Model Optimization in the Presence of Correlations

IV QCDNA, Yale University

Adam C. Lichtl

RIKEN BNL Research Center
Brookhaven National Laboratory

Wednesday, May 2, 2007

Overview

- Motivation
- A detailed look at a simple example
- Estimating the covariance matrix
- Real-world data
- Conclusions

Part I

Motivation

The spectral representation of correlation functions

Consider the vacuum correlation function associated with an operator $\overline{\mathcal{O}}$:

$$C(\tau) \equiv \langle 0|\mathcal{O}(\tau)\overline{\mathcal{O}}(0)|0\rangle.$$

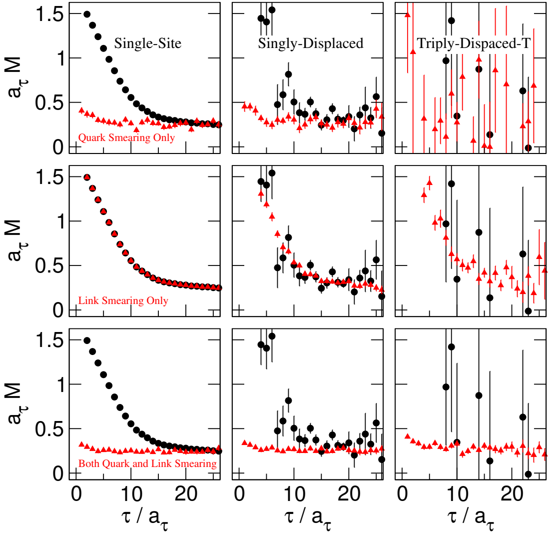
Working in the imaginary time formalism, we may write

$$C(\tau) = \langle 0|e^{+H\tau}\mathcal{O}e^{-H\tau}\overline{\mathcal{O}}|0\rangle,$$

and inserting a complete set of energy eigenstates of the Hamiltonian gives

$$\begin{aligned} C(\tau) &= \langle 0|\mathcal{O}e^{-H\tau}\sum_k |k\rangle\langle k|\overline{\mathcal{O}}|0\rangle \\ &= \sum_k |\langle k|\overline{\mathcal{O}}|0\rangle|^2 e^{-E_k\tau}. \end{aligned}$$

Rich structure available for operator construction



Correlated fitting

- Need to perform fits of the type: (D. Toussaint)

$$C_{\rm fit}(\tau; A, E) = A \exp(-E\tau)$$

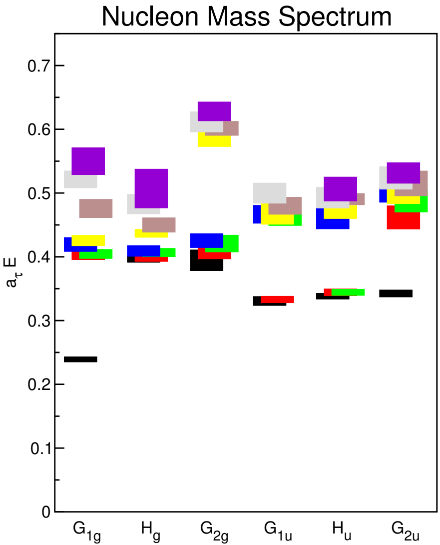
- A and E are the two fit parameters
- Assume no autocorrelations, but take into account cross-correlations on each configuration:

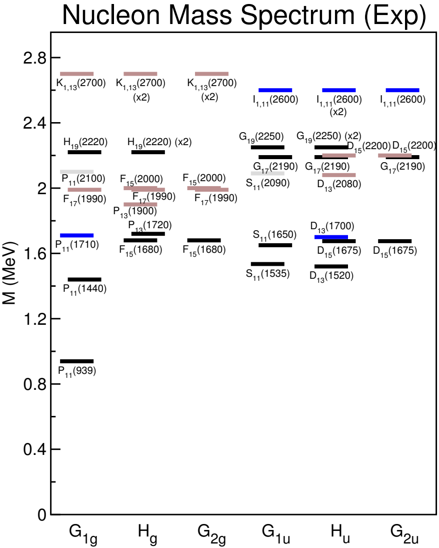
$$\chi^2(A, E) \equiv \sum_{\tau,\tau'} [C(\tau) - C_{\rm fit}(\tau; A, E)] \hat{\sigma}^{-1}_{\tau,\tau'} [C(\tau') - C_{\rm fit}(\tau'; A, E)]$$

- $\hat{\sigma}^{-1}_{\tau,\tau'}$ is the inverse of the estimated covariance matrix:

$$\hat{\sigma}_{\tau,\tau'} \equiv \frac{1}{N(N-1)} \sum_{n=1}^{N} [C_n(\tau) - \bar{C}(\tau)][C_n(\tau') - \bar{C}(\tau')],$$

Nucleon spectroscopy





Correlated χ^2 fitting

- How well do such fits perform?
- How reliable are the quoted errors?
- How reliable is $\chi^2/(dof)$ as a measure of goodness-of-fit?

Part II

A Detailed Look at a Simple Example

Simple example

- Two observables: y_1, y_2, fit to a constant α
- Sample estimates: $\hat{y}_1 = 0.4, \hat{y}_2 = 0.7$
- Correlation matrix known to be

$$\sigma = \frac{1}{\sqrt{2.0 - \delta^2}} \begin{bmatrix} 1.0 & \delta \\ \delta & 2.0 \end{bmatrix}$$

- $|\delta| < \sqrt{2}$
- $\sigma_{11} = 1.0 = \mathrm{Det}(\sigma) > 0$ (positive-definite)
- Can look at

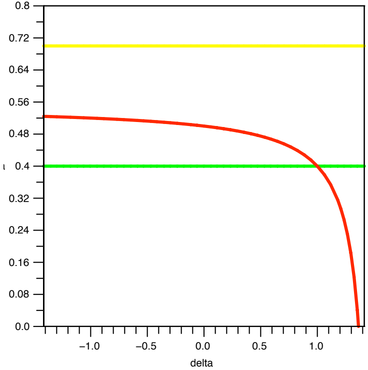
$$\alpha^*(\delta) \leftarrow \min_\alpha \chi^2(\alpha, \delta)$$

where

$$\chi^2(\alpha, \delta) = \sum_{a,b=1}^{2} (\hat{y}_a - \alpha)\sigma^{-1}_{ab}(\delta)(\hat{y}_b - \alpha)$$

Parameter estimate

In the presence of significant positive correlation, the fit value can lie above or below both points!



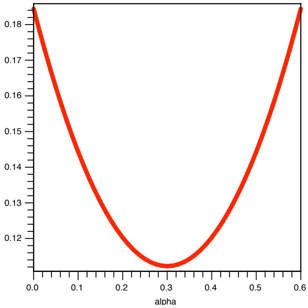
Unfortunately, this is common in LQCD correlation function fitting.

Fun with pathology

- Choose $\delta = 1.2$.

$$\sigma = \frac{1}{\sqrt{2.0 - (1.2)^2}} \begin{bmatrix} 1.0 & 1.2 \\ 1.2 & 2.0 \end{bmatrix}$$

$$\hat{y}_1 = 0.4, \quad \hat{y}_2 = 0.7, \quad \hat{\alpha} = 0.3$$



Is this correct?

- Yes! If the points are strongly correlated, then we are likely to see samples where both sample means fluctuate above or below the true value
- We can simulate this for fixed sample size N
- In general $\sigma \sim 1/N$

$$\hat{y}_a = 0.35 + \sum_{b=1}^{2} [\sigma^{1/2}]_{ab} \tilde{z}_b$$

where

$$\tilde{z}_a \sim N(0,1), \quad \mathrm{Cov}[\tilde{z}_a, \tilde{z}_b] = \delta_{ab}$$

giving

$$\mathrm{E}[\hat{y}_a] = 0.35, \quad \mathrm{Cov}[\hat{y}_a, \hat{y}_b] = \sigma_{ab}$$

Simulation, continued

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}, \quad 1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\chi^2(\alpha) = \frac{1}{N}(\hat{y} - \alpha 1)^T \sigma^{-1}(\hat{y} - \alpha 1)$$

with a minimum at

$$\left. \frac{d}{d\alpha}\chi^2 \right|_{\alpha=\hat{\alpha}} = 0$$

giving the fit value:

$$\hat{\alpha} = \frac{1^T \sigma^{-1} \hat{y}}{1^T \sigma^{-1} 1}$$

- Simple linear fit (don't even need a minimizer), with correlations

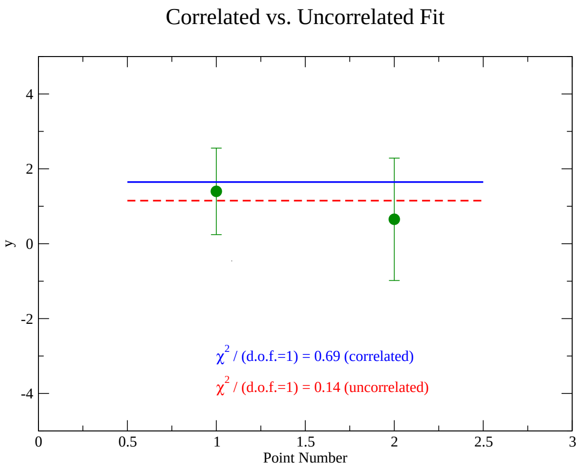
Uncorrelated fit

- Can compare to uncorrelated fit:

$$\hat{\alpha}_{uncorr} = \frac{\hat{y}_1/\sigma_{11} + \hat{y}_2/\sigma_{22}}{1/\sigma_{11} + 1/\sigma_{22}}$$

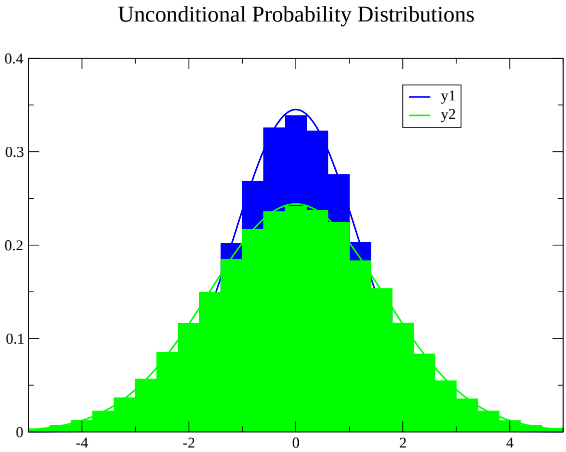
- Note: using $1/\sigma_{aa}$, not $[\sigma^{-1}]_{aa}$ (there is a difference)
- Expect an abnormally small χ^2 because we are neglecting off-diagonal interference in the inverse

Simulation results

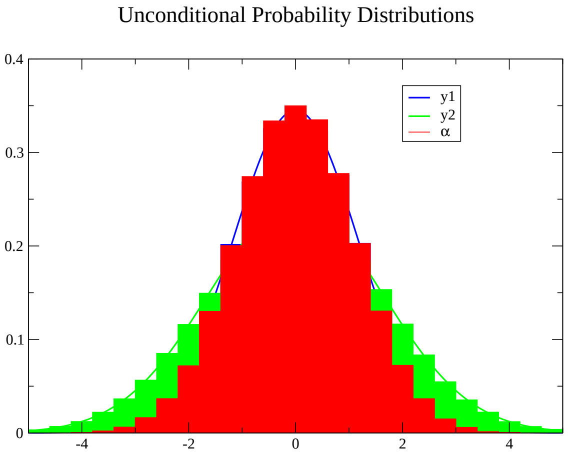


Simulation results

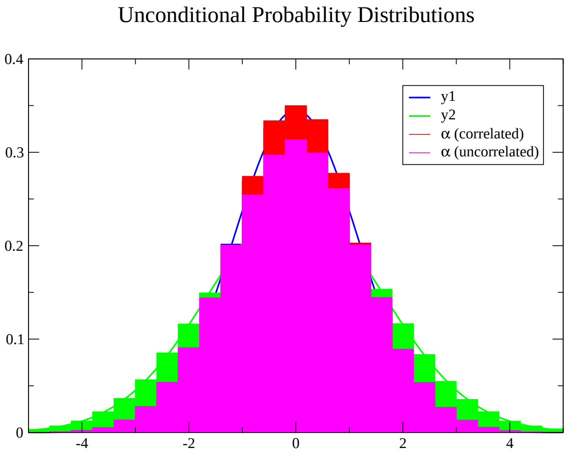
- Independently, the errors appear Gaussian



Simulation results

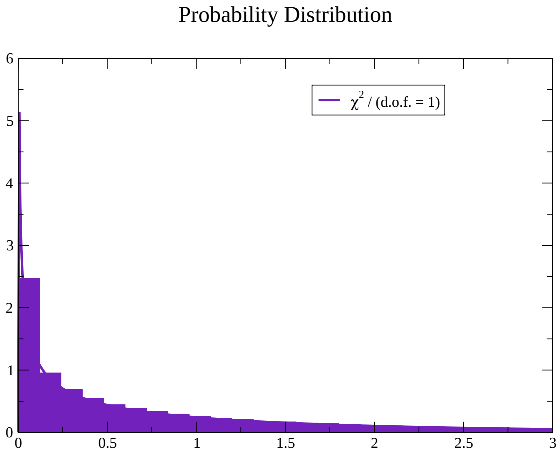


Simulation results



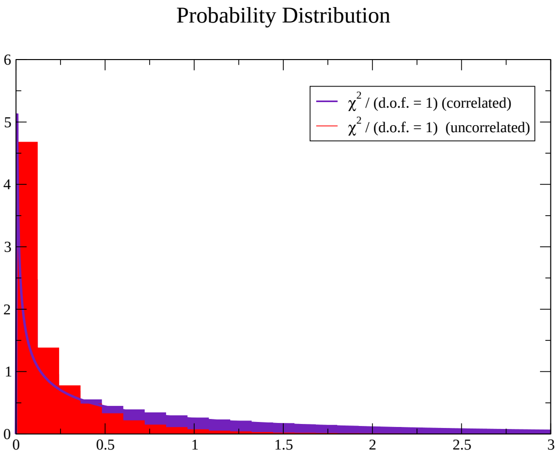
Goodness-of-fit

- The correlated χ^2 can be used for goodness-of fit tests
- σ^{-1} is the metric in the space of D independent variables ($\sim \tilde{z}_a$)



Goodness-of-fit

- The uncorrelated χ^2 is unsuitable for goodness-of-fit tests
- Degrees of freedom are not independent



Part III

Estimating the Covariance Matrix

Serious practical obstacle

- Up to now, we have assumed that we know the covariance matrix σ for our errors
- But we DO NOT know σ
- We must estimate it from the data $\tilde{y}_{ia} \quad (i = 1 \cdots N, a = 1 \cdots D)$

$$\hat{y}_a = \frac{1}{N} \sum_{i=1}^{N} \tilde{y}_{ia}$$

$$\hat{\sigma}_{ab} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\tilde{y}_{ia} - \hat{y}_a)(\tilde{y}_{ib} - \hat{y}_b)$$

- Noisy estimate of something which shifts our parameter estimates on a per-sample basis

Serious practical obstacle

- Up to now, we have assumed that we know the covariance matrix σ for our errors
- But we DO NOT know σ
- We must estimate it from the data $\tilde{y}_{ia} \quad (i = 1 \cdots N, a = 1 \cdots D)$

$$\hat{y}_a = \frac{1}{N} \sum_{i=1}^{N} \tilde{y}_{ia}$$

$$\hat{\sigma}_{ab} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\tilde{y}_{ia} - \hat{y}_a)(\tilde{y}_{ib} - \hat{y}_b)$$

- Noisy estimate of something which shifts our parameter estimates on a per-sample basis
- $\mathrm{E}[\tilde{x}] \neq \mathrm{E}[\tilde{x}^{-1}]$

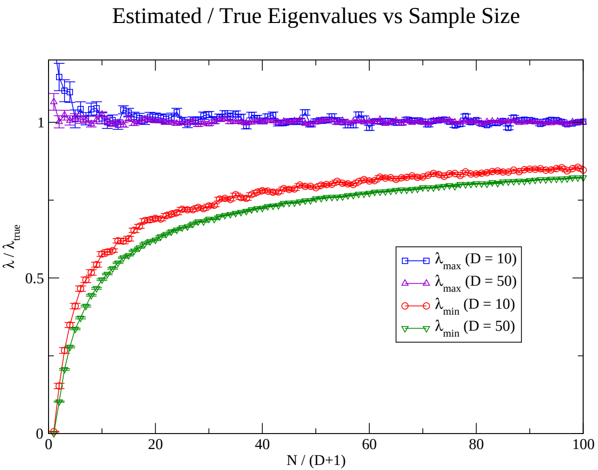
Estimating the covariance matrix

- Let y_{ia} be the elements of the $N \times D$ matrix Y
- $\hat{y} = \frac{1}{N} Y^T 1$ ($D-$dimensional vector)
- $\hat{\sigma} = \frac{1}{N(N-1)} Y^T M Y$ ($D-$dimensional matrix)
- Where $M = (I - \frac{1}{N} 1 1^T)$
- M is idempotent ($M^2 = M$) and of rank $N-1$:

$$M = \frac{1}{N} \begin{bmatrix} N-1 & -1 & \cdots & -1 \\ -1 & N-1 & & \vdots \\ -1 & & \ddots & -1 \\ -1 & \cdots & -1 & N-1 \end{bmatrix}$$

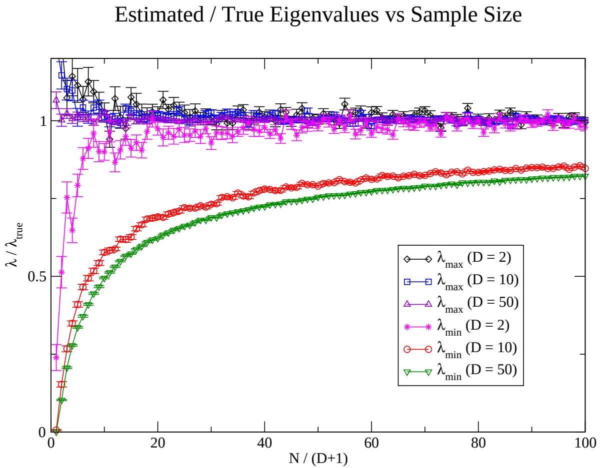
Rank deficiency

If $N < D + 1$, then $\mathrm{Rank}(\hat{\sigma}) < D$, and $\hat{\sigma}$ is not invertible (rank deficient)



Rank deficiency

The lowest eigenvalues are 'repelled' downward, even at $N = D^2$



The Frobenius matrix metric

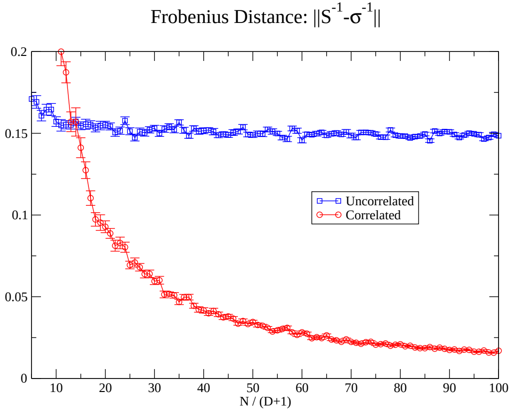
- To quantify how 'far' the estimated covariance matrix is from the true covariance matrix
- Frobenius metric for a $D-$dimensional symmetric matrix:

$$||M|| \equiv \frac{1}{D}\sum_{a=1}^{D}\sum_{b=1}^{D} m_{ab}^2$$

Normalized such that

$$||I|| = 1$$

How good are our estimates?



"Just do an uncorrelated fit if you don't have the statistics."
(or can we do better?)

What happens if the covariance matrix estimate is bad?

- Lattice QCD: "Oh well.. it all averages out. Mumble mumble.."
- Portfolio manager on Wall Street: "Uh, boss? I just lost $4B."
- Pharmaceutical researcher (bioinformatics): "Hmm.. I think that drug will work.

What happens if the covariance matrix estimate is bad?

- Lattice QCD: "Oh well.. it all averages out. Mumble mumble.."
- Portfolio manager on Wall Street: "Uh, boss? I just lost $4B."
- Pharmaceutical researcher (bioinformatics): "Hmm.. I think that drug will work. Let's try it on the new guy."

What happens if the covariance matrix estimate is bad?

- Lattice QCD: "Oh well.. it all averages out. Mumble mumble.."
- Portfolio manager on Wall Street: "Uh, boss? I just lost $4B."
- Pharmaceutical researcher (bioinformatics): "Hmm.. I think that drug will work. Let's try it on the new guy."

Look to other fields for the solution to this well defined problem.

- O. Ledoit: Credit Suisse First Boston
- M. Wolf: Department of Economics and Business, Universitat Pompeu Fabra
- J. Schaefer and K. Strimmer, Department of Statistics, University of Munich

Interpolation of Linear Operators

- E. Stein, American Mathematical Society (1956)
- A linear combination of a quiet biased estimator with a noisy unbiased estimator is superior to both
- Convexity of the Frobenius metric
- Can define our covariance matrix estimate as some interpolation between the sample estimate and the diagonal (uncorrelated) estimate

$$\hat{S} = \delta \hat{V} + (1 - \delta)\hat{\sigma}$$

- We are effectively 'shrinking' the off-diagonal (noisy) elements:

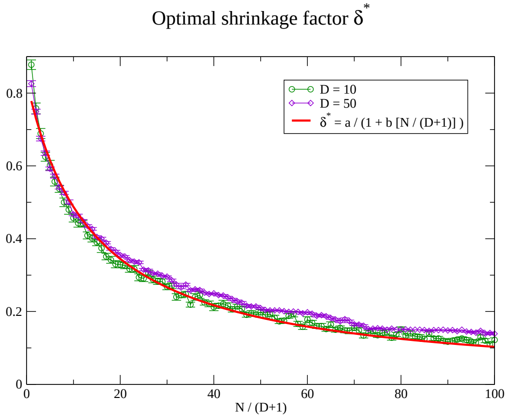
$$\hat{S}_{ab} = \begin{cases} \hat{\sigma}_{ab} & a = b \\ (1 - \delta)\hat{\sigma}_{ab} & a \neq b \end{cases}$$

Choosing the optimal value of δ

- Want to chose δ^* such that the inverse covariance matrix estimate is as close as possible to the true inverse covariance matrix
- $\delta^* \leftarrow \min_\delta ||\hat{S}^{-1} - \sigma^{-1}||$
- $\delta^* \leftarrow \min_\delta ||[\delta \hat{V} + (1-\delta)\hat{\sigma}]^{-1} - \sigma^{-1}||$
- Work in progress: A closed-form expression for δ^* exists
- It is better to estimate δ^* from the data rather than σ

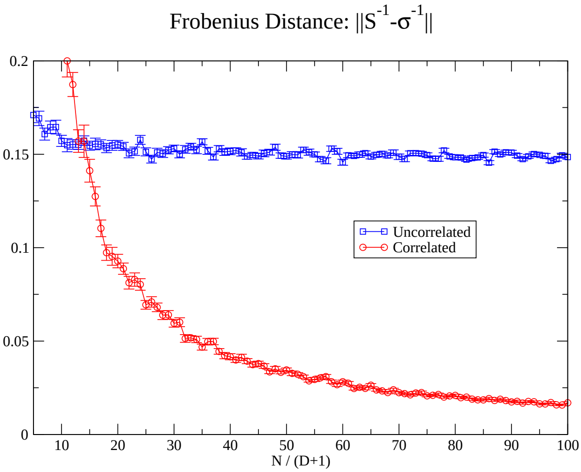
Shrinkage parameter

- $\delta^* \leftarrow \min_\delta ||[\delta \hat{V} + (1-\delta)\hat{\sigma}]^{-1} - \sigma^{-1}||$

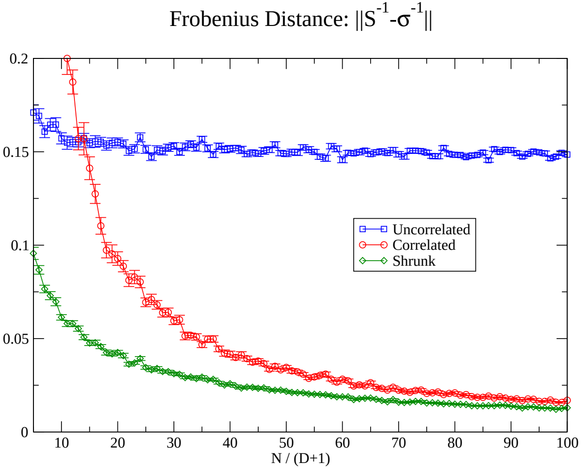


- Goes roughly as $\delta^* = \frac{a}{1+bN/(D+1)} \to O(1/N)$

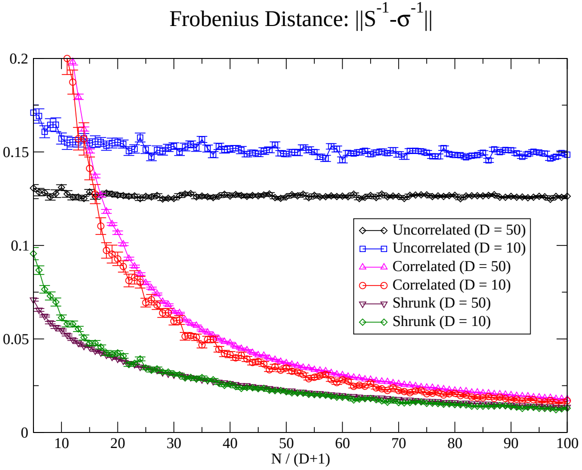
Before shrinking



'Shrinking' the covariance matrix



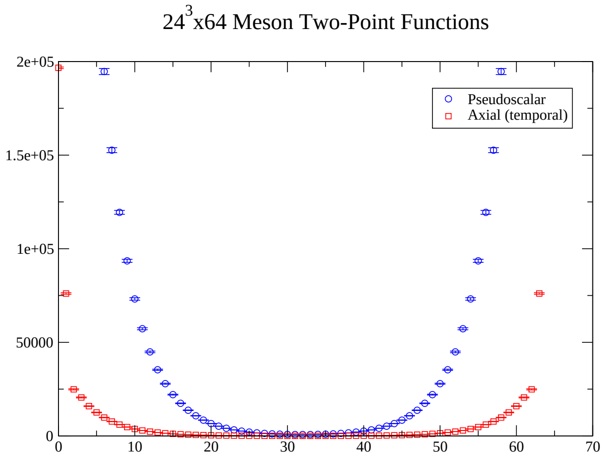
Useful at arbitrary practical D



Part IV

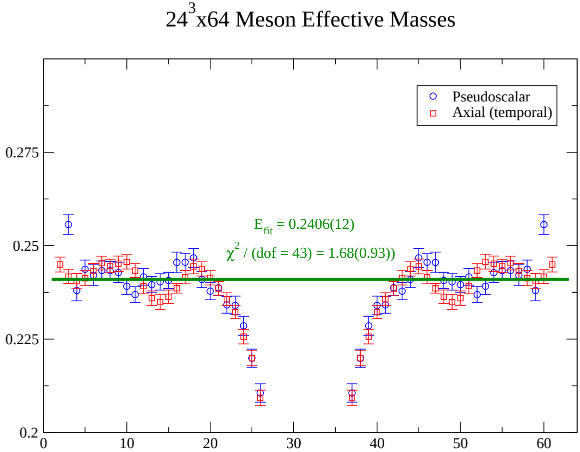
Real-World Data

Meson Data

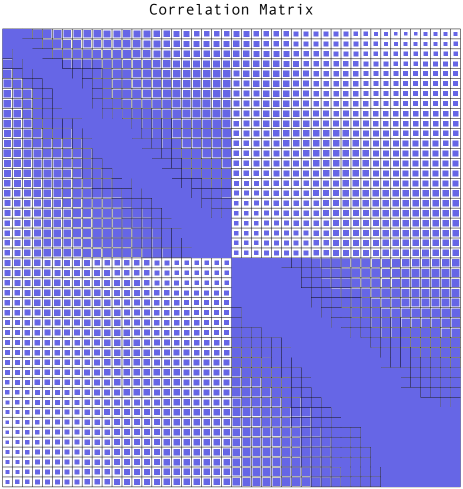


24^3x64 Meson Two-Point Functions
Pseudoscalar
Axial (temporal)
2e+05
1.5e+05
1e+05
50000
0
0
10
20
30
40
50
60
70

Correlations are visible in the data



Visualizing the correlation



Modeling the correlations

- Correlations caused by physical mechanisms such as pion and rho coupling [C. Michael, A, McKerrell]
- Shouldn't look at estimated correlation, should instead look at time-slice coupling

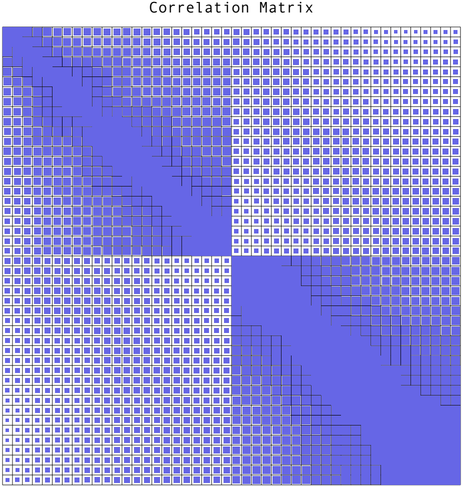
$$\tilde{C}_{n\tau} = C(\tau) + \sum_{\tau'=1}^{D} \tilde{z}_{n\tau'} \sigma^{1/2}_{\tau'\tau}, \qquad \tilde{z}_{n\tau} \sim N(0,1)$$

$$\mathrm{Cov}[\tilde{z}_{n\tau}, \tilde{z}_{m\tau'}] = \delta_{nm}\delta_{\tau\tau'}$$

- Rewrite to see coupling among time-slices

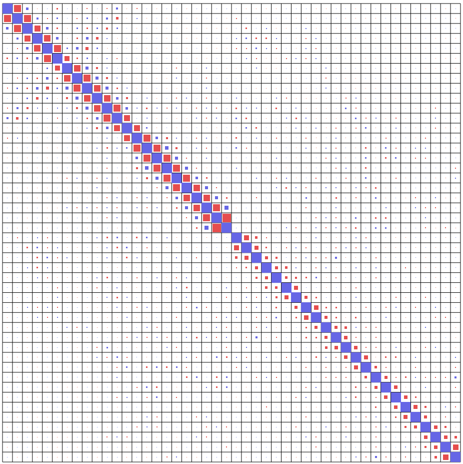
$$\tilde{W}_{n\tau} = \sigma^{-1/2}_{\tau\tau'} C(\tau') + \tilde{z}_{n\tau} \qquad \text{(independent variables)}$$

Don't look at this..



Look at this!

Look at $\hat{\sigma}^{-1/2}$ (This is unaltered sample data!)



Part V

Conclusions

Conclusions

- In the presence of correlations, uncorrelated fits poorly estimate the errors

Conclusions

- In the presence of correlations, uncorrelated fits poorly estimate the errors
- Correlated fitting can introduce large corrections to the estimated central value on a per-sample basis

Conclusions

- In the presence of correlations, uncorrelated fits poorly estimate the errors
- Correlated fitting can introduce large corrections to the estimated central value on a per-sample basis
- The estimated covariance matrix is very noisy but unbiased, while the estimated variances are quiet but biased

Conclusions

- In the presence of correlations, uncorrelated fits poorly estimate the errors
- Correlated fitting can introduce large corrections to the estimated central value on a per-sample basis
- The estimated covariance matrix is very noisy but unbiased, while the estimated variances are quiet but biased
- Stein shrinkage provides a much better estimate of the covariance matrix using Frobenius convexity

Conclusions

- In the presence of correlations, uncorrelated fits poorly estimate the errors
- Correlated fitting can introduce large corrections to the estimated central value on a per-sample basis
- The estimated covariance matrix is very noisy but unbiased, while the estimated variances are quiet but biased
- Stein shrinkage provides a much better estimate of the covariance matrix using Frobenius convexity
- Work in progress: determine $\hat{\delta}$
- Work in progress: shrink to a model, not to a diagonal matrix

Acknowledgements

Special thanks to the following:

- C. Dawson
- G. T. Fleming
- N. Christ, B. Mawhinney
- Meifeng Lin

Any Questions?