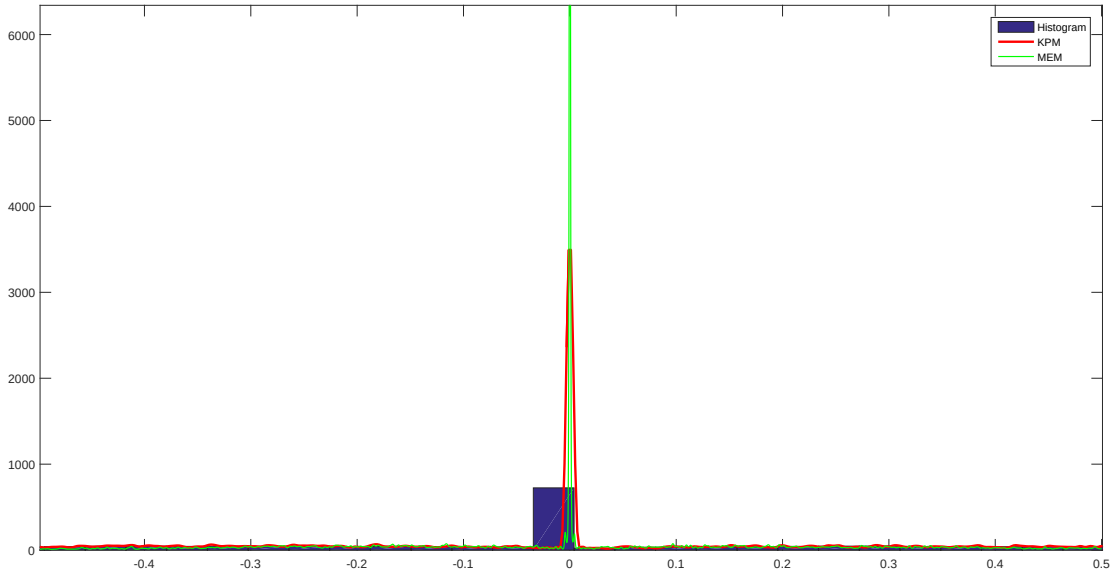

Fig. 9. yeast-cc - K = 4, I = 4

Note that all the simulations were ran with 1000 Chebyshev moments at $K = 4$ and $I = 4$.

V. Conclusion

We have discussed the algorithm that improves resolution of spectral information using maximum entropy method. Estimated Chebyshev density given by the kernel polynomial method is a good starting point for input to the MEM algorithm. In a nutshell, the algorithm balances between the quality of fit (the outer loop) and the information measure of gap (the inner loop) from the default model (e.g. KPM). The parameters can be altered as needed, such as the stopping criteria.
Overall, although not without a few issues (gain difference and negative $D(\phi)$ values), the simulation results shown in Sec. IV agreed the theoretical expectation in Sec. II. In conclusion, MEM is an efficient algorithm that restores the high frequency spectral information, given the initial data at a lower data resolution.

Acknowledgment

This work was supported by Professor David Bindel and based on his prior works.

References

[1] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, "The kernel polynomial method," *Rev. Mod. Phys.*, vol. 78, pp. 275–306, Mar 2006. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.78.275

[2] R. N. Silver and H. Röder, "Calculation of densities of states and spectral functions by Chebyshev recursion and maximum entropy," *Phys. Rev. E*, vol. 56, pp. 4822–4829, Oct 1997. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.56.4822