Efficient computations of correlators with local distillation

LATTICE 2023, Fermilab

Algorithms and Artificial Intelligence

Nicolas Lang*, Mike Peardon, Robert Edwards

August 1st 2023

Why distillation?

Good signal-to-noise ratio

- Most correlators: exponential fall-off of signal-to-noise
- Experience shows: smeared operators improve the signal by increasing overlap on low-momentum modes

Reduce cost of spectroscopy calculations

- Only important degrees of freedom $\rightarrow$ lower-rank space
- Often good results require large variational bases of operators - $\bar{q}q$ and multi-hadron operators $\rightarrow$ large amount of Wick contractions $\rightarrow$ reuse of propagators and operators desirable

Distillation addresses both of these

Recap: smearing

Consider a single-meson operator:

$$\mathcal{O}_M = \bar{q}\Gamma_i q'$$

Smearing is the application of an operator $\square$ to the quark-fields:

$$\tilde{\bar{q}} = \square \bar{q}$$
$$\tilde{q}' = \square q'$$
$$\mathcal{O}_M \to \tilde{\mathcal{O}}_M = \tilde{\bar{q}}\Gamma_i \tilde{q}'$$

Objective:
Maximize $\langle \mathfrak{n} | \mathcal{O}_M | 0 \rangle$ for some low-lying state of interest $|\mathfrak{n}\rangle$.
Empirically Gaussian smearing shapes work well.

Desirable properties:

- Gauge-covariance
- Preservation of other symmetries
- Typically trivial action in time and spin

Recap: distillation (1)

Gauge-covariant Laplace operator:

$$\nabla^2_{xy}(t) = -6\delta_{xy} + \sum_{j=1}^{3} \left(U_j(x,t)\delta_{x+\hat{j},y} + U_j^\dagger(x-\hat{j},t)\delta_{x-\hat{j},y} \right)$$

Gaussian (Jacobi) smearing:

$$J(t;\sigma,n_\sigma) = \left(1 + \frac{\sigma\nabla^2(t)}{n_\sigma}\right)^{n_\sigma}$$

$$\lim_{n_\sigma\to\infty} J(t;\sigma,n_\sigma) = Q(t)\exp\left[\sigma\Lambda(t)\right]Q^\dagger(t)$$

$\Lambda(t)$ is the diagonal matrix of eigenvalues of $\nabla^2(t)$

Distillation operator:

$$\left[\Box(t)\right]_{xy} = \left[V(t)V^\dagger(t)\right]_{xy} = \sum_{k=1}^{N} v_x^{(k)}(t)v_y^{(k)\dagger}(t)$$

$V(t)$: first N_D column vectors of $Q(t)$; $\sigma = 0$

Recap: distillation (2)

Some properties of $\Box(t)$:

- acts in position- and colour space (trivial in time and spin)
- $\left[\Box(t)\right]^2 = \Box(t)$ (projector)
- preserves translation-, rotation and gauge-symmetries

We can now compute correlation functions in *distillation space*:

- Meson correlator:

$$C_M(t',t) = \langle \bar{q}'(t')\Box(t)\Gamma^B(t')\Box(t)q(t') \;\; \bar{q}(t)\Box(t)\Gamma^A(t)\Box(t)q'(t)\rangle$$
$$\to C_M^{\text{conn.}}(t',t) = \text{Tr}\left[\phi^B(t')\tau(t',t)\phi^A(t)\tau(t,t')\right] .$$

- Distillation space objects:

$$\phi^X_{\alpha\beta}(t) = V^\dagger(t)\Gamma^X_{\alpha\beta}(t)V(t) \;\; \text{(elemental)}$$
$$\tau_{\alpha\beta}(t',t) = V^\dagger(t')M^{-1}_{\alpha\beta}(t',t)V(t) \;\; \text{(perambulator)}$$

Meson 2-point function (connected piece):

$$C_M^{\text{conn.}}(t,t') = \text{Tr}\left[\Phi^B(t')\tau(t',t)\Phi^A(t)\tau(t,t')\right]$$
$$= \Phi^B_{\alpha\beta}(t')\tau_{\alpha\bar{\alpha}}(t',t)\Phi^A_{\bar{\alpha}\bar{\beta}}(t)\tau_{\bar{\beta}\beta}(t,t')$$

Computational effort? Produce temporaries:

$\Phi^{A'}{}_{\bar{\alpha}\beta}(t,t') = \Phi^A_{\bar{\alpha}\bar{\beta}}(t)\tau_{\bar{\beta}\beta}(t,t')$ and $\Phi^{B'}{}_{\bar{\alpha}\beta}(t',t) = \Phi^B_{\alpha\beta}(t')\tau_{\alpha\bar{\alpha}}(t',t)$

$$C_M^{\text{conn.}}(t,t') = \Phi^{A'}{}_{\bar{\alpha}\beta}(t,t')\Phi^{B'}{}_{\bar{\alpha}\beta}(t',t)$$

$\to \mathcal{O}(N_D^3)$

Baryon: $\Phi^{B(1)'}{}_{\bar{\alpha}\beta\gamma}(t',t) = \Phi^B_{\alpha\beta\gamma}(t')\tau_{\alpha\bar{\alpha}}(t',t), \ldots$

$$C_B(t,t') = \Phi^{B(3)}_{\bar{\alpha}\bar{\beta}\bar{\gamma}}(t')\Phi^A_{\bar{\alpha}\bar{\beta}\bar{\gamma}}(t)$$

$\to \mathcal{O}(N_D^4)$

In general: $\mathcal{O}(N_D^{(d+1)})$ (for d-quark operator). **Can we do better?**

Local basis of distillation space

Jacobi smearing preserves locality $\rightarrow$ find a local basis of distillation space

- Embed coarse grid $G \subset \Lambda_3$ into lattice
- Place three gauge-covariant sources $q^{(j)}$ at every $x \in G$; $Q_{ij} = q_i^{(j)}$
- These can be constructed from Laplacian eigenvector components
- Project to distillation space: $W = \Box Q$
  $\rightarrow$ bijective map:

$$f : \mathcal{D} \rightarrow G \times \mathcal{C};\ i \mapsto (\boldsymbol{x}, c)$$

  $\mathcal{D} = \{1, \ldots, N_D\}$, $\mathcal{C} = \{1, 2, 3\}$

- Various choices for coarse grid: cubic, face-centred, body-centred

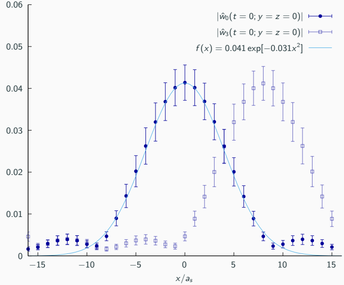
Unitary transformation to new basis

Basis transformation:

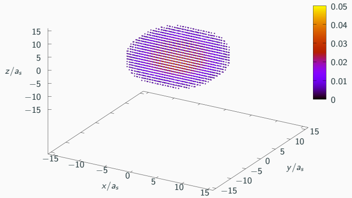
- $A_0 \equiv V^\dagger W = V^\dagger Q$
  $\Leftrightarrow VA_0 = VV^\dagger W = W$
- Would like unitary $\hat{A}^\dagger \hat{A} = \mathbb{1}$
  and $\hat{W} \equiv V\hat{A}$

Permutation-invariant orthogonalization:

- $A(\tau)$ with $\lim_{\tau\to\infty} A(\tau)^\dagger A(\tau) = \mathbb{1}$
  generated by
  $S(A) = \frac{1}{2}\,\mathrm{Tr}[(I - AA^\dagger)^2]$
  $\to \frac{dA}{d\tau} = \frac{\partial S}{\partial [A^\dagger]} = (I - AA^\dagger)A$
- Solve numerically with $A(0) = A_0$
- $\hat{A}$: fixed point of flow



(a) Slice along $\hat{x}$ of $|\hat{w}^{(i=0,3)}(t=0)|$



(b) Scatter plot of $|\hat{w}^{(0)}(t=0)|$ (cut-off applied)

Sampling large sums

Consider sum of complex terms:

$$A = \sum_{i=1}^{N} a_i$$

Draw a sample of indices $S = (s_1, s_2, \ldots, s_n)$ with $s_j = i \in 1, \ldots, N$ with probability p_i and define the *Hansen-Hurwitz estimator*:

$$\hat{A}_{\text{HH}} = \frac{1}{n} \sum_{s \in S} \frac{a_s}{p_s} \Rightarrow \mathbb{E}[\hat{A}_{\text{HH}}] = A$$

which has the variance

$$\mathrm{Var}[\hat{A}_{\text{HH}}] = \frac{1}{n} \sum_{i=1}^{N} p_i \left| \frac{a_i}{p_i} - A \right|^2$$

Variance-minimizing choice of weights: $p_i^* = |a_i| / \sum |a_i|$

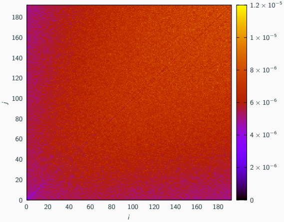
Note: The variance is zero with this choice if all phases are the same

Sparse elementals and weights

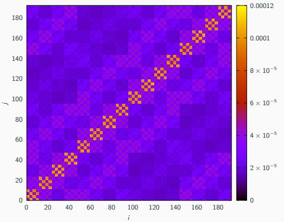
Apply to correlator C:

$$C \to \hat{C} = \frac{1}{n} \sum_{\sigma \in S} \frac{C_\sigma}{p_\sigma}$$

- Ω: index space; $\sigma \in \Omega$
- $\sigma = (\alpha, \beta, \ldots, \bar{\alpha}, \bar{\beta}, \ldots)$
  $\to C_\sigma = \phi^A_{\alpha\beta\ldots} \tau_{\alpha\bar{\alpha}} \tau_{\beta\bar{\beta}} \ldots \phi^B_{\bar{\alpha}\bar{\beta}\ldots}$
  (no summation)
- $p^*_\sigma = |C_\sigma| / \sum_\sigma |C_\sigma|$
- Approximate:
  $p^*_\sigma \approx \langle|\phi^A_{\alpha\beta\ldots}|\rangle\langle|\phi^B_{\bar{\alpha}\bar{\beta}\ldots}|\rangle / \sum_\sigma \ldots$
  where $\langle\ldots\rangle$ is the average over configurations and time-slices



(a) Before unitary transformation



(b) After unitary transformation

Fig: $\langle|\phi^M_{\alpha\beta}(p = 0, t = 0)|\rangle$ (derivative in x-direction)

Sparse contractions

Let's look at baryon contraction again:

$$\Phi^{B(1)}_{\bar{\alpha}\beta\gamma} = \Phi^{B}_{\alpha\beta\gamma}\tau_{\alpha\bar{\alpha}}$$

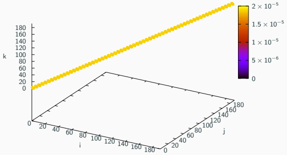
$$\Phi^{B(2)}_{\bar{\alpha}\bar{\beta}\gamma} = \Phi^{B(1)}_{\bar{\alpha}\beta\gamma}\tau_{\beta\bar{\beta}}$$

$$\Phi^{B(3)}_{\bar{\alpha}\bar{\beta}\bar{\gamma}} = \Phi^{B(2)}_{\bar{\alpha}\bar{\beta}\gamma}\tau_{\gamma\bar{\gamma}}$$

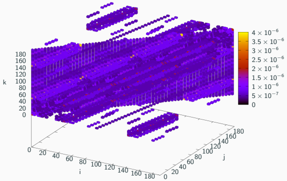
$$C(t,t') = \Phi^{B(3)}_{\bar{\alpha}\bar{\beta}\bar{\gamma}}\Phi^{A}_{\bar{\alpha}\bar{\beta}\bar{\gamma}}(t)$$

Apply sampling:

- ϕ^A and ϕ^B now sparse
- τ is dense
- Only run over indices contained in sample
- **Important**: "Look ahead"
  - barred indices are also sparse!



(a) No derivative



(b) Derivative in z-direction

Fig: $\langle|\phi^{B}_{\alpha\beta\gamma}(p=0,t=0)|\rangle$
(cut-offs applied in plots)

What is the cost?

Number of complex multiplications given by expected number of distinct indices in sample:

$$\bar{\nu}_s \equiv \mathbb{E}[\nu_s] = \mathbb{E}\left[\sum_{\sigma\in\Omega} I_\sigma^s\right] = \sum_{\sigma\in\Omega} \pi_\sigma^s \ .$$

- π_σ^s: inclusion probabilities, $\sum_\sigma \pi_\sigma^s = n$
- For sampling with replacement:

$$\bar{\nu}_s = N - \sum_{\sigma\in\Omega}(1-p_\sigma)^n \leq n$$

- In practice weak dependence on details of p

Consider $\Phi^{B(1)}_{\bar{\alpha}\beta\gamma} = \Phi^B_{\alpha\beta\gamma}\tau_{\alpha\bar{\alpha}}$:
$\rightarrow$Full distribution for $\{\,\alpha,\beta,\gamma\,\}$; *"projected"* distribution $\sum_{\bar{\beta},\bar{\gamma}} p_{\bar{\alpha}\bar{\beta}\bar{\gamma}}$ for $\{\,\bar{\alpha}\,\}$

What can we hope to gain?

The additional variance $\mathrm{Var}_{\mathrm{HH}}\,\hat{C}$ due to distillation-space sampling depends strongly on the structure of the individual operators **and** their correlation.

- Spatial derivatives produce more tensor entries with large magnitudes
- Off-diagonal correlators $\langle \mathcal{O}_i \mathcal{O}_j \rangle$, $i \neq j$, tend to have larger variance
- The variance grows with the time-slice

$\rightarrow$ Need to pick sample-size accordingly
(this is work in progress)

Best case:
$n = \mathcal{O}(N_D)$ is sufficient
Cost for baryon contraction $\rightarrow \mathcal{O}(N_D^2)$ (vs. $\mathcal{O}(N_D^4)$!)

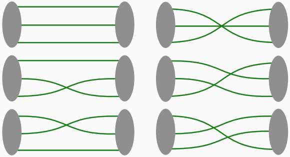
Expectation:
For compact operators the sample size needed to achieve constant variance grows less-than exponentially in the dimension of the operators in distillation space.

Application

As a test case we compute
$C_{N\to N}(t) = \langle \mathcal{O}_N(t)\mathcal{O}_N^\dagger(0)\rangle$ and
$C_{\Delta\to\Delta}(t) = \langle \mathcal{O}_\Delta(t)\mathcal{O}_\Delta^\dagger(0)\rangle$ with

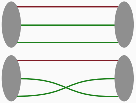
$$O_N^{J=1/2} = \left(N_M \otimes (\tfrac{1}{2}^-)_M \otimes 1_{L=0,S}\right)^{J=\frac{1}{2}}$$

$$O_\Delta^{J=1/2} = \left(\Delta_S \otimes (\tfrac{1}{2}^-)_M \otimes D^{[1]}_{L=1,M}\right)^{J=\frac{1}{2}}$$



(a) $C_{\Delta\to\Delta}$

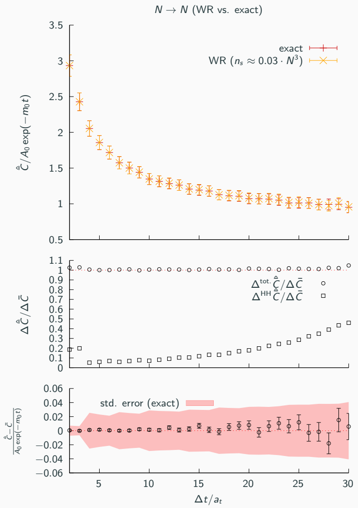
- Follows HadSpec conventions[1] for baryon operators
- Product of flavour, spin and orbital angular momentum representations
- Subscripts S and M indicate *symmetric* and *mixed-symmetric* representations
- Overall anti-symmetric as required



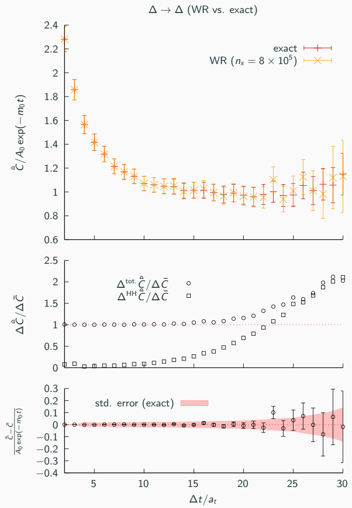
(b) $C_{N\to N}$

[1] arXiv:1104.5152

$C_{N\to N}$ and $C_{\Delta\to\Delta}$: Results



(a) Nucleon correlator



(b) Δ correlator

Summary and outlook

Summary:

- Locality in distillation space allows more efficient Wick contractions
- Method gives sensible results for baryons
- Promising potential especially for high-dimensional compact operators (tetraquarks, ...)

Lots of room for improvements and further study:

- dependence of variance on operator structure and dimensionality $\rightarrow$ recipe to pick a sample size
- adaptive sample size across time slices
- blocked/stratified sampling?
- different grid embeddings
- infer sampling weights from symmetry (needed for large-d)

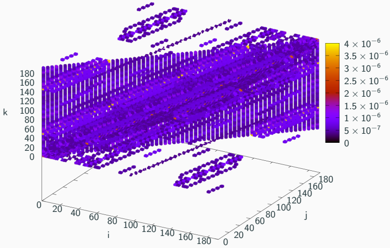
Ultimately: Application to large operator basis with GEV method

References

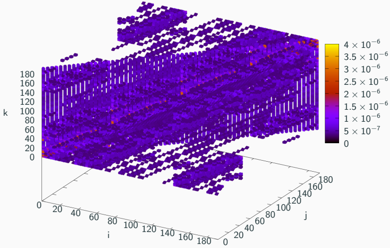
- M. Peardon et al., *A Novel quark-field creation operator construction for hadronic physics in lattice QCD*, 10.1103/PhysRevD.80.054506 (Original distillation paper)
- C. Morningstar et al., *Improved stochastic estimation of quark propagation with Laplacian Heaviside smearing in lattice QCD*, 10.1103/PhysRevD.83.114505 (different approach using sampling but without locality)
- M. H. Hansen and W. N. Hurwitz, *On the Theory of Sampling from Finite Populations*, The Annals of Mathematical Statistics, Vol. 4
- R. Edwards et al., *Excited state baryon spectroscopy from lattice QCD*, 10.1103/PhysRevD.84.074508 (baryon operator construction)

Thank you! Questions?

$\triangle$ weights

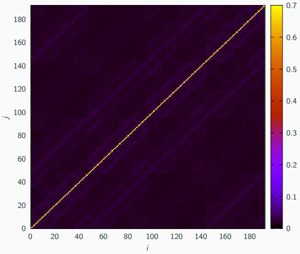


(a) $s = \{\downarrow, \downarrow, \uparrow\}$

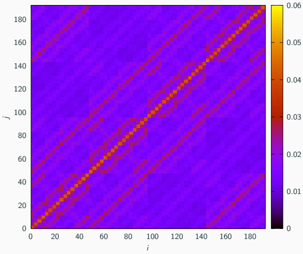


(b) $s = \{\uparrow, \uparrow, \downarrow\}$

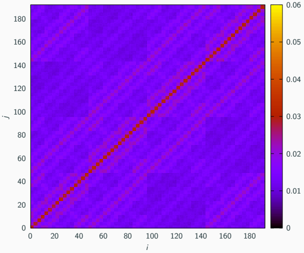
Perambulator weights in local distillation



(a) $\hat{G}(\Delta t = 0)$



(b) $\hat{G}(\Delta t = 8)$



(c) $\hat{G}(\Delta t = 16)$