Stein-optimal transport for the Signal-to-Noise problem

Pietro Butti

with Guilherme Catumba, Alessandro Nada, Louis Spatscheck

Nordic Lattice Meeting
Edinburgh, May 18th, 2026

Introduction:

Signal-to-noise degradation
Source reweighting
Improved reweigthing

Reformulation:

Map in generative modelling
Expanding the KL divergence

Solution:

Neural-network KL minimisation
Stein-Poisson equation and FK formula
Some preliminary results...

Introduction

Signal-to-noise

Parisi-Lepage: Signal and noise decays as states with different mass gaps

$$C(t) = \langle \mathcal{O}(t)\mathcal{O}^\dagger(0)\rangle \xrightarrow{t \gg (\Delta E)^{-1}} Z e^{-M_{\mathrm{signal}} t}$$

Mass of the lightest states of $\mathcal{O}$

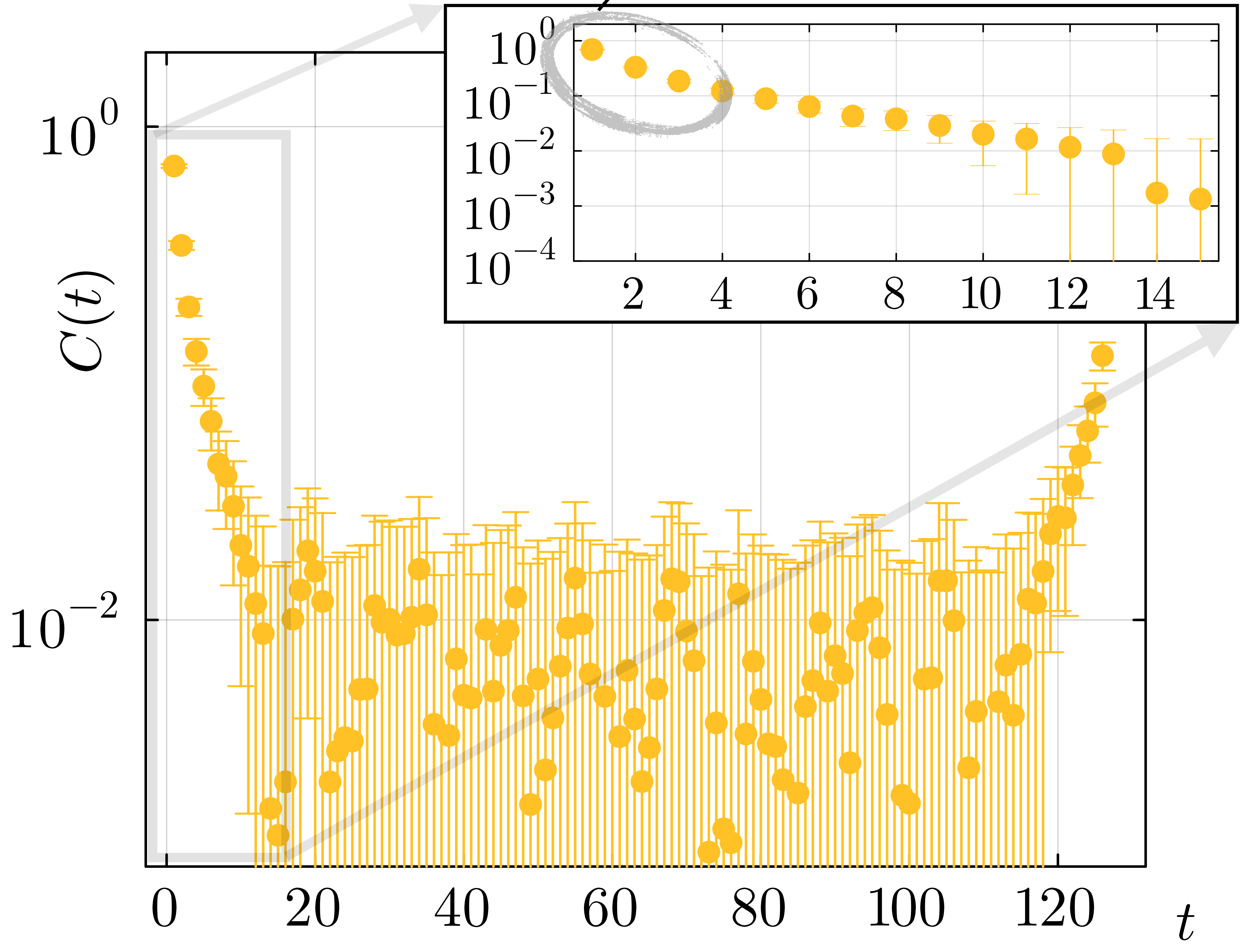
$$\text{noise} \sim \langle |\mathcal{O}(t)\mathcal{O}^\dagger(0)|^2\rangle - |C(t)|^2 \xrightarrow{t \gg (\Delta E)^{-1}} Z e^{-M_{\mathrm{noise}} t}$$

Mass of the lightest state of $\mathcal{O}\mathcal{O}^\dagger$

$$\frac{\mathrm{error}[C(t)]}{C(t)} \xrightarrow{t \gg (\Delta E)^{-1}} \frac{e^{t\Delta}}{\sqrt{N_{\mathrm{conf}}}}$$

Signal dominated by excited states

$Z_0 e^{-E_0 t} + Z_1 e^{-\Delta E t} + \cdots$



A new approach

[Catumba, Ramos] - Eur.Phys.J.C 85 (2025)

Add a source to the action then take derivatives wrt J

$$S_J = S - J\mathcal{O}_0$$

$$\langle \mathcal{O}_t \rangle_{S_J} = \frac{\langle e^{J\mathcal{O}_0}\mathcal{O}_t \rangle_S}{\langle e^{J\mathcal{O}_0} \rangle_S}$$

$$\langle \mathcal{O}_t \mathcal{O}_0 \rangle_S = \frac{\partial}{\partial J}\Big|_{J=0} \langle \mathcal{O}_t \rangle_{S_J}$$

2-point function $C(t)$

$$\mathcal{O}_t[\phi] = \sum_{\vec{x}} \phi(x)|_{x_0=t}$$

Write $J = \varepsilon$, then expand the reweighting formula to $\mathcal{O}(\varepsilon)$

...or use Automatic Differentiation!

Correlators via reweigthing

$\phi \sim e^{-S}$

Configurations

$e^{-(S - J\mathcal{O}_0)}$

Target distribution

Importance weights (normalized):

$$w[\phi] = e^{S[\phi] - (S[\phi] - J\mathcal{O}_0) + \log \frac{Z}{Z_J}}$$

Reweight to compute $\langle\, . \,\rangle_{S_J}$

$$\langle \mathcal{O}_t \rangle_{S_J} = \langle w \mathcal{O}_t \rangle_S$$

Expand in $J = \varepsilon$ to compute d/dJ

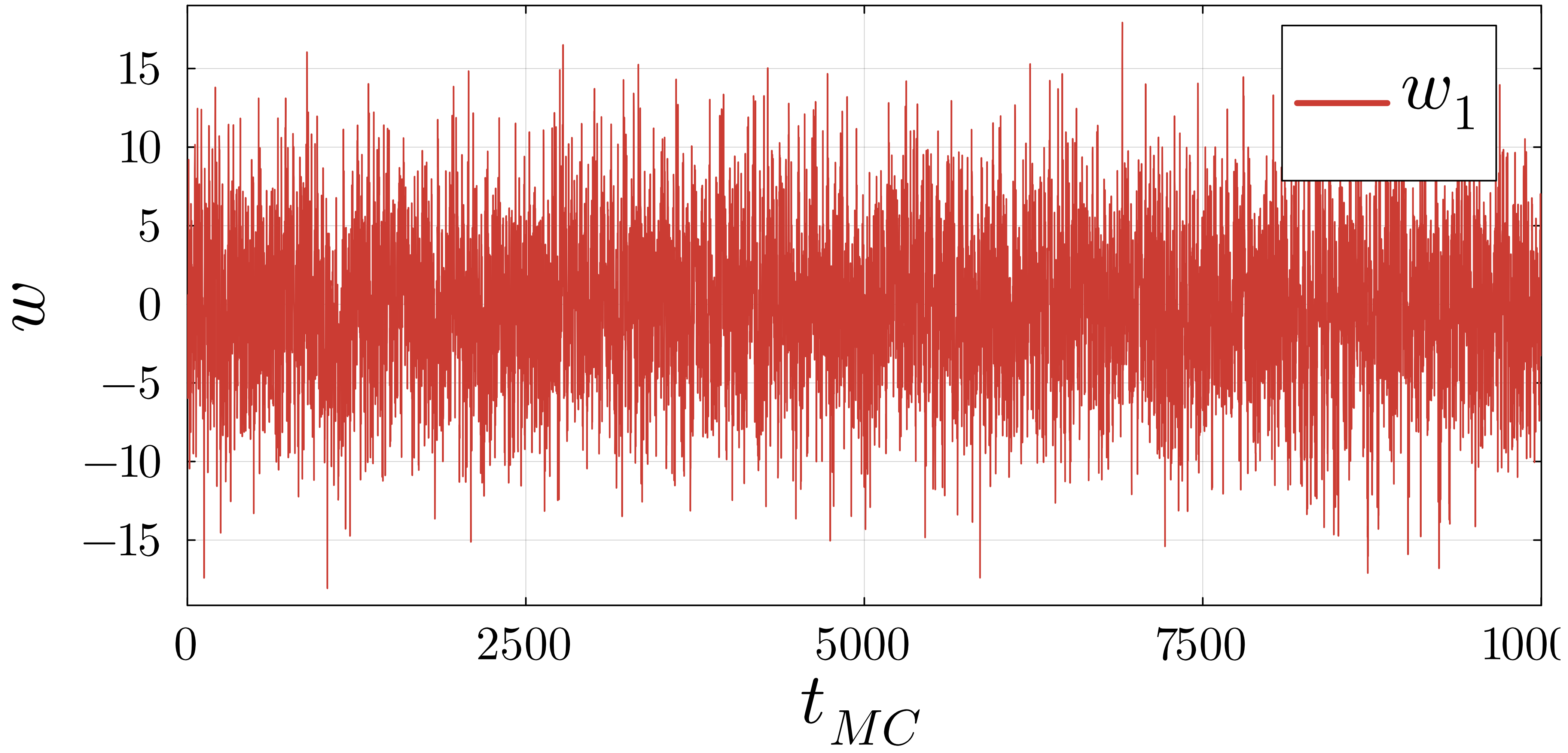
$$= 1 + \varepsilon(\mathcal{O}_0 - \langle \mathcal{O}_0 \rangle) + \ldots$$

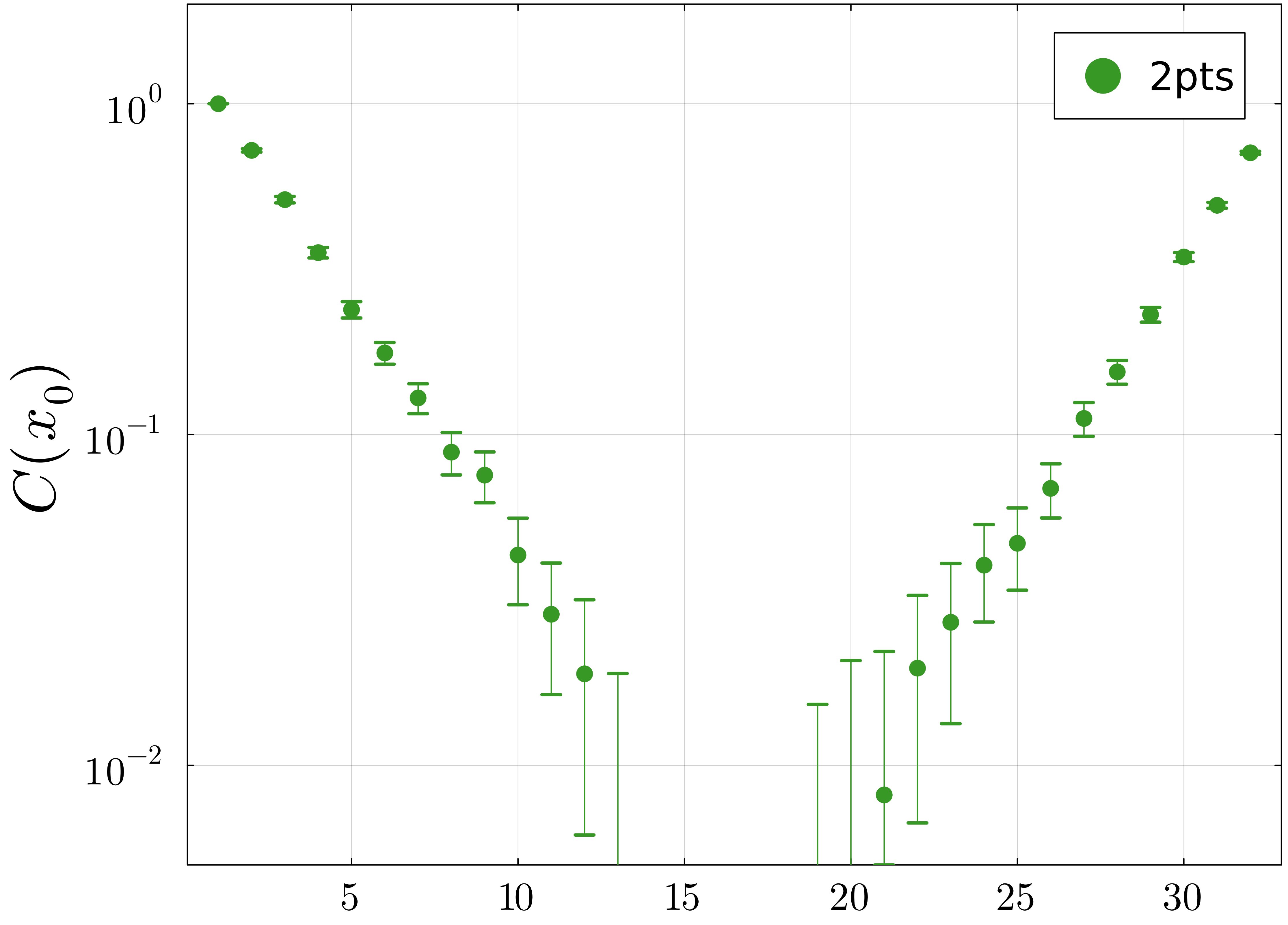
$$= \langle \mathcal{O}_t \rangle_S + \varepsilon\, C(t) + \ldots$$

Big fluctuation in reweighing factors

signal-to-noise degradation

$$\langle w \mathcal{O}_t \rangle_S = \langle \mathcal{O}_t \rangle_S + \varepsilon C(t) + \dots$$





Solution: find sample transformation that renders the w constant!

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

Improved reweigthing

$\phi \sim e^{-S}$
Configurations

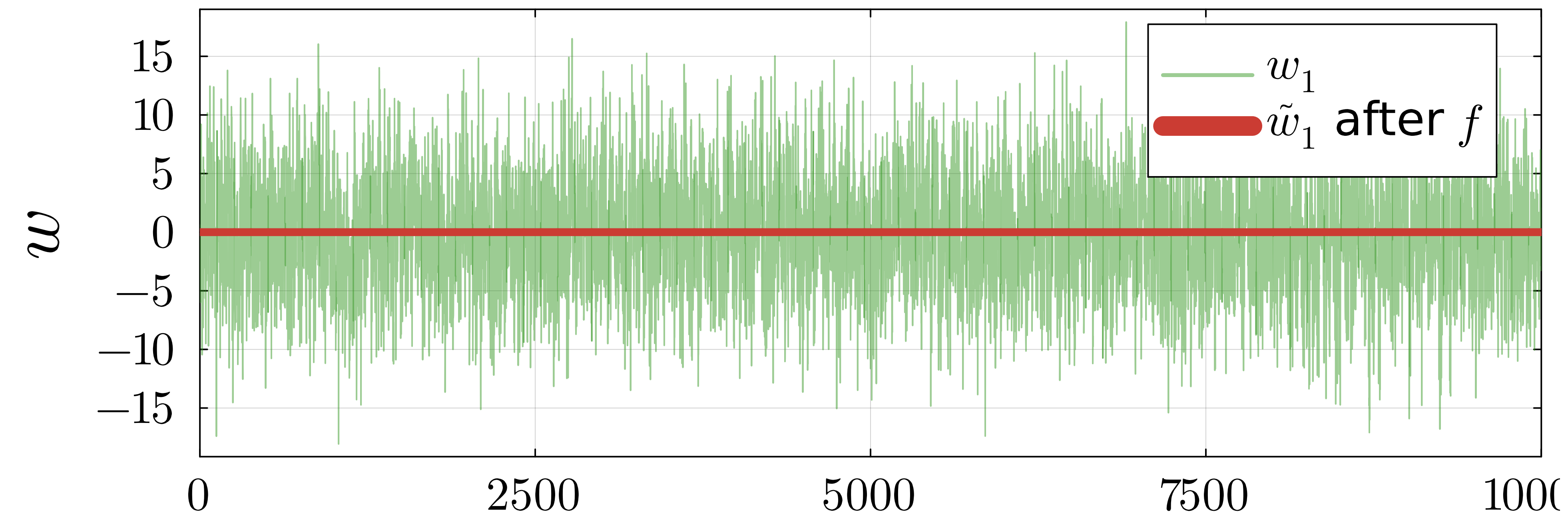
$e^{-(S - J\mathcal{O}_0)}$
Target distribution

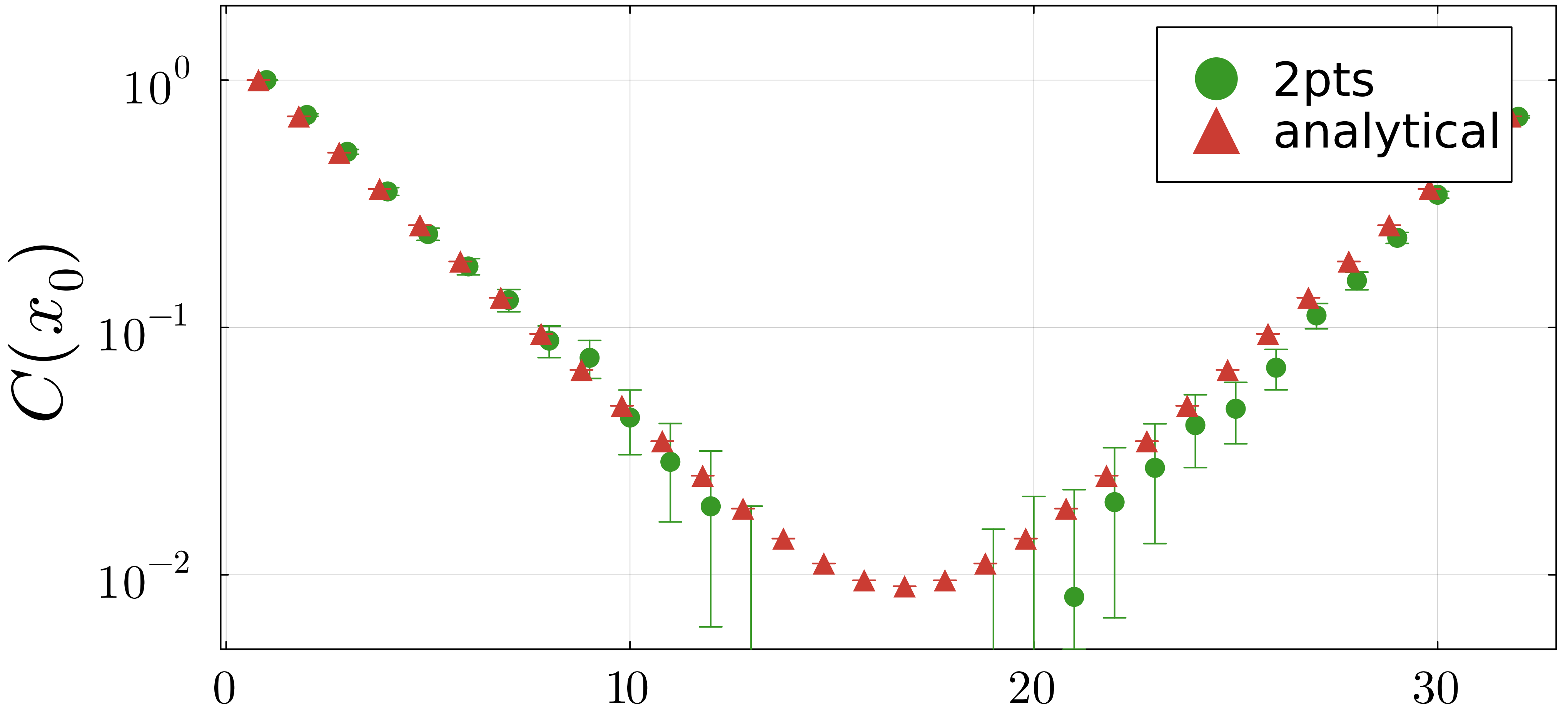
Transform samples before reweigthing

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

Scalar free theory:

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





Improved reweigthing

$\phi \sim e^{-S}$

Configurations

$e^{-(S - J\mathcal{O}_0)}$

Target distribution

Transform samples before reweigthing

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

reweight with...

$$w = \exp\Big[- S_J[\tilde{\phi}] + S[\phi] + \log|1 + \varepsilon J_f| \Big]$$

Scalar ~~free~~ interacting theory:

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

Renormalised mass

Reformulating as generative modelling

Reformulating

Lattice / physics		transport / flow	
Action	Action w/ a source	Base distribution	Target distribution
S	$S - J\mathcal{O}_0$	$r = e^{-S}/Z_0$	$p = e^{-S_J}/Z_J$

Transf. of configs induces effective action

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

Transport/flow map

$$\phi \to T(\phi) \sim q = r|\det J_T|^{-1}$$

Reweighting factors

$$\log w = -\Delta S + \log|1 + \varepsilon J_f|$$

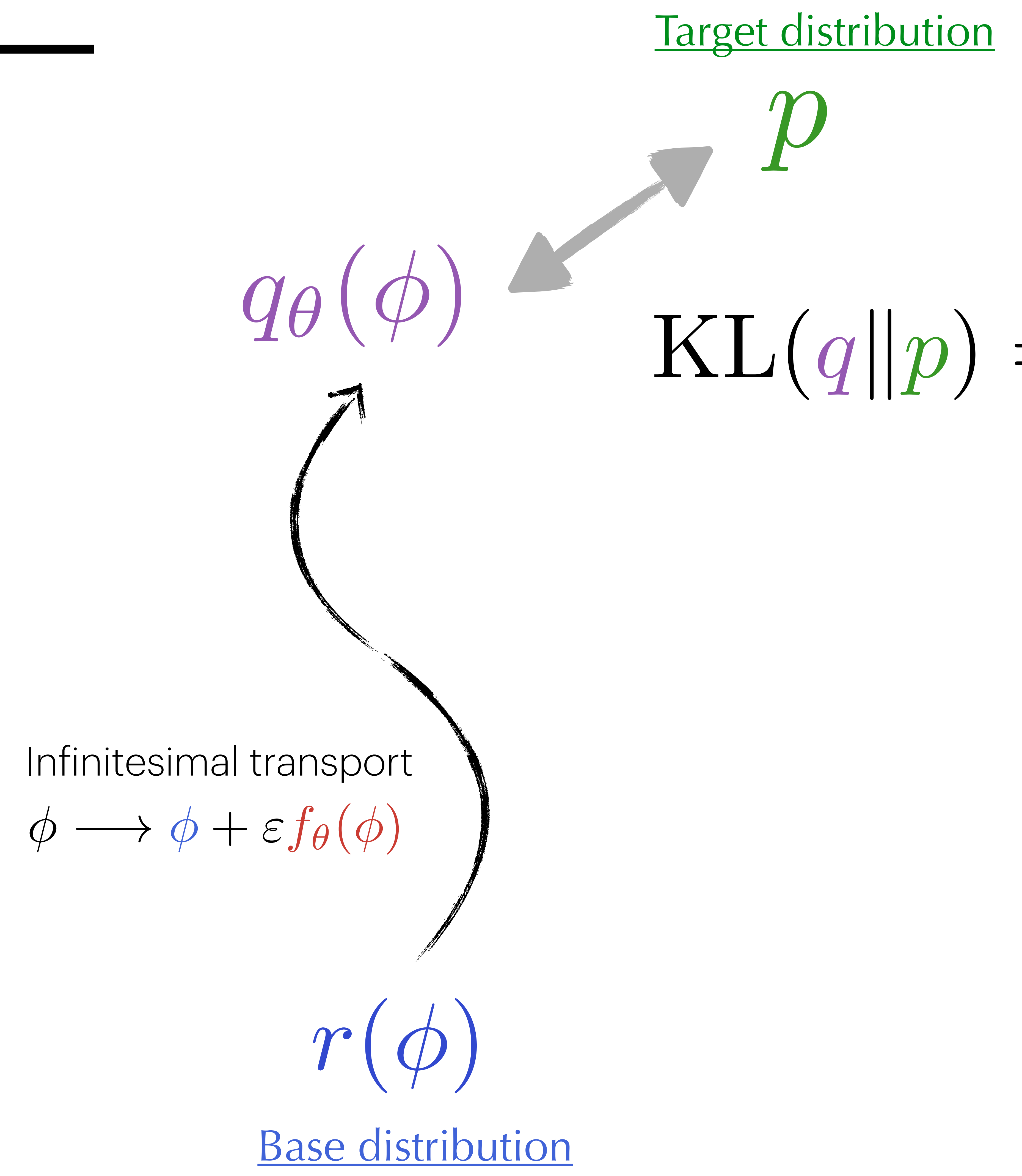
Overlap factor / likelihood ratio

$$\log w = \log p(T(\phi)) - \log r(\phi) + \log\det J_T$$

$$\langle \log w \rangle = -\mathrm{KL}(q\|p_\varepsilon)$$

$$\mathrm{KL}(q\|p_\varepsilon) = \frac{1}{2}\mathrm{Var}[\log w] + \mathcal{O}(\log w)^3$$

Linearised flow maps



$$\mathrm{KL}(q\|p) = \left\langle \log \frac{r}{p} \right\rangle$$

base-target KL divergence (constant)

$$+ \ \varepsilon \left\langle f \cdot \nabla \log \frac{r}{p} \right\rangle$$

Score-mismatch or Stein's discrepancy

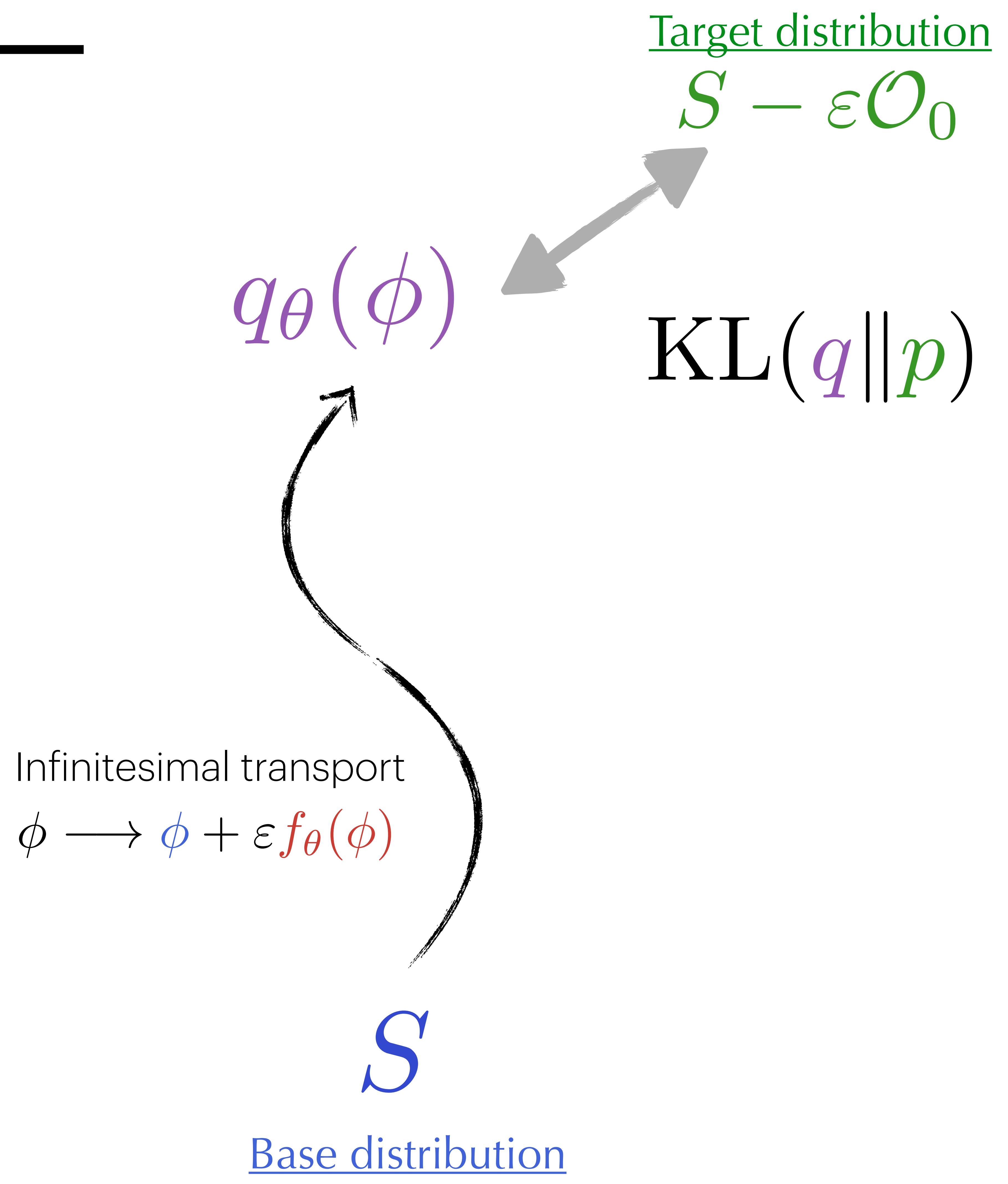
$$+ \ \frac{\varepsilon^2}{2} \left(\left\langle (\mathcal{T}_r f)^2 \right\rangle + \left\langle f H_{\log \frac{r}{p}} f \right\rangle \right)$$

Riemannian "length" of f in Stein's geometry

$$\mathcal{T}_r f = \nabla \cdot f + f \cdot \log r$$

Curvature instabilities (Hessian mismatch)

Linearised flow maps



$$\mathrm{KL}(q\|p) = \left\langle \log \frac{r}{p} \right\rangle + \varepsilon \left\langle f \cdot \nabla \log \frac{r}{p} \right\rangle + \frac{\varepsilon^2}{2}\left(\left\langle (\mathcal{T}_r f)^2 \right\rangle + \left\langle f H_{\log \frac{r}{p}} f \right\rangle \right)$$

base-target KL divergence (constant)

Score-mismatch or Stein's discrepancy

Riemannian "length" of f in Stein's geometry

$$\mathcal{T}_r f = \nabla \cdot f - f \cdot \nabla S$$

Curvature instabilities (Hessian mismatch)

Linearised flow maps

Target distribution

$S - \varepsilon \mathcal{O}_0$

$q_\theta(\phi)$

Infinitesimal transport

$\phi \longrightarrow \phi + \varepsilon f_\theta(\phi)$

S

Base distribution

$$\mathrm{KL}(q\|p) = \frac{\varepsilon^2}{2}\left\langle (\mathcal{T}_r f)^2 - 2f \cdot \nabla \mathcal{O}_0 \right\rangle + \dots$$

Parameterise f with a neural network

$$f_\theta(\phi) = \mathbf{iFFT}\left[\frac{\mathbb{A}_\theta(\tilde{\phi}_p)\delta(\mathbf{p})}{\hat{p}^2 + \mathbb{M}_\theta(\tilde{\phi}_p)}\right]$$

Use KL as a loss function

$$\left\langle (\mathrm{tr} J_{f_\theta} - f_\theta \cdot \nabla S)^2 - 2 f_\theta \cdot \nabla \mathcal{O}_0 \right\rangle$$

Minimise KL w/ variation calculus

$$\mathcal{T}_r f + \Delta \mathcal{O}_0 = 0$$

Solve the Stein-Poisson equation directly
sample per sample via Feynmann-Kac

How source reweighting could solve the StN degradation?

Recap

StN is due to big fluctuations in the reweighting factors

$$\langle w\mathcal{O}_t\rangle = \langle \mathcal{O}_t\rangle + \varepsilon\langle \mathcal{O}_t\mathcal{O}_0\rangle + \dots$$

Overlap can be systematically improved via a transport map that minimises the KL divergence

$$\phi \to \phi + \varepsilon f(\phi)$$

$$\min_f \mathrm{Var}[\log w] \approx \min_T \mathrm{KL}(q\|p_\varepsilon)$$

Minimising KL is equivalent to solve the Stein-Poisson eq.

$$\mathrm{KL}(q\|p) = \frac{\varepsilon^2}{2}\Big((\mathcal{T}_r f)^2 - 2f\cdot\nabla\mathcal{O}_0\Big) + \dots$$

$$\mathcal{T}_r f + \Delta\mathcal{O}_0 = 0$$

Numerical approach

Direct minimisaation via NN

Parameterise f with a NN network

$$f_\theta(\phi) = \mathrm{iFFT}\left[\frac{\mathbb{A}_\theta(\phi)\delta(\mathbf{p})}{\hat{p}^2 + \mathbb{M}_\theta(\phi)} + \mathbb{R}_\theta(\phi)\right]$$

Training: minimise KL to $\mathcal{O}(\varepsilon^2)$

1. Take a batch of samples $\phi \sim e^{-S}$
2. Feedforward through $f_\theta(\phi)$
3. Compute KL loss
   - 3.1 Compute HMC force ∇S
   - 3.2 Estimate $\mathrm{tr} J_f$
   - 3.3 Average over batch

$$\left\langle (\mathrm{tr} J_{f_\theta} - f_\theta \cdot \nabla S)^2 - 2 f_\theta \cdot \nabla \mathcal{O}_0 \right\rangle$$

4. Backprop to update θ

Variance reduction:

1. Improve confs $\tilde{\phi} = \phi + \varepsilon f_\theta(\phi)$
2. Compute importance weigths
   - 2.1 Compute action diff. $\Delta S[\tilde{\phi}]$
   - 2.2 Estimate $\mathrm{tr} J_f$
   - 2.3 Combine for each sample

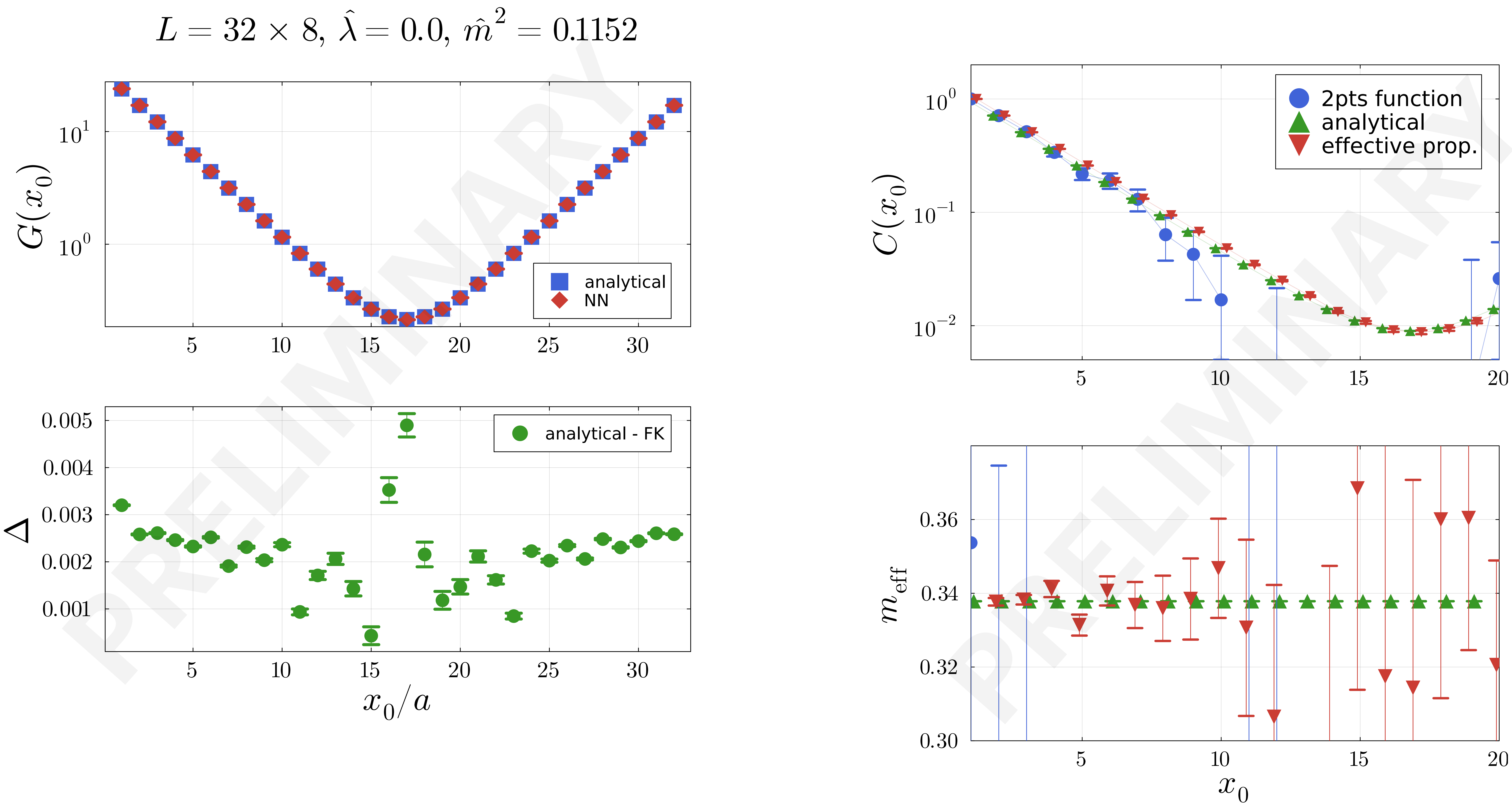
$$w[\phi] = \exp\left[\Delta S[\tilde{\phi}] - \varepsilon\, \mathrm{tr} J_{f_\theta}\right]$$

3. Reweight correlator

$\langle w \mathcal{O}_t \rangle \longrightarrow$ extract $\mathcal{O}(\varepsilon)$

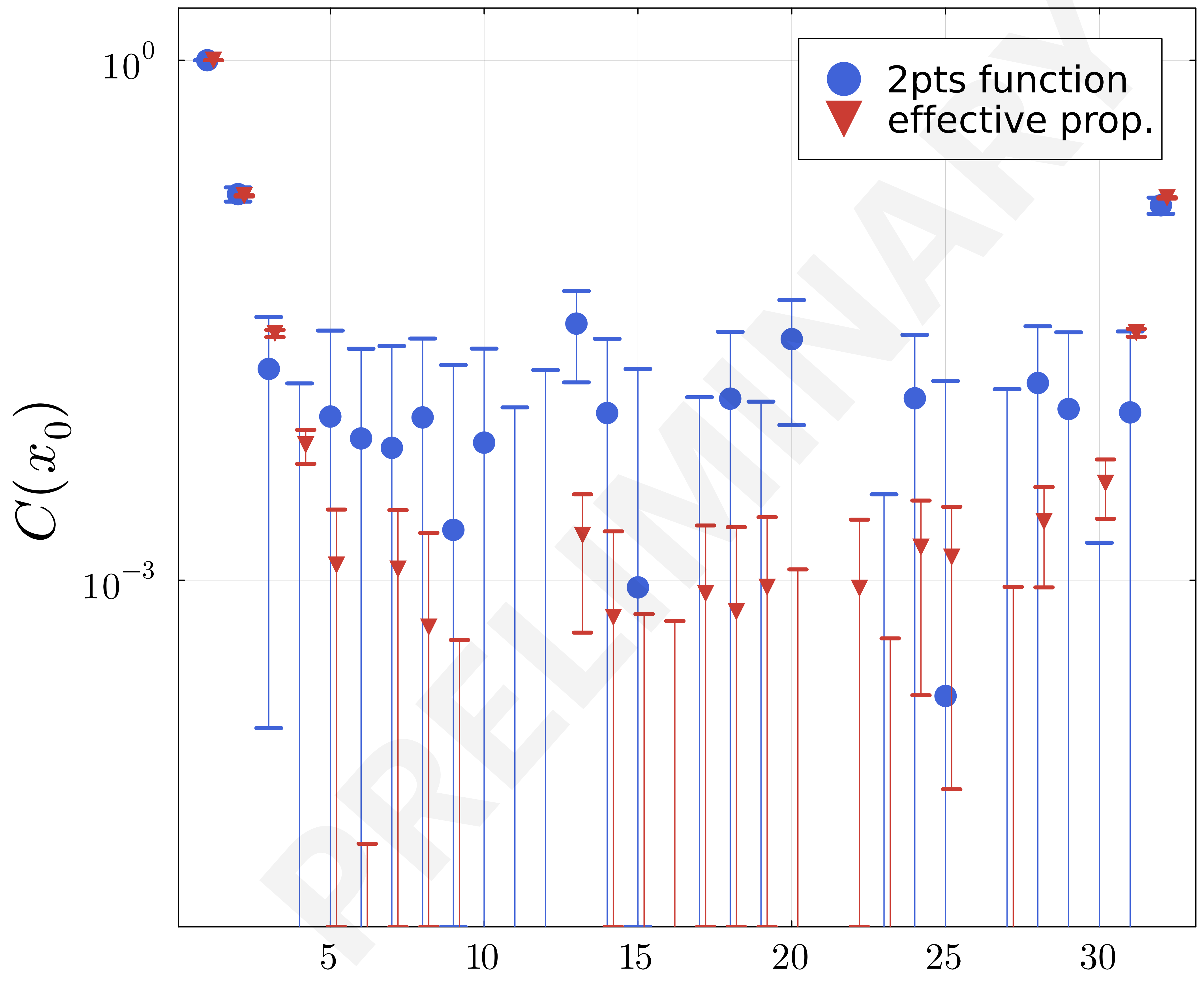
4. Propagate error

Direct minimisation via NN

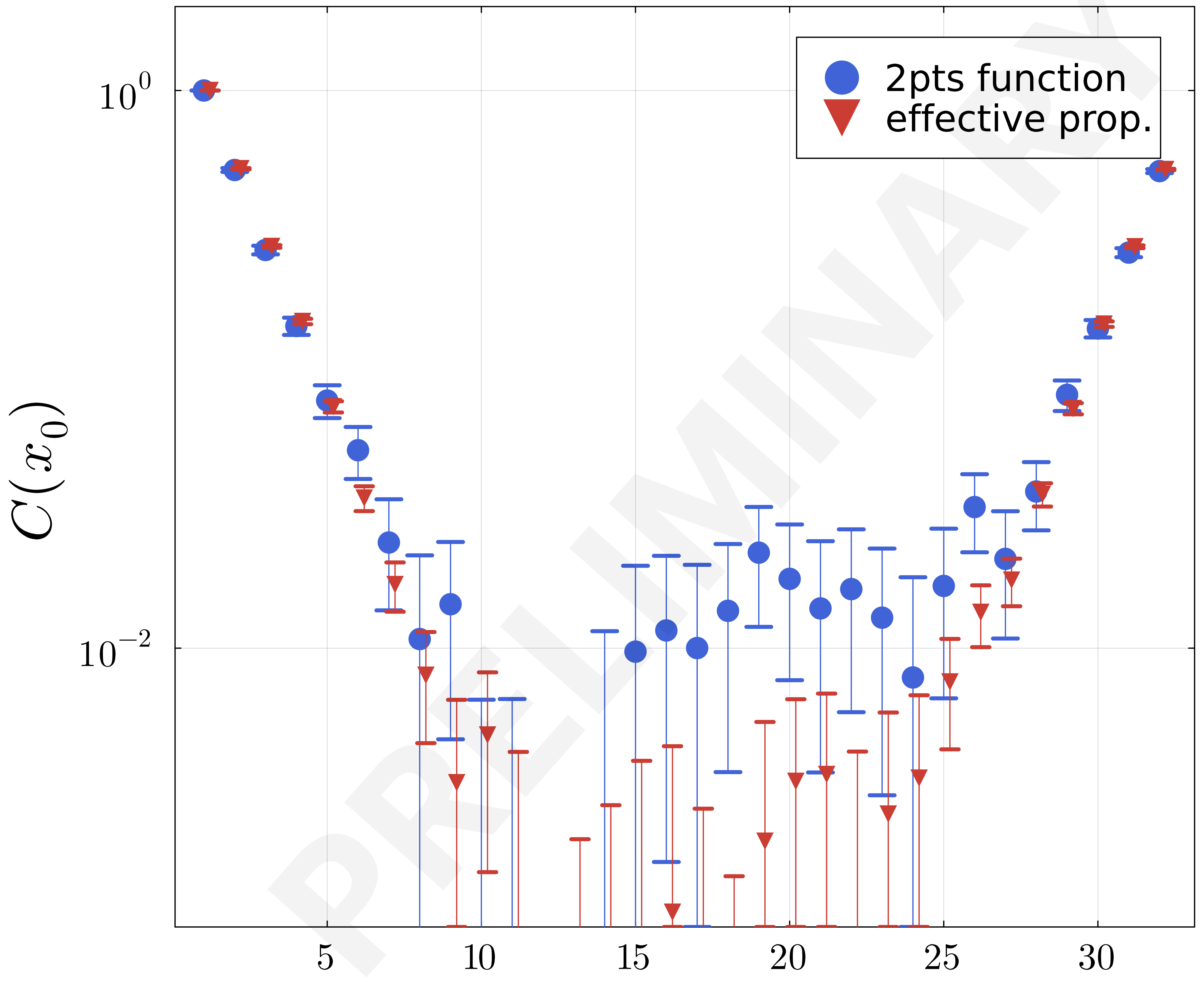


Direct minimisation via NN

$L = 32 \times 8, \lambda = 0.05, \kappa = 0.125$



$L = 32 \times 8, \lambda = 0.05, \kappa = 0.243$



Stein-Poisson equation

Extremal KL divergence ($\delta \mathrm{KL}/\delta f = 0$) implies

$$\mathcal{T}_r f + \Delta\mathcal{O}_0 = 0$$

fluctuation at t=0

$\Delta\mathcal{O}_0 \equiv \mathcal{O}_0 - \langle \mathcal{O}_0 \rangle$

Stein operator

$\mathcal{T}_r f = \nabla \cdot f - f \cdot \nabla S$

How to "invert" the Stein operator?

Restrict to gradient fields $f = \nabla\varphi$

$$\mathcal{T}_r f \equiv \underbrace{(\triangle_2 - \nabla S \cdot \nabla)}\varphi = -\Delta\mathcal{O}_0$$

Generator of (overdamped) Langevin dynamics $\mathcal{L}$

i.e. how a smooth function evolves infinitesimally in (stochastic) time under Langevin dynamics

$$\mathrm{d}Z_\tau = -\nabla S \, \mathrm{d}\tau + \sqrt{2}\, \mathrm{d}W_\tau$$

"inverse" is given by Markov semigroup theory

$$\varphi = -\mathcal{L}^{-1}\Delta\mathcal{O}_0$$

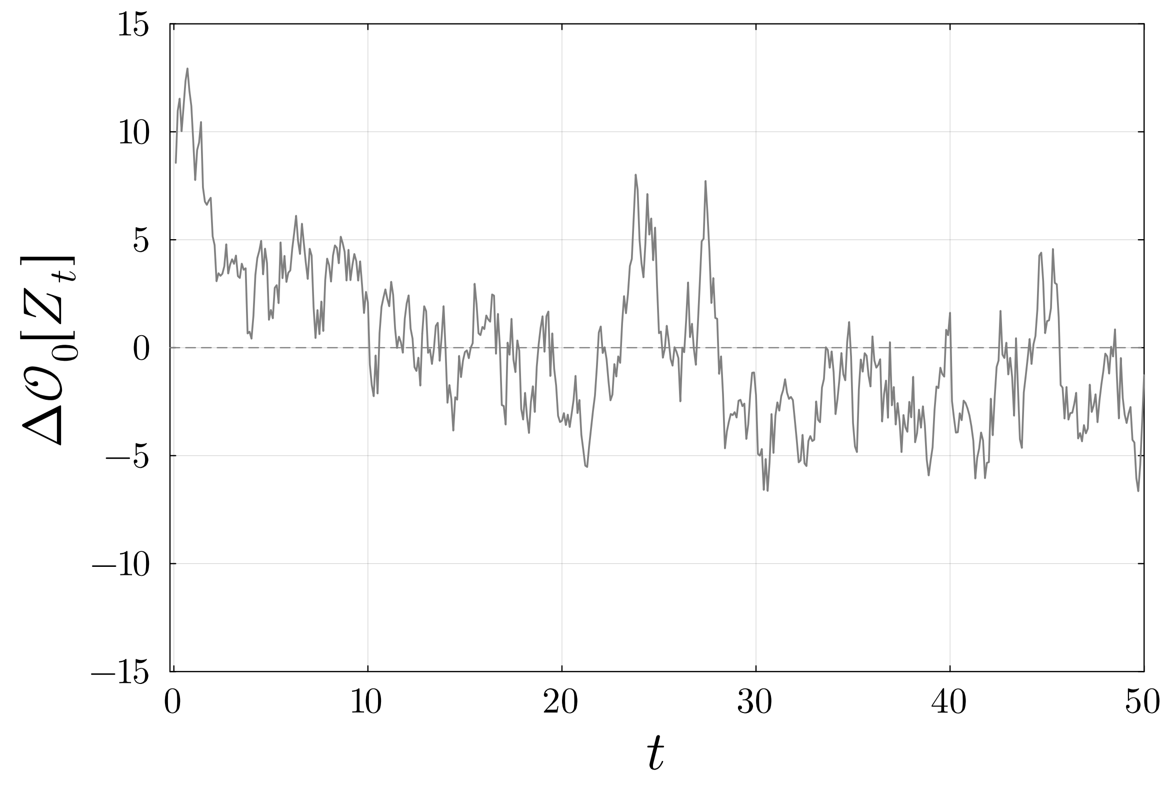
Feynman-Kac formula

$$f(\phi) = \nabla_\phi \varphi$$

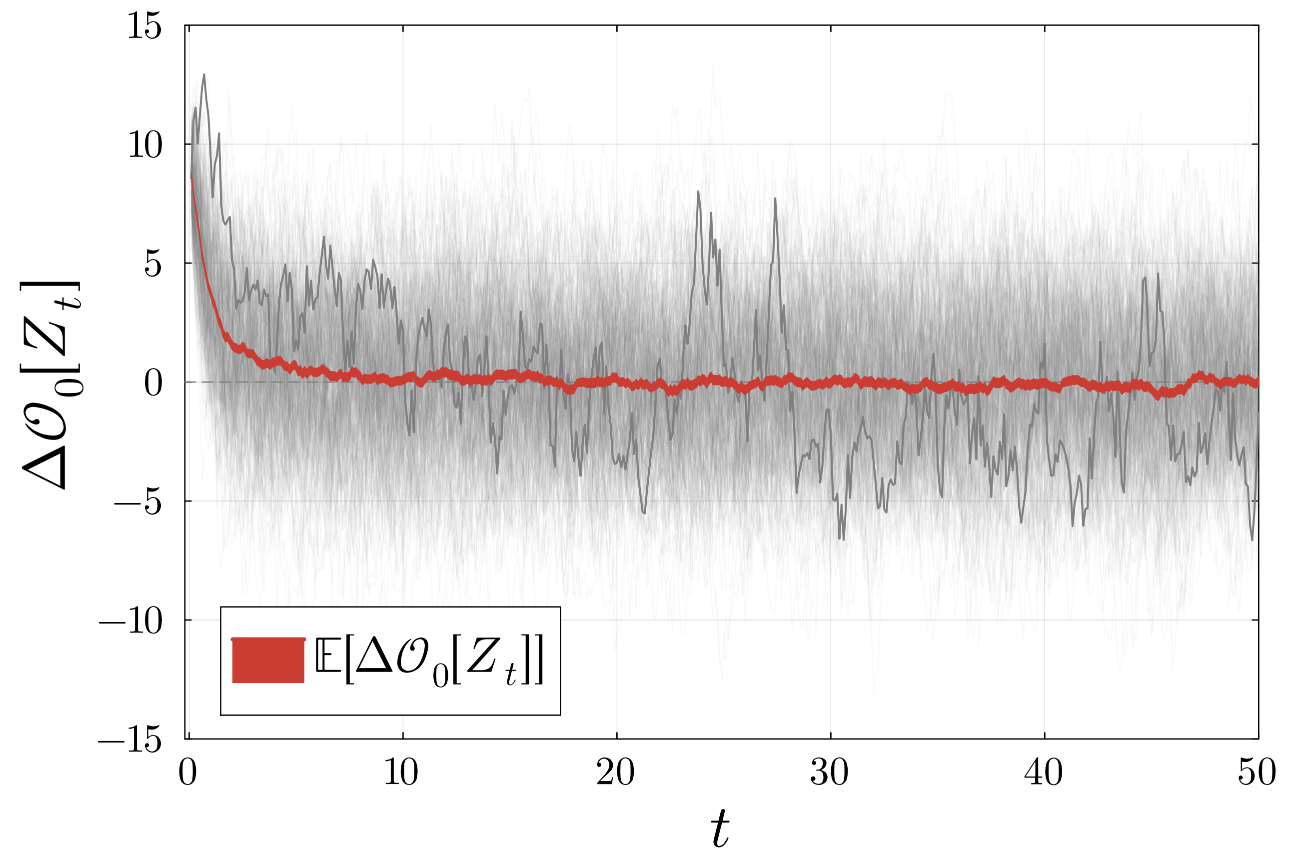
[Kanwar-Albergo] - hep-lat/2603.00252
[Albergo, Eijnden] - 2410.02711

$$\varphi(\phi) = \int_0^\infty \mathrm{d}\tau \, \mathbb{E}\Big[\mathcal{O}_0[Z_\tau] - \langle \mathcal{O}_0 \rangle \Big| Z_0 = \phi\Big]$$

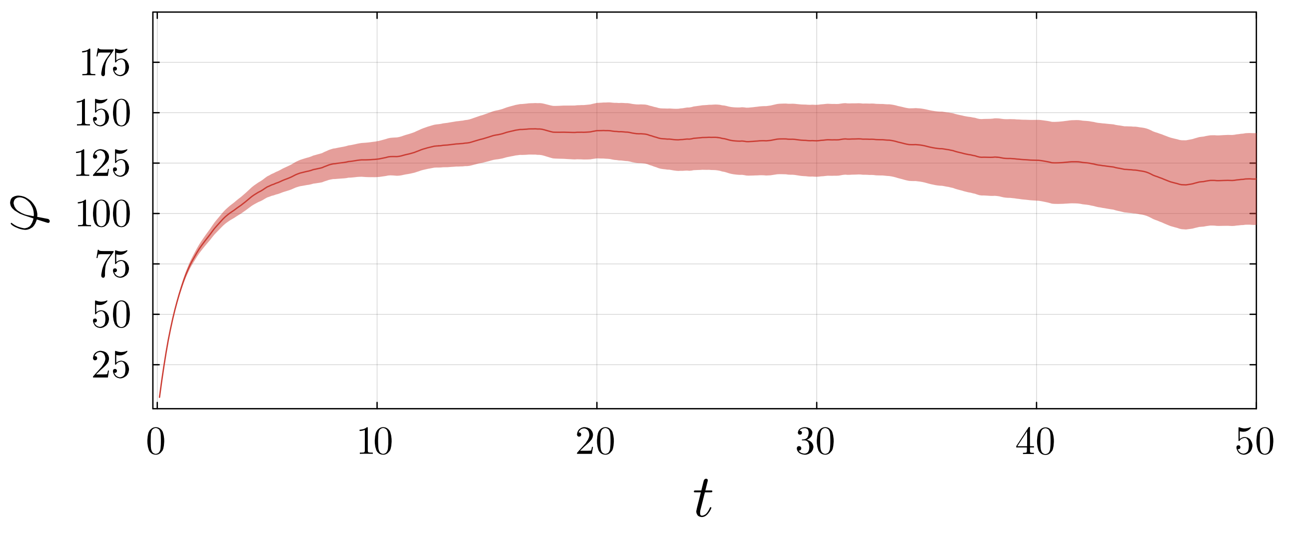
Integrate Langevin dynamics starting from $Z_0 = \phi$



Expectation value over walkers (average over Brownian noise)



Integrate in stochastic time



SDU

ħQTC

Feynman-Kac formula

$f(\phi) = \nabla_\phi \varphi$

[Kanwar-Albergo] - hep-lat/2603.00252
[Albergo, Eijnden] - 2410.02711

$$f(\phi) = \nabla_\phi \int_0^\infty \mathrm{d}\tau\, \mathbb{E}\Big[\mathcal{O}_0[Z_\tau] - \langle \mathcal{O}_0 \rangle \Big| Z_0 = \phi\Big]$$

Swap gradient and noise average

$$\nabla\varphi = \mathbb{E}\Big[\nabla \int_0^\infty \mathrm{d}\tau\ \mathcal{O}_0[Z_\tau | Z_0 = \phi]\Big]$$

$$f = \lim_{T\to\infty} \mathbb{E}[A_T(\phi)]$$

`SDEsolve(…)` forward to obtain

$$\hat{\varphi}^{[0,T]} = \int_0^T \mathrm{d}\tau\ \mathcal{O}_0[Z_t]$$

Backprop to get $\nabla_{Z_0}\hat{\boldsymbol{\varphi}}^{[0,T]}$ or save "every" intermediate Z_t and solve adj. backward

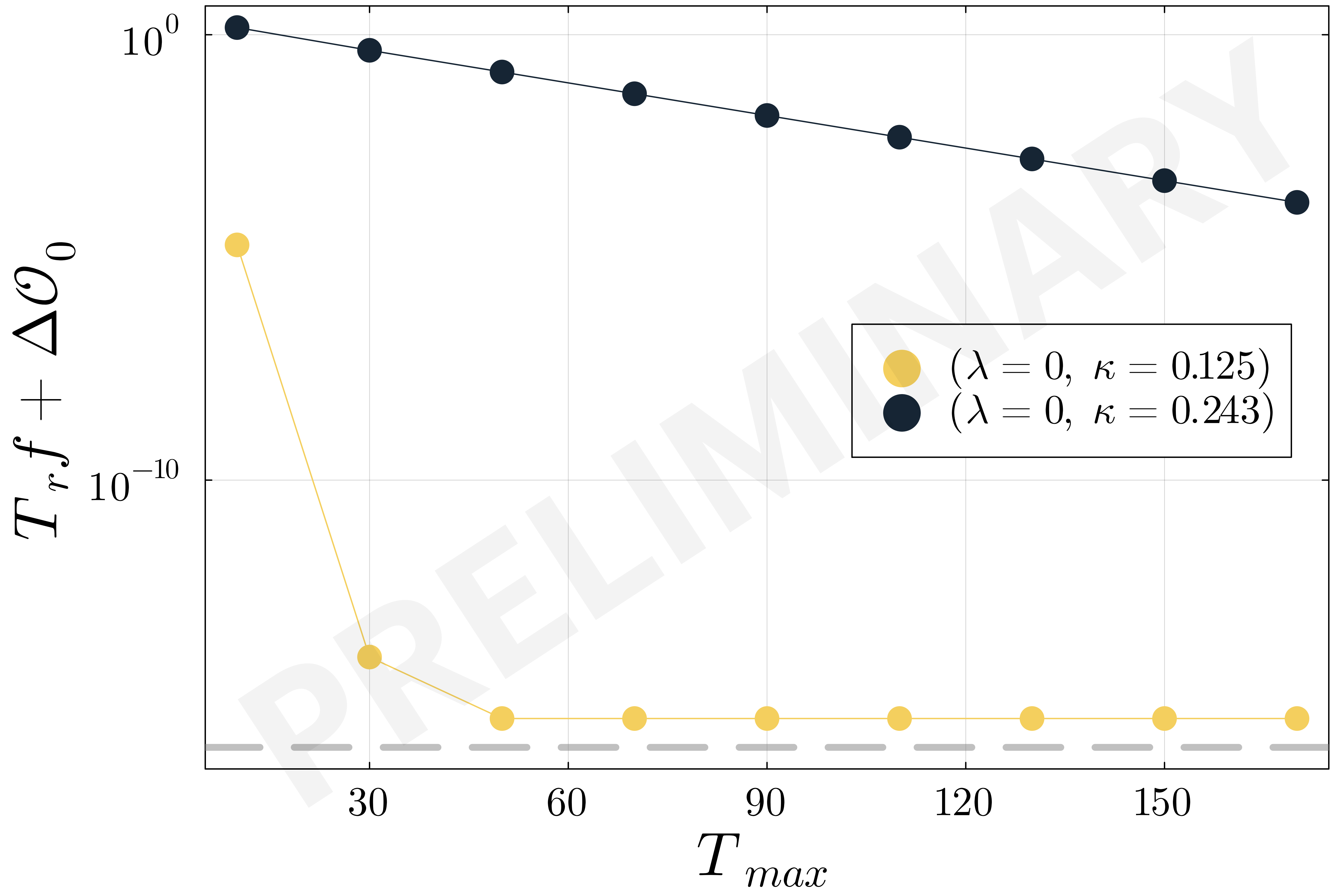
$$\frac{dA_\tau}{d\tau} = -H_S[Z_{t-\tau}]A_\tau + \delta_{x_0,0} \qquad A_0 = 0$$

Works, but reproduce same hessian instabilities of HAD!

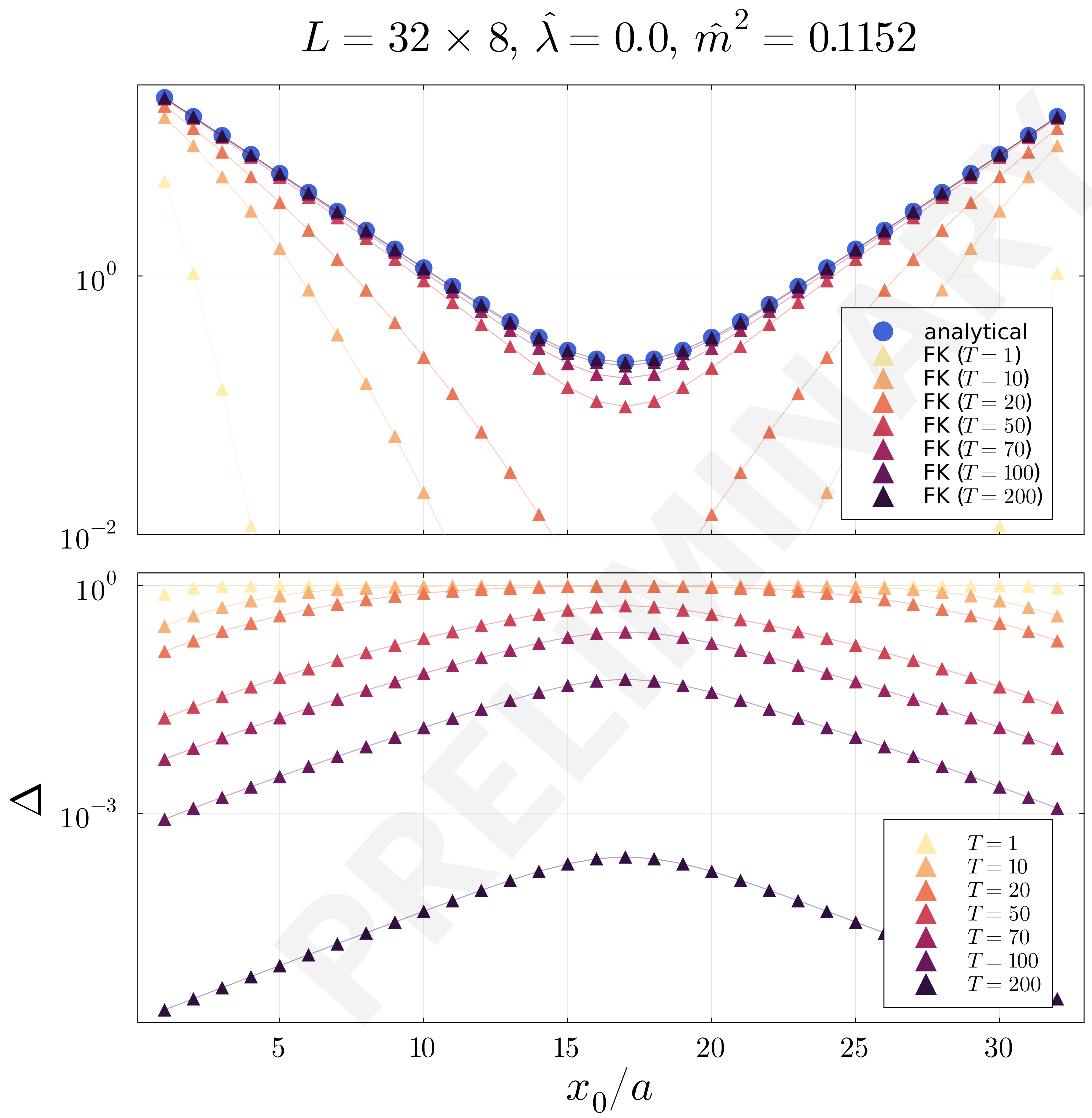
Free theory

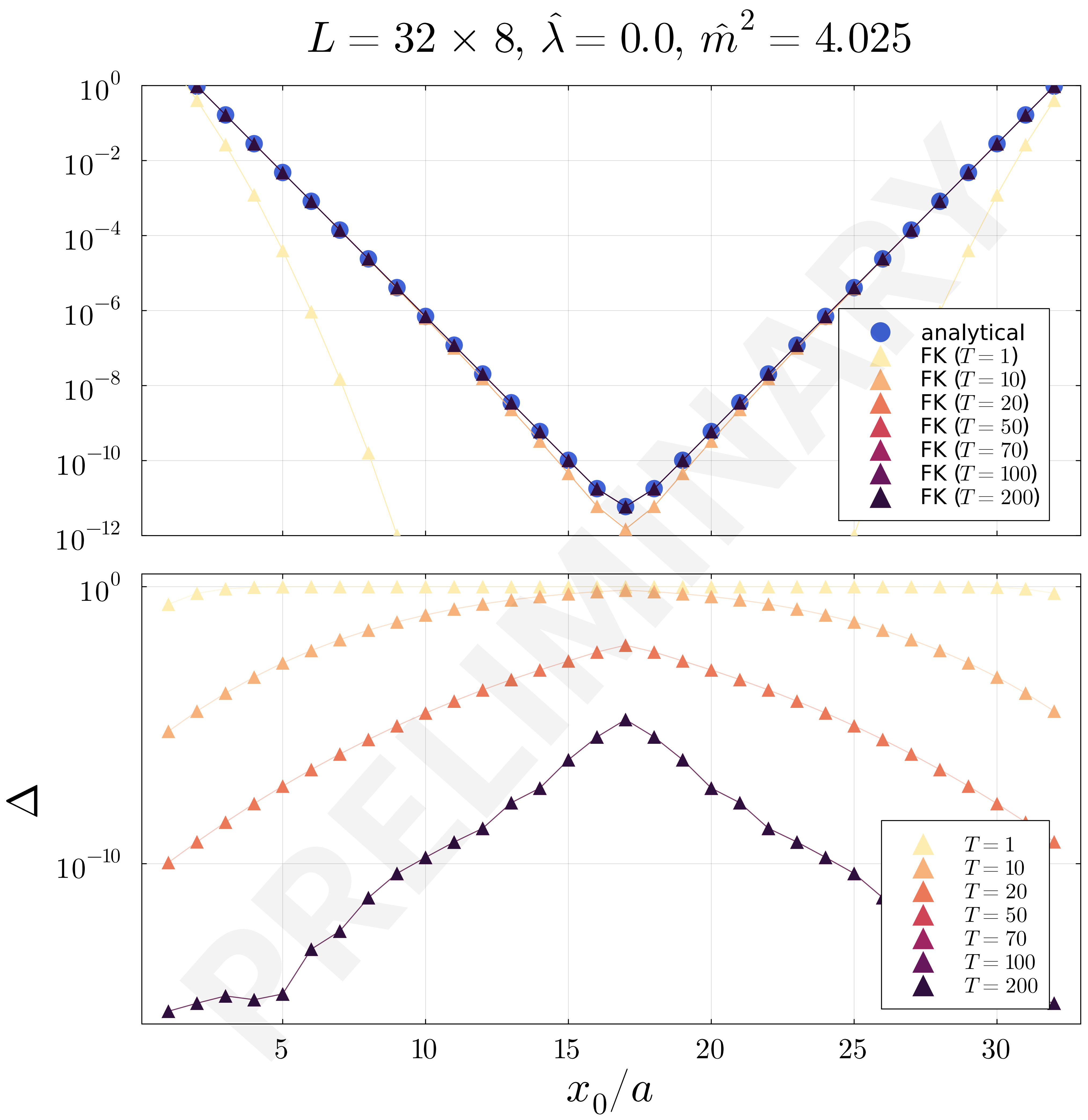
Convergence is increasingly harder towards critical points

$$f = \lim_{T\to\infty} \mathbb{E}[A_T(\phi)]$$



Free propagator

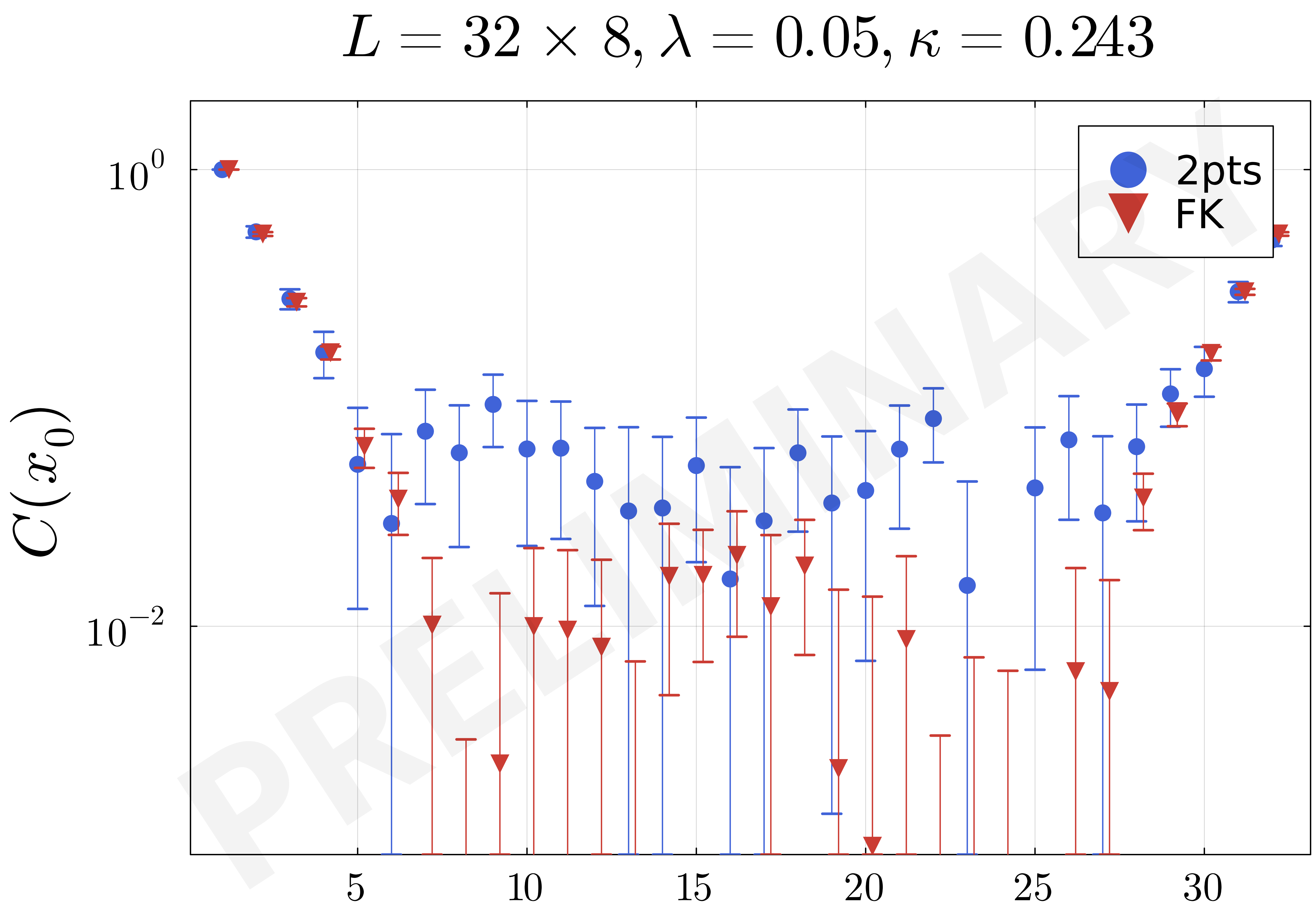




SDU

ħQTC

Correlator (close to criticality)



Thanks for the attention!

- [Catumba, Ramos] - *"Stochastic automatic differentiation and the signal to noise problem"*, Eur.Phys.J.C 85 (2025)
- [Liu, Wang] - *"Stein variational gradient descent: A general purpose bayesian inference algorithm"*, NeurIPS 29 (2016)
- [Albergo, Kanwar] - *"A Monte Carlo estimator of flow fields for sampling and noise problems"*, PoS Lattice 2025
- [Albergo, Eijnden] - *"NETS: A Non-Equilibrium Transport Sampler"*, e-Print: 2410.02711
- [Butti, Catumba, Nada, Spatscheck] - *in preparation*

BACKUP

Linearised flow maps

Target distribution

$S - \varepsilon\mathcal{O}_0$

$q_\theta(\phi)$

$$\mathrm{KL}(q\|p) = \frac{\varepsilon^2}{2}\Big\langle (\mathcal{T}_r f)^2 - 2f\cdot\nabla\mathcal{O}_0 \Big\rangle + \dots$$

Minimise w/ variation calculus

(write $f \to f + \delta f$ then set $\delta\mathrm{KL} = 0$)

Infinitesimal transport

$\phi \longrightarrow \phi + \varepsilon f_\theta(\phi)$

$$\mathcal{T}_r f + \Delta\mathcal{O}_0 = 0$$ *Stein-Poisson equation*

$\Leftrightarrow \min_T \mathrm{KL}(q\|p_\varepsilon)$

S

Base distribution

$\Delta\mathcal{O}_0 \equiv \mathcal{O}_0 - \langle\mathcal{O}_0\rangle$

field fluctuation at t=0

$\mathcal{T}_r f = \nabla\cdot f - f\cdot\nabla S$

Stein operator $\langle\mathcal{T}_r f\rangle \equiv 0$

Improved reweigthing

$\phi \sim e^{-S}$
Configurations

$e^{-(S - J\mathcal{O}_0)}$
Target distribution

What if I don't know the exact f?

Transform samples before reweigthing

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

Find f that renders the (log-)weights constant!

$$\langle \not{w} \mathcal{O}_t \rangle_S$$

Importance weights becomes

$$w = \exp\Big[- S_J[\tilde{\phi}] + S[\phi] + \log|1 + \varepsilon J_f| \Big]$$

$$\tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\vec{p})}{\hat{p}^2 + \hat{\Sigma}_\theta[\phi]}\varepsilon$$

Can we learn an *effective propagator* minimising the variance of the (log) weights?

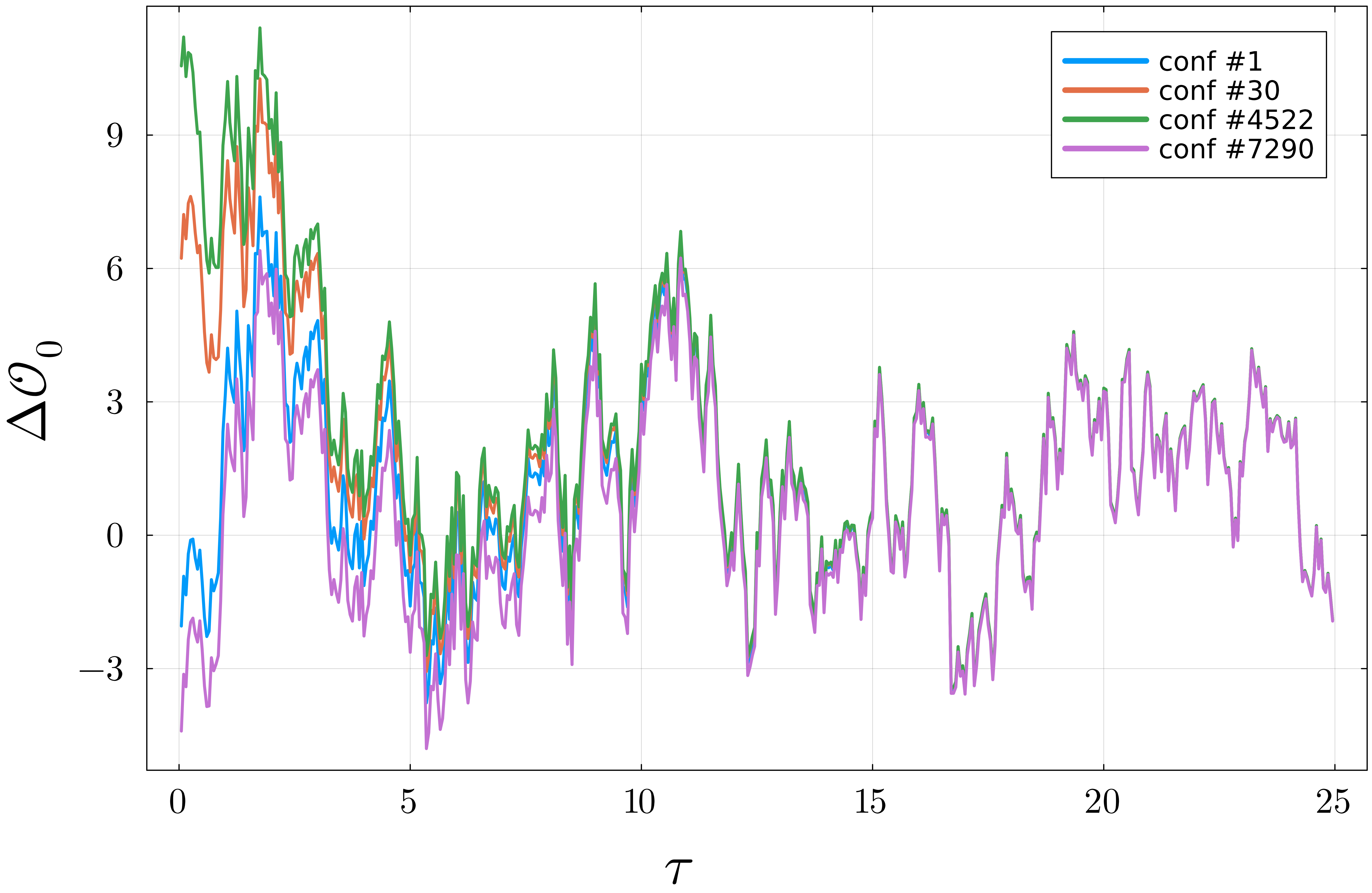
$$\log w = S[\phi] - S_J[\tilde{\phi}] + \log|1 + \varepsilon J_f|$$

Log-det computable with explicit differentiation + backprop

$$\log\det\left(1 + \varepsilon\frac{\mathrm{d}f}{\mathrm{d}\phi}\right) = \varepsilon\,\mathrm{tr}\frac{\mathrm{d}f}{\mathrm{d}\phi} + \mathcal{O}(\varepsilon^2)$$

How do we train $\hat{\Sigma}_\theta$?

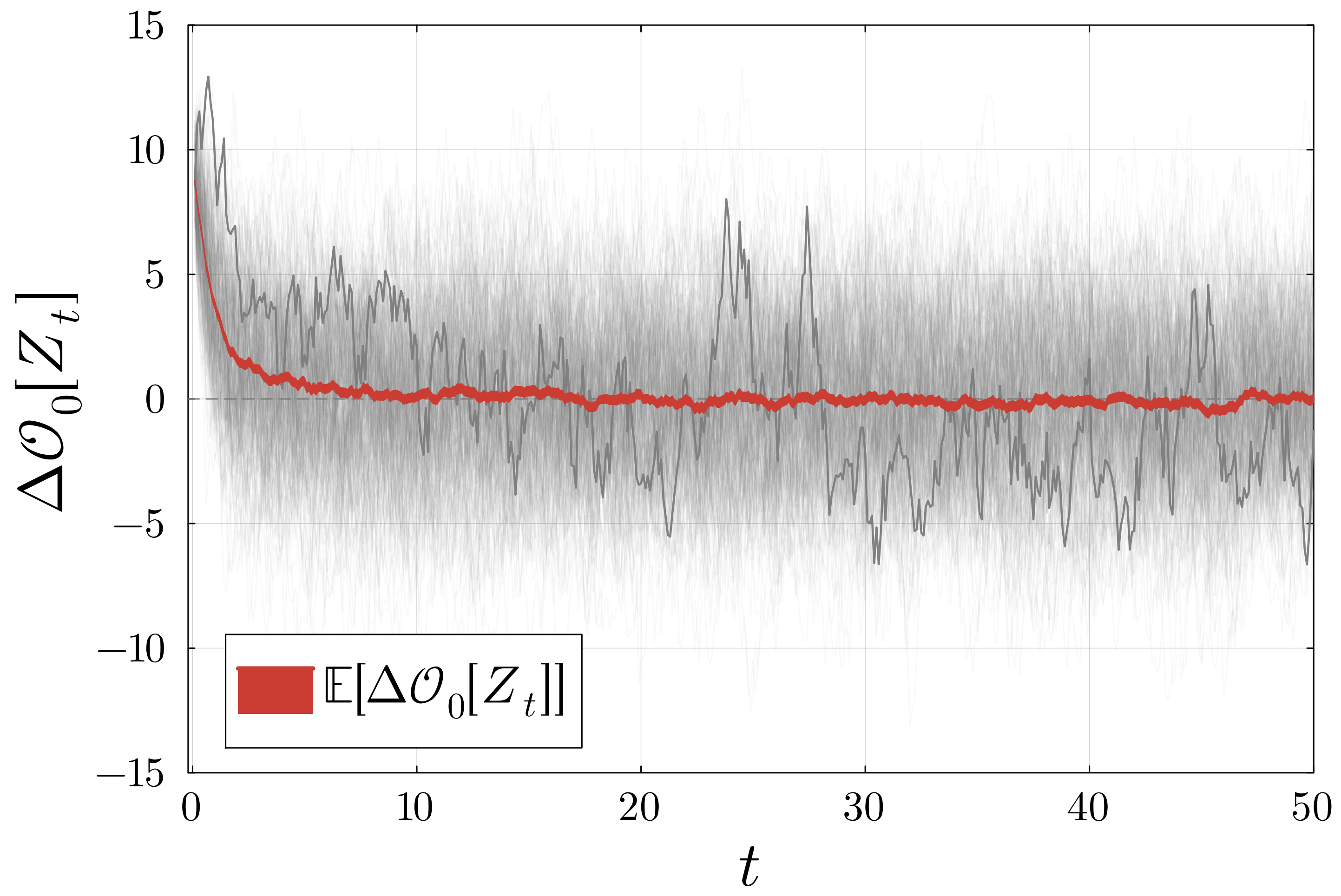
Maximally coupled noise



Feynman-Kac formula

$$\varphi(\phi) = \int_0^\infty \mathrm{d}\tau \, \mathbb{E}\Big[\mathcal{O}_0[Z_\tau] - \langle \mathcal{O}_0 \rangle \Big| Z_0 = \phi\Big]$$

Langevin evolution makes the integrand to converge to zero exponentially in $\boldsymbol{\tau}$



Free-field theory check $S = \phi K \phi$

Langevin dynamics = Ornstein-Uhlenbeck process

$$\mathbb{E}[Z_\tau | Z_0 = \phi] = e^{-K\tau}\phi$$

Feynman-Kac formula gives the *0-mom propagator*

$$\varphi = \sum_{\vec{x}} K^{-1}\phi \qquad f = \nabla\varphi = \sum_{\vec{x}} K^{-1}$$

Define Langevin autocorrelation function

$$\Gamma(\tau) = \Big\langle \Delta\mathcal{O}_0[Z_\tau]\Delta\mathcal{O}_0[Z_0] \Big\rangle_{Z_0 \sim r}$$

Green-Kubo relation:

$$\langle \varphi \Delta\mathcal{O}_0 \rangle = -\int_0^\infty \mathrm{d}\tau \; \Gamma(\tau)$$

$$\varphi(\phi) = \int_0^\infty \mathrm{d}\tau \, \mathbb{E}\Big[\mathcal{O}_0[Z_\tau] - \langle \mathcal{O}_0 \rangle \Big| Z_0 = \phi\Big]$$

Feynman-Kac: numerical implementation

1. Select one sample ϕ. Compute HMC force ∇S

2. Integrate Langevin trajectory (EM scheme, with `dt` and `T`)

   2.1 generate Gaussian noise upfront. `eps = randn(T//dt*size(ϕ))`

   2.2 initial config: `Z[0].= ϕ`

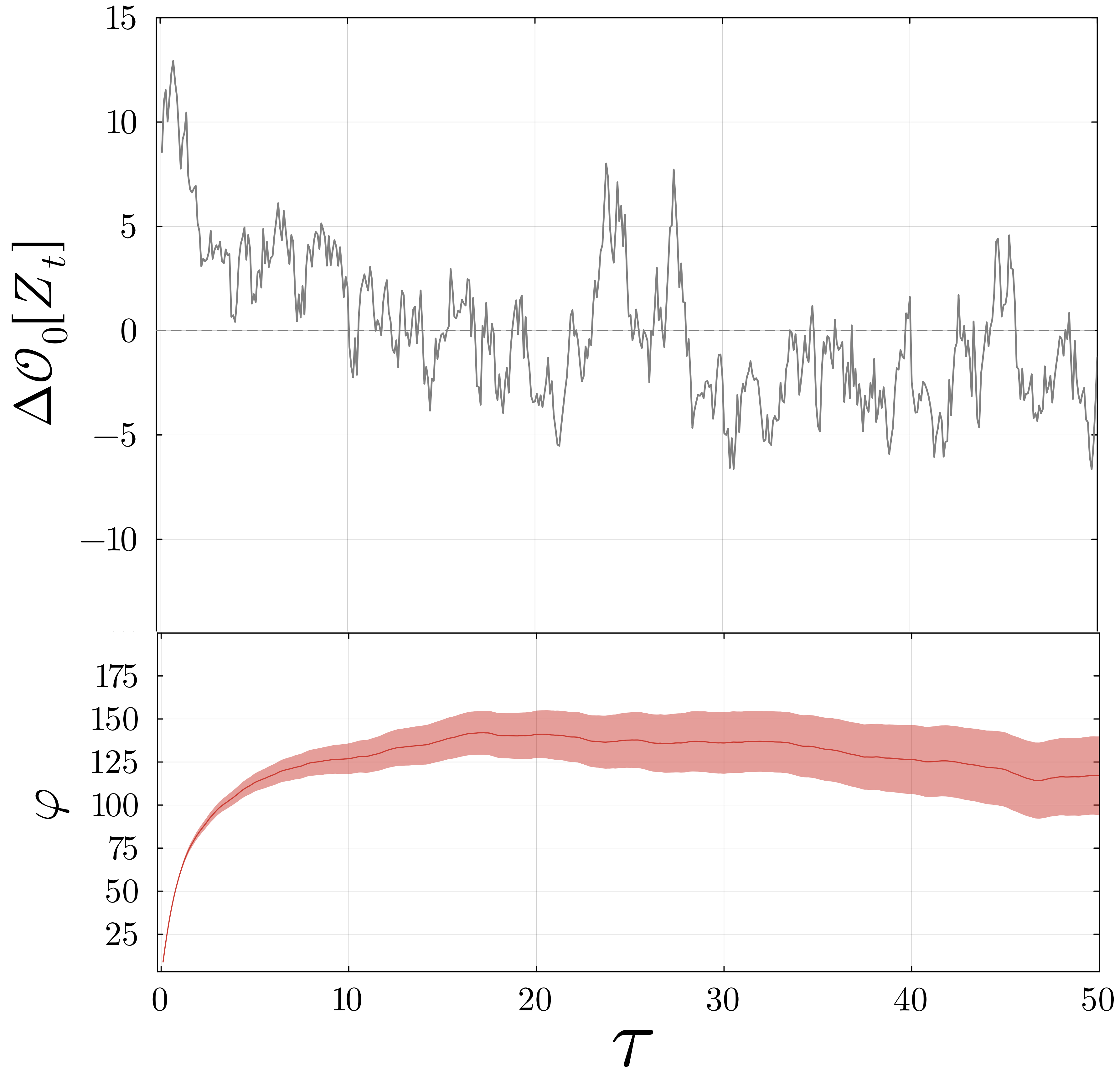
   2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`

3. Repeat for `Nw` walkers

4. Compute integral (correlated cumulative sum in `[0,T]`)

5. Compute the gradient wrt the original conf

6. Repeat for all confgurations



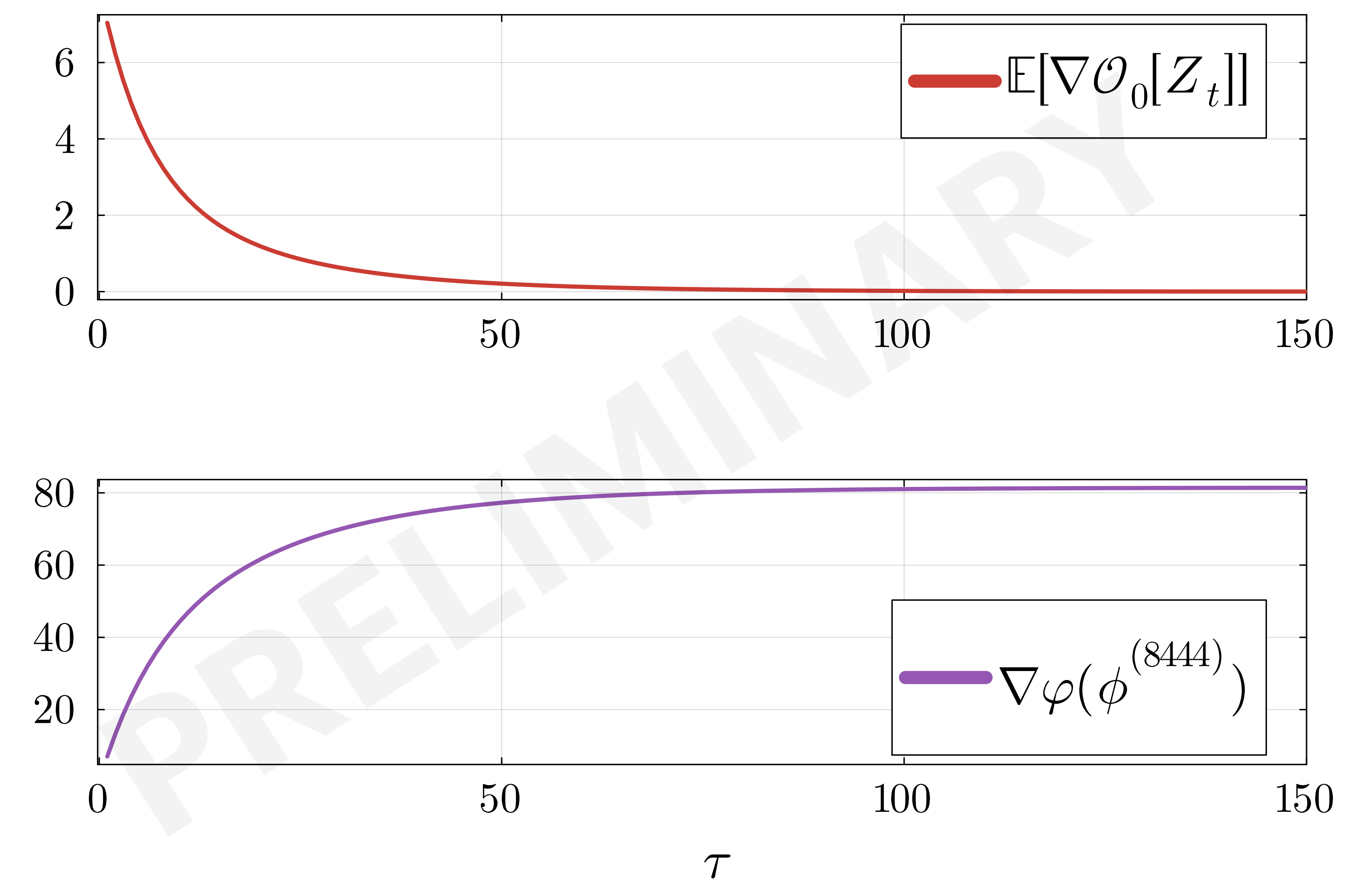
$$\varphi(\phi) = \int_0^\infty \mathrm{d}\tau\, \mathbb{E}\Big[\mathcal{O}_0[Z_\tau] - \langle \mathcal{O}_0 \rangle \Big| Z_0 = \phi\Big]$$

Feynman-Kac: numerical implementation

[Kanwar-Albergo] - hep-lat/2603.00252

1. Select one sample ϕ. Compute HMC force ∇S
2. Integrate Langevin trajectory (EM scheme, with `dt` and `T`)
   - 2.1 generate Gaussian noise upfront. `eps = randn(T//dt*size(ϕ))`
   - 2.2 initial config: `Z[0].= ϕ,  int=0.`
   - 2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`
   - 2.4 accumulate: `int += volume_sum(Z[t+1])`
3. Compute $\hat{\varphi}^{[0,T]}$ *for single walker*
4. Compute gradient $\nabla\hat{\varphi}^{[0,T]}$ *for a single walker* *
5. Repeat for `Nw` walkers with coupled noise
6. Repeat for all confgurations

$$f(\phi) = \mathbb{E}\Big[\nabla \underbrace{\int_0^\infty \mathrm{d}t\, \mathcal{O}_0[Z_t]}_{\hat{\varphi}^{[0,T]}}\Big]$$



Improved reweighting

$\phi \sim e^{-S}$

Configurations

$e^{-(S-J\mathcal{O}_0)}$

Target distribution

Transform confs before reweigthing

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

Weights becomes

$$w = \exp\left[-S_J[\tilde{\phi}] + S[\phi] + \log|1 + \varepsilon J_f| \right]$$

Find f that renders the (log-)weights constant!

$$\langle \not{w} \mathcal{O}_t \rangle_S$$

Scalar free theory:

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

f independent of ϕ, weights are simply

$$w = 1 + \varepsilon \Big(\mathcal{O}_0 - f \cdot \nabla S \Big) + \dots$$

The correlator gets a free control variate...

$$\langle w \mathcal{O}_t \rangle = \langle \mathcal{O}_t \rangle + \varepsilon \Big[C(t) - \langle f \cdot \nabla(\mathcal{O}_t - S) \rangle \Big]$$

zero on average but controls the variance!

(forward) Automatic Differentiation

$$\theta_0$$

$$f$$

$$f(\theta_0)$$

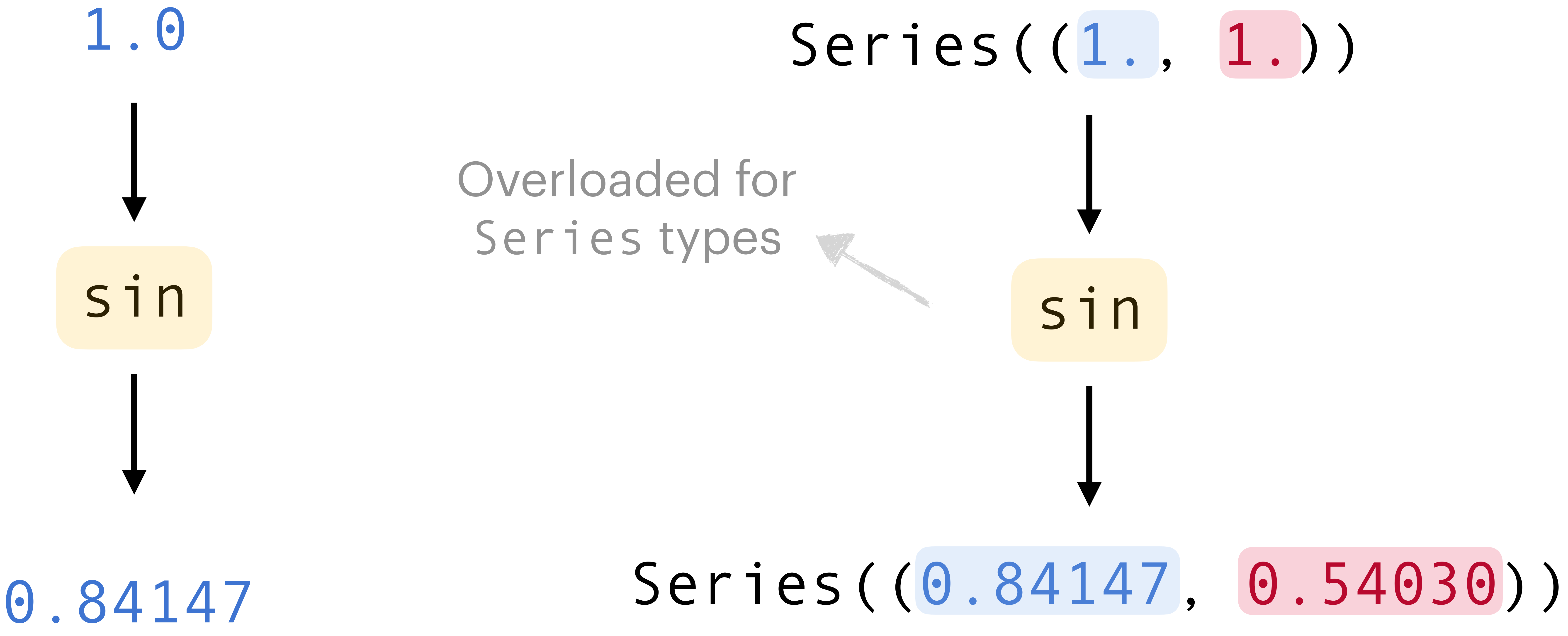
$$\tilde{\theta} = \theta_0 + \varepsilon$$

$$f$$

hard-code

$$\varepsilon^2 = 0$$

$$f(\tilde{\theta}) = f(\theta_0) + \varepsilon \left.\frac{\mathrm{d}f}{\mathrm{d}\theta}\right|_{\theta=\theta_0}$$


1.0
sin
0.84147
Overloaded for
Series types
Series((1., 1.))
sin
Series((0.84147, 0.54030))

MCMC samples with S_β

$\langle \mathcal{O} \rangle_{\beta_0} \; \{\phi\}$

β'

weights

$w \equiv e^{S_{\beta_0} - S_{\beta'}}$

$$\langle \mathcal{O} \rangle_{\beta'} \equiv \frac{\langle w \mathcal{O} \rangle_{\beta_0}}{\langle w \rangle_{\beta_0}}$$

$\tilde{\beta}' = \beta_0 + \varepsilon$

weights

$e^{S_\beta - S_{\tilde{\beta}}} = 1 + (\ldots)\varepsilon$

$$\frac{\langle e^{S_{\beta_0} - S_{\tilde{\beta}'}} \mathcal{O} \rangle}{\langle e^{S_{\beta_0} - S_{\tilde{\beta}'}} \rangle} = \langle \mathcal{O} \rangle + \varepsilon \left. \frac{\mathrm{d}}{\mathrm{d}\beta} \right|_{\beta = \beta_0} \langle \mathcal{O} \rangle$$

(forward) Automatic Differentiation

MCMC samples with S

$$\langle \mathcal{O} \rangle, \{\phi\}$$

$$J$$

weights

$$w \equiv e^{S-(S-J\phi)}$$

$$\langle \mathcal{O} \rangle_J \equiv \frac{\langle w\mathcal{O} \rangle}{\langle w \rangle}$$

$$\tilde{J} = 0 + \varepsilon$$

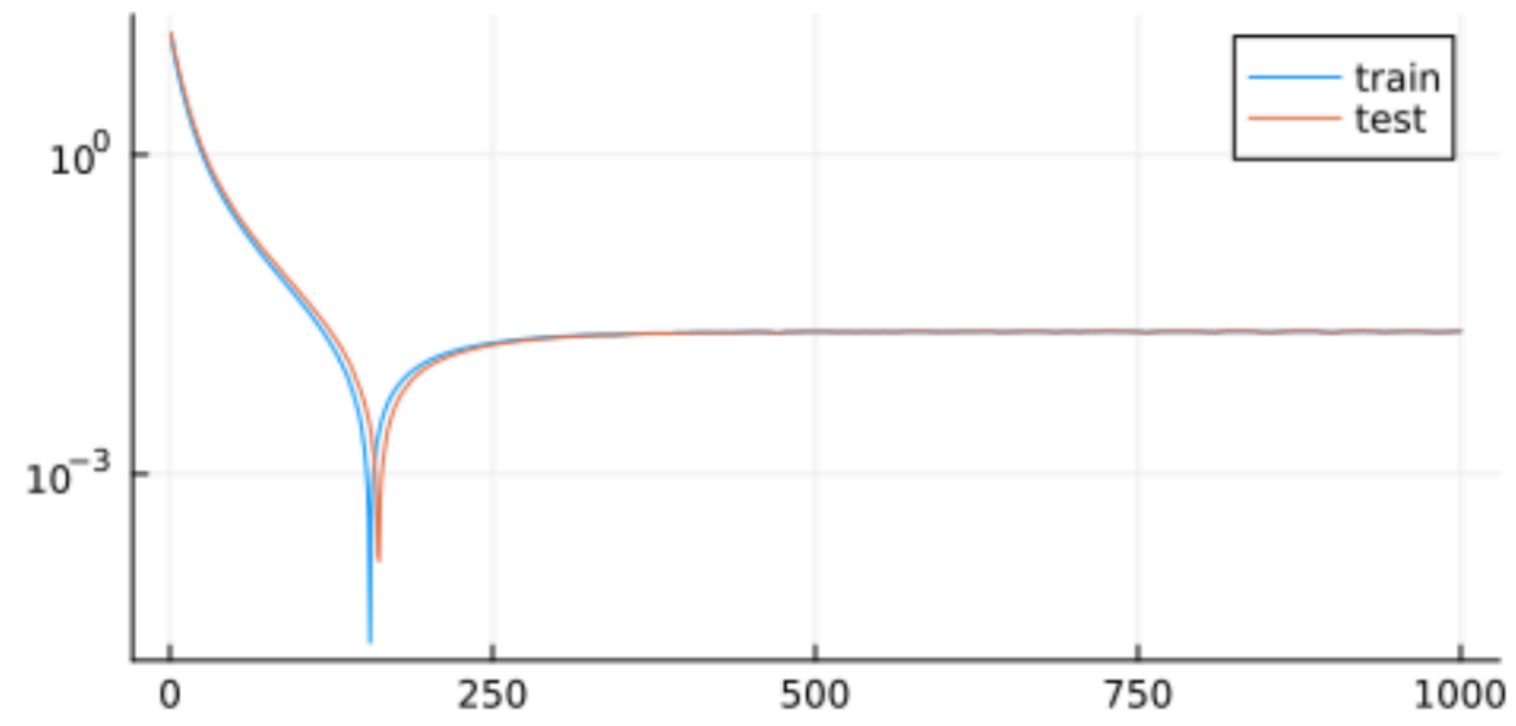
weights

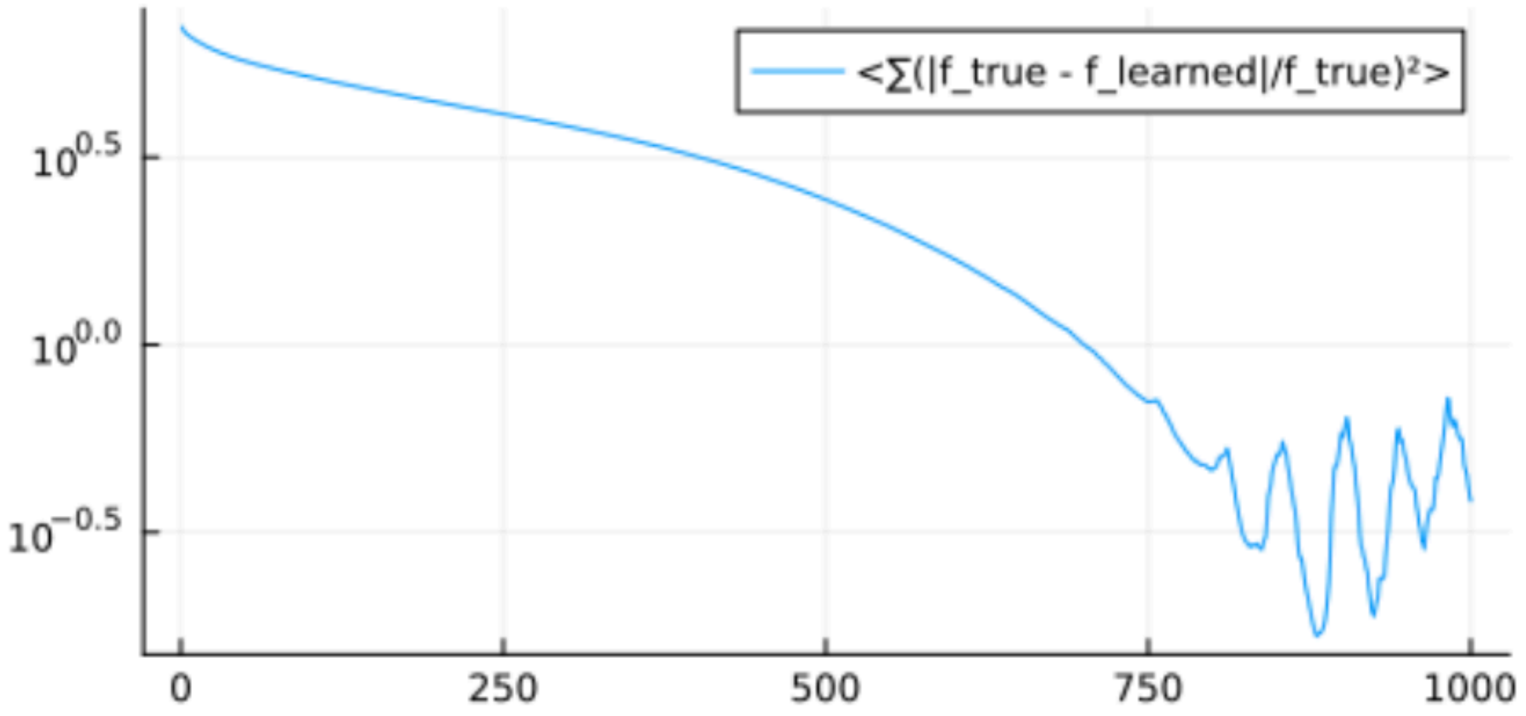
$$e^{S-S_{\tilde{J}}} = 1 + (\ldots)\varepsilon$$

$$\frac{\langle e^{S-S_{\tilde{J}}}\mathcal{O} \rangle}{\langle e^{S-S_{\tilde{J}}} \rangle} = \langle \mathcal{O} \rangle + \varepsilon \frac{\mathrm{d}}{\mathrm{d}J}\Big|_{J=0} \langle \mathcal{O} \rangle$$

Some results

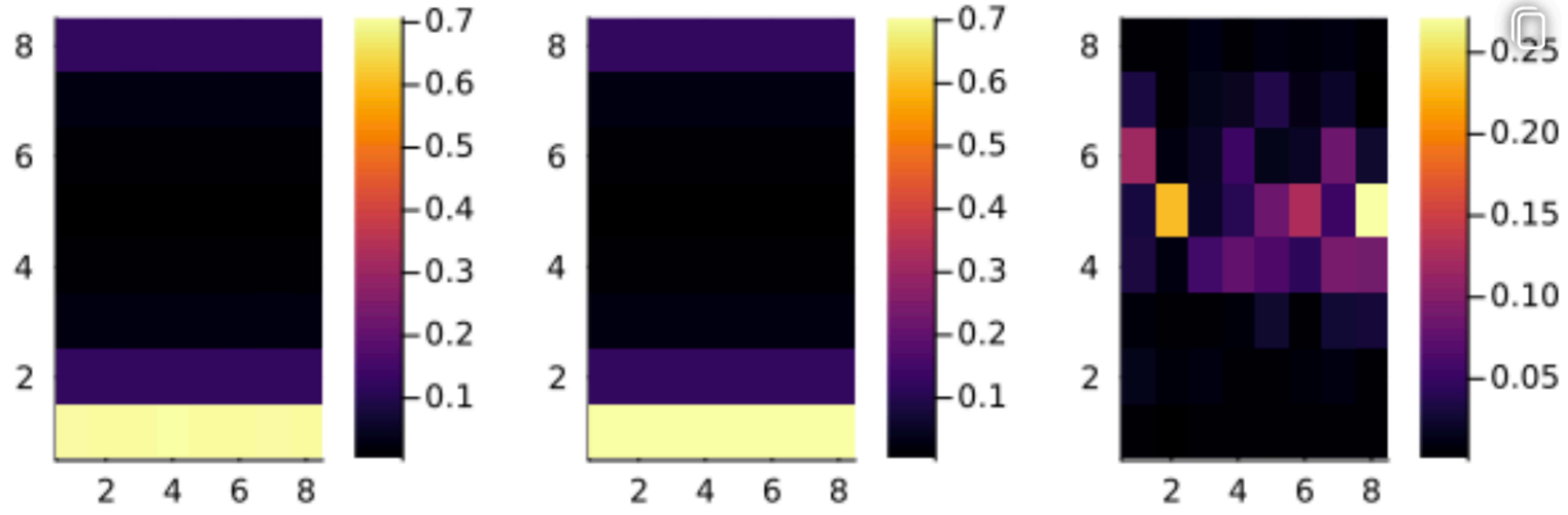
```
net(vol) = Chain(
    FlattenLayer(),
    Dense(prod(vol)=>prod(vol),σ),
    ReshapeLayer((vol...,1))
)
```

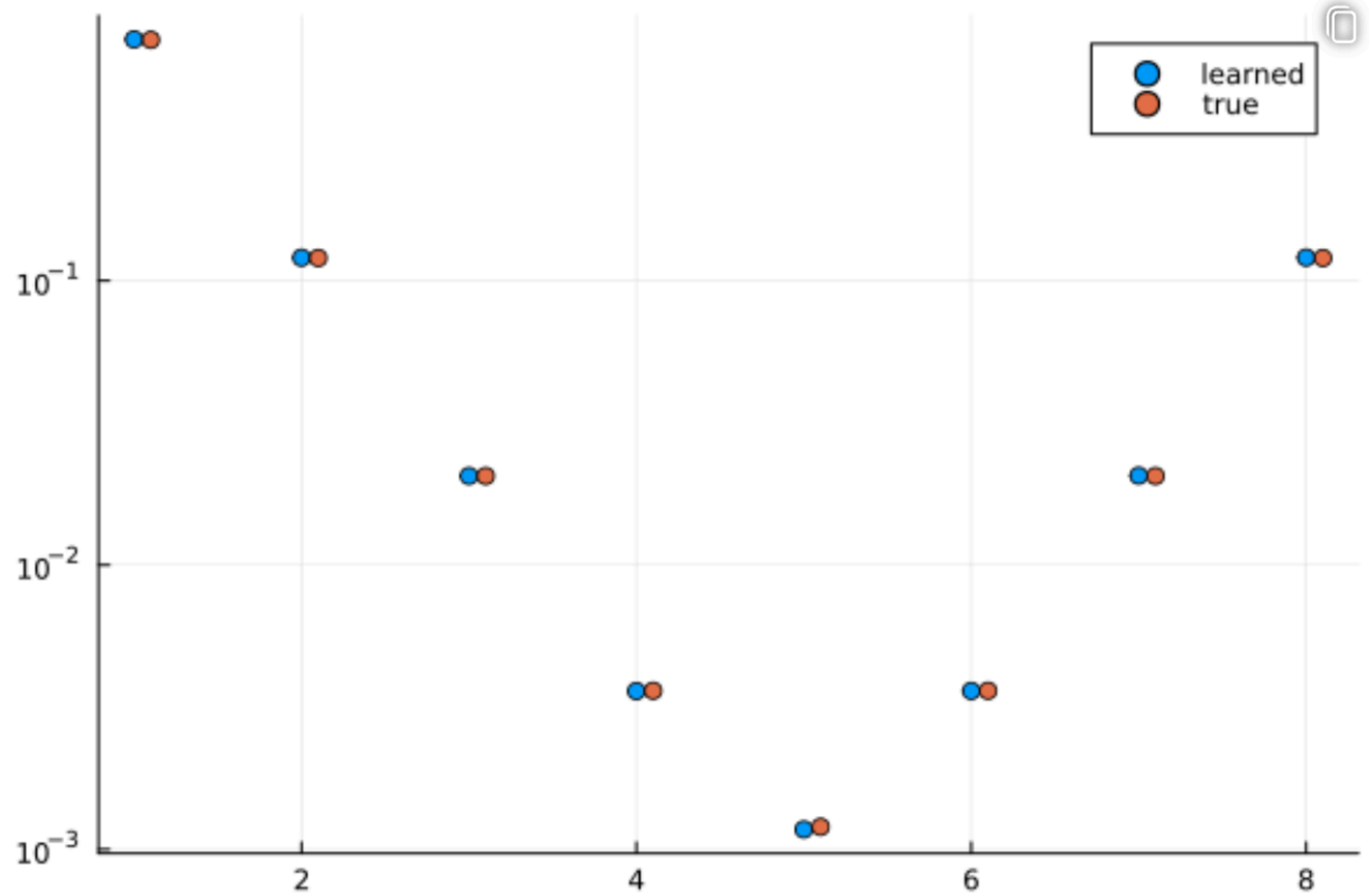




Vol: 8x8, 10K confs, ~30 sec

True - Learned =





Blue points have error bars!!!

Stochastic (H)AD

e^{-S}

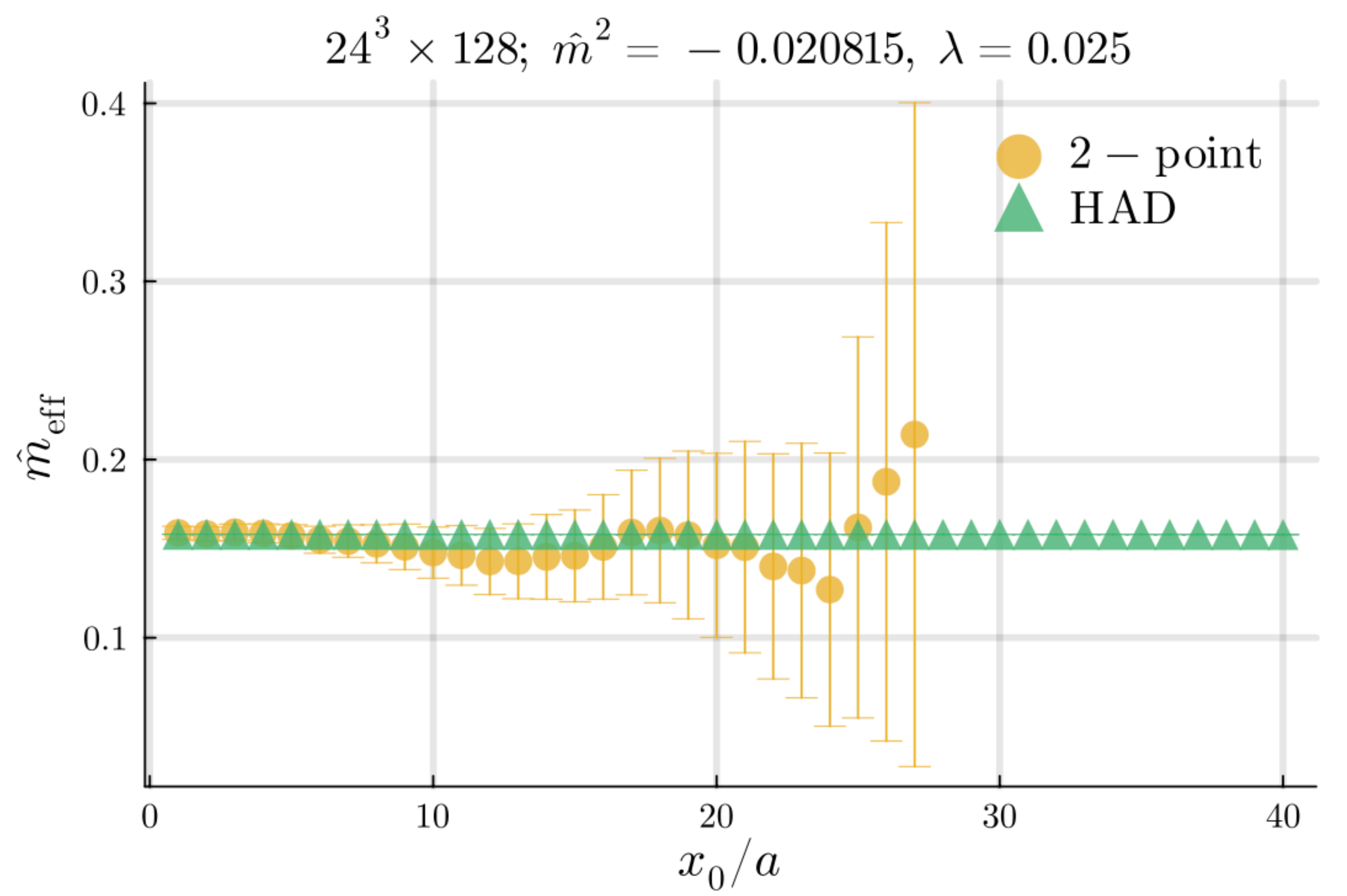
Base distribution

$e^{-S+J\phi}$

Target distribution

[Catumba, Ramos] - *Eur.Phys.J.C* 85 (2025) 9, 1037

[Catumba, Ramos, Zaldivar] - *Comput.Phys.Commun.* 307 (2025) 109396



[Catumba, Ramos] - *Eur.Phys.J.C* 85 (2025) 9, 1037

- Promote also fields and momenta to formal series

$$\phi \to \phi + \varepsilon \qquad \pi \to \pi + \varepsilon$$

- Integrate HMC equations order by order

$$\dot{\phi} = \tilde{\pi} \qquad \dot{\pi} = -\frac{\partial \tilde{S}_{\tilde{J}}}{\partial \tilde{\phi}}$$

- Samples now contains the dependence on the source

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

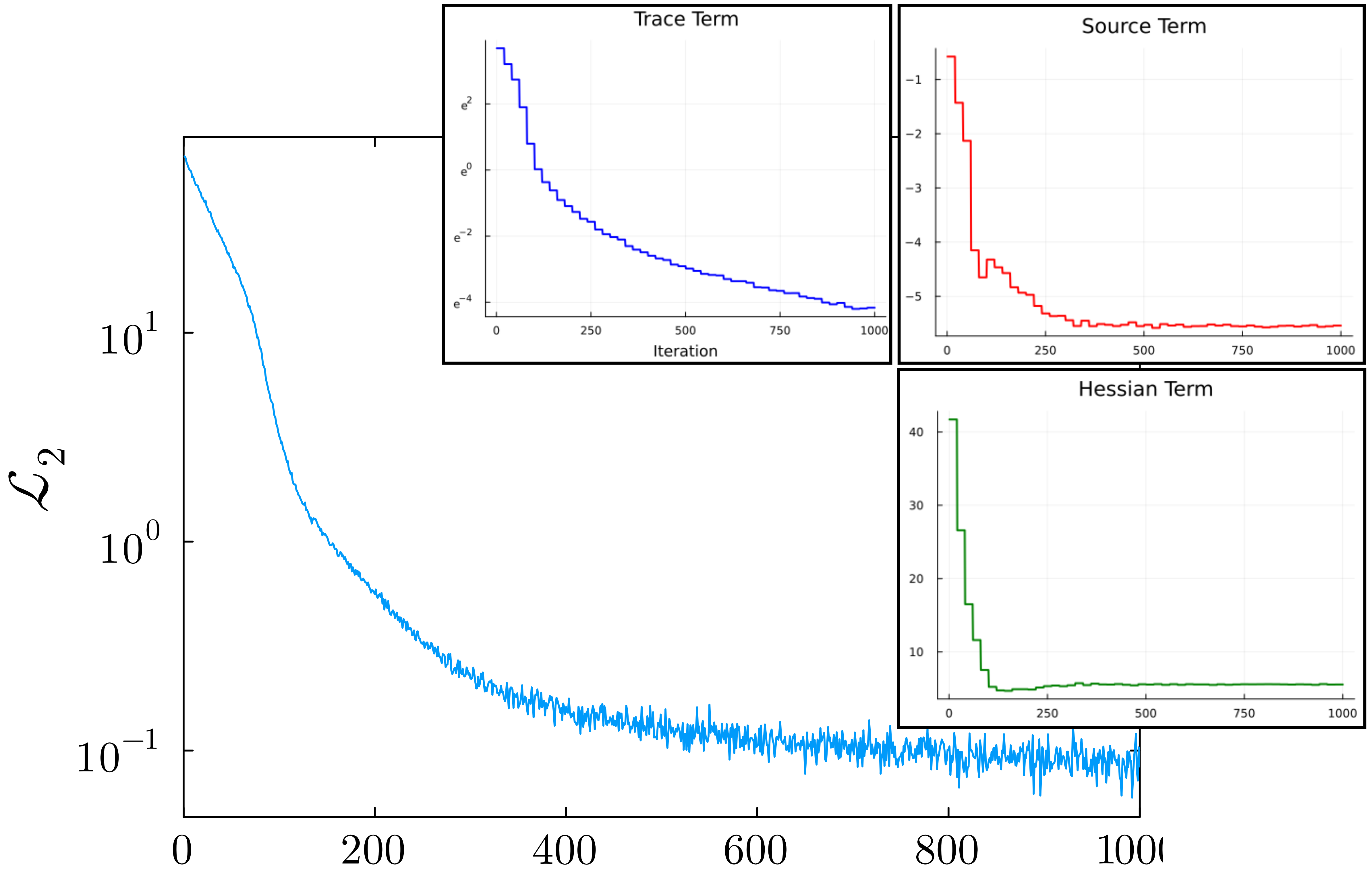
…but exact convergence only for $\partial^2 S > 0$ theories

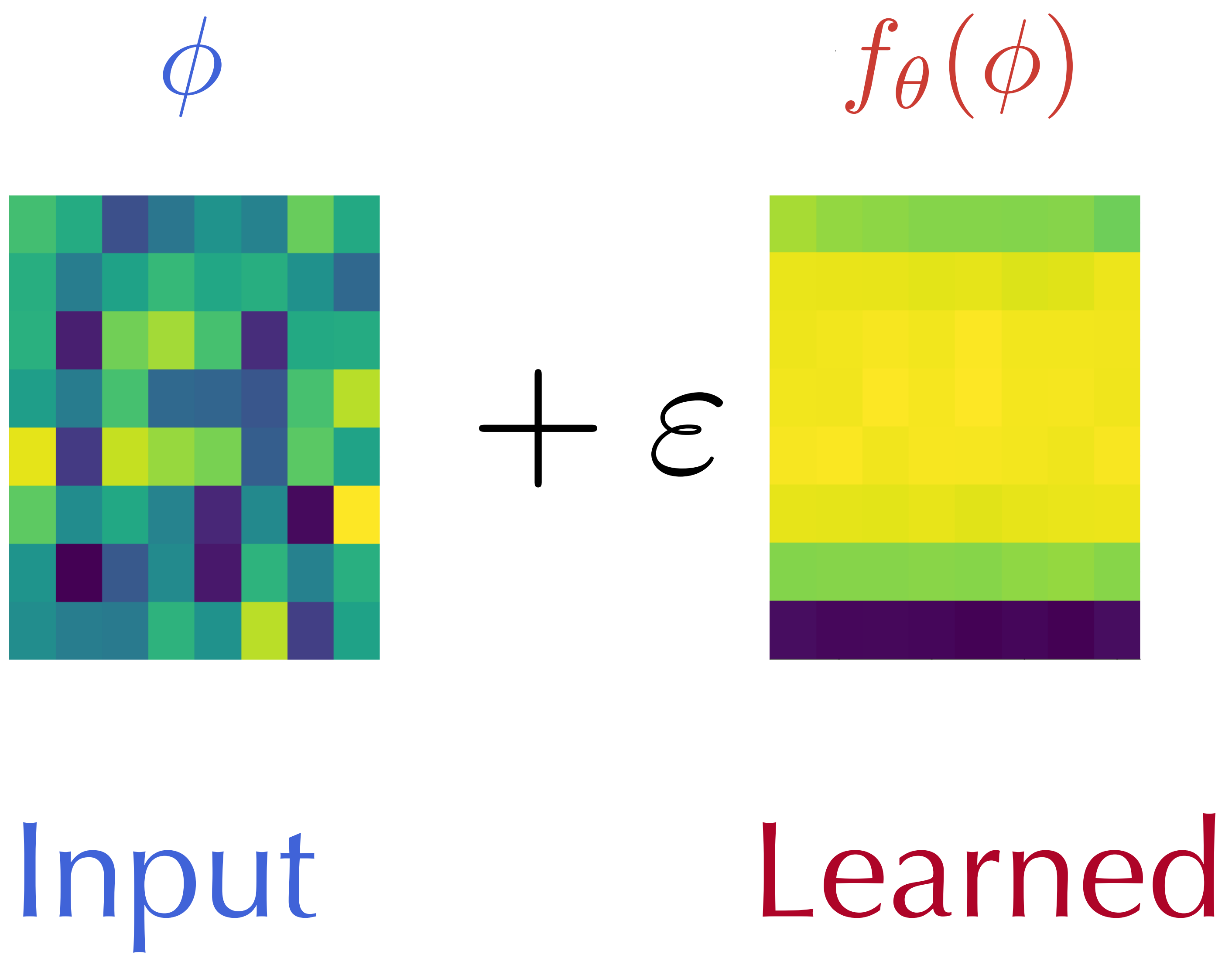
SDU

ħQTC

Some preliminary result

$$\mathcal{L}[f] = \mathbb{E}_r[f] + \frac{1}{2}\mathbb{E}_r\left[f^{\perp} H_S f + \|\nabla f\|_F^2\right]$$





Conclusions

- The signal-to-noise degradation problem can be formulated as a **reweighing error** (overlap problem) in the computation of **derivatives of one-point functions**
- Stochastic Automatic Differentiation allows for the computation of **exact derivatives** of expectation values through reweighing with formal series
- A residual transformation of the configurations allows for systematic improvement of the signal-to-noise ratio in scalar theories
- Linear flows naturally frames the improvement procedure within a geometrical picture in the context of transport maps

Linearised flow

Target distribution

$$e^{-(S+\varepsilon\mathcal{O})}$$

$$q_\theta(\phi)$$

for a generic f
(KL not minimal)

$$\mathcal{O} - \langle\mathcal{O}\rangle_r = -\mathcal{T}_r f + \mathcal{O}_\perp$$

$$\mathrm{Var}_r[\mathcal{O}] = \langle(\mathcal{T}_r f)^2\rangle + \mathrm{Var}_r[\mathcal{O}_\perp]$$

Geometric noise

Parisi-Lepage

Infinitesimal transport

$$\phi \longrightarrow \phi + \varepsilon f_\theta(\phi)$$

$$e^{-S}$$

Base distribution

Automatic Differentiation

- Extend parameters to power series at $\mathcal{O}(\varepsilon^N)$

$$\tilde{x} = x_0 + x_1\varepsilon + x_2\varepsilon^2$$

- Hard-code every elementary operation

$$\tilde{x}\tilde{y} = x_0y_0 + (x_0y_1 + x_1y_0)\varepsilon + (x_0y_2 + x_1y_2 + x_2y_0)\varepsilon^2$$

$$\exp(\tilde{x}) = e^{x_0} + e^{x_0}x_1\varepsilon + e^{x_0}\left(\frac{x_1^2}{2} + x_2\right)\varepsilon^2$$

$$\sin(\tilde{x}) = \dots$$

- Exact derivatives of any function f at x_0 are given by Taylor's theorem

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

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

```
julia> using FormalSeries

julia> f(x) = 1/(1-x)

julia> f(0.)
 1.0

julia> x̃ = Series((1., 1.))
 Series{Float64,2}((1.0,1.0))

julia> f(x̃)
 Series{Float64,2}((1.0,1.0))
```

Linearised flow

Target distribution

$$e^{-(S+\varepsilon\mathcal{O})}$$

$$q_\theta(\phi)$$

$$\mathrm{KL}(q_\varepsilon \| p_J) = \left(\frac{1}{2}\mathbb{E}_r[(\mathcal{T}_r f)^2] - \mathbb{E}_r[f \cdot \nabla\mathcal{O}]\right)\varepsilon^2$$

Minimize $\mathcal{L}_2[f]$

Infinitesimal transport

$$\phi \longrightarrow \phi + \varepsilon f_\theta(\phi)$$

$$\mathcal{O} - \langle\mathcal{O}\rangle_r = -\mathcal{T}_r f$$

Stein-Poisson equation for optimal transport

$$e^{-S}$$

Base distribution

$$\mathrm{Var}_r[\mathcal{O}] = \langle(\mathcal{T}_r f)^2\rangle$$

Conclusions

- The signal-to-noise degradation problem can be formulated as a **reweighing error** (overlap problem) in the computation of **derivatives of one-point functions**
- Stochastic Automatic Differentiation allows for the computation of **exact derivatives** of expectation values through reweighing with formal series
- A residual transformation of the configurations allows for systematic improvement of the signal-to-noise ratio in scalar theories
- Linear flows naturally frames the improvement procedure within a geometrical picture in the context of transport maps