Tackling the Signal to Noise problem with Stochastic Automatic Differentiation

[G.C., A. Ramos, arXiv:2502.15570]

Guilherme Catumba
Alberto Ramos

Università degli Studi di Milano-Bicocca

April 2, 2025

The Signal to Noise problem

- Physical properties – Euclidean time behavior of 2-point functions

$$C(x_0) = \sum_{\vec{x}} \left\langle O^\dagger(x) O(0) \right\rangle_{\mathrm{c}} = \sum_n |\langle 0|O|n\rangle|^2 e^{-E_n x_0} \xrightarrow[x_0 \to \infty]{} |\langle 0|O|i\rangle|^2 e^{-E_i x_0}$$

  - large time behavior
  - energy levels, matrix elements

The Signal to Noise problem

- Physical properties – Euclidean time behavior of 2-point functions

$$C(x_0) = \sum_{\vec{x}} \left\langle O^\dagger(x) O(0) \right\rangle_c = \sum_n |\langle 0|O|n\rangle|^2 e^{-E_n x_0} \xrightarrow[x_0\to\infty]{} |\langle 0|O|i\rangle|^2 e^{-E_i x_0}$$

  - large time behavior
  - energy levels, matrix elements
- Parisi/Lepage argument for the variance [Parisi 1984; Lepage 1989]
  - Case of a scalar interpolator

$$\sigma^2(x_0) \propto \left\langle O^2(x_0) O^2(0) \right\rangle - \left\langle O(x_0) O(0) \right\rangle^2$$

$$\langle A(x_0) B(0) \rangle \xrightarrow[x_0\to\infty]{} \langle A(x_0) \rangle \langle B(0) \rangle = \langle A(0) \rangle \langle B(0) \rangle$$

$$\sigma(x_0) \xrightarrow[x_0\to\infty]{} C(x_0 = 0)/\sqrt{N_{\mathrm{conf}}}$$

The Signal to Noise problem

- Physical properties – Euclidean time behavior of 2-point functions

$$C(x_0) = \sum_{\vec{x}} \left\langle O^\dagger(x) O(0) \right\rangle_c = \sum_n |\langle 0|O|n\rangle|^2 e^{-E_n x_0} \xrightarrow[x_0\to\infty]{} |\langle 0|O|i\rangle|^2 e^{-E_i x_0}$$

  - large time behavior
  - energy levels, matrix elements
- Parisi/Lepage argument for the variance [Parisi 1984; Lepage 1989]
  - Case of a scalar interpolator

$$\sigma^2(x_0) \propto \left\langle O^2(x_0) O^2(0) \right\rangle - \left\langle O(x_0) O(0) \right\rangle^2$$

$$\langle A(x_0) B(0)\rangle \xrightarrow[x_0\to\infty]{} \langle A(x_0)\rangle \langle B(0)\rangle = \langle A(0)\rangle \langle B(0)\rangle$$

$$\sigma(x_0) \xrightarrow[x_0\to\infty]{} C(x_0 = 0)/\sqrt{N_{\text{conf}}}$$

  - Signal to Noise

$$\frac{\text{error}[C(x_0)]}{C(x_0)} \xrightarrow[x_0\to\infty]{} \frac{e^{E_0 x_0}}{\sqrt{N_{\text{conf}}}}$$

The Signal to Noise problem

- More generally: apply the same argument to the variance

$$\left\langle O^\dagger(x)O^\dagger(x)O(0)O(0)\right\rangle = \sum_n \left|\langle 0|O^2|n\rangle\right|^2 e^{-E_n x_0} \xrightarrow[x_0\to\infty]{} |\langle 0|O|j\rangle|^2 e^{-E_j x_0}$$

The Signal to Noise problem

- More generally: apply the same argument to the variance

$$\left\langle O^\dagger(x) O^\dagger(x) O(0) O(0) \right\rangle = \sum_n \left| \langle 0 | O^2 | n \rangle \right|^2 e^{-E_n x_0} \xrightarrow[x_0 \to \infty]{} |\langle 0 | O | j \rangle|^2 e^{-E_j x_0}$$

$$\frac{\text{error}[C(x_0)]}{C(x_0)} \xrightarrow[x_0 \to \infty]{} \frac{e^{(E_i - E_j/2) x_0}}{\sqrt{N_{\text{conf}}}} .$$

The Signal to Noise problem

- More generally: apply the same argument to the variance

$$\left\langle O^\dagger(x)O^\dagger(x)O(0)O(0)\right\rangle = \sum_n \left|\langle 0|O^2|n\rangle\right|^2 e^{-E_n x_0} \xrightarrow[x_0\to\infty]{} |\langle 0|O|j\rangle|^2 e^{-E_j x_0}$$

$$\frac{\mathrm{error}[C(x_0)]}{C(x_0)} \xrightarrow[x_0\to\infty]{} \frac{e^{(E_i - E_j/2)x_0}}{\sqrt{N_{\mathrm{conf}}}}\,.$$

- Scalar: StN $\sim e^{E_0 x_0}$
- Pion: StN $\sim$ const.
- Proton: StN $\sim e^{(m_p - 3m_\pi)x_0}$

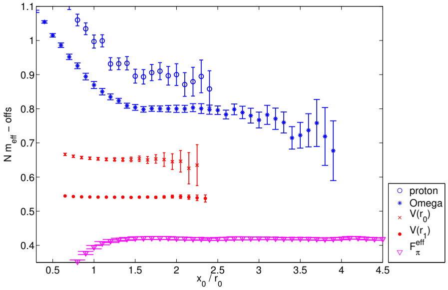
The Signal to Noise problem

- More generally: apply the same argument to the variance

$$\left\langle O^\dagger(x)O^\dagger(x)O(0)O(0)\right\rangle = \sum_n \left|\langle 0|O^2|n\rangle\right|^2 e^{-E_n x_0} \xrightarrow[x_0\to\infty]{} \left|\langle 0|O|j\rangle\right|^2 e^{-E_j x_0}$$

$$\frac{\text{error}[C(x_0)]}{C(x_0)} \xrightarrow[x_0\to\infty]{} \frac{e^{(E_i - E_j/2)x_0}}{\sqrt{N_{\text{conf}}}}.$$

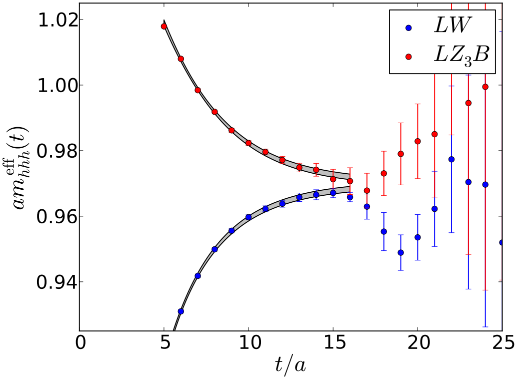
- Scalar: StN $\sim e^{E_0 x_0}$
- Pion: StN $\sim$ const.
- Proton: StN $\sim e^{(m_p - 3m_\pi)x_0}$



[Sommer 2014]

Solutions?

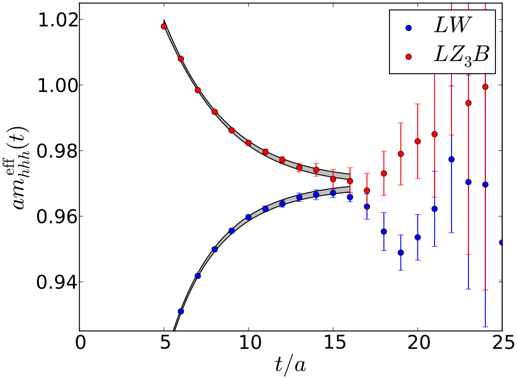
- Longstanding & ubiquitous problem in MC
- Typical solutions (smearing, GEVP, etc.)
  - Allow to extract information at earlier x_0
  - Excited state contamination
  - Do not solve the StN problem



[Blum et al. 2016]

Solutions?

- Longstanding & ubiquitous problem in MC
- Typical solutions (smearing, GEVP, etc.)
  - Allow to extract information at earlier x_0
  - Excited state contamination
  - Do not solve the StN problem



[Blum et al. 2016]

Source:
2-point expectation value

Derivatives & Sources – another perspective

- Do not compute a 2-point function

Derivatives & Sources – another perspective

- Do not compute a 2-point function
- Add source to the action

$$Z[J] = \int D\phi e^{-S[\phi]+JO(x_0=0)} \quad \longrightarrow \quad C(t) = \frac{\partial}{\partial J}\Big|_{J=0} \langle O(x_0) \rangle$$

1-point function!

Derivatives & Sources – another perspective

- Do not compute a 2-point function
- Add source to the action

$$Z[J] = \int D\phi e^{-S[\phi]+JO(x_0=0)} \qquad \longrightarrow \qquad C(t) = \frac{\partial}{\partial J}\bigg|_{J=0} \langle O(x_0)\rangle$$

1-point function!

- 1-point expectation value
- Variance of a derivative depends on how it is computed
  - (usual 2-point function corresponds to a reweighting evaluation of the derivative)

Derivatives & Sources – another perspective

- Do not compute a 2-point function
- Add source to the action

$$Z[J] = \int D\phi e^{-S[\phi]+JO(x_0=0)} \longrightarrow C(t) = \frac{\partial}{\partial J}\bigg|_{J=0} \langle O(x_0)\rangle$$

1-point function!

- 1-point expectation value
- Variance of a derivative depends on how it is computed
  - (usual 2-point function corresponds to a reweighting evaluation of the derivative)
- How to compute derivatives efficiently
  - Not with finite differences [Detmold 2005]
  - Stochastic automatic differentiation [G.C., A. Ramos, B. Zaldivar, 2024]

Automatic Differentiation – truncated polynomials

- Extend AD to MC methods
- Power series $\mathcal{O}\left(\varepsilon^K\right)$

$$\tilde{x} \equiv x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots + x_K\varepsilon^K$$

Automatic Differentiation – truncated polynomials

- Extend AD to MC methods
- Power series $\mathcal{O}\left(\varepsilon^K\right)$

$$\tilde{x} \equiv x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots + x_K\varepsilon^K$$

- Define operations & elementary functions – exact at each order

$$\tilde{x}\tilde{y} = x_0y_0 + (x_0y_1 + x_1y_0)\varepsilon + (x_0y_2 + 2x_1y_1 + x_2y_0)\varepsilon^2 + \ldots$$
$$\exp(\tilde{x}) = e^{x_0} + e^{x_0}x_1\varepsilon + e^{x_0}(x_1^2/2 + x_2)\varepsilon^2 + \ldots$$
$$\ldots$$

Automatic Differentiation – truncated polynomials

- Extend AD to MC methods
- Power series $\mathcal{O}(\varepsilon^K)$

$$\tilde{x} \equiv x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots + x_K\varepsilon^K$$

- Define operations & elementary functions – exact at each order

$$\tilde{x}\tilde{y} = x_0y_0 + (x_0y_1 + x_1y_0)\varepsilon + (x_0y_2 + 2x_1y_1 + x_2y_0)\varepsilon^2 + \dots$$
$$\exp(\tilde{x}) = e^{x_0} + e^{x_0}x_1\varepsilon + e^{x_0}(x_1^2/2 + x_2)\varepsilon^2 + \dots$$
$$\dots$$

- Chain rule

Automatic Differentiation – truncated polynomials

- Extend AD to MC methods
- Power series $\mathcal{O}\left(\varepsilon^K\right)$

$$\tilde{x} \equiv x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots + x_K\varepsilon^K$$

- Define operations & elementary functions – exact at each order

$$\begin{aligned} \tilde{x}\tilde{y} &= x_0y_0 + (x_0y_1 + x_1y_0)\varepsilon + (x_0y_2 + 2x_1y_1 + x_2y_0)\varepsilon^2 + \ldots \\ \exp(\tilde{x}) &= e^{x_0} + e^{x_0}x_1\varepsilon + e^{x_0}(x_1^2/2 + x_2)\varepsilon^2 + \ldots \\ &\ldots \end{aligned}$$

- Chain rule

Evaluate function at $\tilde{x} = x_0 + \varepsilon$ (Taylor theorem)

$$f(\tilde{x}) = f(x_0) + f'(x_0)\varepsilon + \frac{1}{2}f''(x_0)\varepsilon^2$$

Automatic Differentiation – truncated polynomials

- Extend AD to MC methods
- Power series $\mathcal{O}(\varepsilon^K)$

$$\tilde{x} \equiv x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots + x_K\varepsilon^K$$

- Define operations & elementary functions – exact at each order

$$\tilde{x}\tilde{y} = x_0y_0 + (x_0y_1 + x_1y_0)\varepsilon + (x_0y_2 + 2x_1y_1 + x_2y_0)\varepsilon^2 + \dots$$
$$\exp(\tilde{x}) = e^{x_0} + e^{x_0}x_1\varepsilon + e^{x_0}(x_1^2/2 + x_2)\varepsilon^2 + \dots$$
$$\dots$$

- Chain rule

Evaluate function at $\tilde{x} = x_0 + \varepsilon$ (Taylor theorem)

$$f(\tilde{x}) = f(x_0) + f'(x_0)\varepsilon + \frac{1}{2}f''(x_0)\varepsilon^2$$

deterministic function

Numerical Implementation

- “Teach” the computer to operate with truncated polynomials
- Compute Taylor expansion of any deterministic function

Numerical Implementation

- "Teach" the computer to operate with truncated polynomials
- Compute Taylor expansion of any deterministic function

Julia implementation:
https://igit.ific.uv.es/alramos/formalseries.jl

Code (Julia)

```
julia> using FormalSeries
julia> f(x) = 1/(1-x)
julia> xs = Series((0.0,1.0,0.0,0.0,0.0))
julia> f(xs)
Series{Float64, 5}((1.0, 1.0, 1.0, 1.0, 1.0))
```

Numerical Implementation

- "Teach" the computer to operate with truncated polynomials
- Compute Taylor expansion of any deterministic function

Julia implementation:
https://igit.ific.uv.es/alramos/formalseries.jl

Code (Julia)

```
julia> using FormalSeries
julia> f(x) = 1/(1-x)
julia> xs = Series((0.0,1.0,0.0,0.0,0.0))
julia> f(xs)
Series{Float64, 5}((1.0, 1.0, 1.0, 1.0, 1.0))
```

- Formally equivalent to:

$$f(0+\varepsilon) = \frac{1}{1-\varepsilon} = 1 + 1\varepsilon + 1\varepsilon^2 + 1\varepsilon^3 + 1\varepsilon^4 + \dots$$

Numerical Implementation

- "Teach" the computer to operate with truncated polynomials
- Compute Taylor expansion of any deterministic function

Julia implementation:
https://igit.ific.uv.es/alramos/formalseries.jl

Code (Julia)

```
julia> using FormalSeries
julia> f(x) = 1/(1-x)
julia> xs = Series((0.0,1.0,0.0,0.0,0.0))
julia> f(xs)
Series{Float64, 5}((1.0, 1.0, 1.0, 1.0, 1.0))
```

- Formally equivalent to:

$$f(0+\varepsilon) = \frac{1}{1-\varepsilon} = 1 + 1\varepsilon + 1\varepsilon^2 + 1\varepsilon^3 + 1\varepsilon^4 + \dots$$

- Multiple expansions $\varepsilon \longrightarrow \varepsilon_i$
- Basis of Forward Automatic Differentiation

Metropolis & Truncated Polynomials

- Metropolis as basis for sampling algorithms
- Action $S(\theta)$: θ denotes some couplings

Metropolis & Truncated Polynomials

- Metropolis as basis for sampling algorithms
- Action $S(\theta)$: θ denotes some couplings
- Sampling proposal (e.g., HMC)

$$S(\theta) \longrightarrow S'(\theta)$$

Metropolis & Truncated Polynomials

- Metropolis as basis for sampling algorithms
- Action $S(\theta)$: θ denotes some couplings
- Sampling proposal (e.g., HMC)

$$S(\theta) \longrightarrow S'(\theta)$$

- Accept using

$$\min\left\{1, e^{-\Delta S(\tilde{\theta})}\right\}$$

Metropolis & Truncated Polynomials

- Metropolis as basis for sampling algorithms
- Action $S(\theta)$: θ denotes some couplings
- Sampling proposal (e.g., HMC)

$$S(\theta) \longrightarrow S'(\theta)$$

- Accept using

$$\min\left\{1, e^{-\Delta S(\tilde{\theta})}\right\}$$

- Not possible if $\tilde{\theta}$ is a series element

Metropolis & Truncated Polynomials

- Metropolis as basis for sampling algorithms
- Action $S(\theta)$: θ denotes some couplings
- Sampling proposal (e.g., HMC)

$$S(\theta) \longrightarrow S'(\theta)$$

- Accept using

$$\min\left\{1, e^{-\Delta S(\tilde{\theta})}\right\}$$

- Not possible if $\tilde{\theta}$ is a series element

Automatic Differentiation ruled out due to Stochastic elements

How to overcome this?

Reweighting & Automatic Differentiation

[G.C., A. Ramos, B. Zaldivar, 2024]

Samples

$$\{x^\alpha\}_{\alpha=1}^{N} \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta')}$$

Reweighting & Automatic Differentiation

[G.C., A. Ramos, B. Zaldivar, 2024]

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta')}$$

- Conventional Reweighting

$$\langle f(x) \rangle_{S'} = \frac{\left\langle e^{S-S'} f(x) \right\rangle_S}{\langle e^{S-S'} \rangle_S}$$

Reweighting & Automatic Differentiation

[G.C., A. Ramos, B. Zaldivar, 2024]

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta')}$$

- Conventional Reweighting

$$\langle f(x) \rangle_{S'} = \frac{\left\langle e^{S-S'} f(x) \right\rangle_S}{\langle e^{S-S'} \rangle_S}$$

- Introduce truncated polynomials with $\tilde{\theta} = \theta + \varepsilon$

$$w^\alpha = e^{S(x,\theta)-S(x,\tilde{\theta})} = 1 + (\ \dots\ )\varepsilon + (\ \dots\ )\varepsilon^2 + \dots$$

Reweighting & Automatic Differentiation

[G.C., A. Ramos, B. Zaldivar, 2024]

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta')}$$

- Conventional Reweighting

$$\langle f(x)\rangle_{S'} = \frac{\left\langle e^{S-S'} f(x) \right\rangle_S}{\langle e^{S-S'}\rangle_S}$$

- Introduce truncated polynomials with $\tilde{\theta} = \theta + \varepsilon$

$$w^\alpha = e^{S(x,\theta)-S(x,\tilde{\theta})} = 1 + (\ \dots\ )\varepsilon + (\ \dots\ )\varepsilon^2 + \dots$$

- Expectation value an Taylor series coefficients

$$\frac{\sum_\alpha e^{S(x^\alpha,\theta)-S(x^\alpha,\tilde{\theta})} f(x)}{\sum_\alpha e^{S(x^\alpha,\theta)-S(x^\alpha,\tilde{\theta})}} = \sum_{n=0}^{K} f_n \varepsilon^n, \qquad f_n = \frac{1}{n!}\frac{\partial^n}{\partial\theta^n}\langle f(x)\rangle\Big|_\theta$$

Conventional 2-point functions & Reweighting

Samples

$$\{x^\alpha\}_{\alpha=1}^{N} \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta)-JO(0)}$$

Conventional 2-point functions & Reweighting

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta)-JO(0)}$$

- Expansion in $J = 0 + \varepsilon$

$$w^\alpha = e^{S(x,\theta)-S(x,\tilde{\theta})+JO(0)} = 1 + \varepsilon O(0) + \varepsilon^2 O(0)^2/2 + \dots$$

Conventional 2-point functions & Reweighting

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta)-JO(0)}$$

- Expansion in $J = 0 + \varepsilon$

$$w^\alpha = e^{S(x,\theta)-S(x,\tilde{\theta})+JO(0)} = 1 + \varepsilon O(0) + \varepsilon^2 O(0)^2/2 + \dots$$

- Reweighting to compute d/dJ

$$\langle O(x_0)\rangle_{S'} = \frac{\langle w^\alpha O(x_0)\rangle_S}{\langle w^\alpha\rangle_S} = \langle O(x_0)\rangle + \varepsilon\langle O(x_0)O(0)\rangle_c + \mathcal{O}\left(\varepsilon^2\right)$$

  - Conventional computation of a 2-point function can be seen as reweighting

Conventional 2-point functions & Reweighting

Samples

$$\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$$

Expectation values w.r.t

$$e^{-S(x;\theta)-JO(0)}$$

- Expansion in $J = 0 + \varepsilon$

$$w^\alpha = e^{S(x,\theta)-S(x,\tilde{\theta})+JO(0)} = 1 + \varepsilon O(0) + \varepsilon^2 O(0)^2/2 + \dots$$

- Reweighting to compute d/dJ

$$\langle O(x_0)\rangle_{S'} = \frac{\langle w^\alpha O(x_0)\rangle_S}{\langle w^\alpha\rangle_S} = \langle O(x_0)\rangle + \varepsilon\langle O(x_0)O(0)\rangle_c + \mathcal{O}\left(\varepsilon^2\right)$$

  - Conventional computation of a 2-point function can be seen as reweighting
  - Signal to Noise can be seen as a reweighting variance problem

HMC & Automatic Differentiation

1. Fictitious momenta π conjugate to ϕ

$$H(\phi, \pi) = \frac{1}{2}\pi^2 + S(\phi; \theta)$$

2. Solve EoM with initial random momenta $\pi(t=0) \sim N(0,1)$

$$\dot{\phi} = \frac{\partial H}{\partial \pi} = \pi, \qquad \dot{\pi} = -\frac{\partial H}{\partial \phi}$$

$$(\phi(0), \pi(0)) \longrightarrow (\phi(t), \pi(t))$$

3. Metropolis: Acc./Rej. with probability $e^{-\Delta H}$
4. Repeat

HMC & Automatic Differentiation

1. Fictitious momenta π conjugate to ϕ

$$H(\phi, \pi) = \frac{1}{2}\pi^2 + S(\phi; \theta)$$

2. Solve EoM with initial random momenta $\pi(t=0) \sim N(0,1)$

$$\dot{\phi} = \frac{\partial H}{\partial \pi} = \pi, \qquad \dot{\pi} = -\frac{\partial H}{\partial \phi}$$

$$(\phi(0), \pi(0)) \longrightarrow (\phi(t), \pi(t))$$

3. Metropolis: Acc./Rej. with probability $e^{-\Delta H}$
4. Repeat

Promote $\theta \longrightarrow \tilde{\theta} = \theta + \varepsilon$
(also $\pi,\ \phi$)

HMC & Automatic Differentiation

1. Fictitious momenta π conjugate to ϕ

$$H(\phi, \pi) = \frac{1}{2}\pi^2 + S(\phi; \theta)$$

2. Solve EoM with initial random momenta $\pi(t=0) \sim N(0,1)$

$$\dot{\phi} = \frac{\partial H}{\partial \pi} = \pi, \qquad \dot{\pi} = -\frac{\partial H}{\partial \phi}$$

$$(\phi(0), \pi(0)) \longrightarrow (\phi(t), \pi(t))$$

3. Metropolis: Acc./Rej. with probability $e^{-\Delta H}$
4. Repeat

Promote $\theta \longrightarrow \tilde{\theta} = \theta + \varepsilon$ (also $\pi,\ \phi$)

Solve EoM at each order

$\phi \longrightarrow \tilde{\phi}$

HMC & Automatic Differentiation

1. Fictitious momenta π conjugate to ϕ

$$H(\phi, \pi) = \frac{1}{2}\pi^2 + S(\phi; \theta)$$

2. Solve EoM with initial random momenta $\pi(t=0) \sim N(0,1)$

$$\dot{\phi} = \frac{\partial H}{\partial \pi} = \pi, \qquad \dot{\pi} = -\frac{\partial H}{\partial \phi}$$

$$(\phi(0), \pi(0)) \longrightarrow (\phi(t), \pi(t))$$

3. Metropolis: Acc./Rej. with probability $e^{-\Delta H}$
4. Repeat

Promote $\theta \longrightarrow \tilde{\theta} = \theta + \varepsilon$ (also π, ϕ)

Solve EoM at each order

$\phi \longrightarrow \tilde{\phi}$

ΔH as Taylor series
Cannot Acc./Rej.

HMC & Automatic Differentiation

1. Fictitious momenta π conjugate to ϕ

$$H(\phi, \pi) = \frac{1}{2}\pi^2 + S(\phi; \theta)$$

2. Solve EoM with initial random momenta $\pi(t = 0) \sim N(0, 1)$

$$\dot{\phi} = \frac{\partial H}{\partial \pi} = \pi, \qquad \dot{\pi} = -\frac{\partial H}{\partial \phi}$$

$$(\phi(0), \pi(0)) \longrightarrow (\phi(t), \pi(t))$$

3. Metropolis: Acc./Rej. with probability $e^{-\Delta H}$
4. Repeat

Promote $\theta \longrightarrow \tilde{\theta} = \theta + \varepsilon$ (also π, ϕ)

Solve EoM at each order

$\phi \longrightarrow \tilde{\phi}$

ΔH as Taylor series
Cannot Acc./Rej.

MC average as Taylor series

$$\langle f(\phi) \rangle = \frac{1}{N} \sum_{\alpha} f(\tilde{\phi}) = \sum_{n=0}^{K} f_n \varepsilon^n$$

HMC & Automatic Differentiation

- HMC samples are truncated polynomials

$$\tilde{\phi} = \phi + \varepsilon\phi_{(1)}$$
$$\{\tilde{\phi}^\alpha\}_{\alpha=1}^{N_{\text{conf.}}} = \{\phi^\alpha + \varepsilon\phi_{(1)}^\alpha\}_{\alpha=1}^{N_{\text{conf.}}} \sim e^{-S(\phi,\tilde{\theta})}$$

  - Samples carry information about the dependence on θ

HMC & Automatic Differentiation

- HMC samples are truncated polynomials

$$\tilde{\phi} = \phi + \varepsilon\phi_{(1)}$$

$$\{\tilde{\phi}^{\alpha}\}_{\alpha=1}^{N_{\text{conf.}}} = \{\phi^{\alpha} + \varepsilon\phi_{(1)}^{\alpha}\}_{\alpha=1}^{N_{\text{conf.}}} \sim e^{-S(\phi,\tilde{\theta})}$$

  - Samples carry information about the dependence on θ

- Usual MC averages

$$\left\langle \tilde{O}(\phi) \right\rangle = \frac{1}{N_{\text{conf}}} \sum_{\alpha} O(\tilde{\phi}^{\alpha})$$
$$= \langle O(\phi) \rangle + \frac{\mathrm{d}}{\mathrm{d}\theta} \langle O(\phi) \rangle \varepsilon$$

  - No RW factors – no disconnected contributions

HMC & Automatic Differentiation

- HMC samples are truncated polynomials

$$\tilde{\phi} = \phi + \varepsilon\phi_{(1)}$$
$$\{\tilde{\phi}^{\alpha}\}_{\alpha=1}^{N_{\text{conf.}}} = \{\phi^{\alpha} + \varepsilon\phi_{(1)}^{\alpha}\}_{\alpha=1}^{N_{\text{conf.}}} \sim e^{-S(\phi,\tilde{\theta})}$$

  - Samples carry information about the dependence on θ

- Usual MC averages

$$\left\langle \tilde{O}(\phi) \right\rangle = \frac{1}{N_{\text{conf}}} \sum_{\alpha} O(\tilde{\phi}^{\alpha})$$
$$= \langle O(\phi) \rangle + \frac{\mathrm{d}}{\mathrm{d}\theta} \langle O(\phi) \rangle \varepsilon$$

  - No RW factors – no disconnected contributions

- Hamiltonian AD finds the exact transformation that leads to constant RW factors [G.C., A. Ramos, B. Zaldivar, 2024; G.C. 2025 (Thesis)]

Field transformations & Reweighting

- Reweighting generally leads to large variance
  - weights w^α

Field transformations & Reweighting

- Reweighting generally leads to large variance
  - weights w^α
- Consider a change of variables before reweighting

$$\phi \longrightarrow \tilde{\phi} = \phi + \varepsilon f(\phi)$$

Field transformations & Reweighting

- Reweighting generally leads to large variance
  - weights w^α
- Consider a change of variables before reweighting

$$\phi \longrightarrow \tilde{\phi} = \phi + \varepsilon f(\phi)$$

- Reweighting factors become

$$\tilde{w}[\tilde{\phi}] = \exp\left[-\tilde{S}_J[\tilde{\phi}] + \log\left|\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}\phi}\right| + S[\phi]\right]$$

Field transformations & Reweighting

- Reweighting generally leads to large variance
  - weights w^α
- Consider a change of variables before reweighting

$$\phi \longrightarrow \tilde{\phi} = \phi + \varepsilon f(\phi)$$

- Reweighting factors become

$$\tilde{w}[\tilde{\phi}] = \exp\left[-\tilde{S}_J[\tilde{\phi}] + \log\left|\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}\phi}\right| + S[\phi]\right]$$

- Find transformation that eliminates RW factors
  - Exact – RW factors drop from the computation!

$$-\tilde{S}_J[\tilde{\phi}] + \log\left|\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}\phi}\right| + S[\phi] = \text{const.}$$

  - Approximate – partially eliminate RW factors

4D Scalar field theory

$$S^{4\mathrm{D}}_{\mathrm{latt}}(\phi; m, \lambda) = \sum_x \left\{ \frac{1}{2} \sum_\mu [\phi(x+\mu) - \phi(x)]^2 + \frac{m^2}{2}\phi^2(x) + \lambda\phi^4(x) \right\}$$

$$\text{Operator:} \quad O(x_0) = \sum_{\vec{x}} \phi(x_0, \vec{x})$$

- Good testbed – No excited state contamination

4D Scalar field theory

$$S^{4D}_{\text{latt}}(\phi; m, \lambda) = \sum_x \left\{ \frac{1}{2} \sum_\mu [\phi(x+\mu) - \phi(x)]^2 + \frac{m^2}{2}\phi^2(x) + \lambda\phi^4(x) \right\}$$

$$\text{Operator:} \quad O(x_0) = \sum_{\vec{x}} \phi(x_0, \vec{x})$$

- Good testbed – No excited state contamination

- HAD (HMC + AD):

$$S(\tilde{\phi}) = S_{\text{latt}}(\tilde{\phi}) + J O(0)$$
$$J = 0 + \varepsilon$$

4D Scalar field theory

$$S^{4D}_{\text{latt}}(\phi; m, \lambda) = \sum_x \left\{ \frac{1}{2} \sum_\mu [\phi(x+\mu) - \phi(x)]^2 + \frac{m^2}{2}\phi^2(x) + \lambda\phi^4(x) \right\}$$

Operator: $O(x_0) = \sum_{\vec{x}} \phi(x_0, \vec{x})$

- Good testbed – No excited state contamination

- HAD (HMC + AD):

$$S(\tilde{\phi}) = S_{\text{latt}}(\tilde{\phi}) + J O(0)$$
$$J = 0 + \varepsilon$$

$$\dot{\tilde{\phi}}(x) = \tilde{\pi}(x)\,,$$
$$\dot{\tilde{\pi}}(x) = \frac{1}{2}\sum_\mu \left[\tilde{\phi}(x+\mu) + \tilde{\phi}(x-\mu)\right] - (4 + \hat{m}^2)\tilde{\phi}(x) - 4\lambda\tilde{\phi}^3(x) + \epsilon\delta_{x_0,0}$$

HAD solution for the Signal to Noise problem

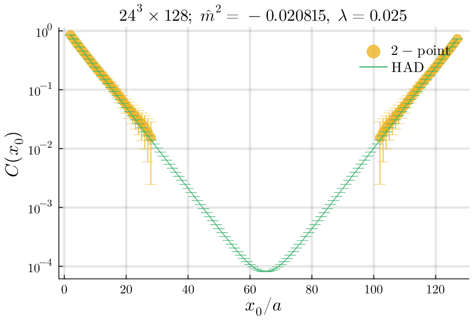
$$C(x_0)^{2-\text{point}} = \langle O(x_0) O(0) \rangle$$

$$C(x_0)^{\text{HAD}} = \left\langle \tilde{O}(x_0) \right\rangle_{(1)}$$

HAD solution for the Signal to Noise problem

$$C(x_0)^{\text{2-point}} = \langle O(x_0) O(0) \rangle$$

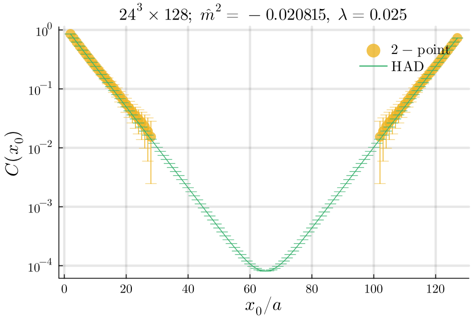
$$C(x_0)^{\text{HAD}} = \left\langle \tilde{O}(x_0) \right\rangle_{(1)}$$



HAD solution for the Signal to Noise problem

$$C(x_0)^{2-\text{point}} = \langle O(x_0) O(0) \rangle$$
$$C(x_0)^{\text{HAD}} = \left\langle \tilde{O}(x_0) \right\rangle_{(1)}$$

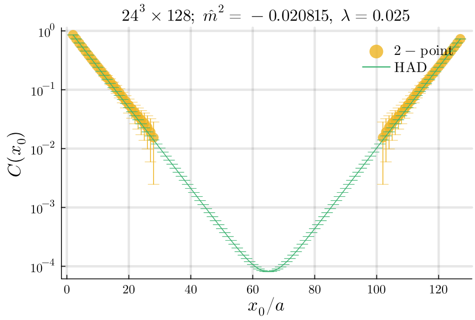
$$\hat{m}_{\text{eff}}(x_0) = -\log \frac{C(x_0 + a)}{C(x_0)} \,.$$

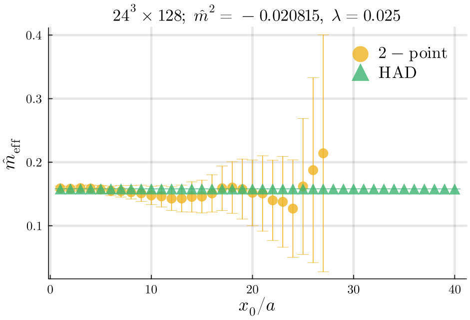


HAD solution for the Signal to Noise problem

$$C(x_0)^{2-\text{point}} = \langle O(x_0)O(0)\rangle$$
$$C(x_0)^{\text{HAD}} = \left\langle \tilde{O}(x_0)\right\rangle_{(1)}$$

$$\hat{m}_{\text{eff}}(x_0) = -\log\frac{C(x_0+a)}{C(x_0)}\,.$$

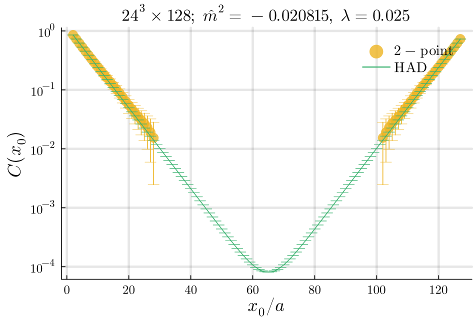


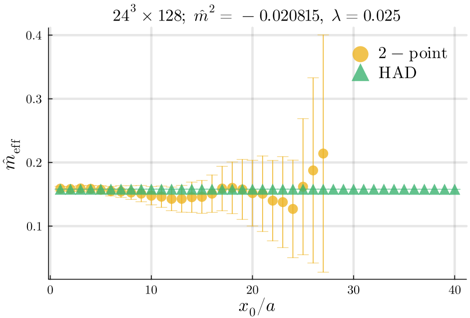


HAD solution for the Signal to Noise problem

$$C(x_0)^{\text{2-point}} = \langle O(x_0) O(0) \rangle$$
$$C(x_0)^{\text{HAD}} = \left\langle \tilde{O}(x_0) \right\rangle_{(1)}$$

$$\hat{m}_{\text{eff}}(x_0) = -\log \frac{C(x_0 + a)}{C(x_0)} \, .$$





- HAD solves StN problem completely
  - convergence
  - compact variables, complex interpolators

Transformed Reweighting (TRW)

- Transformation?

$$S_{\text{latt}}(\phi_p; m) = \sum_p \phi_p^* \left[\sum_\mu \hat{p}_\mu^2 + m^2 \right] \phi_p + \mathcal{O}(\lambda), \qquad \hat{p} = 2\sin(ap/2)$$

Transformed Reweighting (TRW)

- Transformation?

$$S_{\text{latt}}(\phi_p; m) = \sum_p \phi_p^* \left[\sum_\mu \hat{p}_\mu^2 + m^2 \right] \phi_p + \mathcal{O}(\lambda), \qquad \hat{p} = 2\sin(ap/2)$$

- Free Theory case – exact transformation

$$\sum_{\vec{x}} \phi(0, \vec{x}), \qquad \phi_p \longrightarrow \tilde{\phi}_p = \phi_p + \frac{1}{2} \frac{\delta(\vec{p})}{\hat{p}^2 + m_{\text{transf}}^2} \varepsilon$$

  - $\hat{m}_{\text{transf.}} = \hat{m}$
  - Transforms $S^{\text{free}}(\phi)$ into $\tilde{S}^{\text{free}}_{\tilde{J}}(\tilde{\phi})$
  - Constant reweighting factors

Transformed Reweighting (TRW)

- Transformation?

$$S_{\text{latt}}(\phi_p; m) = \sum_p \phi_p^* \left[\sum_\mu \hat{p}_\mu^2 + m^2 \right] \phi_p + \mathcal{O}(\lambda), \qquad \hat{p} = 2\sin(ap/2)$$

- Free Theory case – exact transformation

$$\sum_{\vec{x}} \phi(0, \vec{x}), \qquad \phi_p \longrightarrow \tilde{\phi}_p = \phi_p + \frac{1}{2} \frac{\delta(\vec{p})}{\hat{p}^2 + m_{\text{transf}}^2} \varepsilon$$

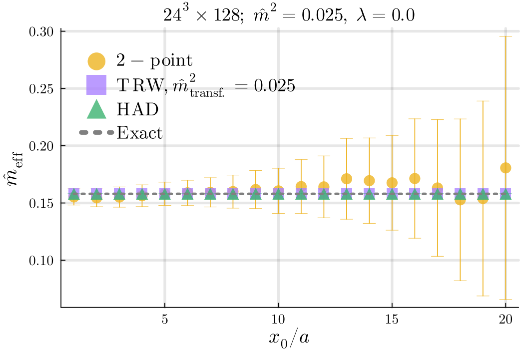
  - $\hat{m}_{\text{transf.}} = \hat{m}$
  - Transforms $S^{\text{free}}(\phi)$ into $\tilde{S}_{\tilde{J}}^{\text{free}}(\tilde{\phi})$
  - Constant reweighting factors
- $\lambda \neq 0$: use the same (approximate) transformation
  - $\hat{m}_{\text{transf.}} = \hat{m}_R$

Exact & approximate transformations

- Conventional samples tuned to have $(am_R)^2 \sim 0.025$

Exact & approximate transformations

- Conventional samples tuned to have $(am_R)^2 \sim 0.025$

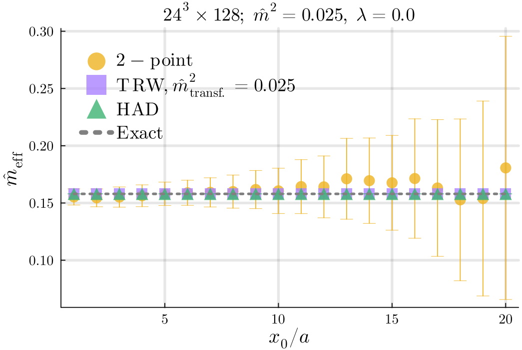


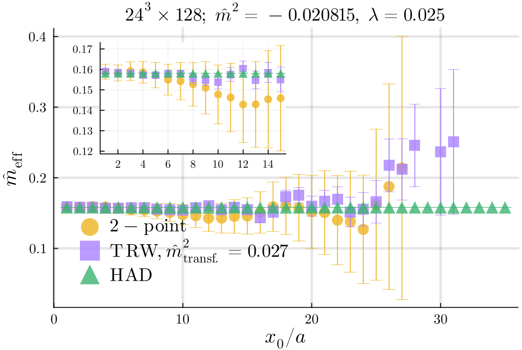
- Exact for $\lambda = 0$
- HAD = TRW
  - Confirms reparametrization argument

Exact & approximate transformations

- Conventional samples tuned to have $(am_R)^2 \sim 0.025$

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$



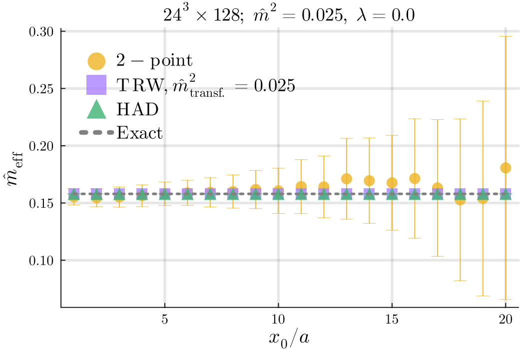


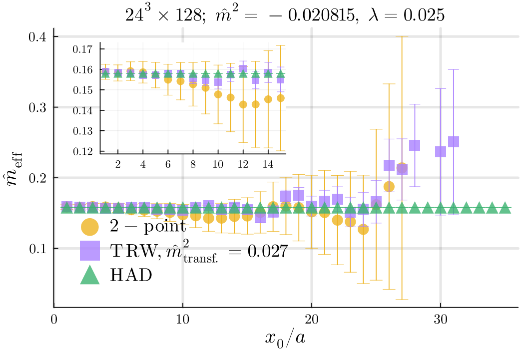
- Exact for $\lambda = 0$
- HAD = TRW
  - Confirms reparametrization argument

Exact & approximate transformations

- Conventional samples tuned to have $(am_R)^2 \sim 0.025$

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$





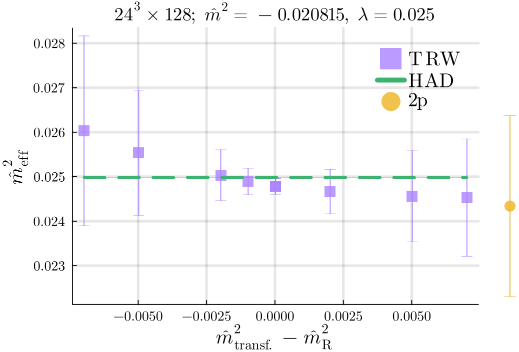
- Exact for $\lambda = 0$
- HAD = TRW
  - Confirms reparametrization argument

- Approximate for $\lambda \neq 0$
- Still provides an improvement
  - But how do we choose $\hat{m}_{\text{transf.}}$?

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m_{\text{transf}}^2}\varepsilon$$

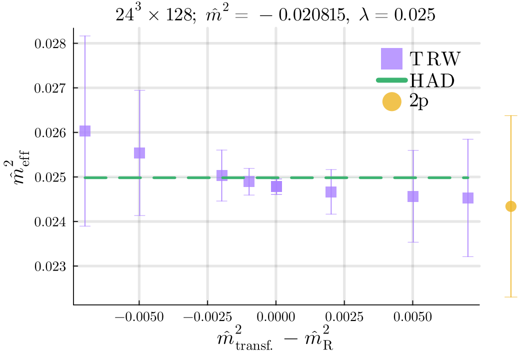


- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$

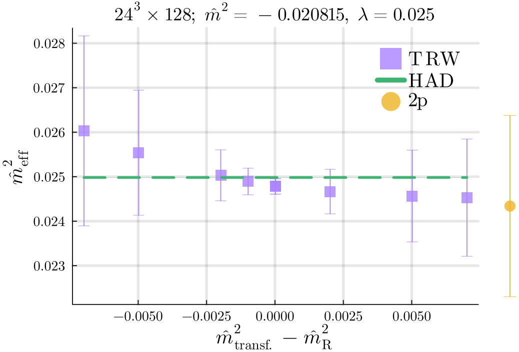


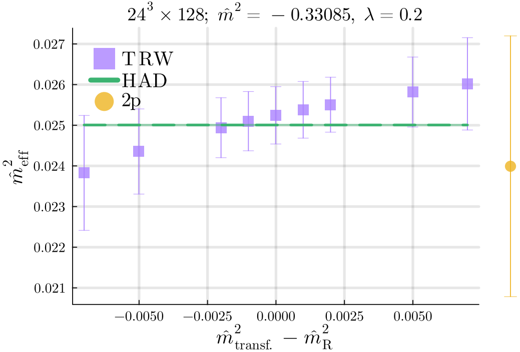
- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$
- Nonetheless, large range of $\hat{m}_{\text{trans.}}$

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$



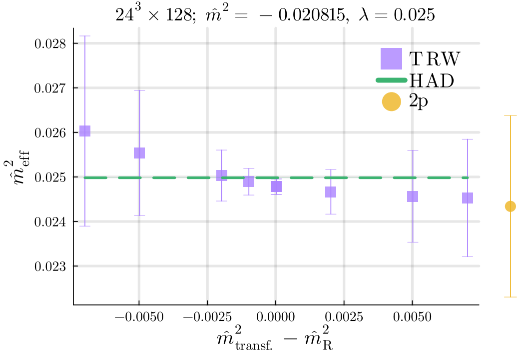


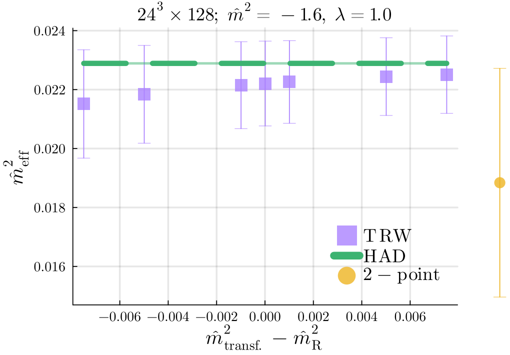
- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$
- Nonetheless, large range of $\hat{m}_{\text{trans.}}$

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$



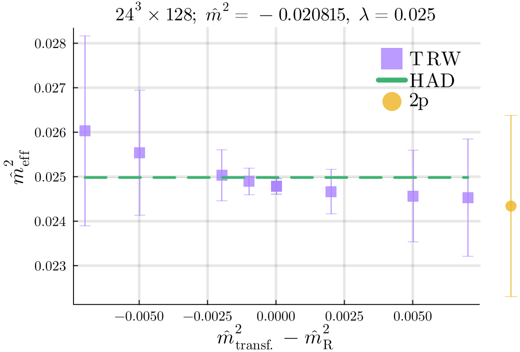


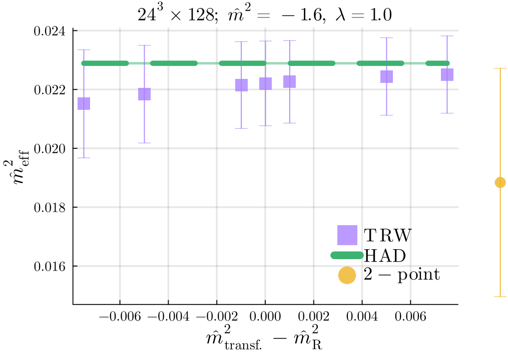
- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$
- Nonetheless, large range of $\hat{m}_{\text{trans.}}$

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2} \frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}} \varepsilon$$



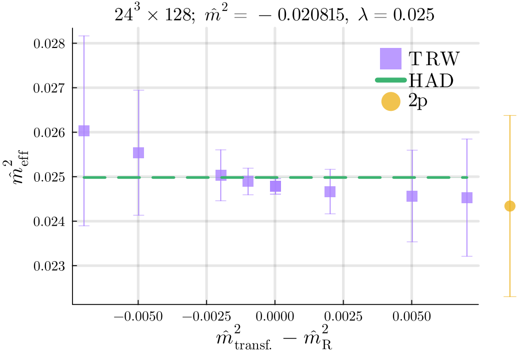


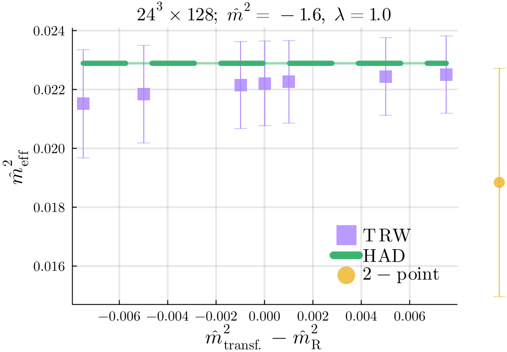
- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$
- Nonetheless, large range of $\hat{m}_{\text{trans.}}$
- Precision degrades with λ

Interacting case – transformation dependence

- Scan over $\hat{m}_{\text{trans.}}$ (cheap)

$$\phi_p \to \tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + m^2_{\text{transf}}}\varepsilon$$





- Optimal $\hat{m}_{\text{trans.}} = \hat{m}_R$
- Nonetheless, large range of $\hat{m}_{\text{trans.}}$

- Precision degrades with λ
- $\hat{m}_{\text{trans.}} \neq \hat{m}_R$ still provides an improvement w.r.t. 2-point

Conclusions

- StN is an important problem for lattice field theory
- A change of perspective of the StN is possible:
  - The usual 2-point computation is a reweighting computation of a derivative
  - the StN problem stems from the variance of the RW factors
  - Correlator as derivative of a 1-point function – variance depends on how it is computed (RW is just one possible way)
- Key ingredient: Stochastic AD
  - Computes derivatives exactly
  - Single simulation, no need for finite-differences
- Hamiltonian method solves the StN exactly
  - not generally applicable (convergence, complex interpolators, etc.)
- Field transformations can reduce (or eliminate) the variance – solves the StN problem
  - Re-utilize samples – TRW is cheap
  - General (no problems with convergence, etc)
  - Normalizing flows/Trivializing maps – attack the StN and not the sampling

Comparison & Change of variables

Reweighting

- $\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$
- Weights $w^\alpha = e^{S(x^\alpha;m,\lambda)-S(x^\alpha;\tilde{m},\tilde{\lambda})}$ take into account dependence on parameters θ

HMC

- $\{\tilde{x}^\alpha\}_{\alpha=1}^N \sim e^{-S(\tilde{x};\tilde{\theta})}$
- Samples carry dependence on the parameters $\tilde{\theta}$

How are these methods related?

Comparison & Change of variables

Reweighting

- $\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$
- Weights $w^\alpha = e^{S(x^\alpha;m,\lambda)-S(x^\alpha;\tilde{m},\tilde{\lambda})}$ take into account dependence on parameters θ

HMC

- $\{\tilde{x}^\alpha\}_{\alpha=1}^N \sim e^{-S(\tilde{x};\tilde{\theta})}$
- Samples carry dependence on the parameters $\tilde{\theta}$

How are these methods related?

Toy model: $$p_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}, \qquad \{x^\alpha\} \sim p_{\sigma^*=1},$$

Comparison & Change of variables

Reweighting

- $\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$
- Weights $w^\alpha = e^{S(x^\alpha;m,\lambda)-S(x^\alpha;\tilde{m},\tilde{\lambda})}$ take into account dependence on parameters θ

HMC

- $\{\tilde{x}^\alpha\}_{\alpha=1}^N \sim e^{-S(\tilde{x};\tilde{\theta})}$
- Samples carry dependence on the parameters $\tilde{\theta}$

How are these methods related?

Toy model: $p_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$, $\{x^\alpha\} \sim p_{\sigma^*=1}$,

$\sigma^* = 1 \longrightarrow \sigma$ $w^\alpha(x) = x-\text{dependent}$

Comparison & Change of variables

Reweighting

- $\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$
- Weights $w^\alpha = e^{S(x^\alpha;m,\lambda)-S(x^\alpha;\tilde{m},\tilde{\lambda})}$ take into account dependence on parameters θ

HMC

- $\{\tilde{x}^\alpha\}_{\alpha=1}^N \sim e^{-S(\tilde{x};\tilde{\theta})}$
- Samples carry dependence on the parameters $\tilde{\theta}$

How are these methods related?

Toy model: $p_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$, $\{x^\alpha\} \sim p_{\sigma^*=1}$,

$\sigma^* = 1 \longrightarrow \sigma$ $w^\alpha(x) = x - \text{dependent}$

Transformation: $y^\alpha = \sigma x^\alpha$, $\{y^\alpha\} \sim p_\sigma$

Comparison & Change of variables

Reweighting

- $\{x^\alpha\}_{\alpha=1}^N \sim e^{-S(x;\theta)}$
- Weights $w^\alpha = e^{S(x^\alpha;m,\lambda)-S(x^\alpha;\tilde{m},\tilde{\lambda})}$ take into account dependence on parameters θ

HMC

- $\{\tilde{x}^\alpha\}_{\alpha=1}^N \sim e^{-S(\tilde{x};\tilde{\theta})}$
- Samples carry dependence on the parameters $\tilde{\theta}$

How are these methods related?

Toy model:	$p_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$,	$\{x^\alpha\} \sim p_{\sigma^*=1}$,
$\sigma^* = 1 \longrightarrow \sigma$	$w^\alpha(x) = x -$ dependent	
Transformation:	$y^\alpha = \sigma x^\alpha$,	$\{y^\alpha\} \sim p_\sigma$
$\sigma^* = 1 \longrightarrow \sigma$	$w^\alpha(y) = y -$ independent	'No Reweighting'

Comparison & Change of variables

$$y^\alpha = \sigma x^\alpha$$

$$\varepsilon = \sigma - 1$$

Comparison & Change of variables

$$y^\alpha = \sigma x^\alpha \qquad \tilde{y}^\alpha = x^\alpha + x^\alpha \varepsilon \qquad \varepsilon = \sigma - 1$$

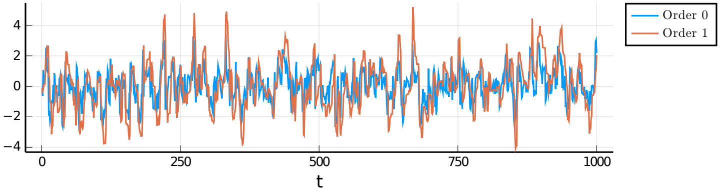
Comparison & Change of variables

$$y^\alpha = \sigma x^\alpha \qquad \tilde{y}^\alpha = x^\alpha + x^\alpha \varepsilon \qquad \varepsilon = \sigma - 1$$

What about HMC?
Equations of motion

$$\ddot{x}_0 = -\frac{x_0}{\sigma^2}, \qquad \langle x_0 \rangle$$

$$\ddot{x}_1 = -\frac{x_1}{\sigma^2} + 2\frac{x_0}{\sigma^3}, \qquad \langle x_1 \rangle = \langle x_0 \rangle$$



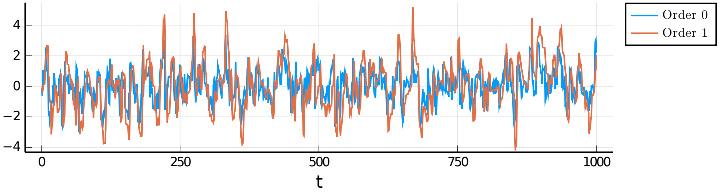
Comparison & Change of variables

$$y^\alpha = \sigma x^\alpha \qquad \tilde{y}^\alpha = x^\alpha + x^\alpha \varepsilon \qquad \varepsilon = \sigma - 1$$

What about HMC?
Equations of motion

$$\ddot{x}_0 = -\frac{x_0}{\sigma^2}, \qquad \langle x_0 \rangle$$

$$\ddot{x}_1 = -\frac{x_1}{\sigma^2} + 2\frac{x_0}{\sigma^3}, \qquad \langle x_1 \rangle = \langle x_0 \rangle$$



- Hamiltonian method finds the change of variables $x \to \tilde{y}$ that lead to constant reweighting factors