Computational strategies for hadron matrix elements

Sara Collins

RQCD Collaboration: Gunnar Bali, Daniel Jenkins, Andreas Schäfer, Jakob Simeth, Wolfgang Söldner, Simon Weishäupl, ...

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 813942.

Numerical Challenges in Lattice QCD, Meinerzhagen August 17, 2022

Overview

★ Preliminaries

★ Efficient calculation of connected three-point functions. Stochastic estimation provides flexibility in evaluating three-point functions for multiple baryons.

★ Estimation of the disconnected three-point functions: truncated solver method, partitioning, hopping parameter expansion (HPE), stochastic noise vs gauge noise, extending the HPE, cluster decomposition error reduction, one-end trick.

★ Summary

Preliminaries

We assume Wilson-type fermions with a discretised Euclidean Dirac operator $M = a\not{D} + am_j$:

$$M^j = M(\kappa_j) = \frac{1}{2\kappa_j}\left(\mathbb{1} - \kappa_j D\right), \quad am_j = \left(\frac{1}{\kappa_j} - \frac{1}{\kappa_c}\right), \quad \kappa_c = \frac{1}{8} + \mathcal{O}(g^2)$$

of a quark with the mass m_j. As $am_j \searrow 0$, $\kappa_j \nearrow \kappa_c > 1/8$.

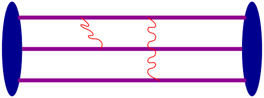
Most methods mentioned are applicable to different fermion formulations.

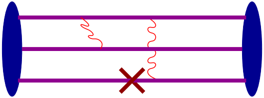
However: the hopping parameter expansion (HPE) requires an ultra-local action (all entries of M that are not near its diagonal vanish).

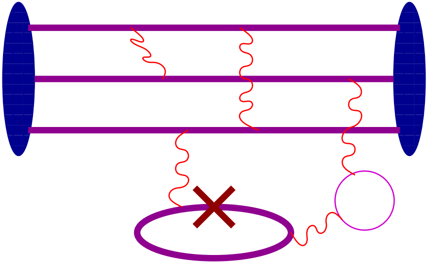
M has the position indices x', x, colour indices i', i and spin indices α', α at its "sink" and "source": $M_{xi\alpha,x'i'\alpha'}$: sparse $12V \cdot 12V$ matrix.

γ_5-Hermiticity: $M^\dagger = \gamma_5 M \gamma_5$, i.e. $M^*_{x'i'\alpha',xi\alpha} = \sum_{\beta\beta'} \gamma_5^{\alpha'\beta} M_{xi\beta,x'i'\beta'} \gamma_5^{\beta'\alpha}$.

Point-to-all and all-to-all propagators







Hadron structure: $\langle \boldsymbol{\Omega} | \mathbf{N}(\mathbf{t}) \boldsymbol{J}(\boldsymbol{\tau}) \overline{\mathbf{N}}(\mathbf{0}) | \boldsymbol{\Omega} \rangle$: (Example: $\boldsymbol{J = q^\dagger \Gamma q}$)

Propagator: $G = M^{-1}$. Often only $G_{x'i'\alpha',x_0 i\alpha}$ is needed for a fixed source position x_0 (**point-to-all propagator**, vector of V $12 \cdot 12$ spin-colour matrices). This can be obtained by solving the 12 linear systems

$$\sum_{x'i'\alpha'} M_{yj\beta,x'i'\alpha'} G_{x'i'\alpha',x_0 i\alpha} = \delta_{yx_0}\delta_{\beta\alpha}\delta_{ji}, \quad x_0 \text{ fixed.}$$

Sometimes **all-to-all propagators** are needed e.g. $\mathrm{tr}\,\Gamma G_{xx} = \mathrm{tr}\,\Gamma M^{-1}_{xx}$, where the trace is over spin and colour.

$\longrightarrow$ unbiased stochastic estimate [K Bitar et al,NPB 313 (89) 348].

Stochastic estimation of $G = M^{-1}$

Generate a set of random noise vectors $|\eta^\ell\rangle$, $\ell = 1, \ldots, n$. Define

$$\frac{1}{n}\sum_\ell |\eta^\ell\rangle\langle\eta^\ell| = \overline{|\eta\rangle\langle\eta|}_n = \overline{|\eta\rangle\langle\eta|} = \mathbb{1} + \mathcal{O}(1/\sqrt{n}),$$

$$\overline{\langle\eta|} = \mathcal{O}(1/\sqrt{n}).$$

Often: $\eta^\ell_{xi\alpha} \in \mathbb{Z}_2 \otimes i\,\mathbb{Z}_2/\sqrt{2}$ [S Dong, K-F Liu,PLB 328 (94) 130].

Other choices: $\mathbb{Z}_2, \mathbb{Z}_3, \mathrm{U}(1), \mathrm{SU}(3)$

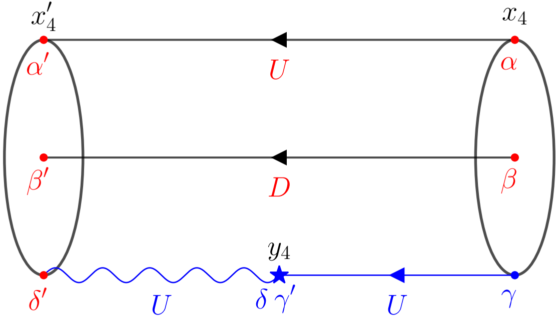
Solve $M|s^\ell\rangle = |\eta^\ell\rangle$ for each $\ell \in \{1, \ldots, n\}$ and construct an unbiased estimate:

$$M_E^{-1} = \overline{|s\rangle\langle\eta|} = M^{-1}\overline{|\eta\rangle\langle\eta|} = M^{-1}\underbrace{\left(\overline{|\eta\rangle\langle\eta|} - \mathbb{1}\right)}_{\mathcal{O}(1/\sqrt{n})} + M^{-1}$$

$\Rightarrow n \ll 12V$

Noise $\propto 1/\sqrt{n}$. Can be large, depending on the observable.

Evaluation of connected baryon 3-point functions



Factorization of connected 3-point functions $C = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{c=1}^{3} \mathbf{S}_{\ell c} \mathbf{I}_{\ell c}$ into

- **a "spectator" part *S*,**
- **an "insertion" part *I*,**

leaving all eight spin-indices open. (ℓ: stochastic index, c: colour index.)

Advantage: big saving in computer time, great flexibility.

Disadvantage: stochastic noise, storage.

(However, storing all the possible 3-point functions would require even more disk/tape space.)

Details

$$C(\mathbf{p}', \mathbf{q}, x_4', y_4, x_4)^{\alpha'\alpha\beta'\beta\delta'\delta\gamma'\gamma}_{\mathbf{UDUU}} = \frac{1}{n}\sum_{\ell=1}^{n}\sum_{c=1}^{3}\left(S_{\mathbf{UD}}(\mathbf{p}', x_4', x_4)^{\alpha'\,\alpha\,\beta'\,\beta\,\delta'}_{\ell c} \cdot I_{\mathbf{UU}}(\mathbf{q}, y_4, x_4)^{\delta\,\gamma'\,\gamma}_{\ell c}\right).$$

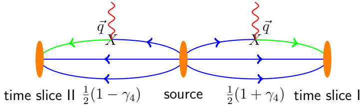
$$S_{\mathbf{UD}}(\mathbf{p}', x_4', x_4)^{\alpha'\alpha\beta'\beta\delta'}_{\ell c} = \sum_{a'b'd'}\sum_{ab}\varepsilon_{a'b'd'}\varepsilon_{abc}\left[\sum_{\mathrm{x}'} G_{\mathbf{U}}{}^{\alpha',\alpha}_{x'a',xa} \cdot G_{\mathbf{D}}{}^{\beta',\beta}_{x'b',xb} \cdot (\gamma_5\eta_\ell)^{\delta'}_{x'd'} \cdot e^{-i\mathbf{p}'\cdot\mathbf{x}'}\right]$$

$$I_{\mathbf{UU}}(\mathbf{q}, y_4, x_4)^{\delta\gamma'\gamma}_{\ell c} = \sum_{c'}\left\{\sum_{\mathrm{y}}\left[(\gamma_5 s_{\mathbf{D}\,\ell})^{\delta}_{yc'}\right]^* \cdot G_{\mathbf{D}}{}^{\gamma',\gamma}_{yc',xc} \cdot e^{i\mathbf{q}\cdot\mathbf{y}}\right\}, \quad \sum_{\alpha'} M^{q}_{\alpha\alpha'} s^{\alpha'}_{Q\ell} = \eta^{\alpha}_{\ell}.$$

The propagators $G_U = G_D$, G_S are obtained from 12 (smeared) point-sources at $(\mathbf{0}, x_4)$.

The noise η_ℓ is time-partitioned (support only at $t = x_4'$).

Seed the noise at two well-separated time slices and vary x_4 inbetween them, simultaneously obtaining forward- and backward-propagating 3-point functions.



"Smearing" is applied to the G-propagator sources and sinks. η is smeared (different for an u/d and s quark) $\rightarrow$ 8 different flavour combinations for S, 4 for I.

Computational and storage cost

Four source positions $\rightarrow$ 48 solves (96 including strange).
100 stochastic solves (200 including strange).

The stochastic method provides all octet (N, Λ, Σ, Ξ) and decuplet (Δ, Σ^*, Ξ^*, Ω) baryons, many sink momenta, negative parity etc. for (almost) free.

Standard method: 12 solves for the source (propagator),
$2 \cdot 4 \cdot 4 \cdot 12 = 384$ sequential solves
(2 Wick contractions, 4 polarizations, 4 sink positions).

Computational and storage cost

Storage:

Spectator part: stochastic (100), colour (3), 5 spins (1024), 4 source positions, 8 flavour combinations, double complex = **157 MB (× # of sink momenta)**.

Insertion part: stochastic (100), colour (3), 3 spins (64), 4 source positions, 4 flavour combinations, 5 non-local (1 and D_μ), double complex = **25 MB (× # of momentum transfers × # of insertion times (forward plus backward))**.

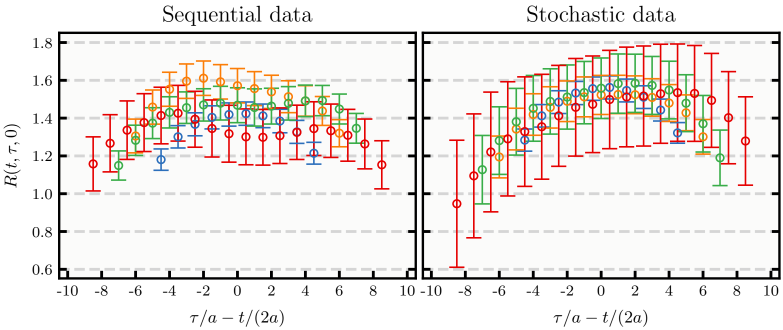
If needed, additional insertion parts can be computed subsequently, e.g., for additional flavours, 1st/2nd derivative (or additional meson spectator).

The resulting hd5-files are large!

SymPy/Python code to generate and extract 3-point functions for arbitrary Wick contractions and baryon interpolators.

Isovector scalar charge: nucleon

N200 ($a \approx 0.064\,\mathrm{fm}$, $M_\pi \approx 285\,\mathrm{MeV}$)

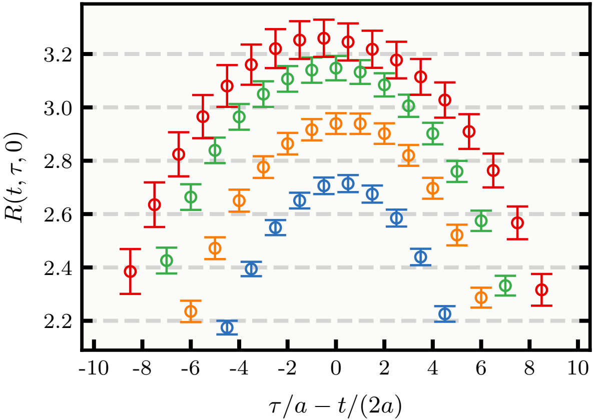


Ratio: 3-point over 2-point function (unrenormalized).

Sequential method: $1+2+3+4$ measurements (depending on the distance).
$\rightarrow 4 + (4+3+2+1)\cdot 8 =$ **84 propagators.**

Stochastic method: $2\cdot 2\cdot 4$ measurements
$\rightarrow 8 + 200/12 \approx$ **25 propagators** (only counting the light quarks).

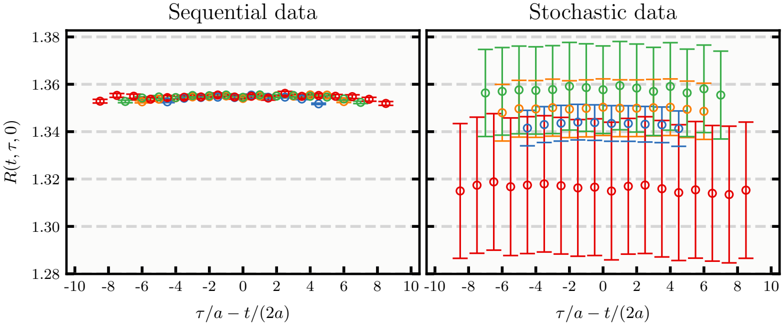
Isovector scalar charge: cascade hyperon



Quite precise results for the Ξ (and also the Σ).

Isovector vector charge: nucleon

N200 ($a \approx 0.064\,\mathrm{fm}$, $M_\pi \approx 285\,\mathrm{MeV}$)

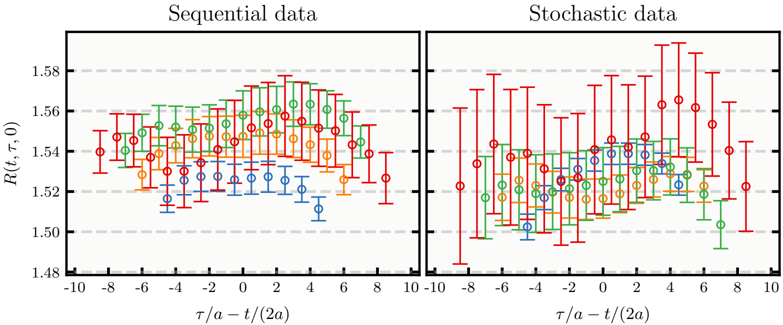


Ratio: 3-point over 2-point function (obviously unrenormalized).

Problem for the vector current: strong correlation between 2- and 3-point function (charge conservation) destroyed by the stochastic noise.

Isovector axial charge: nucleon

N200 ($a \approx 0.064\,\mathrm{fm}$, $M_\pi \approx 285\,\mathrm{MeV}$)

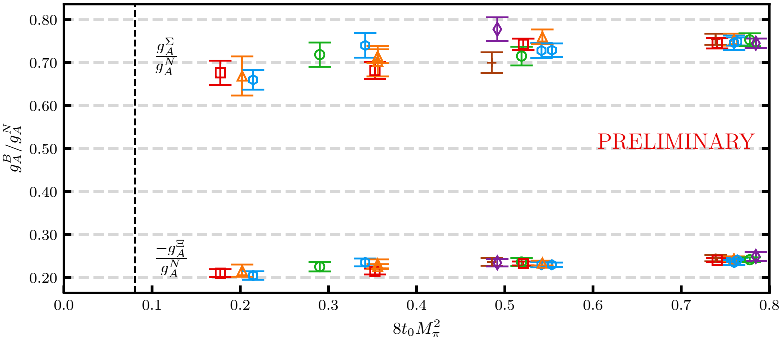


Ratio: 3-point over 2-point function (unrenormalized).

Also for the axial charge, the stochastic method destroys correlations but not as bad as for the vector current.

Strategy: combine axial charge for the nucleon that we have from the sequential method with the stochastic method for the hyperons.

Results 1



SU(3) symmetry: $g_A^p = \overline{D} + \overline{F}$, $g_A^{\Sigma^+} = 2\overline{F}$, $g_A^{\Xi^0} = \overline{F} - \overline{D}$

Replacing $\overline{F}$ and $\overline{D}$ by their values in the SU(3) chiral limit

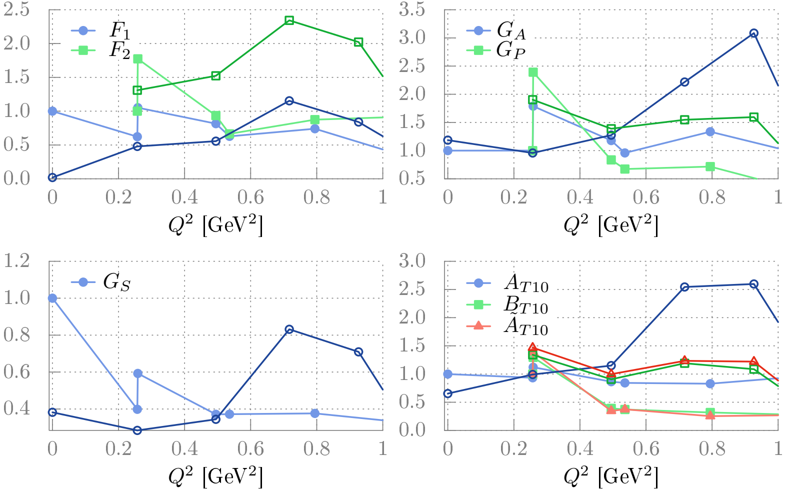
$F = 0.447(7)$ and $D = 0.730(11)$ [RQCD: S Weishäupl et al,2201.05591] gives

$$\frac{g_A^\Sigma}{g_A^N} \approx 0.76, \quad -\frac{g_A^\Xi}{g_A^N} \approx 0.24, \quad \frac{g_A^\Sigma - g_A^\Xi}{g_A^N} = 1,$$

where the latter equation holds exactly as long as $m_{ud} = m_s$.

Results 2

[RQCD,1311.1718]



Gain at finite momentum transfer $Q^2 = (p_f - p_i)^2$. Many equivalent momentum combinations available for the same cost.

Literature

[G Bali et al,1008.3293] (for mesons)

[ETM,1302.2608] (for the nucleon)

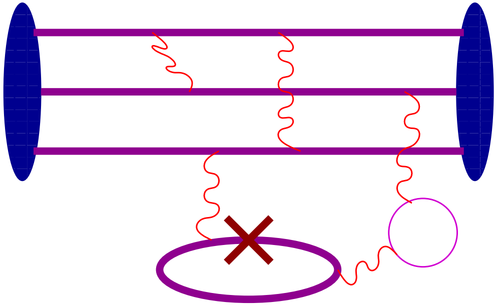
[G Bali et al,1311.1718] (factorization and T symmetry for the nucleon)

[χQCD: Y-B Yang et al,1509.04616] (for the nucleon)

[RQCD: M Löffler et al,1711.02384] (implementation: baryons and mesons with open indices)

[RQCD: S Weishäupl et al,1907.13454] (application)

Disconnected three-point functions



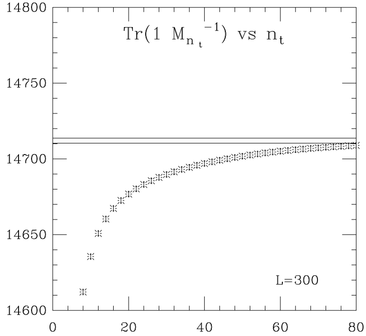
We use: TSM, HPE, partitioning.

The Truncated Solver Method (TSM): loop

Obtain approximate solutions $|s^{\ell}_{n_t}\rangle$ after n_t solver iterations (before convergence) and estimate the difference stochastically to obtain an unbiased estimate of M^{-1} [S Collins et al,PoS(LAT2007)141]:

$$M_E^{-1} = \overline{|s_{n_t}\rangle\langle\eta|}_{n_1} + \overline{(|s\rangle - |s_{n_t}\rangle)\langle\eta|}_{n_2} \quad \text{with} \quad n_2 \ll n_1\,.$$

n_2/n_1 can be optimized to minimize the cost for a given error via Lagrange multipliers. See also [G Bali et al,CPC 181 (2010) 1570].



Also studied in

[C Alexandrou et al,CPC 183 (2012) 1215],

[T Blum et al,PRD 88 (2013) 094503].

Other factorizations of M^{-1} into an expensive part with a small error and a cheap part with a larger error:

e.g., frequency splitting, low modes, ...

Efficacy depends on the solver, e.g. $O(50)$ iterations for IDFLS solve to convergence for $m_\pi = 130$ MeV, $V = 96^3 \times 192$.

TSM 2: two-point function

[T Blum et al,PRD 88 (2013) 094503] [E Shintani et al,1402.0244]:
"Covariant approximation averaging": use of TSM with different numbers of point sources, exploiting translational invariance of $A_x = \langle C_{2pt}(x,t)\rangle$ to remove the bias.

"All mode averaging": combining this with low mode averaging (LMA).
TSM+low modes for the loops also studied in [G Bali et al,CPC 181 (2010) 1570].

Decompose

$$A = A_{\mathrm{approx}}|_Z + [A_{\mathrm{exact}} - A_{\mathrm{approx}}]|_{Z_0}\,, \quad \dim Z > \dim Z_0$$

A_{approx} may be computed for many source points $\in Z$.

Again, problem: more efficient solver $\longrightarrow$ less gain, other overheads, e.g., smearing.

Care must be taken with non-linear applications.

Gauge vs stochastic noise

On each configuration an estimate A_E of A has a stochastic error $\Delta_{\mathrm{stoch}} A = \mathcal{O}(1/\sqrt{n})$. We define its ensemble average over N independent configuration:

$$\sigma^2_{A,\mathrm{stoch}} := \frac{\langle (\Delta_{\mathrm{stoch}} A)^2 \rangle}{N} \propto \frac{1}{Nn} \quad \text{for} \quad n,\ N \ \text{large.}$$

The ensemble average $\langle A_E \rangle$ carries the statistical error ΔA. We define

$$\Delta A^2 = \sigma^2_{A,\mathrm{gauge}} + \sigma^2_{A,\mathrm{stoch}} \quad \propto \frac{1}{N}\left[1 + O\left(\frac{1}{n}\right)\right].$$

(1) $\sigma_{A,\mathrm{gauge}} < \sigma_{A,\mathrm{stoch}} \longrightarrow$ **increase n.**

(2) $\sigma_{A,\mathrm{gauge}} \gg \sigma_{A,\mathrm{stoch}} \longrightarrow$ **reduce n and increase N (or # of source positions).**

The optimal choice depends on the observable A.

Instead of (1) can reduce the coefficient of the $1/\sqrt{n}$ term.

Stochastic noise

$$M_E^{-1} = M^{-1} \underbrace{\left(\overline{|\eta\rangle\langle\eta|} - \mathbb{1}\right)}_{\mathcal{O}(1/\sqrt{n})} + M^{-1} \qquad \left[\Delta M_{XZ}^{-1}\right]^2 \propto \frac{1}{n} \sum_{Y \neq X,Z} M_{XY}^{-1} M_{YZ}^{-1\dagger}.$$

Off-diagonal entries of M^{-1} will determine the stochastic error.

For the disconnected loop:

$$\left[\Delta \left(\mathrm{tr}\ \Gamma M^{-1}\right)\right]^2 \propto \frac{1}{n} \sum_{x,y,x \neq y} C_\Gamma(y-x) + (\text{x=y, non-diagonal terms in the spin and colour})$$

$C_\Gamma(y-x)$ is the point-point meson correlation function for $O_M = \bar{q}\Gamma\gamma_5 q$.

Biggest contributions are from the "neighbourhood", where $C_\Gamma(y-x)$ is large.

Avoidance of short distance noise:

Partitioning (= dilution) [S Bernardson et al,CPC 78 (1993) 256] [J Viehoff et al,NPPS 63 (1998) 269] [W Wilcox,hep-lat/9911013].

Hopping parameter expansion [C Thron et al,PRD 57 (1998) 1642] [C Michael et al,NPPS 83 (2000) 185].

One-end-trick [R Sommer,NPPS 42 (1995) 186] [M Foster, C Michael,PRD 99 (1999) 074503] [C McNeile, C Michael,PRD 73 (2006) 074506].

Noise reduction methods: partitioning

... also known as spin-explicit method (SEM) or dilution.
Decompose $\mathcal{R} =$ volume $\otimes$ colour $\otimes$ spin into n_{p} subspaces:

$$\mathcal{R} = \oplus_{j=1}^{n_{\mathrm{p}}} \mathcal{R}_j \,.$$

Set components of $|\eta^{\ell}_{|j}\rangle$ to zero outside of the supporting domain $\mathcal{R}_j$.
Calculate restricted solutions

$$M|s^{\ell}_{|j}\rangle = |\eta^{\ell}_{|j}\rangle \,.$$

Now: $M_E^{-1} = \sum_j \overline{|s_{|j}\rangle\langle\eta_{|j}|}$

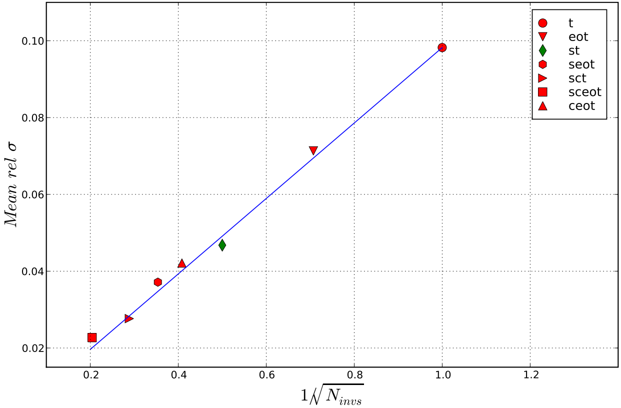
This can be used to black out large off-diagonal error terms.

Can choose the same random vector components within each subspace (if these share the same dimension).

Example: spin-explicit method. Same noise for each spin-component.

Partitioning 2

Higher # of subsets $\rightarrow$ higher minimal # of inversions. Over-partitioning: danger of carrying out more inversions than necessary for given gauge error.



Comparison of partitioning patterns pseudoscalar three point functions [R Evans et al,PRD 82 (10) 094501] 1 config., vector current, $t \rightarrow 100$ estimates.

Partitioning 3

Clear gain when not all columns of M^{-1} are required (e.g. time partitioning for 3-point functions).

Spin partitioning for non-pseudoscalars can give an error reduction larger than two if the same noise vectors are used
(cost increase by a factor of four.. so worth doing).

The partitioning pattern can be adapted to the problem
([C Ehmann et al,0903.2947]).

Also possible to increase # of partitions subsequently by choosing recursive binary pattern. See “Hierarchical probing” [A Strathopoulos et al,1302.4018].

Noise reduction methods: hopping parameter expansion

Exploits ultra-locality of the action (if ultra-local). [C Thron et al,PRD 57 (98) 1642] [C Michael et al,NPPS 83 (00) 185]. For separated source and sink (together with eigenmodes) [GB et al,PRD 71 (05) 114513].

$$M^{-1} = 2\kappa \left(\mathbb{1} - \kappa D\right)^{-1} = 2\kappa \sum_j (\kappa D)^j$$

$$= 2\kappa \sum_{j=0}^{n-1} (\kappa D)^j + (\kappa D)^n M^{-1}$$

The first terms of the HPE contribute most to the noise.

These may vanish identically:

- $\mathrm{Tr}(\Gamma M^{-1}) = \mathrm{Tr}(\Gamma \kappa^n D^n M^{-1})$, where n depends on Γ and the action.

Provides an improved estimate for small κ, large quark mass (e.g. strange).

The next few terms can in principle be computed analytically (cumbersome and implementation can be costly).

NB: a clover-specific implementation exists [V Gülpers et al,1309.2104], however, the resulting n is smaller or equal to that of the above naive method.

“Colouring” of source positions

Obtaining subsets of lattice points that are separated by a minimal number of “hops”, e.g., for a nearest-neighbour action like clover-Wilson can be very useful:

Possible application 1: partitioning/dilution pattern.

Possible application 2: increase the HPE order from n to $n + m$ by computing the first m non-vanishing terms exactly.

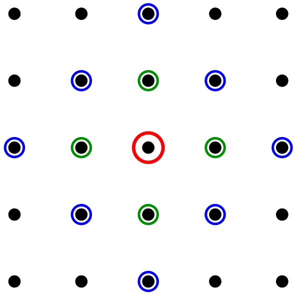
This can be done, applying $(\kappa D)^k$ for $k = 1, \ldots, m$ to a point source. Carrying this out for V points is of course prohibitively expensive.

If, however, k different sets exist that can only be connected by $m + 1$ hops then this can be achieved in parallel for each set (k times rather than V times).

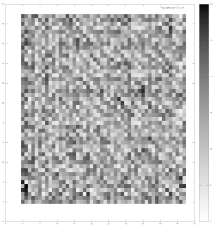
Note that multiplication with κD is an inexpensive operation (sparse matrix times vector).

Work in progress

Implementation of colouring



iterate:
- loop over spatial lattice sites:
  - loop over 1-hop neighbours $\vec{x}_{\mathrm{neigh}}$:
    - if colour$[\vec{x}]$ $==$ colour$[\vec{x}_{\mathrm{neigh}}]$
      - colour$[\vec{x}]$++
    - loop over 1-hop neighbours of $\vec{x}_{\mathrm{neigh}}$:
      - if colour$[\vec{x}]$ $==$ colour$[\vec{x}_{\mathrm{neigh+1hop}}]$
      - colour$[\vec{x}]$++



+2 hops: 32 colours (0 16, 1 4104, 2 2621, 3 5939,..., 20 2956, 21 2304, 22 1615, 23 1010, 24 582, 25 293, 26 132, 27 55, 28 21, 29 10, 30 3, 31 1)

+4 hops: 107 colours

+5 hops: 148 colours

+6 hops: 248 colours

Colouring the source positions

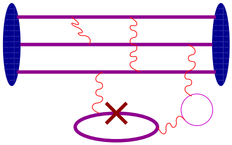
Better patterns (smaller number of colours) possible.

Compute non-zero contribution to $Tr(\Gamma\kappa^p D^p M^{-1})$, $p = n$ to $n + m - 1$ for each colour source (populated with 1s)

Compute $Tr(\Gamma\kappa^{n+m} D^{n+m} M^{-1})$ stochastically.

Cluster decomposition error reduction: getting rid of large distances

Nucleon at rest:

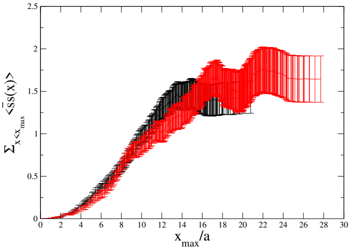


$$R^{\mathrm{dis}}(t,t_{\mathrm{f}}) = -\frac{\left\langle \Gamma^{\alpha\beta}_{\mathrm{2pt}} C^{\beta\alpha}_{\mathrm{2pt}}(t_{\mathrm{f}}) \sum_{\mathbf{x}} L(\mathbf{x},t;\mathbf{x},t) \right\rangle_c}{\left\langle \Gamma^{\alpha\beta}_{\mathrm{unpol}} C^{\beta\alpha}_{\mathrm{2pt}}(t_{\mathrm{f}}) \right\rangle_c}$$

where

$$L(\mathbf{x},t;\mathbf{x},t) = \mathrm{Tr}\left(M^{-1}(\mathbf{x},t;\mathbf{x},t)\Gamma_{\mathrm{loop}}\right)$$

$$C^{\beta\alpha}_{\mathrm{2pt}}(t_{\mathrm{f}}) = \sum_{\mathbf{y}} C^{\beta\alpha}_{\mathrm{2pt}}(\mathbf{y},t_{\mathrm{f}};\mathbf{0},0)$$



Numerator $\sim c(1 - e^{-mx}(mx+1))$. Assume $L(\mathbf{x},t;\mathbf{x},t) \sim e^{-mx}/x$

Large $x = |\mathbf{x}|$ (2pt source at $\mathbf{0}$) only contribute to the noise.

$\Rightarrow$ restrict the sum over x.

[QCDSF,0911.2407] $M_\pi = 290$ MeV, $a = 0.076$ fm, $L = 24a$, $32a$.

Expectation: need to sum up to $x \sim 1/m$, $m \geq m_\pi$

Remove bias: fit to finite x and extrapolate.

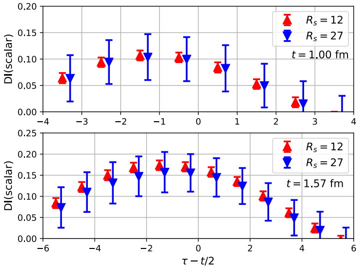
Likely to obtain gains for lattices with large LM_π

Challenges: dependence on the fit form and correlations between results at different x.

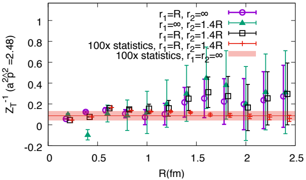
Again, work in progress.

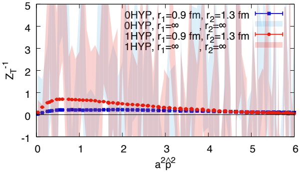
[χQCD,1705.06358]

$g_S^{s,dis} = 0.160(15)(15)$ (CDER)
cf. $\quad 0.143(45)$ (no CDER)



[χQCD,1805.00531] (including P. Shanahan)





Noise reduction methods: OET

Define noise $\eta^\ell_{xj\alpha} \in Z$ that is zero for any timeslice $x_4 \neq t_0 = 0$.

$$\frac{1}{n}\sum_{\ell=1}^{n} |\eta^\ell\rangle\langle\eta^\ell| = \mathbb{1}_{t_0} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \approx \sum_{\mathbf{x}j\alpha} |(\mathbf{x}, t_0)j\alpha\rangle\langle(\mathbf{x}, t_0)j\alpha|\,,$$

Consider the (not ensemble averaged) pion two-point function ($y = (\mathbf{y}, t)$),

$$\begin{aligned} C_\pi(t) &= \sum_{\mathbf{xy}} \mathrm{tr}\; M^{-1}_{yx}[M^{-1}_{xy}]^\dagger \approx C_{\pi E}(t) \\ &= \sum_{\mathbf{y}} \frac{1}{n}\sum_{\ell=1}^{n} \mathsf{Tr}\,\langle y|M^{-1}|\eta^\ell\rangle\langle\eta^\ell|M^{-1\dagger}|y\rangle \\ &= \sum_{\mathbf{y}} \frac{1}{n}\sum_{\ell=1}^{n} \mathrm{tr}\,\langle y|s^\ell\rangle\langle s^\ell|y\rangle = \sum_{\mathbf{y}k\beta} \frac{1}{n}\sum_{\ell=1}^{n} |s^\ell_{yk\beta}|^2, \end{aligned}$$

where $M|s^\ell\rangle = |\eta^\ell\rangle$. $C_{\pi E}(t)$ differs from $C_\pi(t)$ by terms of $\mathcal{O}(1/\sqrt{n})$.
Since the noise is unbiased: $\langle C_\pi(t)\rangle = \langle C_{\pi E}(t)\rangle$.

OET 2

Without the OET we would have needed two sets of sources $|\eta_1^\ell\rangle$ and $|\eta_2^\ell\rangle$:

$$C_{\pi E}^{\rm naive}(t) = \sum_{\mathbf{y}} \frac{1}{n^2} \sum_{\ell,k=1}^{n} \mathrm{tr}\,\langle y|s_1^\ell\rangle\langle\eta_1^\ell|\eta_2^k\rangle\langle s_2^k|y\rangle$$

$$= \sum_{\mathbf{y}} \frac{1}{n^2} \sum_{\ell,k=1}^{n} \mathrm{tr}\,\langle y|M^{-1}\overline{|\eta_1\rangle\langle\eta_1|}\,\overline{|\eta_2\rangle\langle\eta_2|}M^{-1\dagger}|y\rangle.$$

Each outer product $\overline{|\eta\rangle\langle\eta|}$ involves a sum over $12V_3$ randomly oscillating contributions.

Therefore the error is $\propto \sqrt{V_3^2/n}$. Self-averaging over the source positions gives a factor $1/\sqrt{V_3}$, relative to the point-to-all method: for a constant error, an increase $n \propto V_3$ is needed!

The OET removes one $|\eta\rangle\langle\eta|$ product and therefore a factor $\sqrt{V_3}$.

$\longrightarrow$ **The OET stochastic error scales $\propto 1/\sqrt{n}$!**

The prefactor can be reduced via "thinning" or (at the cost of more inversions) "partitioning". Too much "thinning" reduces the self-averaging effect.

OET 3

In general: **the number of different stochastic propagators must be kept small.**

$\longrightarrow$ **combine OET with sequential propagators for n-point functions**
[S Aoki et al,PRD 76 (2007) 094506] [S Simula et al,PoS (LAT2007) 371]
[P Boyle et al,JHEP 0807 (2008) 112]

Often $n = 1$ is sufficient $\longrightarrow$ cost smaller than point-to-all.

NB: "recycling" trick if independent stochastic sources are needed:
[J Foley et al,CPC 172 (2005) 145]

$$\frac{1}{n^2}\sum_{\ell,k}^{n}\langle\eta_1^\ell|\eta_2^k\rangle \mapsto \frac{1}{n(n-1)}\sum_{\ell\neq k}^{n}\langle\eta^\ell|\eta^k\rangle.$$

(Reduction of # of inversions by almost a factor of two.)

Summary

★ Tremendous progress in the computation of hadron structure observables due (in part) to the development of stochastic methods and associated variance reduction techniques:

— Flexible determination of hadron (N, Λ, Σ, Ξ, Δ, Σ^*, Ξ^*, Ω) matrix elements

— Largest improvements come from combining several methods for variance reduction (TSM, HPE, partitioning, . . .).

— Reducing gauge error: TSM extensions (AMA), low mode averaging, . . .

— Smallest error for given cost/effort: need to balance stochastic error vs gauge error.

— Highly efficient solvers mean other aspects of the calculation need to be improved, e.g. reducing the cost of the smearing.

— Further methods being developed, frequency splitting, . . .

Outlook

★ One of the main challenges is dealing with the signal/noise problem (which reduces exponentially) and the contribution from excited states.

★ Designing better hadron interpolators, including those for multiparticles (e.g. $N\pi$).

★ Dealing with a large interpolator basis (contractions and methods to evaluate the associated diagrams).