# State of the Art of Time Series Analysis in LQFT

George T. Fleming

# The Big Picture

- Tremendous advances in algorithms (thanks QCDNA) and computing have made generating large lattices $O(100^4)$ feasible.

- Latest preconditioners enable $O(10^5)$ solves per lattice. New observables can be constructed.

- Getting to a physics result usually requires one more step: analysis.

- Despite all the effort on the first two steps, often the bottle neck to publication is the third step.

# The Big Picture

- The LHC discovered a Higgs boson. Nobel prizes were awarded.
- Some theorists imagine that the Higgs boson could be a composite particle made of something like quarks and gluons of QCD.
- To compute the properties of such a Higgs boson would be a very hard observable calculation, naively involving O(Vol) linear solves per config, on an ensemble of O(1,000) independent configs.

# Computing the Composite Higgs Mass

- The Higgs mass would be extracted from a correlation function with “disconnected contributions

$$C(\vec{p}, |t-t'|) = \int d^3x dt d^3x' dt' \langle \bar{\psi}\psi(\vec{x}, t)\bar{\psi}\psi(\vec{x}', t')\rangle e^{-i\vec{p}\cdot\vec{x}} e^{i\vec{p}\cdot\vec{x}'}$$

- Written in terms of Dirac matrices

$$C(\vec{p}, |t-t'|) = \int d^3x dt d^3x' dt' \langle Tr[M^{-1}(x,t;x',t')M^{-1}(x',t';x,t)]$$
$$-Tr[M^{-1}(x,t;x,t)]Tr[M^{-1}(x',t';x',t')]\rangle e^{-i\vec{p}\cdot\vec{x}} e^{i\vec{p}\cdot\vec{x}'}$$

- Translation invariance symmetry allows us to compute the first term from a single point but the second term is O(Vol) harder.

# Fast Convolution Theorem

- 1965: Cooley-Tukey (FFT), 1966: Stockham, Helms, Sande

$$C_{ij}(x-y)_c = \langle O_i(x) O_j(y) \rangle_c = \langle O_i(x) O_j(y) \rangle - \langle O_i(x) \rangle \langle O_j(y) \rangle$$

- Use FFT on $O_i(x)$

$$\widetilde{C}_{ij}(q) = \langle \widetilde{O}_i^*(q) \widetilde{O}_j(q) \rangle = \widetilde{C}_{ij}(q)_c + c_0 \delta_{q,0}$$

- Disconnected piece only at q=0, easily subtracted.

- For composite Higgs $O(x) = [M^{-1}]_{x,x}$. Is there an efficient way to estimate just the diagonal part of the inverse without doing O(Vol) linear solves?

- Perfect for autocorrelation. Why isn't it used more?

# Got $C_{ij}$(x-y), now what?

- Euclidean time hadron correlator in infinite box

$$C(t_n) = \sum_{m=1}^{M} A_m e^{-anE_m}, \quad n \geq 0, A_m \in \mathbb{R}, 0 \leq E_1 \leq \cdots$$

- Vandermonde system of $N = 2M$ time slices given $M$ states,
  $\mathbf{y} = \mathbf{V}(x)\mathbf{a}$, $y_n \equiv C(t_n)$, $x_m \equiv e^{-aE_m}$, $a_m \equiv A_m$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{2M} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_M \\ x_1^2 & x_2^2 & \cdots & x_M^2 \\ x_1^3 & x_2^3 & \cdots & x_M^3 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{2M-1} & x_2^{2M-1} & \cdots & x_M^{2M-1} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix}$$

- Could use NLLS to fit, but exponential fits can be unstable. Or solve the Vandermonde system.

# Vandermonde solution

- Successful collaboration from previous QCDNA.
- Find the roots of the polynomial below.
- Easy to get many solutions.
- Note the left block is a Hankel matrix.

$$\left|\begin{array}{cccc:c} y_1 & y_2 & \cdots & y_M & 1 \\ y_2 & y_3 & \cdots & y_{M+1} & x_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{M+1} & y_{M+2} & \cdots & y_{2M} & x_1^M \end{array}\right| = 0.$$

# Even More Solutions

- We can get lots of solutions.  But, most are spurious due to overfitting noise.

- The challenge is to find the few physically relevant states among many irrelevant ones.

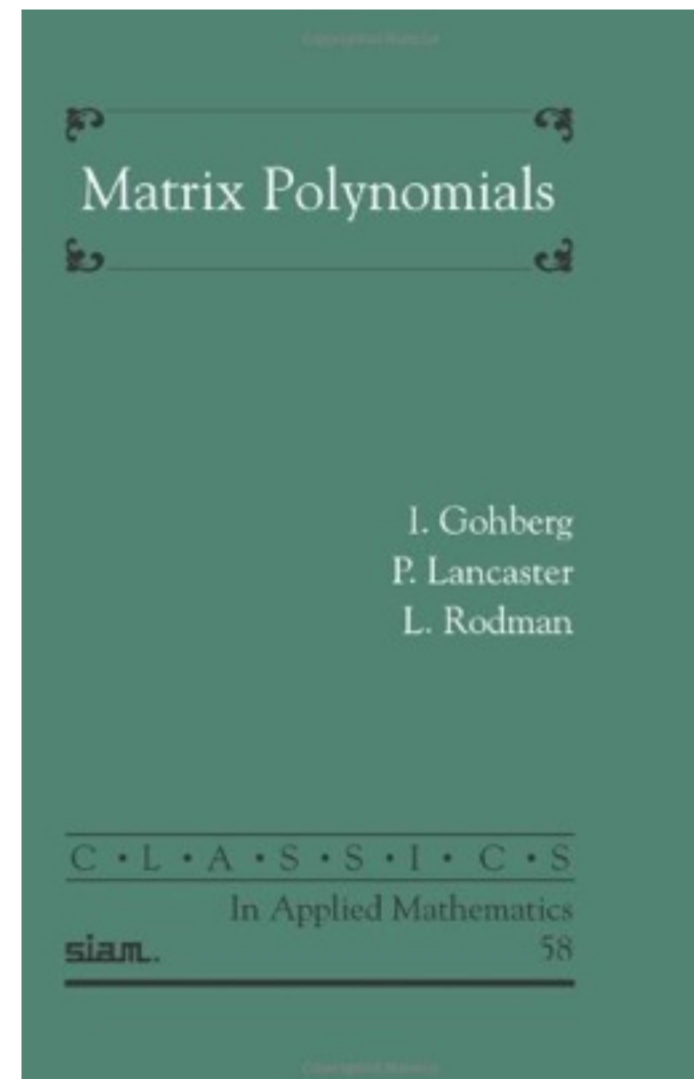
- The general polynomial equation for $K$ correlators:

$$\left| \begin{array}{c:c:c:c:c} \mathbf{H}_1^{N\times M_1} & \mathbf{H}_2^{N\times M_2} & \cdots & \mathbf{H}_K^{N\times M_K} & \begin{matrix} 1 \\ x_1 \\ x_1^2 \\ \vdots \\ x_1^M \end{matrix} \end{array} \right| = 0.$$

- A determined system requires $M = \sum_{k=1}^{K} M_k$ and $N = M + 1$ as before.

# Correlator Matrices, Matrix Polynomials

- Treat $C_{ij}(t)$ as a NxN Hermitian matrix $\mathbf{C}(t)$.
- Matrix pencil: $\mathbf{C}(t+1) = x_n \mathbf{C}(t)$.
- Why stop there:
  $2\,\mathbf{C}(t+2) - x\,\mathbf{C}(t+1) - x^2\,\mathbf{C}(t) = 0$.
- In general can always form a matrix polynomial equation of order $N_t$.



# Yet Even More solutions

- If the $y_m$=**C**(t) are matrix-valued, one can form many polynomials, what's a good suggestion?
- The Vandermonde polynomial with Matrix-valued coefficients.
- Obviously, more statistics helps because it suppresses the spurious solutions.

$$\left|\begin{array}{cccc:c} y_1 & y_2 & \cdots & y_M & 1 \\ y_2 & y_3 & \cdots & y_{M+1} & x_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{M+1} & y_{M+2} & \cdots & y_{2M} & x_1^M \end{array}\right| = 0.$$