Numerical methods for lattice field theory

Mike Peardon

Trinity College Dublin

August 16, 2007

Monte Carlo measurements of correlation functions

- The mass (for example) of a QCD state can be computed from the Euclidean matrix element $(t' > t)$

$$\langle \Phi(t')|\Phi^\dagger(t)\rangle = \langle \Phi|e^{-(t'-t)\hat{H}}|\Phi^\dagger\rangle = \sum_n \langle \Phi|n\rangle\langle n|e^{-(t'-t)\hat{H}}|n\rangle\langle n|\Phi^\dagger\rangle$$

$$C_\Phi(t'-t) = \langle \Phi(t')|\Phi^\dagger(t)\rangle = \sum_n |\langle \Phi|n\rangle|^2 e^{-(t'-t)\hat{E}_n}$$

- Recall the Euclidean path integral representation of QCD on a lattice relates a matrix element to an integral on a finite lattice:

$$\langle \Phi(t')|\Phi^\dagger(t)\rangle = \frac{1}{Z}\int \mathcal{D}\phi \;\; \Phi[\phi](t')\Phi^\dagger[\phi](t) \;\; e^{-S(\phi)}$$

- And the ergodic theorem related this integral to the expected value of our Markov chain ensemble average:

$$C_\Phi(t'-t) = E[\Phi(t')\Phi^\dagger(t)]_{S(\phi)}$$

- A reweighting function might be needed, for example with $N_f = 1$ light quarks.

Monte Carlo measurements of correlation functions (2)

- A useful means of visualising this data is via the "effective mass"

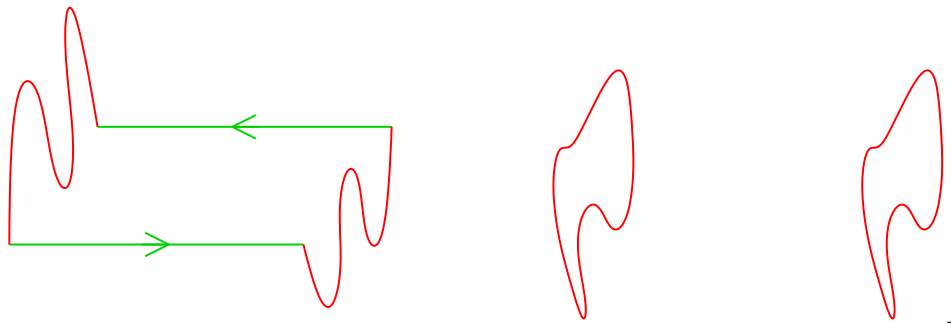
$$am_{\text{eff}}(t) = -\log \frac{C_\Phi(t+1)}{C_\Phi(t)}$$

$$\lim_{t\to\infty} C_\Phi(t+1) = |Z|^2 e^{-E_0 t} \quad \text{so} \quad \lim_{t\to\infty} m_{\text{eff}}(t) = E_0$$

- We look for a "plateau" in the effective mass, which indicates when the correlation function is dominated by the lowest state with significant overlap with Φ.

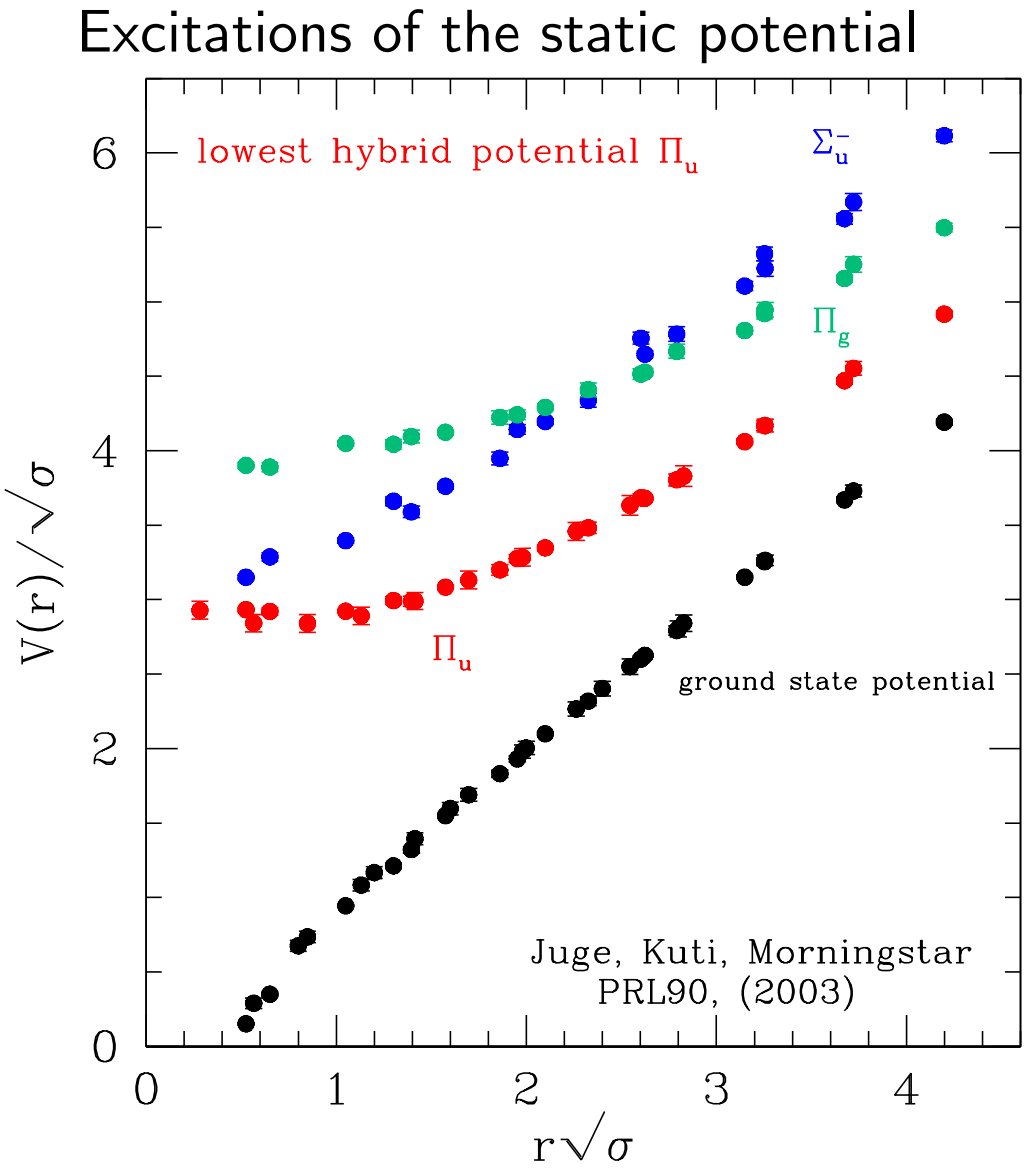
Bosonic fields

- For the bosons, we have direct access to the fields in the ensemble, and can make direct measurements of operators.
- Examples from QCD: The potential between static colour sources, the glueball spectrum.
- A static colour source Q has a propagator defined as the Wilson line at a site x. Two static sources can be combined into a colour singlet by adding a path-ordered product of gauge fields and so the energy of this system (with gluonic excitations too) can be extracted from the time-like decay of Wilson loops:
- For glueballs, an operator can be formed from a linear combination of Wilson loops on a time-slice. The correlator that must be measured is then the product of two of these operators.



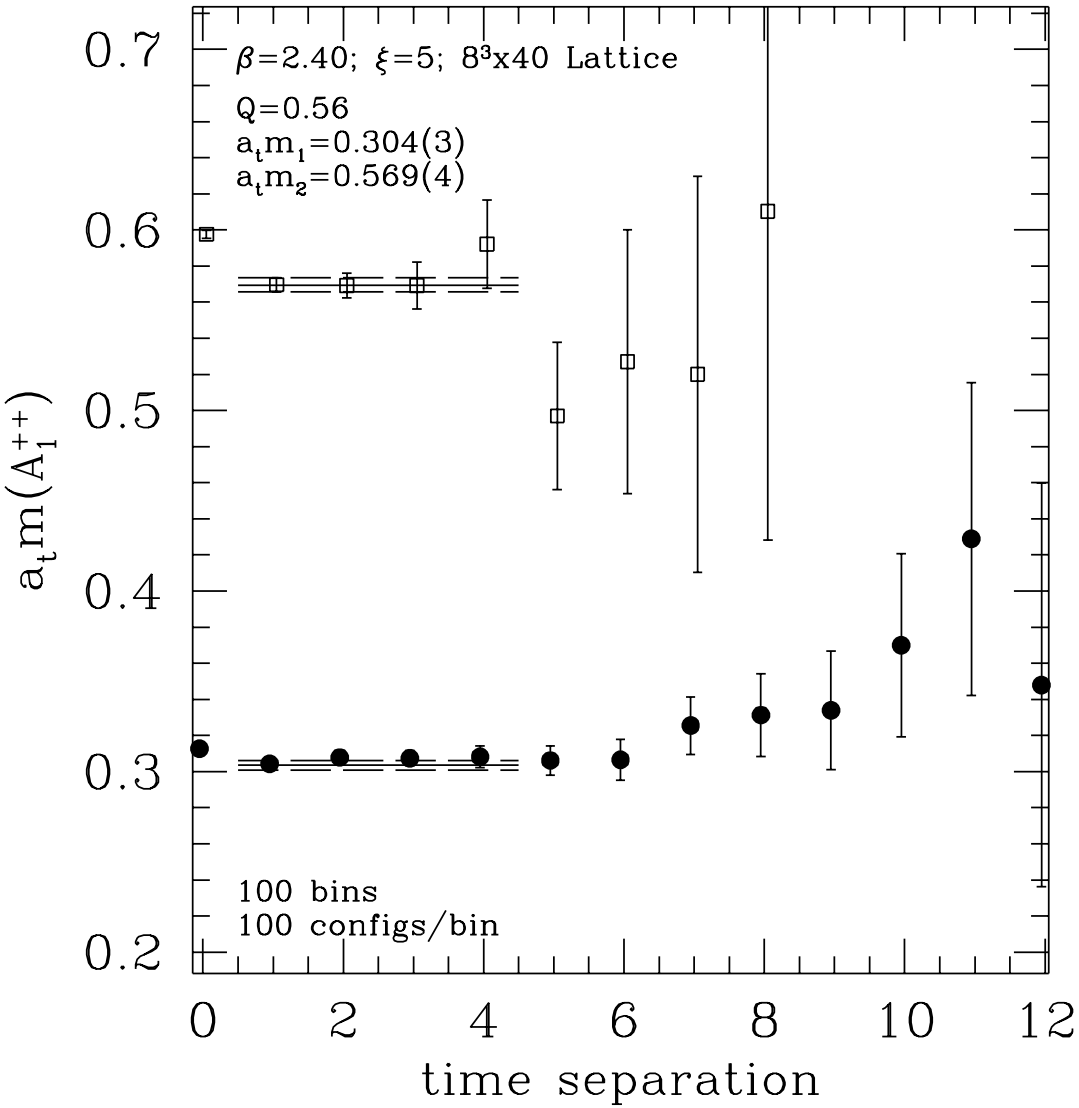
The static potential $\langle$ $\rangle$ The glueball $\langle$ $\rangle$

Bosonic fields



Bosonic fields

An effective mass - the scalar glueball



Quark propagators

- Consider a computation of a two-point correlation function suitable to measure the rho meson energy

$$C_i(t'-t) = \sum_{x,y} \langle \mathrm{Tr}\ M^{-1}(x,t;y,t')\gamma_i M^{-1}(y,t';x,t)\gamma_i \rangle$$

- Computing all elements of the quark propagator would require full knowledge of the inverse - prohibitively expensive
- One solution is to rely on translational invariance of the QCD vacuum and restrict the computation to a few columns of the inverse.

$$C_i(t) = V \sum_{y} \langle \mathrm{Tr}\ M^{-1}(0,0;y,t)\gamma_i M^{-1}(y,t;0,0)\gamma_i \rangle$$

- Then γ_5-hermiticity gives us $M^{-1}(y,t;0,0) = \gamma_5 M^{\dagger -1}(0,0;y,t)\gamma_5$ so

$$C_i(t) = V \sum_{y} \langle \mathrm{Tr}\ M^{-1}(0,0;y,t)\gamma_i\gamma_5 M^{\dagger -1}(0,0;y,t)\gamma_5\gamma_i \rangle$$

Quark propagators (2)

- Now computing this requires just a small subset of elements of M^{-1}
- These elements can be computed using sparse matrix solvers

$$M^{-1}_{ij}(y,t';0,0) = M^{-1}_{ik}(y,t';x,t)\delta_{kj}(x,t;0,0)$$

  acting on 12 appropriate Kronecker delta function sources, so 12 matrix inversions are required for all spin and colour components at a site.
- Simple baryon operators (with three quark fields at a common point) can be constructed in the same way.
- If more complicated diagrams are required, then more sources must be included and more inversions performed.

Smearing

- Point-like operators have overlap with a broad spectrum of states, not just the ground-state. Different operators will have a different pattern of overlaps. Empirically, it is seen that using a **smeared** operator helps to improve ground-state overlap dramatically.
- Gauge-invariant smearing is performed by the repeated application of a covariant, 3d laplace operator to the quark fields in operators, so an operator to create a rho meson at rest would become

$$\rho_i(t) = \sum_x \bar{u}(x,t)\gamma_i \Box^n(x,y;t) d(y,t)$$

with

$$\Box(x,y) = \delta_{x,y} + \alpha \sum_\mu U_\mu(x)\delta_{x+\hat{\mu},y} + U^\dagger_\mu(x-\hat{\mu})\delta_{x-\hat{\mu},y}$$

- The smeared field inherits the same transformation properties under rotations on the lattice so the quantum numbers of the state created are unchanged. $\{\alpha, n\}$ are free parameters that can be tuned to optimise overlap.

Smearing (2)

- Alternatively, smearing can be performed by first fixing to a smooth gauge and convolving the fields with a well-chosen wave-function.
- The new, smeared operator $\tilde{\rho}$ now couples fields at more than one point. How can the correlation function be computed with point propagators?
- For the sink operator, this is straightforward; the smearing can be applied to the result vectors coming from the inversion. This leads to a "smeared-local" correlator;

$$\tilde{C}_i(t) = V \sum_{y,z} \langle \mathrm{Tr}\ M^{-1}(0,0;z,t) \overleftarrow{\Box}\,(z,y;t)\gamma_i\gamma_5 M^{\dagger -1}(0,0;y,t)\gamma_5\gamma_i \rangle$$

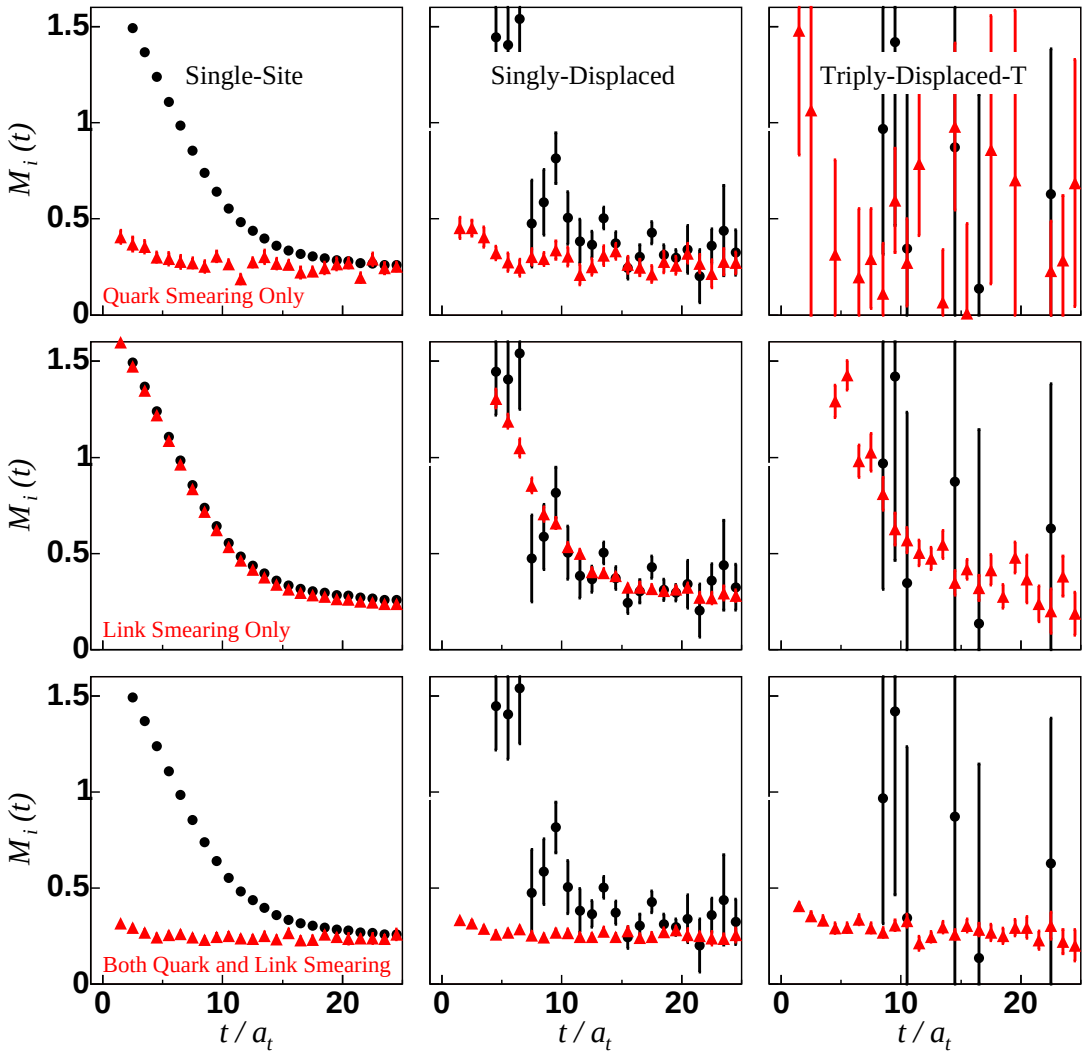
- Smearing at the source will require applying the operator to the Kronecker delta source before applying the matrix inverse. For each different smearing, a new inversion is required.

Smearing (3)

- Often, better operators can be constructed by also applying a smearing to the gauge fields in the operator too. Operators will include gauge fields once invariant smearing is applied, or if displacements are included.
- APE smearing is the most used. Here, a link is added to a weighted average of the four (no time) staples. The resulting variable is projected back into $SU(3)$ in a way that maintains its transformation properties.
- There are a number of ways of applying this projection. The projection improves the effectivness of the smearing.

An illustration: the nucleon

The effective mass of a nucleon created by operators with different quark and gauge smearings (Top - quark smearing, middle - gauge smearing, bottom - both) Basak et. al. hep-lat/0601034



Variational techniques

- Suppose we can construct a (small) set of operators, $\phi_i, i = 1 \ldots m$ with the quantum numbers of interest. Any linear combination of these operators has the same transformation properties; $\Phi = d_i \phi_i$
- Can we determine the linear combination that maximises the contribution to the correlation function from the ground-state?
- The correlation function of Φ is (assuming ϕ are real operators)

$$C_\Phi(t) = d_i C_{ij}(t) d_j \quad \text{with} \quad C_{ij}(t) = \langle \phi_i(t) | \phi_j(t) \rangle$$

- **IF** the action has a positive-definite transfer matrix, then the energies of all states are real and positive and so

$$C_{ij} = \langle \phi_i | n \rangle \langle n | e^{-tH} | n \rangle \langle n | \phi_j \rangle$$

is a positive-definite symmetric matrix and the lowest eigenstate of the generalised eigenvalue problem

$$C(t_1)\vec{v} = e^{-\lambda(t_1 - t_0)} C(t_0)\vec{v}$$

has $\lambda > E_0$. This eigenvector then gives the linear combination of operators with the lowest energy and thus the "best" ground-state

All-to-all quark propagators

- Computing all elements of the quark propagator would require full knowledge of the inverse - prohibitively expensive
- If we are satisfied with an unbiased *estimator* of all elements then sparse matrix methods can be used. Quark propagation is being measured on a random ensemble of gauge field backgrounds.
- Variance reduction will be crucial

All-to-all quark propagators (2)

- Start with a spectral representation of $Q = \gamma_5 M$ (choose Q here because it is hermitian so eigenvalues are easier to compute).
- If we can compute all the eigenvectors and eigenvalues, $\{\lambda^{(i)}, v^{(i)}\}$ of

$$Q = \sum_{i=1}^{N} \lambda^{(i)} v^{(i)} \otimes v^{*(i)} \quad \text{and} \quad Q^{-1} = \sum_{i=1}^{N} \frac{1}{\lambda^{(i)}} v^{(i)} \otimes v^{*(i)}$$

- Unfortunately, finding even a small sub-set of eigenvectors is computationally expensive, so we are forced to truncate this representation at $N_{\mathrm{ev}} \ll N$

All-to-all quark propagators (3)

- Start again with a stochastic representation of Q. Fill a vector, η with random noise, such as $Z_4 = \{1, i, -1, -i\}$. Then

$$E[\eta_i \eta_j^*] = \delta_{ij}$$

and Z_N noise has the useful property that

$$\eta_i \eta_i^* = 1 \text{ (no sum)}$$

- Now applying the fermion solver gives $\psi = Q^{-1}\eta$ and

$$E[\psi_i \eta_j^*] = Q_{ij}^{-1}$$

and we have an estimator for all elements of Q^{-1}.

- Unfortunately, the variance of this estimator is large and any signal will be swamped by noise.

All-to-all quark propagators (4)

- The variance can be reduced by recalculating this estimator for m different random sources but errors fall like $1/\sqrt{m}$. Can we do better?
- The exact propagator can be computed with a finite (but large!) amount of effort; use point-propagator methods with Kronecker delta sources put everywhere.
- This suggests a trick; break the vector space of the quark fields, V into d smaller sub-spaces $V = V_1 \oplus V_2 \oplus \dots$ spanned by sub-sets of the basis vectors. For example, even-odd partitioning:

$$V_1 = \left\{ e^{(1)} = (1, 0, 0, 0, \dots), e^{(3)} = (0, 0, 1, 0 \dots) \right\}$$

$$V_2 = \left\{ e^{(2)} = (0, 1, 0, 0, \dots), e^{(4)} = (0, 0, 0, 1, \dots) \right\}$$

This partitioning ("dilution") is arbitrary. A useful example is "time dilution", where N_T sub-spaces are defined, with support on one time-slice only.

All-to-all quark propagators (5)

- The basis is complete, so if $\mathcal{S}_i$ is a projector into space V_i and $\eta^{(i)} = \mathcal{S}_i \eta$ then $\eta = \sum_{i=1}^{d} \eta^{(i)}$. Since $\mathcal{S}_i^2 = \mathcal{S}_i$, we can write an identity

$$1 = \sum_{i=1}^{d} \mathcal{S}_i = \sum_{i=1}^{d} \mathcal{S}_i^2 = \sum_{i=1}^{d} \mathcal{S}_i E[\eta \otimes \eta^*] \mathcal{S}_i = \sum_{i=1}^{d} E[\eta^{(i)} \otimes \eta^{*(i)}]$$

  and another representation of the propagator can be written as

$$Q^{-1} = \sum_{i=1}^{d} E[\psi^{(i)} \otimes \eta^{*(i)}] \quad \text{where} \quad \psi^{(i)} = Q^{-1} \eta^{(i)}$$

- The variance in this estimator is reduced by explicit cancellation of terms that vanished before only as $m \to \infty$. If $d = N$, the **exact** propagator is recoved.
- A good choice of dilution should beat statistics.

All-to-all quark propagators (6)

- Most of the physics is contained in the lowest few eigenvectors of Q. A hybrid method can be constructed that uses an exact representation of the lowest few eigenvectors and corrects for the truncation using a stochastic estimator. Break V into two subspaces, V_L and V_H, with V_L the space spanned by the lowest N_{ev} eigenvectors. Since Q leaves this space invariant, we have

$$Q^{-1} = \bar{Q}_L + \bar{Q}_H = Q^{-1}\mathcal{P}_L + Q^{-1}\mathcal{P}_H$$

- $\bar{Q}_L$ is the truncated eigenvector representation, and $\bar{Q}_H$ can be estimated with the dilution method. The action of $\bar{Q}_H$ is

$$\bar{Q}_H = Q^{-1}\mathcal{P}_H = Q^{-1}(1 - \mathcal{P}_L)$$

  which is a Gram-schmidt orthogonalisation against the known eigenvectors followed by application of the inverse.

All-to-all quark propagators (7)

The Hybrid all-to-all method

- Compute N_{ev} eigenvectors and eigenvalues, $\{\lambda^{(i)}, v^{(i)}\}$
- Generate **one** noise vector, and dilute $\{\eta^{(1)}, \eta^{(2)}, \dots\}$
- For each dilute vector, compute $\psi^{(i)} = Q^{-1}(1 - \mathcal{P}_L)\eta^{(i)}$
- Now Q^{-1} is estimated as

$$Q^{-1} = \sum_{i=1}^{N_{ev}} \frac{1}{\lambda^{(i)}} v^{(i)} \otimes v^{*(i)} + \sum_{j=1}^{N_d} \psi^{(j)} \otimes \eta^{*(j)}$$

- Since both terms are sums of outer products, they can be packed into a single sum over $j = 1 \dots N_{ev} + N_d$. The "hybrid list" representation becomes $Q^{-1} = \sum_{j=1}^{N_{ev}+N_d} u^{(j)} \otimes w^{*(j)}$

All-to-all quark propagators (8)

- We don't directly want the quark propagator; we are using it in correlation functions. Take the simplest case (one quark line) of a static-light meson

$$C_{\mathrm{stat}}(t) = \mathrm{Tr}\ \Gamma M^{-1}(x,t;x,t')U_t(x;t',t)$$

with $U_t(x;t',t)$ the Wilson line at site x from time-slice t to t' (the colour structure of a propagator of an infinitely heavy quark).

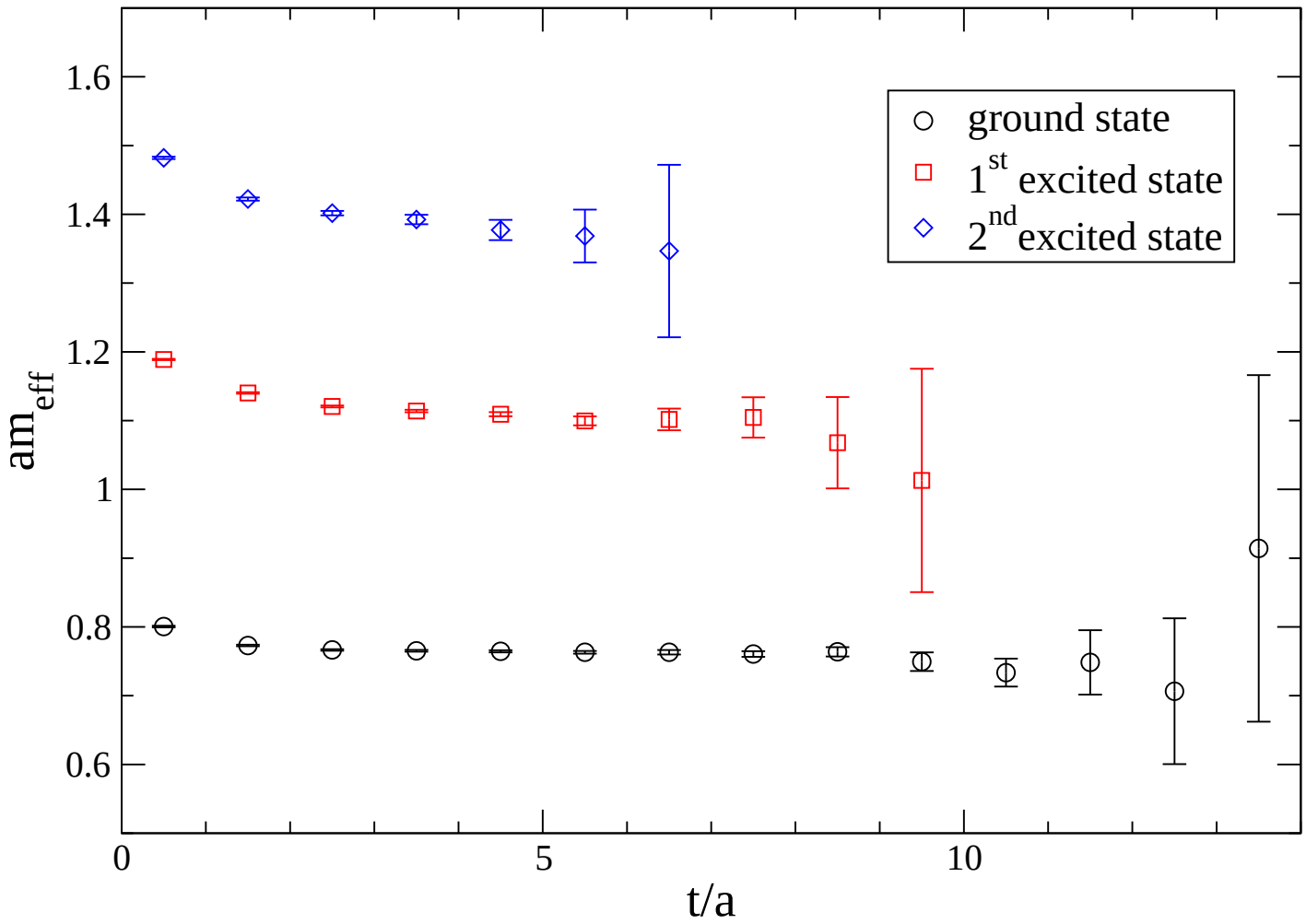
- This becomes

$$C_{\mathrm{stat}}(t) = \sum_{j=1}^{N_H} w^{*(j)}(x,t')U_t(x;t',t)\gamma_5\Gamma u^{(j)}(x,t)$$

- Measurements resemble their original form in terms of quark fields. Operator construction is simplified and more complicated operators can be built.

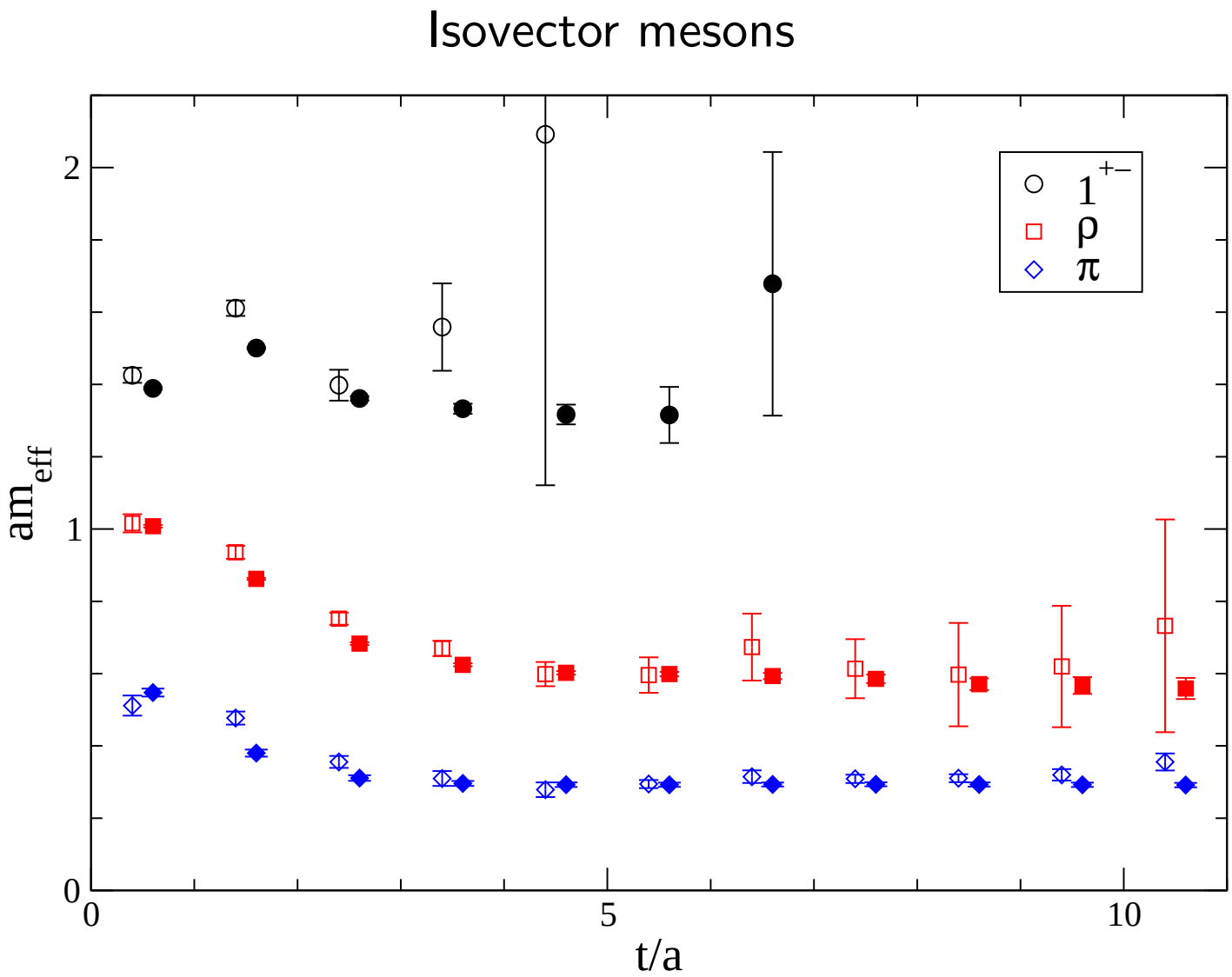
Results from all-to-all quark propagators (1)

Static-light S-wave - variational calculation



$\beta = 5.7, 12^3 \times 24$ lattice Wilson fermions, $\kappa = 0.1675 (m_\pi / m_\rho = 0.50)$ 75 configurations. 100 eigenvectors.

Results from all-to-all quark propagators (2)



$\beta = 5.7, 12^3 \times 24$ lattice Wilson fermions, $\kappa = 0.1675 (m_\pi/m_\rho = 0.50)$ 75 configurations. 100 eigenvectors. Time/even-odd/colour/spin dilution.

Data analysis

- Since the measurements are just a stochastic estimator of the "true" answer, careful interpretation is required.
- Statistical analysis is not the end of the story. There are errors from using a finite lattice spacing, finite volume, incorrect physical parameters (such as quark mass), ...
- Often a physical model of the expected behaviour is known (perhaps with unknown parameters) and a test of the validity of the model (with a determination of the "best-fit" model parameters) is needed. The χ^2 statistic provides the usual discriminant between good and bad models (and parameter choices).
- For our Markov chain Monte Carlo data, the problem of testing and fitting the model to our data is compounded by the fact that many **different measurements** to be compared have been made on the **same ensemble** of field configurations. We need information on the data covariance.

Data analysis

- For correlated data, the measured mis-match between the model $\tilde{y}$ and the data y_i is

$$\chi^2(\alpha) = \sum_{i,j=1}^{m} (\tilde{y}_i(\alpha) - y_i) C_{ij}^{-1} (\tilde{y}_j(\alpha) - y_j)$$

  where C_{ij} is a positive-definite **covariance matrix**

$$C_{ij} = \frac{N}{N-1} \sum_k (y_i^{(k)} - \bar{y}_i)(y_j^{(k)} - \bar{y}_j)$$

  and $y_i^{(k)}$ is the measurement of observable i on configuration k.
- Here we have assumed the model is being compared to simple averages of of observables (for example, a fit to a correlation function for mass extraction).
- In more complicated cases, the covariance matrix must be computed through a jack-knife resampling.

Data analysis (3)

- This estimate of the data covariance matrix is often unreliable. It contains many (highly correlated) entries and can be close to being singular. Techniques to improve estimators and control singularities have been developed. See for example C.Michael and A.McKerrell (`hep-lat/9412087`) for a discussion.
- A "small" value of χ^2 indicates a model that can not be ruled out by the data. The "best-fit" parameters, α^* define the model that minimises $\chi^2(\alpha)$.
- The model would not be ruled out by the data if $\chi^2 \approx m$ when the model parameters are fixed or $\chi^2 \approx N_{df} = m - n_\alpha$ if they are determined
- The "goodness of fit" provides an estimate of the probability a correct model would have χ^2 as large as that measured. It is

$$Q(\chi^2, N_{df}) = \frac{1}{\Gamma(N_{df}/2)} \int_0^{\chi^2/2} e^{-t} t^{N_{df}/2-1} dt$$

Data analysis (4)

- For a linear model;

$$\tilde{y}_i = \sum_{a=1}^{n_\alpha} \alpha_a f_{i,a}$$

  with $f_{i,a}$ constants then the best-fit model is unique and finding this model is just linear algebra.
- For non-linear models (such as a fit to an exponential fall-off; $\tilde{y}_k = \alpha_1 e^{-k\alpha_2}$), finding the global minimum can be difficult. A commonly used algorithm is the Levenberg-Markhardt algorithm (see eg. Numerical Recipes).

Jack-knife resampling

- Suppose we want to compute a (non-linear) function of some observables;

$$Q = q(E[y_1], E[y_2], \dots)$$

a (biased) estimate of Q is just given by applying the function to the sample averages of y. The bias falls like $1/N$. Remember it is incorrect (and possibly ill-defined) to compute

$$Q_{\mathrm{bad}} = E[q(y_1, y_2, \dots)] \neq Q$$

and this estimator is always biased.

- Similarly, a fit result must be assigned an uncertainty. Often it is a complicated function of the underlying data; the expected value of this parameter determined by the fitting procedure - how can an uncertainty be given?

Jack-knife resampling (2)

- An uncertainty can estimated by resampling the data using either the **jack-knife** or **boot-strap** methods.
- Consider the original ensemble of field configurations, E and a set of N new ensembles, $\{E_j\}$ each made by removing configuration j from E. On each ensemble, a new estimate of Q can be made;

$$Q_j = q(E[y_1]_j, E[y_1]_j, \dots)$$

- The uncertainty can be estimated by computing

$$\sigma_Q^2 = \frac{N}{N-1} \sum_{j=1}^{N} (Q - Q_j)^2$$

- The ($1/N$ biased) mean $\sum_j Q_j$ can be used to estimate the bias. See e.g. B. Berg "Markov Chain Monte Carlo Simulations and their Statistical Analysis"

Boot-strap resampling (3)

- In the jack-knife method, we still assume the sample mean of Q is normally distributed and estimate its variance to give an uncertainty.
- A means of estimating the distribution of Q requires another Monte Carlo calculation! As with jack-knife, the ensemble of field configurations is resampled B times where ensemble E_B contains N entries randomly selected with replacement from E.
- Now Q is recomputed on each of these resamplings. If Q is a model parameter determined by a best-fit procedure, this means re-finding the minimum of χ^2 on all ensembles. After this process, we have a set $\{Q_b\}$ of B estimators for Q. These values are sorted numerically and then a given confidence interval (95% for example) can be quoted by reading the $(0.05 \times B)^{\text{th}}$ entry (as a lower limit) and the $(0.95 \times B)^{\text{th}}$ entry (as an upper limit).

Summary

- The Monte Carlo method can be applied to quantum field theory on a finite lattice.
- Algorithms to define irreducible, positive recurrent Markov chains of configurations suitable for importance sampling can be constructed.
- Fermion fields can be handled by replacing them with auxiliary fields (the pseudofermions).
- This leads to non-local actions on bosonic fields. At present, the best methods for dealing with these actions is to use methods based on molecular dynamics or diffusion equations.
- Manipulating quarks requires good algorithms for solving sparse linear systems.
- Measurements on these configurations can be related to any function of the quantum fields. For quarks, this involves computing propagators.
- There is a robust set of statistical methods for analysing Monte Carlo data and its autocorrelations