

Using evolutionary algorithms to extract field theory mass spectra

Georg M. von Hippel*, **Randy Lewis**, and **Robert G. Petry[†]**

Department of Physics, University of Regina, Regina, SK, Canada, S4S 0A2

E-mail: rob.petry@uregina.ca

The spectrum of masses from a lattice QCD simulation may be found by fitting exponential functions to correlators of operators possessing the quantum numbers of the particles of interest. The ability of evolutionary algorithms to find globally optimized solutions containing a variable number of states across multiple data sets is exploited to provide a promising solution to the problem of finding these fits.

The XXV International Symposium on Lattice Field Theory

July 30 - August 4 2007

Regensburg, Germany

*New address: DESY, Platanenallee 6, 15738 Zeuthen, Germany.

[†]Speaker.

1. Introduction

To calculate the hadronic energy spectrum using lattice QCD the procedure is in principle straightforward. First one must identify quantum numbers of the channel S of interest which may include the particle's intrinsic angular momentum¹ J , parity P , charge conjugation C , isospin I , etc. Next one constructs the corresponding operators Φ^S and $\Phi^{S\dagger}$ which will destroy and create a state with this symmetry. If one calculates the two-point correlator²

$$G(t) = \langle 0 | T \Phi^S(t) \Phi^{S\dagger}(0) | 0 \rangle , \quad (1.1)$$

then the energy state spectra must be extracted from the result. Assuming periodic boundary conditions and a mesonic operator, the fit function is of the form³

$$G(t) = \sum_{n=0}^{n_{max}} Z_n \left(e^{-E_n t} + e^{-E_n(T-t)} \right) , \quad (1.2)$$

where here T is the temporal extent of the lattice. The problem addressed by this paper is how to find the coefficients and energy states

$$G = \{ (Z_n, E_n) : n = 1, \dots, n_{max} \} , \quad (1.3)$$

which minimize $\chi^2(G)/n_{dof}(G)$. Here, due to timestep correlations, we have the correlated χ^2 involving the covariance matrix σ_{ij} defined by

$$\chi^2(G) = \sum_{t_i, t_j} (\overline{G}_i - G(t_i)) (\sigma^{-1})_{ij} (\overline{G}_j - G(t_j)) , \quad (1.4)$$

$$\sigma_{ij} = \overline{G}_i \overline{G}_j - \overline{G}_i \overline{G}_j . \quad (1.5)$$

Critically one notes that the number of degrees of freedom,

$$n_{dof}(G) = (t_{max} - t_{min} + 1) - 2n_{max} , \quad (1.6)$$

depends on n_{max} , the number of terms in a given fit. Since the latter is unknown, one has a discontinuous optimization problem having a solution space spanning multiple dimensions. We propose the use of an evolutionary algorithm to solve the problem. A complementary discussion of our approach may be found in [6].

2. Evolutionary Algorithms

Evolutionary or *genetic* algorithms use the concept of *natural selection* to solve function optimization problems. (See [6] and references therein.) The terminology reflects this. Candidate

¹More accurately, since one works on the lattice one is interested in the corresponding lattice quantum number Λ which labels an irreducible representation of the octahedral group which corresponds to the quantum number J of the broken continuous rotational symmetry. For instance, $J = 0 \rightarrow \Lambda = A_1$, $J = 1 \rightarrow \Lambda = T_1$. Higher J correspond often to multiple octahedral irreps. See, for example, [1, 2] for mesons and [3, 4] for baryons and references therein.

²For the sake of the discussion assume translationally invariant zero momentum operators [5].

³The asymmetries which arise in baryon correlation functions require only slight modification of this discussion which is restricted for simplicity to meson spectra.

solutions such as $G(t)$ are *organisms*. The internal encoding of the solution is its *genotype*, here $G = \{(Z_n, E_n) : n = 1, \dots, n_{max}\}$. The target function, in our case $f(G) = -\chi^2(G)/n_{dof}(G)$, is the *fitness* of the organism.⁴ Each step in the algorithm produces a new *generation* P_τ of individuals.

There are many ways of implementing an evolutionary algorithm. What we use is representative:

1. Create the first generation P_0 with N randomly generated individuals.⁵
2. Derive $P_{\tau+1}$ from P_τ as follows:
 - (a) *Mutate* each member of the population P_τ with a fixed (small) probability.⁶
 - (b) Select the fittest N_{elite} organisms and a further $N_{diversity}$ random organisms placing all their pairwise *offspring* into $P_{\tau+1}$.⁷
 - (c) Add N_{mutant} forced mutations of random elements in the elite to $P_{\tau+1}$ to explore the solution space around the elite.
3. Repeat until a suitable termination criterion is reached.⁸

In addition to these generic steps one must specify how mutation and breeding are accomplished within the population. For the case of a fit to a single correlator, *mutation* of an individual may include:

- Adding or removing a random element (Z_n, E_n) from the genotype's list.
- Replacing each (Z_n, E_n) by $(Z_n + \Delta Z_n, E_n + \Delta E_n)$ where $(\Delta Z_n, \Delta E_n)$ are random Gaussian deviates.⁹
- Doing a local (e.g. Levenberg-Marquardt) optimization of the fit.¹⁰

Breeding or *crossover* of distinct parent organisms *par1* and *par2* produces two *child* organisms. To produce each child, take corresponding ordered pairs in the parent¹¹ and generate independent

⁴We introduce a minus sign in $f(G)$ to ensure a higher value indicates a fitter organism.

⁵Here $N = (N_{elite} + N_{diversity})^2 + N_{mutant}$, with the latter constants defined in the algorithm.

⁶We do not mutate the fittest organism to ensure that it will survive to the next generation.

⁷Here the offspring of the diagonal "pairs" between identical organisms is considered just a copy of the original organisms themselves and we thereby are including the elite in the next generation.

⁸For instance, one may require a minimum number of generations be exceeded and that a fixed number of generations pass with no improvement in fitness of the best organism. A limit on the maximal number of generations may also be imposed.

⁹Here the standard deviation of the added noise can be tuned to the fitness of our genotype G by making it proportional to $(1 - e^{-\alpha\chi^2(G)/n_{dof}(G)})$ for some fixed α .

¹⁰Inclusion of such Newtonian optimizations into evolutionary algorithms is often found to be useful [7]. We restrict the mutation to a fixed number of steps of the local optimization for the sake of efficiency. Greater efficiency might be achieved by noting that the linearity of a known fit function of the form (1.2) admits local optimization with numerical methods which exploit the ability to separate linear and non-linear parameters [8]. However such methods will not further aid in the discontinuous general problem we are solving of finding said function since here the parameter space is not fixed.

¹¹For parents having different numbers of ordered pairs we copy the extra pairs of the longer parent into the child.

uniformly distributed random numbers $(x, y) \in [-\delta; 1 + \delta]$.¹² The child element becomes:

$$(Z^{child}, E^{child}) = (xZ^{par1} + (1-x)Z^{par2}, yE^{par1} + (1-y)E^{par2}). \quad (2.1)$$

Sample fits to single correlators may be found in [6].

3. Fitting Multiple Correlators

Having shown how to fit a single correlator to extract the spectrum it contains, we now generalize this to discuss the more practical problem of fitting several correlators. Fitting multiple correlators representing a single channel S is desirable as more data allows the resolution of a larger number of states with greater accuracy. One typically creates *many* operators for the channel S through group theory methods¹³ or by transforming existing operators in ways which conserve their channel properties. An example of the former may be seen in figure 1, while quark and link smearing of operators which map $\Phi^S \xrightarrow{\text{smeared}} \Phi'^S$ is an example of the latter. The entire set of operators

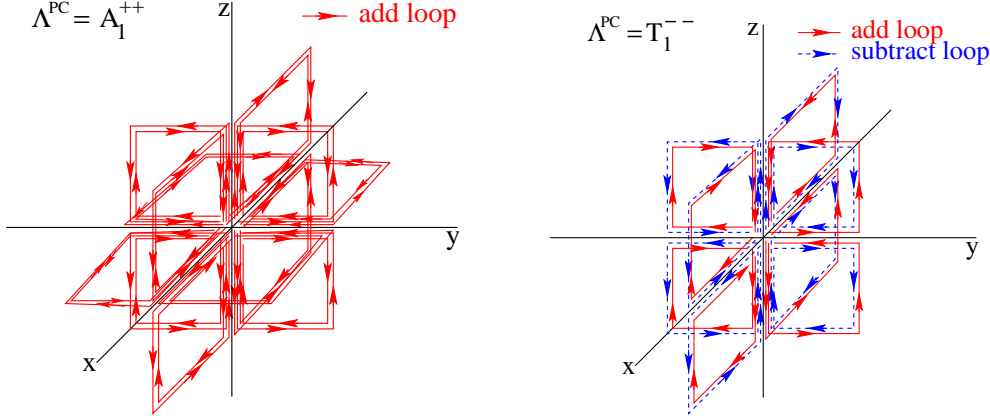


Figure 1: Extended gauge field structures contributing the Λ^{PC} shown may be combined via octahedral Clebsch-Gordan coefficients to quark operators at the origin to create a greater set of operators which couple to desired meson channels. The T_1 “vector” structure on the right has three components; only the one symmetric about the z axis is shown. See [2] and references therein.

corresponding to S allows one in principle to evaluate an entire *correlator matrix* $G_{ij}(t)$ between them:

$$\{\Phi_i^S : i = 1, \dots, i_{max}\} \Rightarrow G_{ij}(t), \quad (3.1)$$

all or some subset of which is to be evaluated and fit.

Since the correlator matrix grows as the number of operators squared, and because the off-diagonal entries have slightly different functional forms, consider the special case of fitting multiple diagonal correlators $\{G_{ii}, i = 1, \dots, i_{max}\}$ with an evolutionary algorithm. This will require changes

¹²The use of δ allows for the possibility of extrapolation in addition to interpolation between the parents’ parameters, thereby avoiding an unwanted rapid contraction to a central point [9].

¹³This may be done from the top down by creating an operator space and using group theory projection to extract operators with quantum numbers S or from the bottom up by using Clebsch-Gordan coefficients to construct the desired operators [3, 4].

to our single-correlator problem. First we modify the genotype with $(Z, E) \rightarrow (Z, I)$, where index I points to a state list $(E_1, \dots, E_{m_{max}})$ common to all the correlators.¹⁴ The full genotype becomes

$$\begin{aligned} \text{Fit Genotype} &= (\text{Dataset coefficients, Energy state list}) \\ &= ((\text{Dataset 1 coefficients, } \dots), \text{Energy state list}) \\ &= (((Z_1^{(1)}, I_1^{(1)}), \dots, (Z_{n_{max}}^{(1)}, I_{n_{max}}^{(1)})), \dots), (E_1, \dots, E_{m_{max}})). \end{aligned} \quad (3.2)$$

The fitness function $f(G) = -\chi^2(G)/n_{dof}(G)$ is modified due to having multiple datasets to

$$\chi^2(G) = \sum_{i=1}^{i_{max}} \chi^2(i), \quad (3.3)$$

$$n_{dof}(G) = n_{data} - m_{max} - \sum_{i=1}^{i_{max}} n_{max}^{(i)}, \quad (3.4)$$

where n_{data} is the product of the number of timesteps fit and the number of correlators. The complexity of the genotype permits enhanced evolutionary operations. As a nested hierarchy of lists, (3.2) admits more complicated list-based mutations and breeding. Integer indices may be bred and mutated bitwise. Finally a reduction mutation which orders masses and coefficients is useful to encourage the algorithm to converge to a single representation of the solution.

In figure 2 the effectiveness of the algorithm to find a known solution is shown. Four synthetic correlators, each with 48 timesteps, were created by adding noise to the model function depicted on the right of the plot. There one sees four masses displaced horizontally with the corresponding coefficients in each respective dataset plotted vertically above the corresponding mass. The left side of the plot shows the best fit of each generation. The plot also displays χ^2/n_{dof} of the best fit (circles) and one sees it converge to 1 as expected as the fit improves.¹⁵ A simultaneous fit to actual data of eight diagonal ρ meson (i.e. $\Lambda^{PC} = T_1^{-}$) correlators¹⁶ is shown in figure 3. Only the energy states are shown of the best fit of each generation up to generation 600. The coefficients in each dataset, a further 30 parameters in the final fit, are not shown. The last column depicts the best fit found with bootstrap errors produced via Levenberg-Marquardt fits to bootstrap configurations with its fixed functional form.¹⁷

4. Advantages of Evolutionary Algorithm Fitting

To conclude, we present advantages of the evolutionary algorithm fitting method. For one it is a global optimization method which is furthermore independent of initial conditions. The solution space is discontinuous as it spans multiple dimensions, arising from the fact that one does not know the exact functional form *a priori*. The evolutionary algorithm fitting method is capable of handling this problem; by minimizing χ^2/n_{dof} it finds the number of states in the data in a natural manner. As well, the ability to identify whether a state exists or not in an individual correlator in a

¹⁴Here the integer index I is taken modulo m_{max} to ensure the coefficient points to an actual energy state.

¹⁵See [6] for further discussion of this plot.

¹⁶Simulation details: Wilson quarks, $\beta = 6.0$, $\kappa = .1554$, $20^3 \times 48$, 600 configurations, quenched.

¹⁷Note that as well a Levenberg-Marquardt optimization was done on the best fit found to produce the final result.

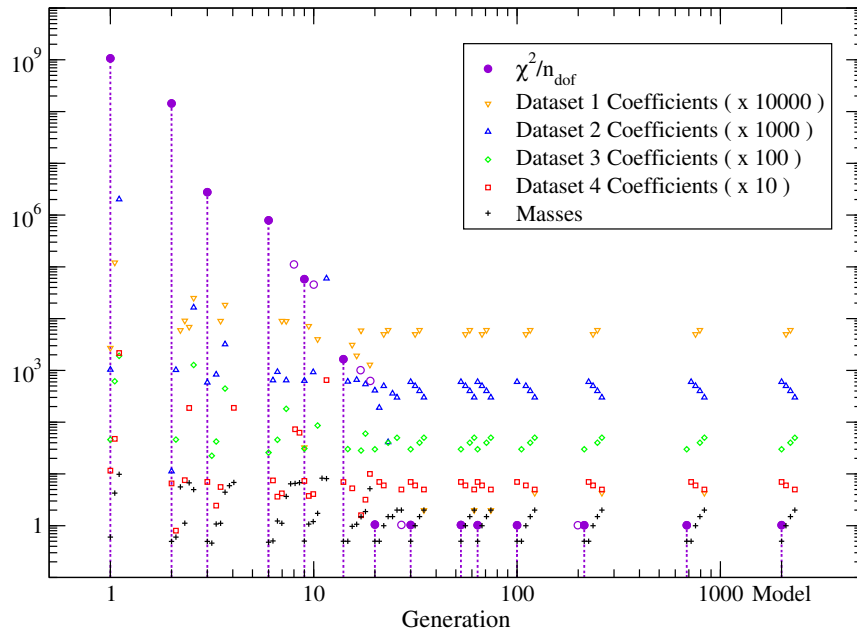


Figure 2: Masses and coefficients of simultaneous fit to four synthetic correlators.

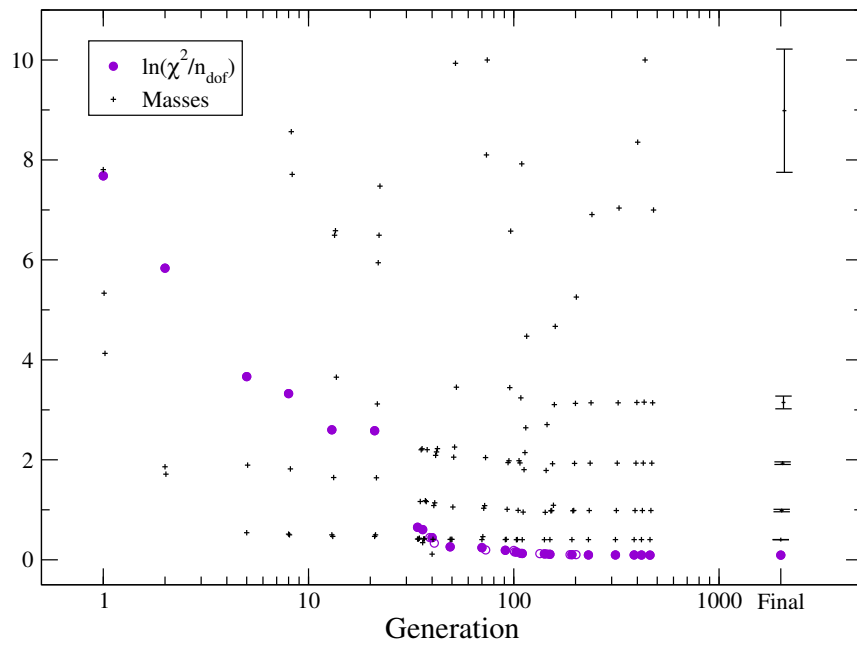


Figure 3: Masses of simultaneous fit to eight ρ meson (T_1^{--}) correlators.

discrete way means that evolutionary fitting can in principle identify which orthogonal irreps Λ a state straddles and hence aids in the identification of its continuous angular momentum J [2].

Computationally, evolutionary algorithms are inherently parallelizable. One can break populations into *islands* breeding on different nodes/CPU's largely independently with only occasional migration between them. Large datasets can also be partitioned with sub-genotypes being initially evaluated and then stitched together for further evaluation on the entire dataset. Evolutionary fitting does not require evaluation of the full correlator matrix, which allows for inexpensive asymmetrical smearing between the sink and source operators.¹⁸ This also means one can restrict oneself to evaluating only the diagonal correlators of the correlator matrix where one expects to have the strongest signals. This in turn allows a wider assortment of operators to be evaluated. Finally, there is the potential for combination with established methods.¹⁹ We are currently applying this evolutionary fitting method to analyze all the operators detailed in [2] with promising results.

Acknowledgements

We thank George T. Fleming for drawing our attention to reference [8], and Richard M. Woloshyn for providing gauge field configurations and propagators [11]. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, the Canada Research Chairs Program and the Government of Saskatchewan.

References

- [1] P. Lacock, C. Michael, P. Boyle, and P. Rowland. Phys. Rev. **D54**, 6997 (1996) [arXiv:hep-lat/9605025].
- [2] D. Harnett, R. Lewis, and R. G. Petry. PoS **LAT2006**, 194 (2006) [arXiv:hep-lat/0609071].
- [3] S. Basak *et al.* Phys. Rev. **D72**, 094506 (2005) [arXiv:hep-lat/0506029].
- [4] S. Basak *et al.* Phys. Rev. **D72**, 074501 (2005) [arXiv:hep-lat/0508018].
- [5] R. Edwards *et al.* Nucl. Phys. Proc. Suppl. **129**, 236 (2004) [arXiv:hep-lat/0309079].
- [6] G. M. von Hippel, R. Lewis, and R. G. Petry (2007) [arXiv:0707.2788 [hep-lat]].
- [7] P. Moscato. *On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms*. Technical Report C3P 826, Caltech, Pasadena, CA (1989).
- [8] G. H. Golub and V. Pereyra. SIAM Journal on Numerical Analysis **10**(2), 413 (1973).
- [9] B. C. Allanach, D. Grellscheid, and F. Quevedo. JHEP **07**, 069 (2004) [arXiv:hep-ph/0406277].
- [10] T. Burch *et al.* Phys. Rev. **D73**, 094505 (2006) [arXiv:hep-ph/0601026].
- [11] A. M. Abdel-Rehim, R. Lewis, and R. M. Woloshyn. Phys. Rev. **D71**, 094505 (2005) [arXiv:hep-lat/0503007]; A. M. Abdel-Rehim, R. Lewis, R. M. Woloshyn, and J. M. S. Wu. Phys. Rev. **D74**, 014507 (2006) [arXiv:hep-lat/0601036].

¹⁸Indeed, our algorithm permits consideration of additional correlators which would not even be allowed in a correlator matrix, namely those for which only a single operator at the sink or source has the definite quantum numbers of channel S . The other operator could potentially couple to many channels beyond that of interest.

¹⁹One can determine priors for Bayesian fits using distributions of parameters spread across population islands [6]. In the variational method, one can diagonalize the operators in the usual way and then fit only the new diagonal correlators with our fitting algorithm. This would maximize the information in a minimized amount of data for the subsequent evolutionary fit. See [10], for example, on the variational method used with meson correlators.