

Prony methods for extracting excited states

Kimmy K. Cushman*

Yale University

E-mail: kimmy.cushman@yale.edu

George T. Fleming

Yale University

E-mail: george.fleming@yale.edu

We propose an algebraic method for extracting excited states from lattice gauge theory correlation functions. Instead of fitting to a sum of decaying exponentials, we adopt a variant of Prony's method to obtain M energies (exponential decay rates) by finding the roots of an M^{th} order polynomial, and then solving for the amplitudes linearly. The resulting states tend to have overlapping error ellipses, making identification of states ambiguous. This is especially problematic at large Euclidean times where the signal to noise may be low, as well as when many states are extracted. We propose a variation of K-means clustering to identify each extracted state.

The 36th Annual International Symposium on Lattice Field Theory - LATTICE2018
22-28 July, 2018
Michigan State University, East Lansing, Michigan, USA.

*Speaker.

1. Introduction

Many lattice gauge theory analyses involve excited state extraction. Modeling the first excited state well helps produce a better variational estimate of the ground state, regardless of whether the excited states are the focus of the study. Additionally, the excited state spectrum encodes two particle scattering phase shifts through the Luscher method [1–3]. By extracting the spectrum of two particles in a box, this formalism provides a connection to the infinite volume continuum of two particle elastic scattering. Also, a robust method for extracting excited states will be valuable to studies of nuclei because their spectrum will be denser compared to hadrons [4, 5]. Additionally, extracting excited states from confining gauge theories is crucial for beyond the standard model lattice explorations such as composite Higgs or composite dark matter [6, 7].

Excited state extraction often relies on non-linear least squares fitting to decaying exponentials. Finding an acceptable solution can be hard, requiring experienced user input to guide the minimization, particularly when extracting excited states [8, 9]. Also, plateau finding in effective mass plots with one or more extracted states is inherently subjective. We present a black-box method which requires no user input. Instead of fitting to exponentials, it algebraically extracts a fixed number of states proportional to the number of input data points [10]. Contrary to a least squares fit, this black-box method is exactly constrained. Even for a ground state extraction, our method may be preferable because it provides an earlier time approximation to the ground state compared to a typical effective mass plateau. Furthermore, since this algorithm only requires that the signal be decomposed into a sum of exponentials, our method may be useful to any application where such a signal must be analyzed.

2. State Extraction Method

In this section we outline the algebraic method for extracting M states from correlation function data. The method can be generalized to a correlation matrix using Matrix Prony (see [11–13]), but here we focus on extraction from a single correlation function in order to address the challenges of overlapping errors. First we set up notation by reviewing the ground state effective mass. We then introduce Prony’s method and apply it to bootstrap samples of lattice correlation function data.

2.1 Effective Masses

Consider a two point function (one element of a correlation matrix) of the following form :

$$C_{ij}(t) = \langle 0 | \mathcal{O}_i^\dagger(t) \mathcal{O}_j(0) | 0 \rangle \quad (2.1)$$

$$= \sum_{m=0}^{\infty} \langle 0 | \mathcal{O}_i^\dagger | E_m \rangle e^{-E_m t} \langle E_m | \mathcal{O}_j | 0 \rangle \quad (2.2)$$

$$\Rightarrow C_{ij}(t) = \sum_{m=0}^{\infty} a_{ij,m} e^{-E_m t}. \quad (2.3)$$

Note that the sum is invariant under permutation of labels of states m . As we will see, this is problematic when the error regions of pairs (a_m, E_m) overlap because the labeling can be ambiguous.

However, a simple method to find the ground state energy, E_0 is

$$E_0 = \log\left(\frac{C(t)}{C(t+1)}\right). \quad (2.4)$$

Although this approximation becomes more accurate at larger times, for particles more massive than the pion, the signal to noise ratio decreases exponentially [14, 15]. Therefore, the ground state energy is typically found by first finding an effective mass plateau, but this has inherent systematic bias, and the result will change depending on who selects the plateau.

2.2 Multiple State Extraction

To extract M states, one must find the $2M$ parameters a_m and E_m in Eq. 2.3, so one needs a minimum of $2M$ timeslices from the correlation function. The minimal set of consecutive timeslices of the data can be written as a length $2M$ vector $y_n(t)$:

$$\begin{aligned} y_n(t) &\equiv C(t+n) \\ &= \sum_m^M a_m e^{-E_m t} e^{-E_m n} \end{aligned} \quad (2.5)$$

$$\Rightarrow y_n(t) = \sum_m^M A_m(t) z_m^n, \quad (2.6)$$

where we define $A_m(t) = a_m e^{-E_m t}$ and $z_m = e^{-E_m}$, and the data is modeled as a sum of only the first M states. We can write this as the following matrix equation:

$$\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{2M-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_M \\ z_1^2 & z_2^2 & \cdots & z_M^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{2M-1} & z_2^{2M-1} & \cdots & z_M^{2M-1} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{pmatrix}. \quad (2.7)$$

In this algebraic algorithm, we will solve for the z 's algebraically with Prony's method and then solve this simple linear problem as usual. Prony's original result [16, 17] can be written in the following way: the non-linear step to find the z 's can be done by solving an M^{th} order polynomial from the so-called Hankel matrix determinant

$$0 = \begin{vmatrix} y_0 & y_1 & \cdots & y_{M-2} & y_{M-1} & 1 \\ y_1 & y_2 & \cdots & y_{M-1} & y_M & z \\ y_2 & y_3 & \cdots & y_M & y_{M+1} & z^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ y_M & y_{M+1} & \cdots & y_{2M-2} & y_{2M-1} & z^M \end{vmatrix}. \quad (2.8)$$

The problem of solving for the energies is then to solve for the roots of a polynomial. This black-box method is exactly constrained and we do not input any initial guesses to the energy and amplitude parameters, as one must do when fitting to exponentials.

2.3 Excited State Extraction on Real Data

We show our use of Prony's method on actual correlation function data obtained from the LSD collaboration's investigation of an $SU(3)$ gauge theory with $N_f = 8$ degenerate flavors [7]. Figs. 1, 2, and 3 show the extraction of z_m and a_m for two and three state extractions for the pseudoscalar point-point correlator. Note that the linear Eq. 2.7 is solved for A_m after the roots z_m are found. Here we are showing the extraction of the original amplitudes $a_m = A_m e^{E_m t}$ from Eq. 2.3.

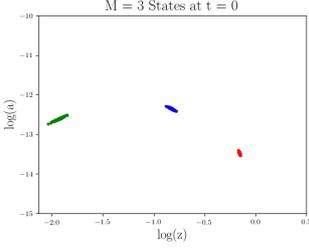


Figure 1: Three state extraction for $y(t=0)$ with well separated error ellipses.

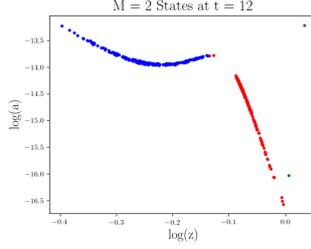


Figure 2: Two state extraction for $y(t=12)$ with overlapping error ellipses.

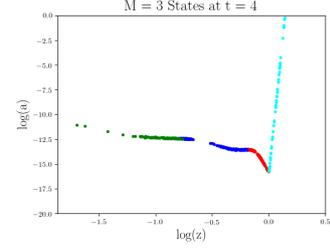


Figure 3: Three state extraction for $y(t=4)$ with overlapping error ellipses.

In this extraction, two (three) states were identified from each of 232 bootstrap samples of the $C(t)$ data, and assigned to two (three) distinct states. The state assignment here is determined by the extracted values of z_m . The red points are identified as the ground state, and the blue points are identified as the first excited state. The two green points in the two state extraction are labeled as backwards propagating ground states. The green and cyan points in the three state extraction correspond to the second excited and backwards propagating ground states, respectively. These appear to be the correct assignments for $t=0$ in Fig. 1 because the distributions do not overlap. However, the signal to noise ratio drops as time goes on, so the distribution of the extracted parameters z_m and a_m start to overlap (see Figs. 2 and 3). Hence, the initial state assignments fail. In order to resolve the best fit parameters a_m and z_m and their errors, we must find the most likely state assignment for each bootstrap sample.

3. Clustering

In order to re-cluster the points into their most likely states, we will use a variation of the K-means cluster algorithm with expectation maximization. Each of the initial clusters as above is fit to a bivariate Gaussian with probability distribution

$$p(\vec{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})} \quad (3.1)$$

where $\vec{\mu}$ is the mean location of the cluster, and Σ is the covariance matrix. Using expectation maximization, a point at \vec{x} would be assigned to cluster 1 instead of cluster 2 if

$$p_1(\vec{x}) > p_2(\vec{x}) \quad (3.2)$$

$$\Rightarrow d_1(\vec{x}) + \log|\Sigma_1| < d_2(\vec{x}) + \log|\Sigma_2|, \quad (3.3)$$

where $d(\vec{x}) = (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})$ is the “distance” in standard deviations squared between \vec{x} and $\vec{\mu}$. The log term is a distance penalty for a cluster being larger; a point the same number of standard deviations away from two clusters would be assigned to the smaller cluster. Here, we use a clustering algorithm that looks at each bootstrap sample individually. For M states extracted from the correlations functions, Prony’s method provides M points which come from M unique clusters. Therefore our algorithm is as follows:

1. Identify initial clusters.
2. Compute the mean and covariance matrix for each cluster.
3. For each bootstrap sample, reassign the M points into M distinct clusters by minimizing the total distance as in Eq. 3.3.
4. Steps 2 and 3 are repeated until the clusters no longer change.

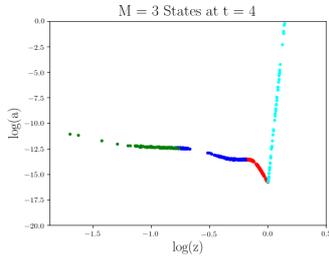


Figure 4: Initial cluster assignment as in Fig. 3.

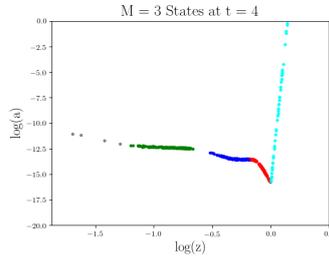


Figure 5: After three iterations of clustering

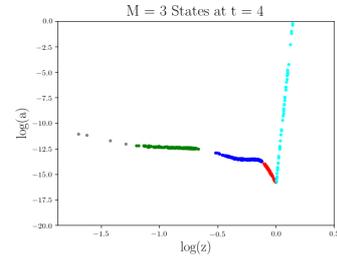


Figure 6: After 13 iterations of clustering

Figs. 4-6 show our implementation of the above algorithm on the $M = 3, t = 4$ states. This set of clusters was chosen as an instructive example because the “correct” cluster assignments can be determined by observation due to the small gaps between clusters. The algorithm appears to recover the clusters correctly. After three iterations, the second excited state in green is separated from the first excited state, and after 13 iterations, the first excited state is separated from the ground state.

Misestimation of the bivariate Gaussian parameters is highly susceptible to outliers, and the distributions that we are modeling have large tails. In the above cluster re-assignments, we also introduced a “background cluster” to account for this effect. The probability density to be in the background is uniform and simply the inverse of the area, A , of a rectangular region containing the data. In the algorithm, this simply becomes another cluster, and the “distance” is computed as twice the negative log likelihood as in Eq. 3.3. That is, $d(\vec{x}) = -2\log(1/A)$. The points assigned to the background cluster are interpreted as noise that has a higher probability of being something we are not modeling. The grey points in Figs. 5 and 6 have been assigned to the background cluster.

4. Results and Discussion

In this section we report the effective masses of the two and three state extractions and compare them to the simpler one-state extraction. We then discuss how the algorithm may be improved.

4.1 Preliminary state extractions

In Figs. 7 and 8 the two and three state effective masses are plotted as a function of time. For $M = 2$, the ground state has very little excited state contamination; it approaches its infinite time or

true value much earlier than a typical ground state effective mass. The extracted first excited state is absorbing most of the excited state contamination. Similarly, in the $M = 3$ extraction, we see most of the excited state contamination has been removed from the first excited state. In both plots we can see the backwards propagating state in cyan.

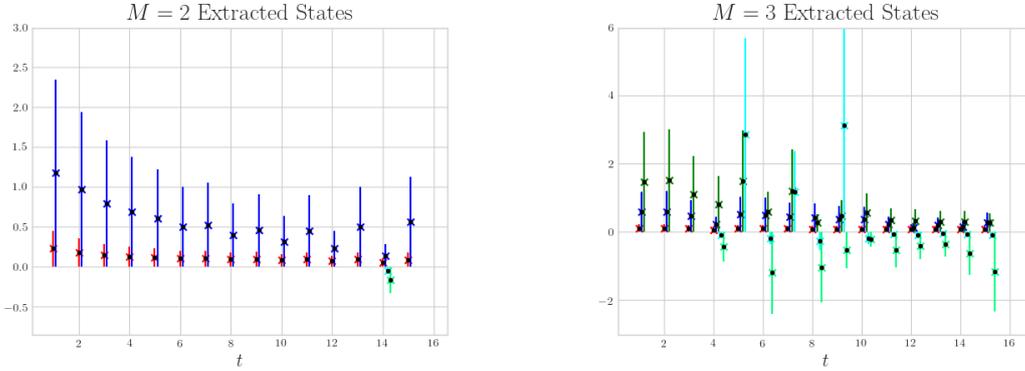


Figure 7: Two state extraction results for $y(t = 12)$. Figure 8: Three state extraction results for $y(t = 4)$.

In Fig. 9 we compare the extraction of the ground state as a function of time for different numbers of total extracted states. As expected, the more states we account for, the faster the mass approaches its infinite time or true value. This is because the ground state has less excited state contamination for two and three state extractions, so a reasonable plateau could be assigned to these states at much earlier times as compared to the single state extraction.

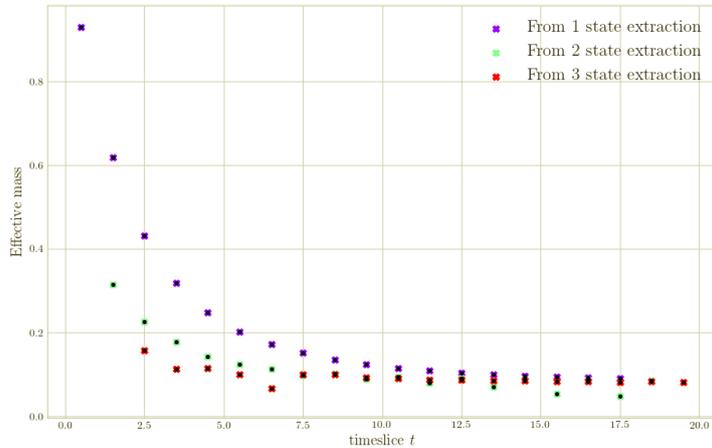


Figure 9: Ground state extractions with $M = 1, 2, 3$. Bootstrap errors are smaller than the size of the points.

4.2 Discussion

One could do studies on fake data. However, using a simple approach to add noise to fake data, we found it was too easy to cluster overlapping distributions of the resulting parameters. Producing fake *but realistic* lattice correlation function data is very difficult because lattice noise is highly correlated, so Gaussian clusters do not capture the complexity of real data.

To improve this method, we would like to account for the fact that these clusters are non-Gaussian. One method we would like to try to account for the non-Gaussian shaped of the clusters is by using the so-called Tukey depth [18] as a measure of distance from a cluster. As a generalization of percentiles to multiple dimensions, the Tukey depth has the benefit of being a non-parametric statistic, so a model for the distribution is not required. The Tukey distance is defined by a curve such that all of its tangent planes have a certain fraction of the data on either side. Tukey states in his original work that this set of planes forms a closed curve in two dimensions. The Tukey depths of Gaussian distributed data would replicate their standard deviation ellipses.

The Tukey distance may be a better metric for this data set. Thus, it may improve the efficiency of the clustering algorithm and increase our ability to extract more states. The Prony method allows us to extract M states from $2M$ pieces of data. For N_t timeslices, we should be able to extract up to $N_t/2$ states, and we hope to obtain estimates for several of the first excited states.

5. Conclusion

We have presented an analytic method for extracting excited states from lattice correlation function data which in the absence of noise would yield Prony's exact result. This algorithm has the benefit of being an exact algebraic extraction, rather than being a best-fit search in a high-dimensional parameter space. However, with realistic, noisy correlation function data, to do any kind of error analysis, we must assign a consistent set of label across a set of ensembles, and hence clustering the distributions of parameters is required. With these challenges in mind, this method may still be preferable to conventional approaches because it has the possibility to remove all user input to the process of extracting excited states.

6. Acknowledgements

This material is based upon work that is supported by the Visiting Scholars Award Program of the Universities Research Association. KC acknowledges support from the United States Department of Energy through the Computational Sciences Graduate Fellowship (DOE CSGF) through grant number DE-SC0019323. GF acknowledges support from the United States Department of Energy through grant number DE-SC0019061. We thank the Lawrence Livermore National Laboratory (LLNL) Multiprogrammatic and Institutional Computing program for Grand Challenge allocations and time on the LLNL BlueGene/Q supercomputer. We also thank Argonne National Laboratory (ANL) for allocations through the ALCC and INCITE programs. Computations for this work were carried out in part on facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy (DOE).

References

- [1] Zohreh Davoudi. The path from finite to infinite volume: Hadronic observables from lattice QCD. In *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018*, 2018.
- [2] Martin Lüscher. Volume dependence of the energy spectrum in massive quantum field theories. *Communications in Mathematical Physics*, 105:153–188, 1986.

- [3] Martin Lüscher. Two-particle states on a torus and their relation to the scattering matrix. *Nuclear Physics B*, 354:531–578, 1991.
- [4] Emanuele Mereghetti. Lattice QCD and Nuclear Physics for BSM searches. In *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018*, 2018.
- [5] Ning Li. New lattice interaction and the spectrum of light and medium mass nuclei. In *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018*, 2018.
- [6] Oliver Witzel. Review on composite higgs models. In *Lattice field theory. Proceedings, 36th International Symposium, Lattice 2018, East Lansing MI, USA, July 22-28, 2018*, 2018.
- [7] T. Appelquist et al. Nonperturbative investigations of SU(3) gauge theory with eight dynamical flavors. *Phys. Rev.*, D99(1):014509, 2019.
- [8] George T. Fleming, Saul D. Cohen, Huey-Wen Lin, and Victor Pereyra. Excited-State Effective Masses in Lattice QCD. *Phys. Rev.*, D80:074506, 2009.
- [9] George Tamminga Fleming. What can lattice QCD theorists learn from NMR spectroscopists? In *QCD and numerical analysis III. Proceedings, 3rd International Workshop, Edinburgh, UK, June 30-July 4, 2003*, pages 143–152, 2004.
- [10] Huey-Wen Lin Saul Cohen, George Fleming. Exponential time series in lattice quantum field theory. In V. Pereyra and G. Scherer, editors, *Exponential Data Fitting and Its Applications*, chapter 4, pages 71–93. Bentham eBooks, 2010.
- [11] Evan Berkowitz et al. Progress in Multibaryon Spectroscopy. In *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018*, 2019.
- [12] Evan Berkowitz, Amy Nicholson, Chia Cheng Chang, Enrico Rinaldi, M. A. Clark, Bálint Joó, Thorsten Kurth, Pavlos Vranas, and André Walker-Loud. Calm Multi-Baryon Operators. *EPJ Web Conf.*, 175:05029, 2018.
- [13] S. R. Beane, W. Detmold, K. Orginos, and M. J. Savage. Nuclear Physics from Lattice QCD. *Prog. Part. Nucl. Phys.*, 66:1–40, 2011.
- [14] G. Peter Lepage. The Analysis of Algorithms for Lattice Field Theory. In *Boulder ASI 1989:97-120*, pages 97–120, 1989.
- [15] Michael G. Endres, David B. Kaplan, Jong-Wan Lee, and Amy N. Nicholson. Listening to Noise. *PoS, LATTICE2011:017*, 2011.
- [16] G. C. F. M. R. de Prony. *J. Ecole Poly*, 1:24, 1795.
- [17] David Vandevoorde. *A fast exponential decomposition algorithm and its applications to structured matrices*. PhD thesis, Rensselaer Polytechnic Institute, Troy, New York, USA, ProQuest Dissertations Publishing, 1996.
- [18] J. Tukey. Mathematics and picturing data. In *Proceedings of the 1975 International Congress of Mathematics*, volume 2, pages 523–531, 1975.