

A variational framework for variance reduction in lattice field theory

Pietro Butti

with Guilherme Catumba, Alessandro Nada, Louis Spatscheck



Lattice2026 - Maryland
College Park, Jul 31st, 2026

$$\mathcal{O}_t = \sum_{\vec{x}} \phi(t, \vec{x})$$

Computing the correlator

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

Connected 2-point function

$$C(t) = \langle \mathcal{O}_t \mathcal{O}_0 \rangle_r - \langle \mathcal{O} \rangle_r^2$$

Monte Carlo average over base Boltzmann distribution with action S

$$r = \frac{1}{Z} e^{-S}$$

Derivative of 1-point function

$$C(t) = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \langle \mathcal{O}_t \rangle_{r_\varepsilon}$$

MC average with respect to a distr. with action deformed by a source term

$$r_\varepsilon = \frac{1}{Z_\varepsilon} e^{-S_\varepsilon} \quad \text{with} \quad S_\varepsilon = S - \varepsilon \mathcal{O}_0$$

[Abbot et al.] - 2606.15986

[Abbot et al.] - PRD 114 (2026)

$$\mathcal{O}_t = \sum_{\vec{x}} \phi(t, \vec{x})$$

In practice, one computes the weights

$$w(\phi) = \frac{e^{\varepsilon \mathcal{O}_0(\phi)}}{\langle e^{\varepsilon \mathcal{O}_0} \rangle} = 1 + \varepsilon (\mathcal{O}_0 - \langle \mathcal{O}_0 \rangle) + \dots$$

...and reweight a 1 point function

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

Fluctuation in
the weights

Signal-to-noise
degradation

Computing the correlator

Derivative of 1-point function

$$C(t) = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \langle \mathcal{O}_t \rangle_{r_\varepsilon}$$

MC average with respect to a distr. with
action deformed by a source term

$$r_\varepsilon = \frac{1}{Z_\varepsilon} e^{-S_\varepsilon} \quad \text{with} \quad S_\varepsilon = S - \varepsilon \mathcal{O}_0$$

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

[Abbot et al.] - 2606.15986

[Abbot et al.] - PRD 114 (2026)

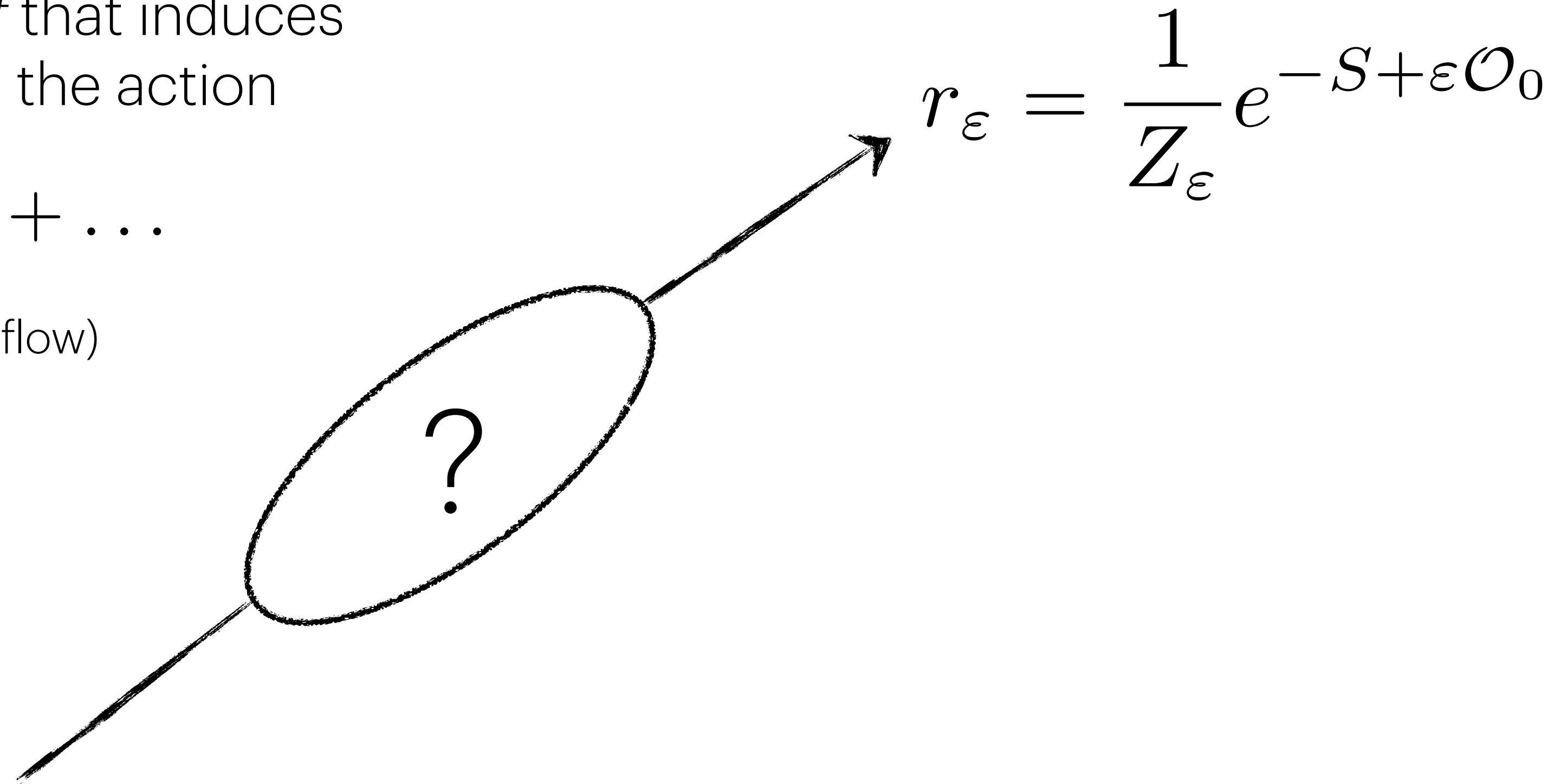
Source deformations

Goal: find the vector field f that induces a source deformation in the action

$$\phi \rightarrow \phi + \varepsilon f(\phi) + \dots$$

(or more generally with a flow)

$$r = \frac{1}{Z} e^{-S}$$


$$r_\varepsilon = \frac{1}{Z_\varepsilon} e^{-S + \varepsilon \mathcal{O}_0}$$

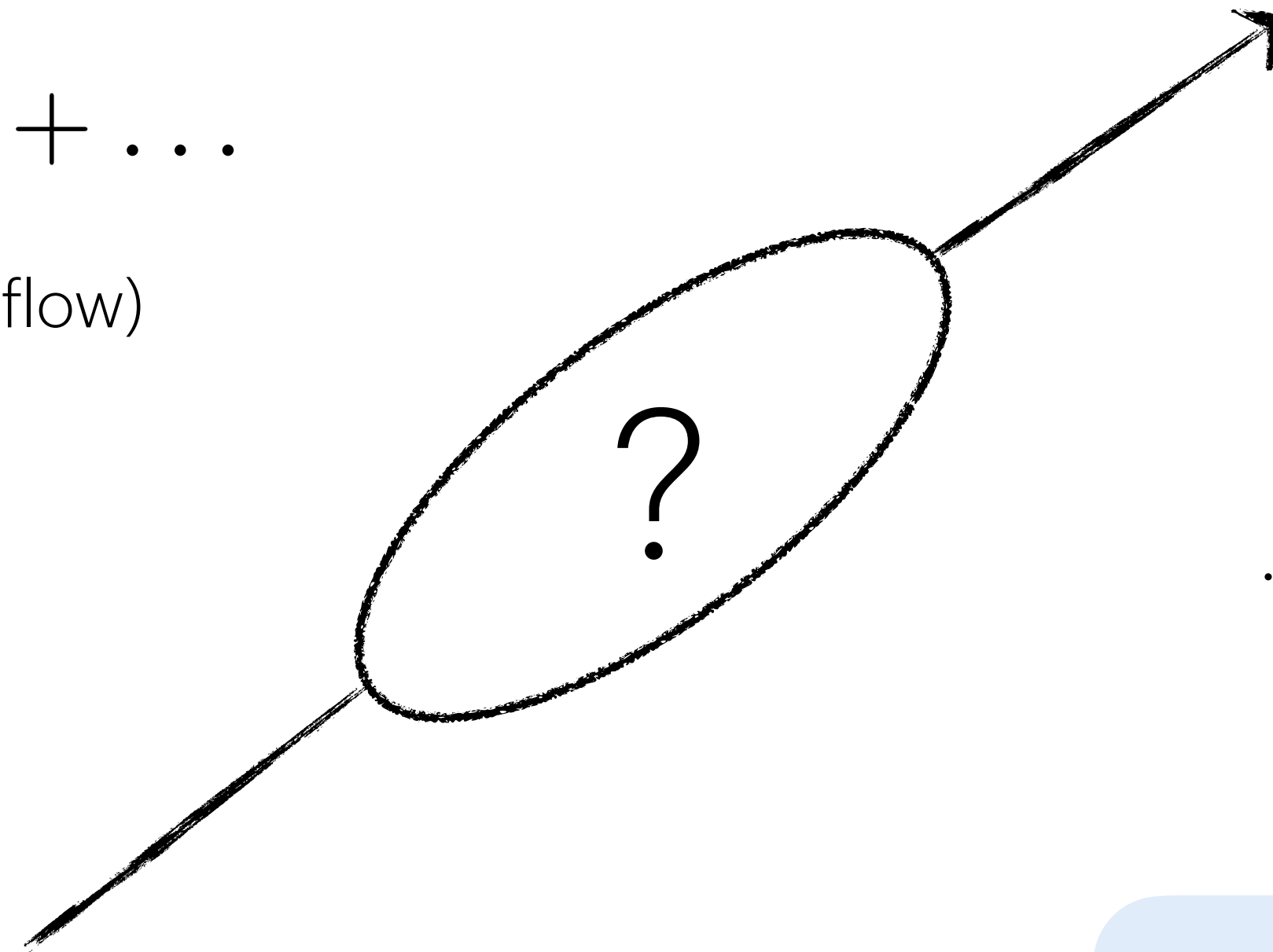
Source deformations

Goal: find the vector field f that induces a source deformation in the action

$$\phi \rightarrow \phi + \varepsilon f(\phi) + \dots$$

(or more generally with a flow)

$$r = \frac{1}{Z} e^{-S}$$



$$r_\varepsilon = \frac{1}{Z_\varepsilon} e^{-S + \varepsilon \mathcal{O}_0}$$

...aka, find the flowed distribution by minimising the KL divergence (variance of the weights)

$$\text{KL}(q \| r_\varepsilon) = \frac{1}{2} \text{Var}[\tilde{w}] + \mathcal{O}(\langle (\tilde{w} - 1)^3 \rangle)$$

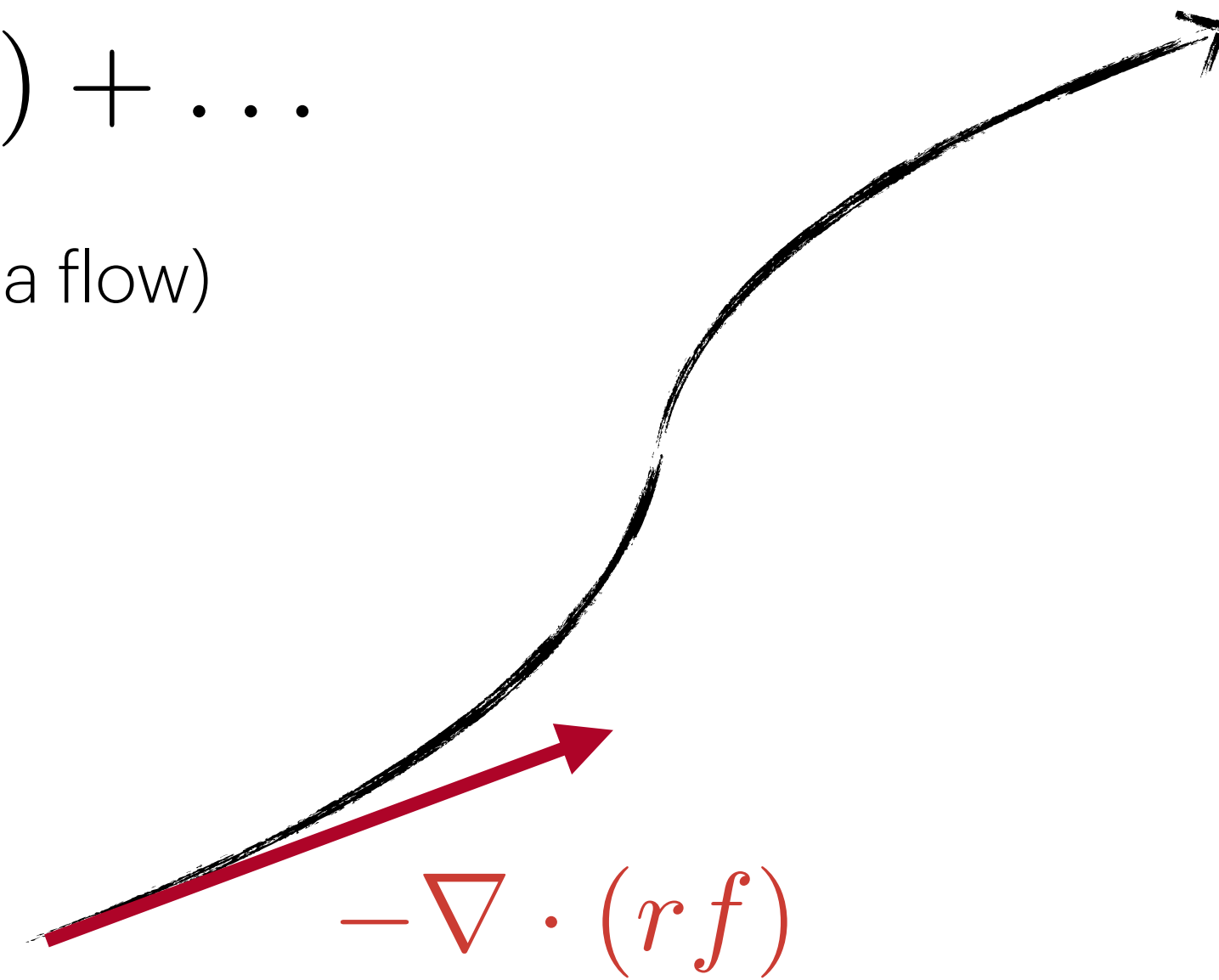
Source deformations

Goal: find the vector field f that induces a source deformation in the action

$$\phi \rightarrow \phi + \varepsilon f(\phi) + \dots$$

(or more generally with a flow)

$$r = \frac{1}{Z} e^{-S}$$



$$r_\varepsilon = \frac{1}{Z_\varepsilon} e^{-S + \varepsilon \mathcal{O}_0}$$

...aka, find the flowed distribution by minimising the KL divergence (variance of the weights)

$$\text{KL}(q \| r_\varepsilon) = \frac{1}{2} \text{Var}[\tilde{w}] + \mathcal{O}(\langle (\tilde{w} - 1)^3 \rangle)$$

$$\frac{d\phi_\varepsilon}{d\varepsilon} = f(\phi_\varepsilon) \quad \phi_0 \equiv \phi$$

The Stein operator

Stein operator: generator of the action deformation under the flow

$$\frac{dS_\varepsilon}{d\varepsilon} = -\mathcal{T}_r f$$

Stein's identity

Conservation of probability along the flow

$$\langle \mathcal{T}_r f \rangle = \int D\phi \nabla \cdot (r f) = 0$$

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

Hutchinson
O(V) scaling

HMC force

Define a Riemannian length over the manifold of probability densities

$$\langle (\mathcal{T}_r f)^2 \rangle \equiv \|f\|_{\mathcal{T}_r}^2$$

The KL expansion

Given a vector field $f(\phi)$, evolve the configurations via flow equation in time ε

$$\frac{d\phi_\varepsilon}{d\varepsilon} = f(\phi_\varepsilon) \quad \text{with} \quad \phi_0 \equiv \phi \sim r$$

The Lie derivative along the flow $L = f \cdot \nabla$ dictates the evolution of the Jacobian

$$\log \det \mathcal{J}(\varepsilon) = \sum_{n \geq 1} \frac{\varepsilon^n}{n!} (\mathcal{L}^{n-1} \nabla \cdot f)(\phi)$$

The KL divergence expansion reads

$$\text{KL}(q \| r_\varepsilon) = \text{KL}(r \| r_\varepsilon) - \sum_{n \geq 1} \frac{\varepsilon^n}{n!} \langle \mathcal{L}^{n-1} (\mathcal{T}_{r_\varepsilon} f) \rangle$$

$$\text{KL}(q \| p) = \left\langle \log \frac{p}{r} \right\rangle - \varepsilon \langle \mathcal{T}_p f \rangle + \frac{\varepsilon^2}{2} \langle (\mathcal{T}_p f)(\mathcal{T}_r f) \rangle + \mathcal{O}(\varepsilon^3)$$

Controlled by the Stein operator

$$\mathcal{T}_r f \equiv \text{tr} J_f - f \cdot \nabla S$$

For source deformation:

$$\text{KL}[f] = \frac{\varepsilon^2}{2} \left\langle (\text{tr} J_f - f \cdot \nabla S)^2 - 2f \cdot \nabla \mathcal{O}_0 \right\rangle + \mathcal{O}(\varepsilon^3)$$

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 change of variable that minimises the KL divergence

$$\phi \rightarrow \phi + \varepsilon f(\phi) + \dots \quad \text{s.t.} \quad \min_f \text{KL}(q_f || r_\varepsilon)$$

Expanding KL divergence in ε , reduces the problem to minimise

$$\text{KL}[f] = \frac{\varepsilon^2}{2} \left\langle (\text{tr} J_f - f \cdot \nabla S)^2 - 2f \cdot \nabla \mathcal{O}_0 \right\rangle + \mathcal{O}(\varepsilon^3)$$

Find field $f(\phi)$ that minimises the following scalar functional

If you're not convinced, do this

$$\text{KL}[f] = \left\langle (\text{tr} J_f - f \cdot \nabla S)^2 - 2f \cdot \nabla \mathcal{O}_0 \right\rangle$$

Stein operator
 $\mathcal{T}_r f \equiv \text{tr} J_f - f \cdot \nabla S$

Solve Euler-Lagrange equation of KL

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

“Exact” solution in terms of the stochastic resolvent of the theory

Direct ML minimisation:

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

Reweighting method gives the following improved estimator

Once you're done, compute this

$$\hat{A}_t = \mathcal{O}_0 \mathcal{O}_t + (\mathcal{T}_r f) \mathcal{O}_t + f_t$$

$$\langle \hat{A}_t \rangle \equiv C(t)$$

(ensured by Stein identity)

The variance of the improved estimator can be easily decomposed as

$$\text{Var}[\hat{A}_t] = \sigma_C^2 + \text{Var}[(\mathcal{T}_r f) \mathcal{O}_t + f_t] - \langle \mathcal{O}_t^2 f_0 + \mathcal{O}_0 \mathcal{O}_t f_t \rangle$$

2pt func. Variance
(StN degradation)

Reduced during
minimisation

Pure variance
reduction term

Stochastic approach

Stein-Poisson equation

Euler-Lagrange equation for the KL divergence gives the Stein-Poisson equation

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

fluctuation at t=0
 $\Delta \mathcal{O}_0 \equiv \mathcal{O}_0 - \langle \mathcal{O}_0 \rangle$

Stein operator

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

Restrict to gradient fields $f = \nabla g$

$$(\underbrace{\Delta_2 - \nabla S \cdot \nabla}_{\text{Generator of (overdamped) Langevin dynamics}})g + \Delta \mathcal{O}_0 = 0$$

Generator of (overdamped)
Langevin dynamics

The Stein-Poisson equation has a solution in via the stochastic resolvent (*Feynman-Kac*)

$$g(\phi) = \int_0^\infty d\tau \left\langle \mathcal{O}_0[Z_\tau(\phi)] \right\rangle_\eta$$

A practical recipe

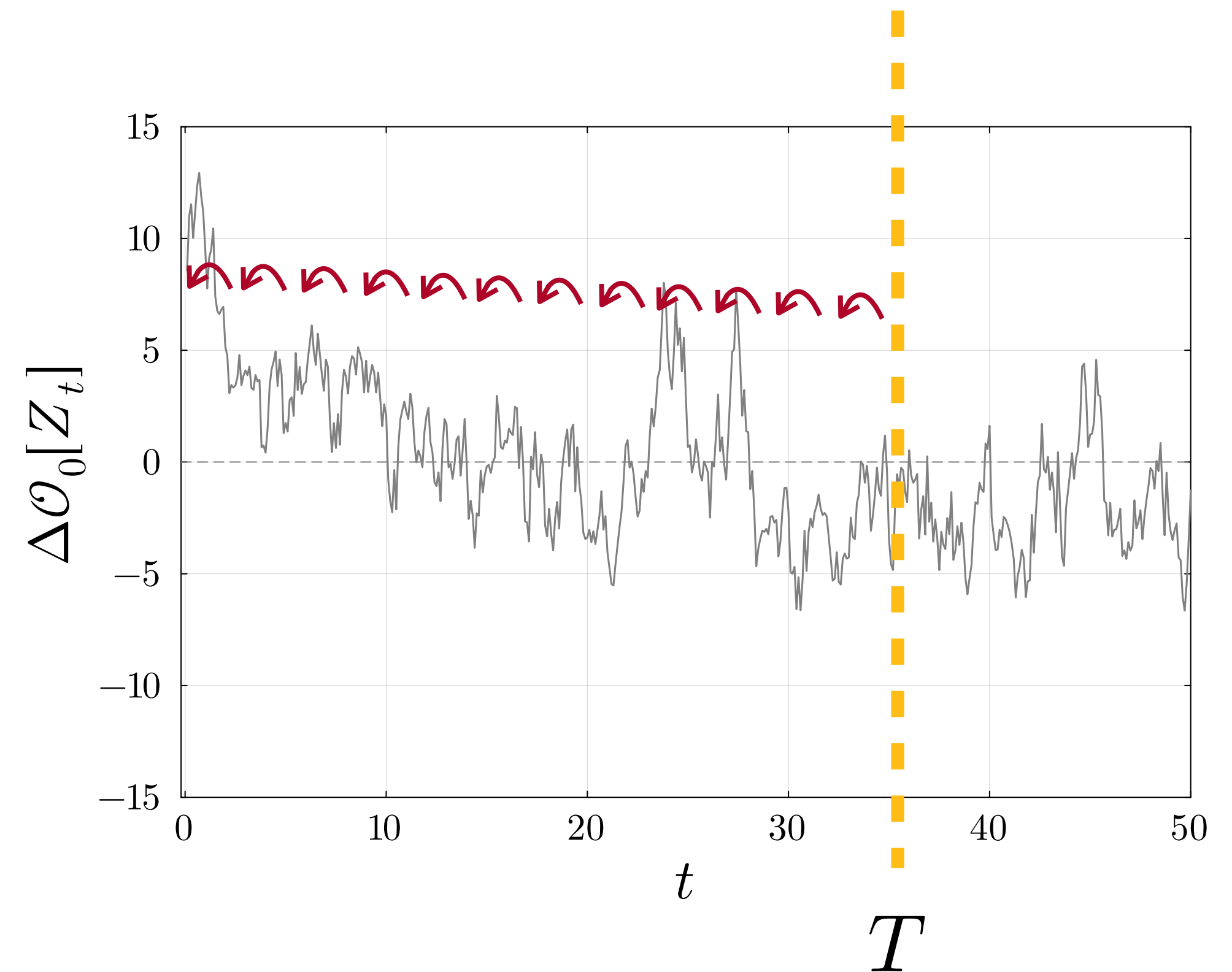
1. Integrate overdamped Langevin equation up to T

$$dZ_\tau = -\nabla S d\tau + \sqrt{2} dW_\tau \quad \text{with} \quad Z_0 = \phi$$

2. Integrate backward adjoint equation (deterministic)

$$\frac{dA_\tau}{d\tau} = -H_S[Z_{T-\tau}]A_\tau + \delta_{x_0,0} \quad \text{with} \quad A_0 = 0$$

3. Repeat for N_w walkers



$$f(\phi) = \lim_{T \rightarrow \infty} \langle A_T \rangle_\eta$$

Once you done, compute this

$$\hat{A}_t = \mathcal{O}_0 \mathcal{O}_t + (\mathcal{T}_r f) \mathcal{O}_t + f_t$$

Once you done, compute this

$$\hat{A}_t = \mathcal{O}_0 \mathcal{O}_t + (\mathcal{T}_r f) \mathcal{O}_t + f_t$$

Main signal.
Biased, no variance

Once you done, compute this

$$\hat{A}_t = \mathcal{O}_0 \mathcal{O}_t + (\mathcal{T}_r f) \mathcal{O}_t + f_t$$

Main signal.
Biased, no variance

$$\rho_T \mathcal{O}_t$$

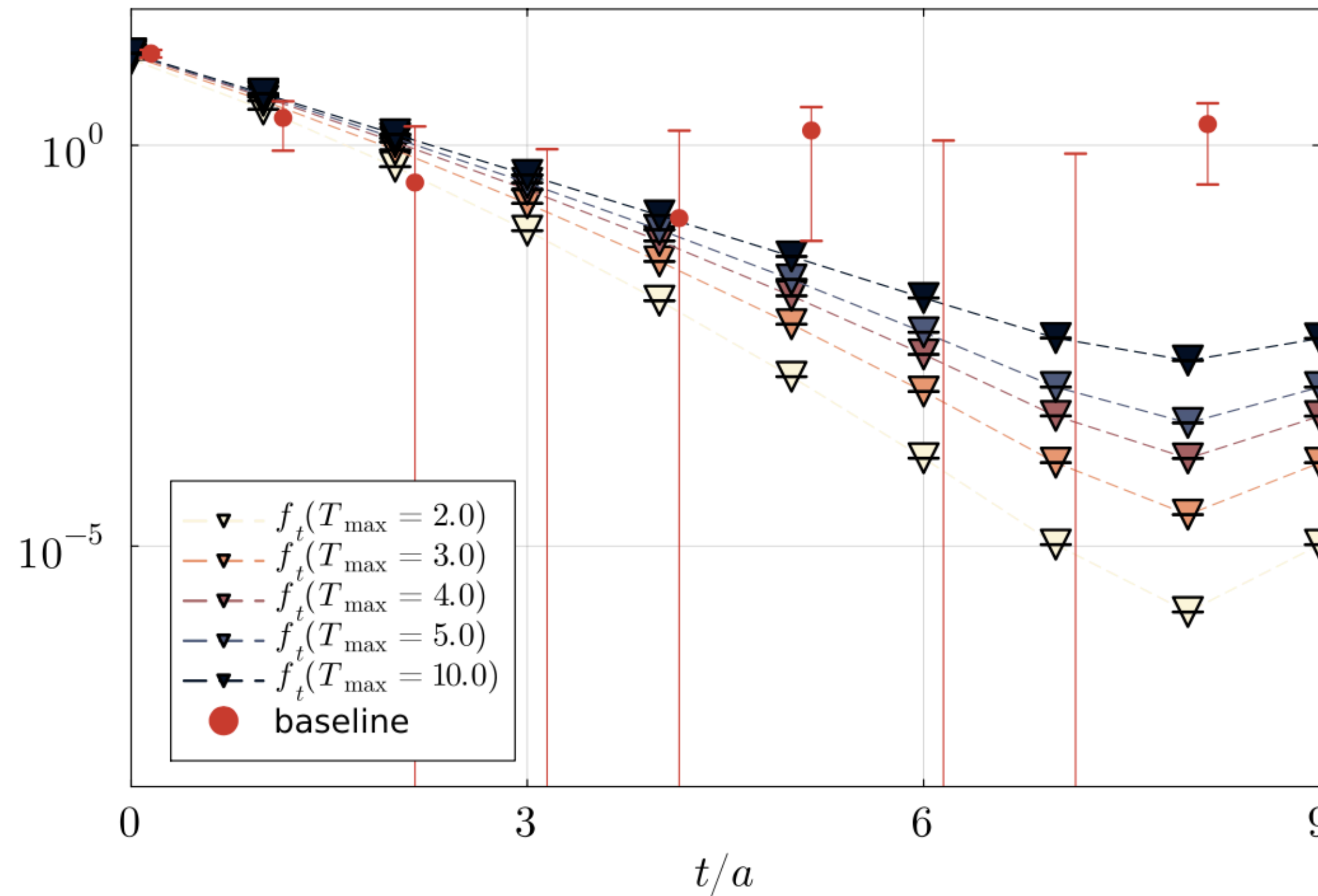
Bias. (residual Stein-Poisson)
controls the variance

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

$$\hat{A}_t = \rho_T \mathcal{O}_t + f_T$$

The biased estimator

$\lambda = 0.1, \kappa = 0.2, \text{stepsize} = 0.05$

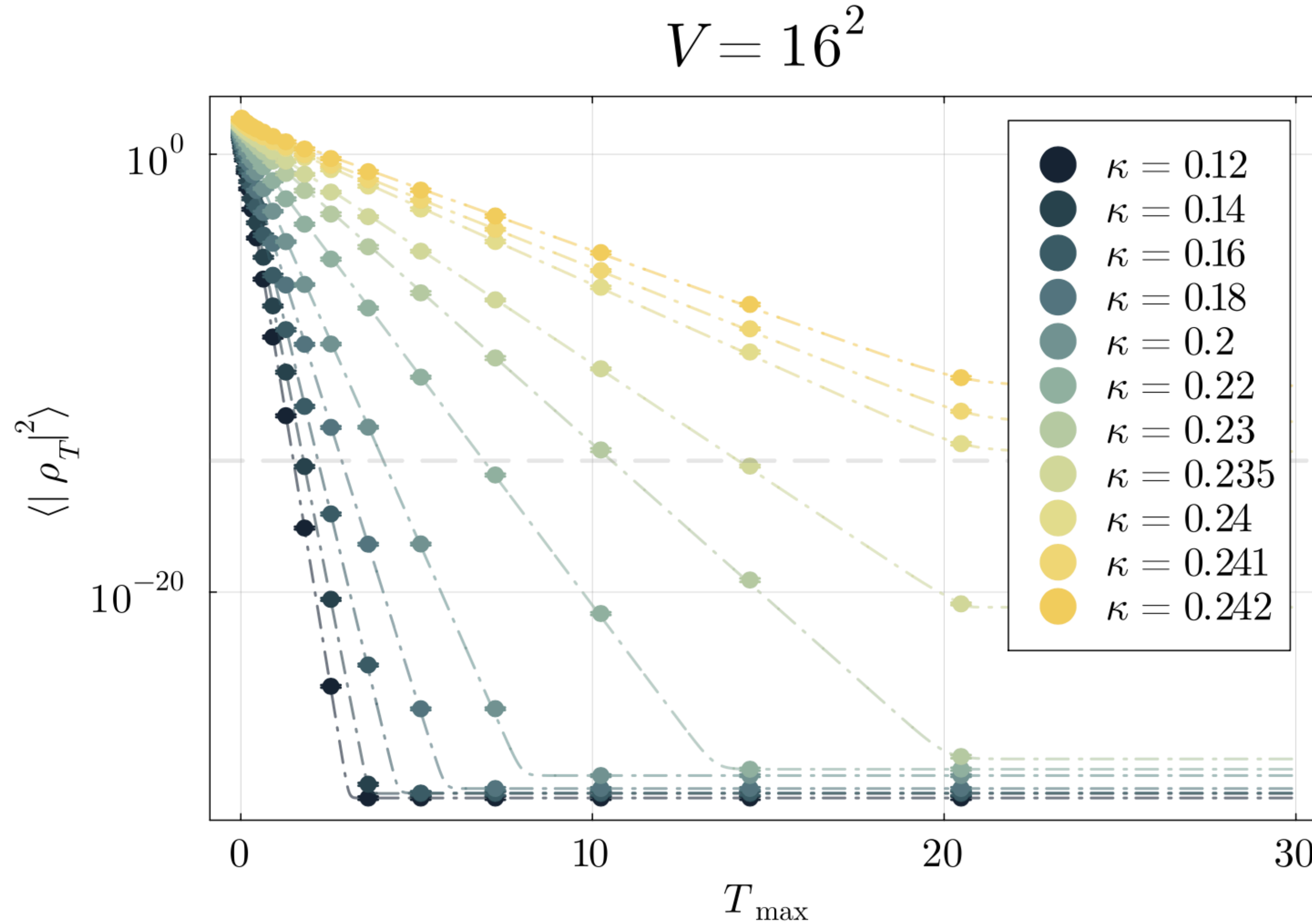


The estimator has no variance but is biased

The bias is proportional to T, t and the stepsize

$$\hat{A}_t = \rho_T \mathcal{O}_t + f_T$$

The bias

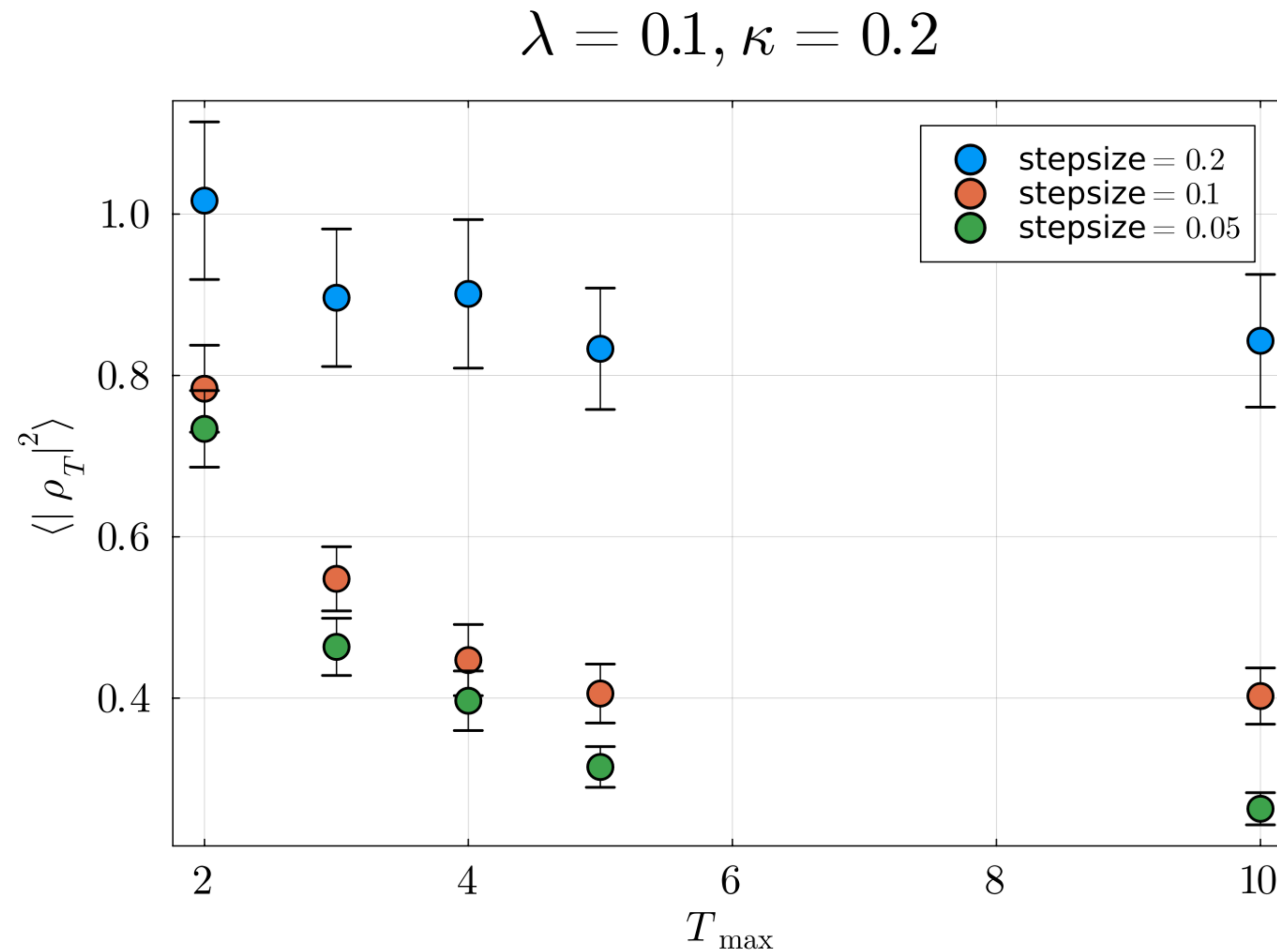


$$\langle \rho_T \mathcal{O}_t \rangle \sim e^{-\lambda_* T}$$

Spectral gap of the theory

$$\hat{A}_t = \rho_T \mathcal{O}_t + f_T$$

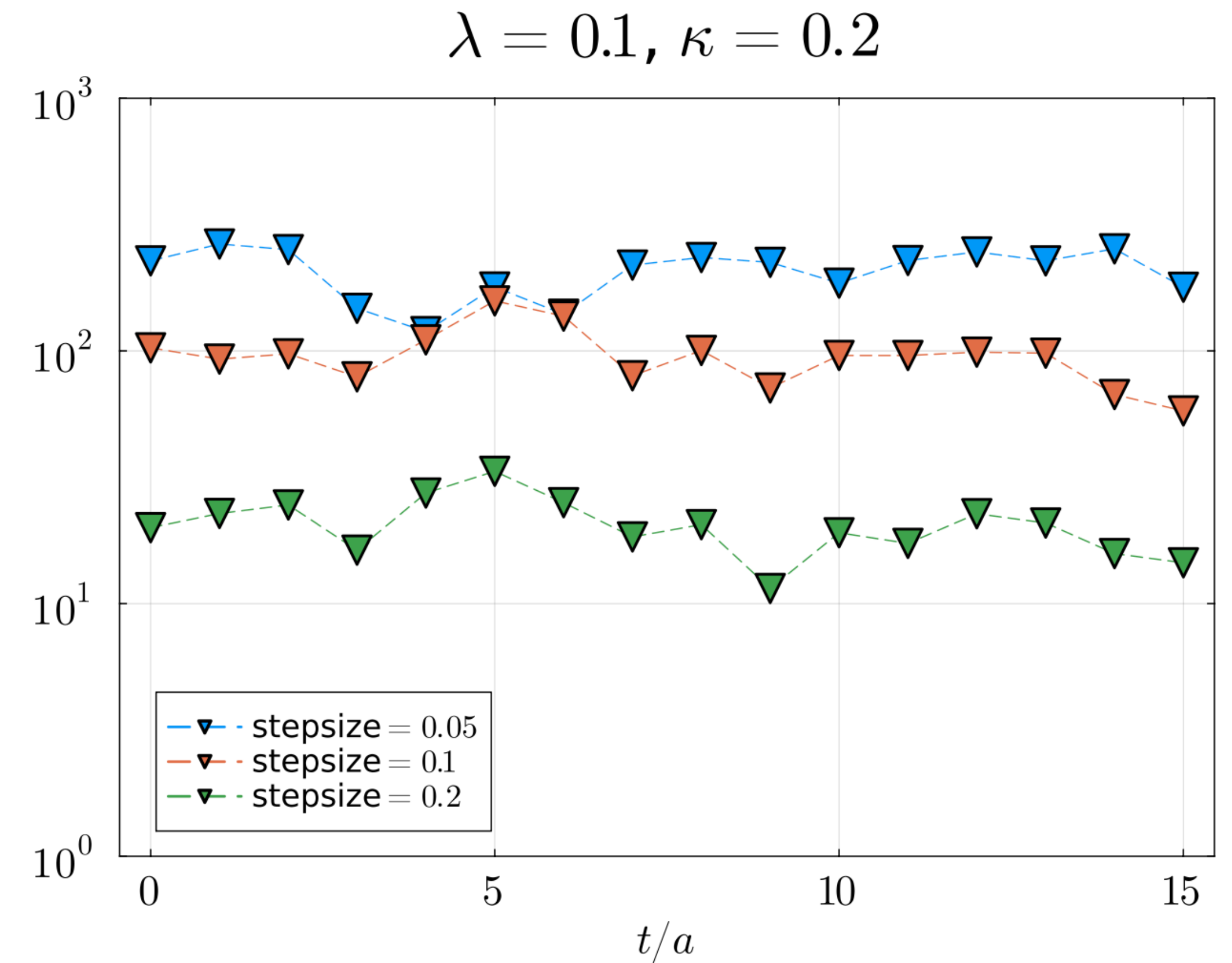
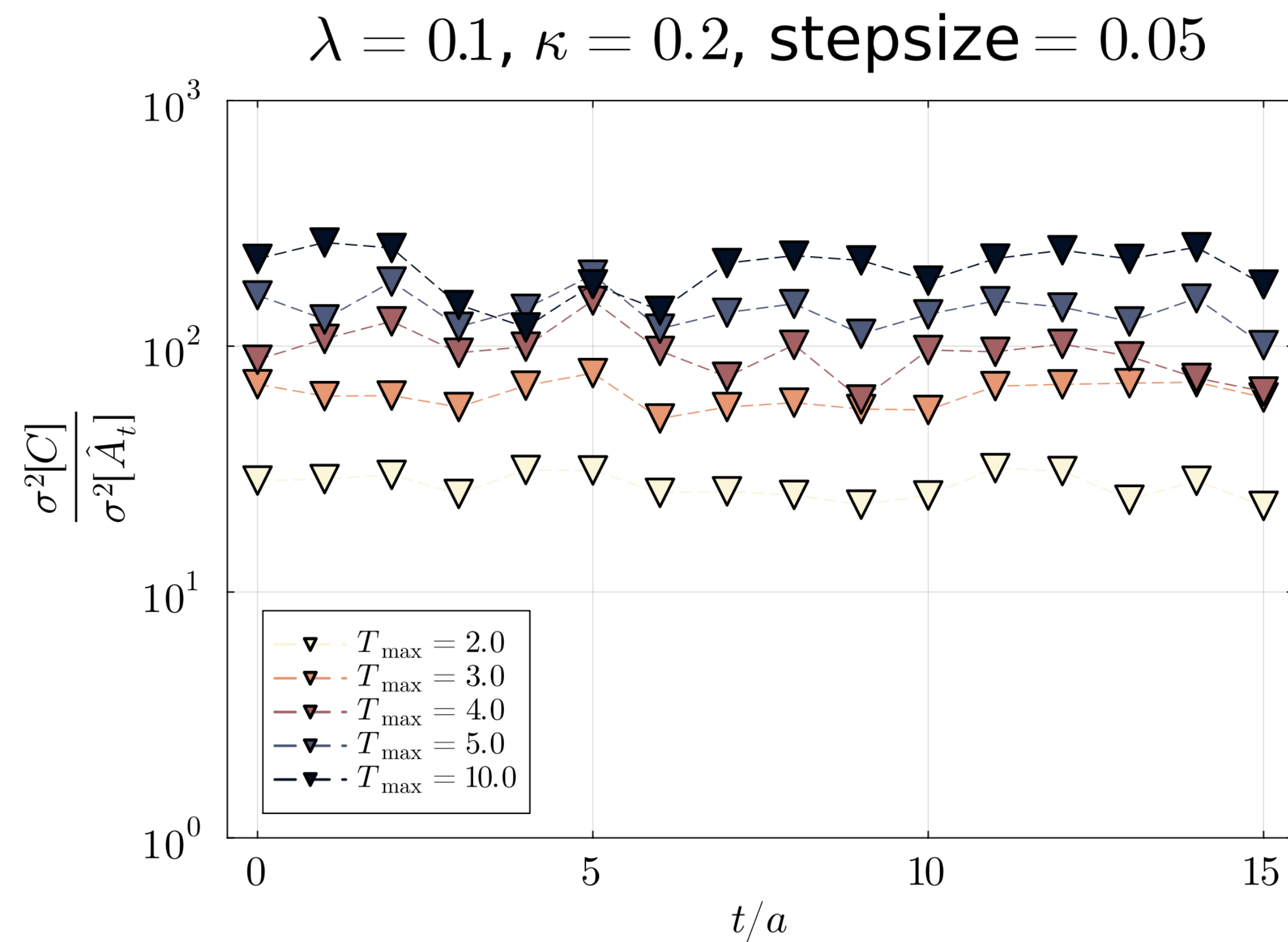
The bias



$$\langle \rho_T \mathcal{O}_t \rangle \sim e^{-\lambda_* T} + c(\delta\tau)$$

$$\hat{A}_t = \rho_T \mathcal{O}_t + f_T$$

Variance reduction



ML minimisation

Direct ML minimisation:

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

Training

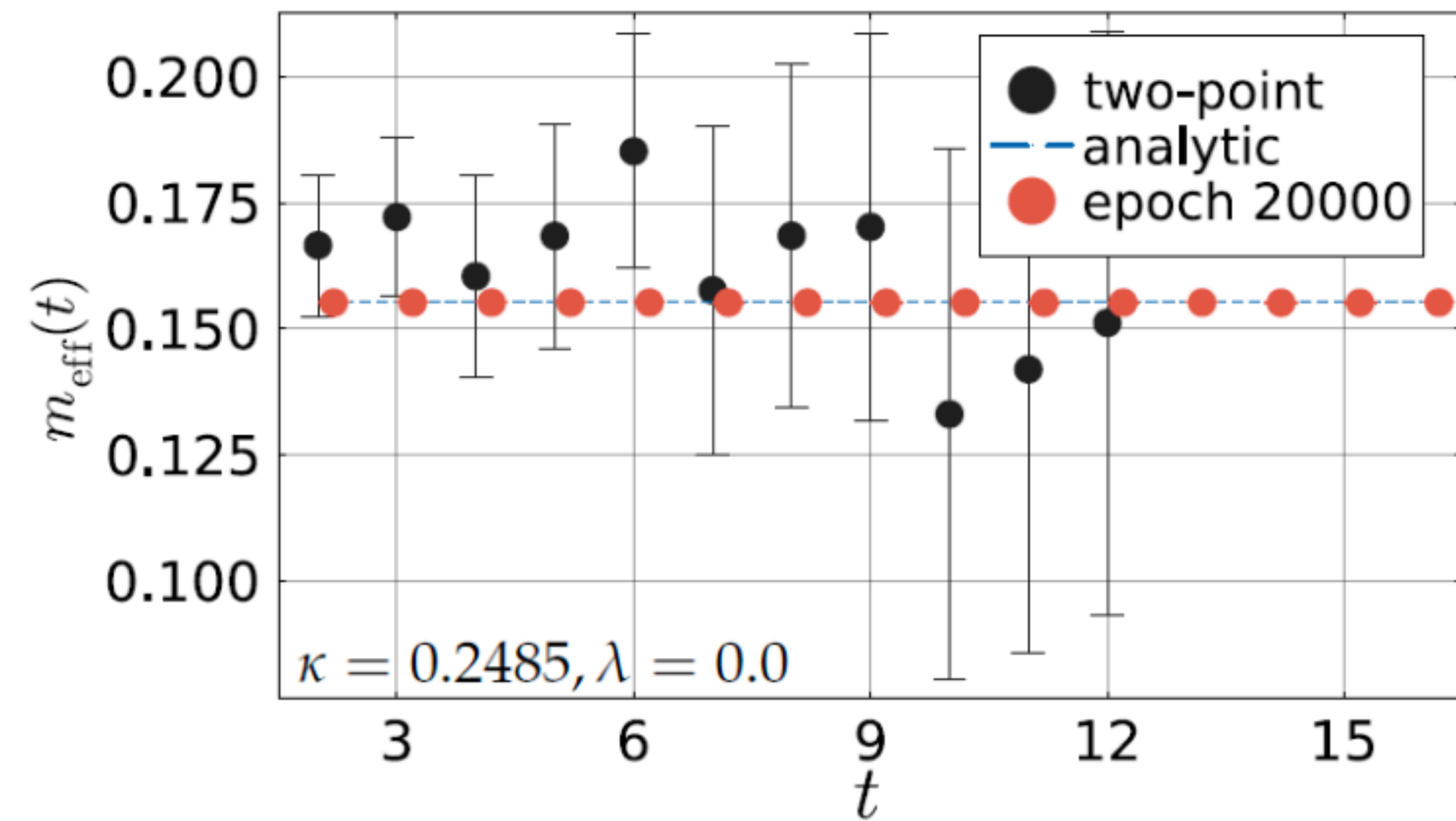
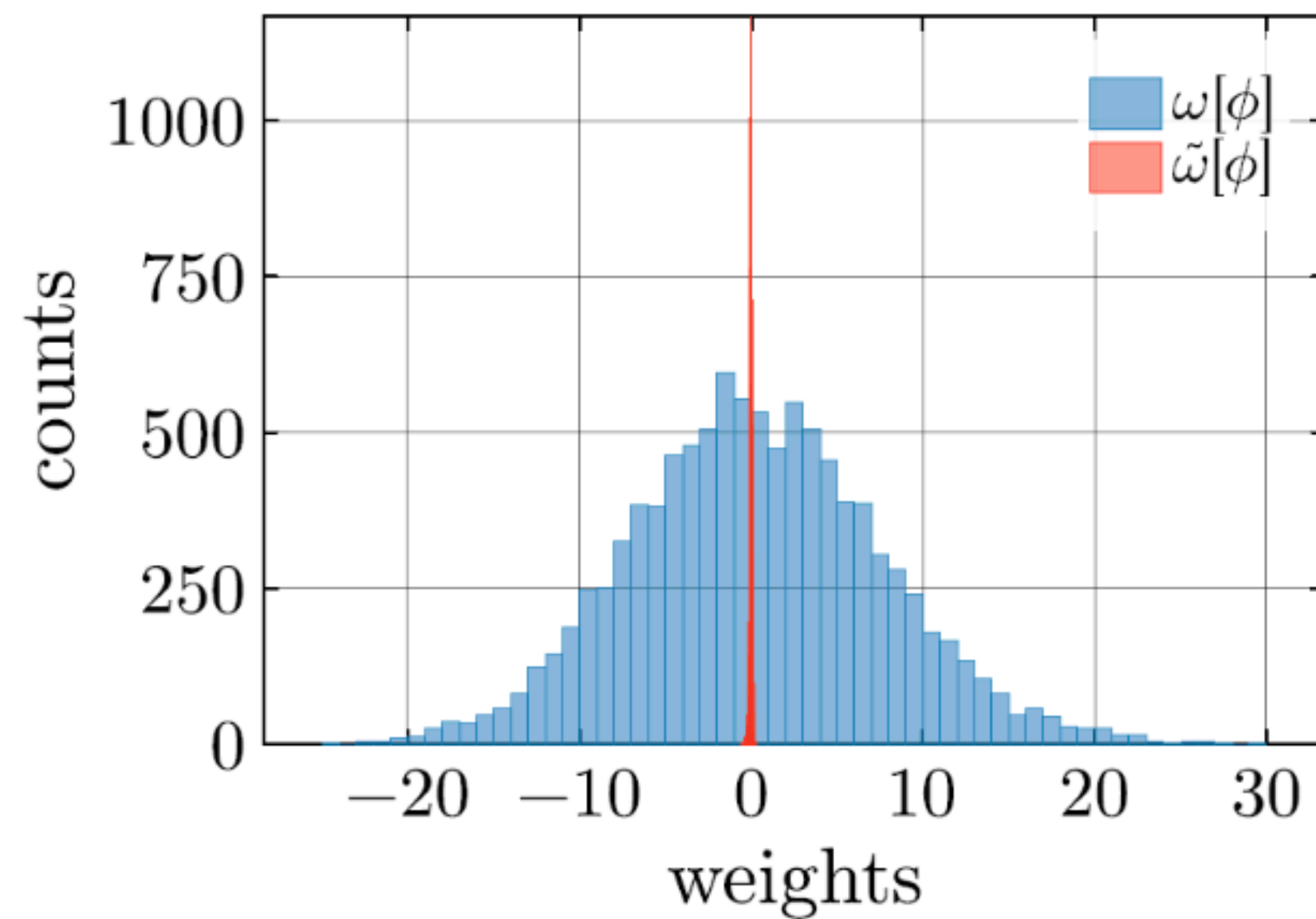
- sample ϕ from $r \sim e^{-S}$
- flow through f_{θ}
- compute KL divergence at $\mathcal{O}(\varepsilon^2)$
- minimize using standard ML toolbox (and repeat as necessary)

Sampling

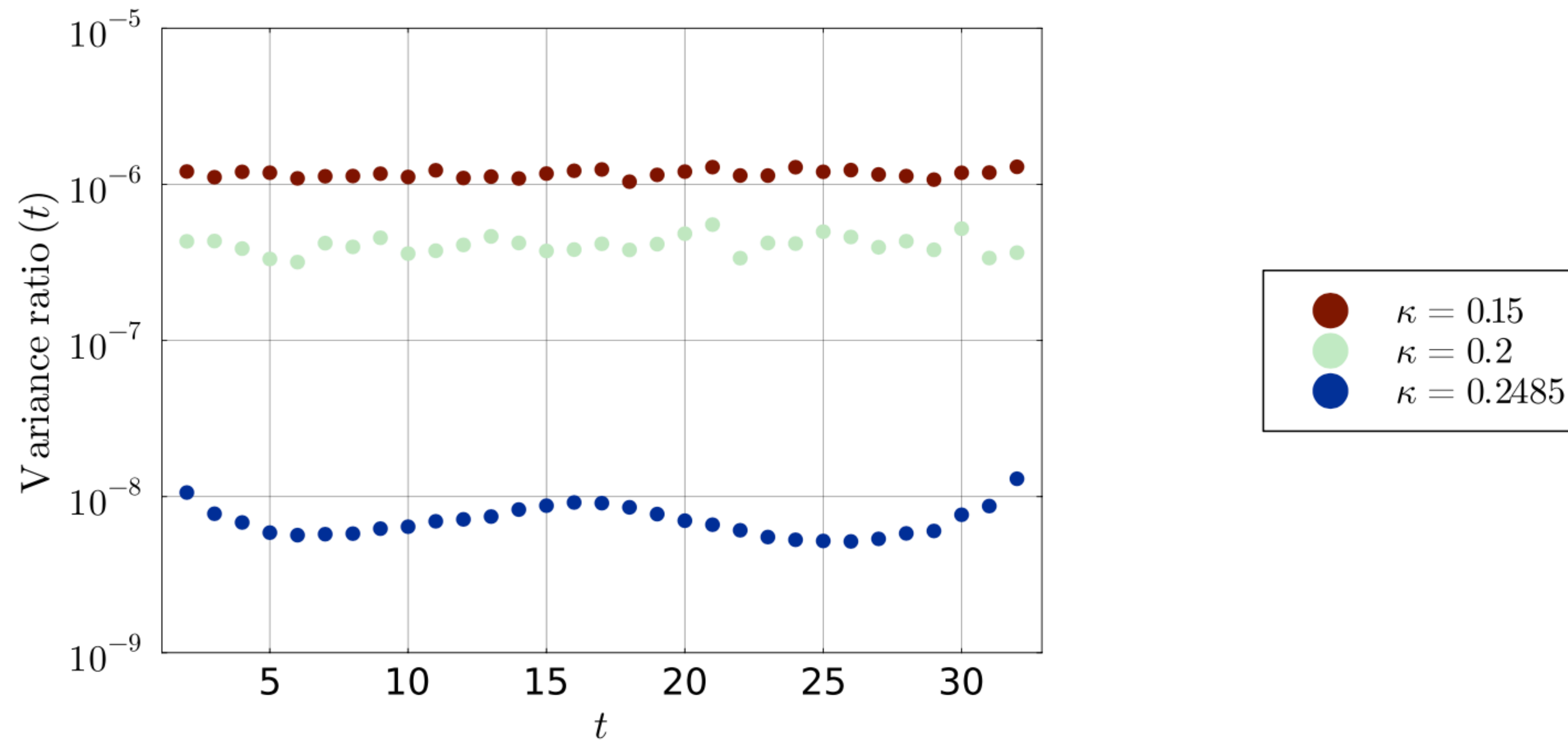
- sample ϕ from $r \sim e^{-S}$
- flow through f_{θ}
- compute weights $\tilde{w}$ at $\mathcal{O}(\varepsilon)$
- extract $C(t)$ analytically or using [Catumba+Ramos; 2502.15570]

Main limiting factor is
Hutchinson trace to
compute Jacobian trace!
 $\mathcal{O}(V)$ scaling in cost

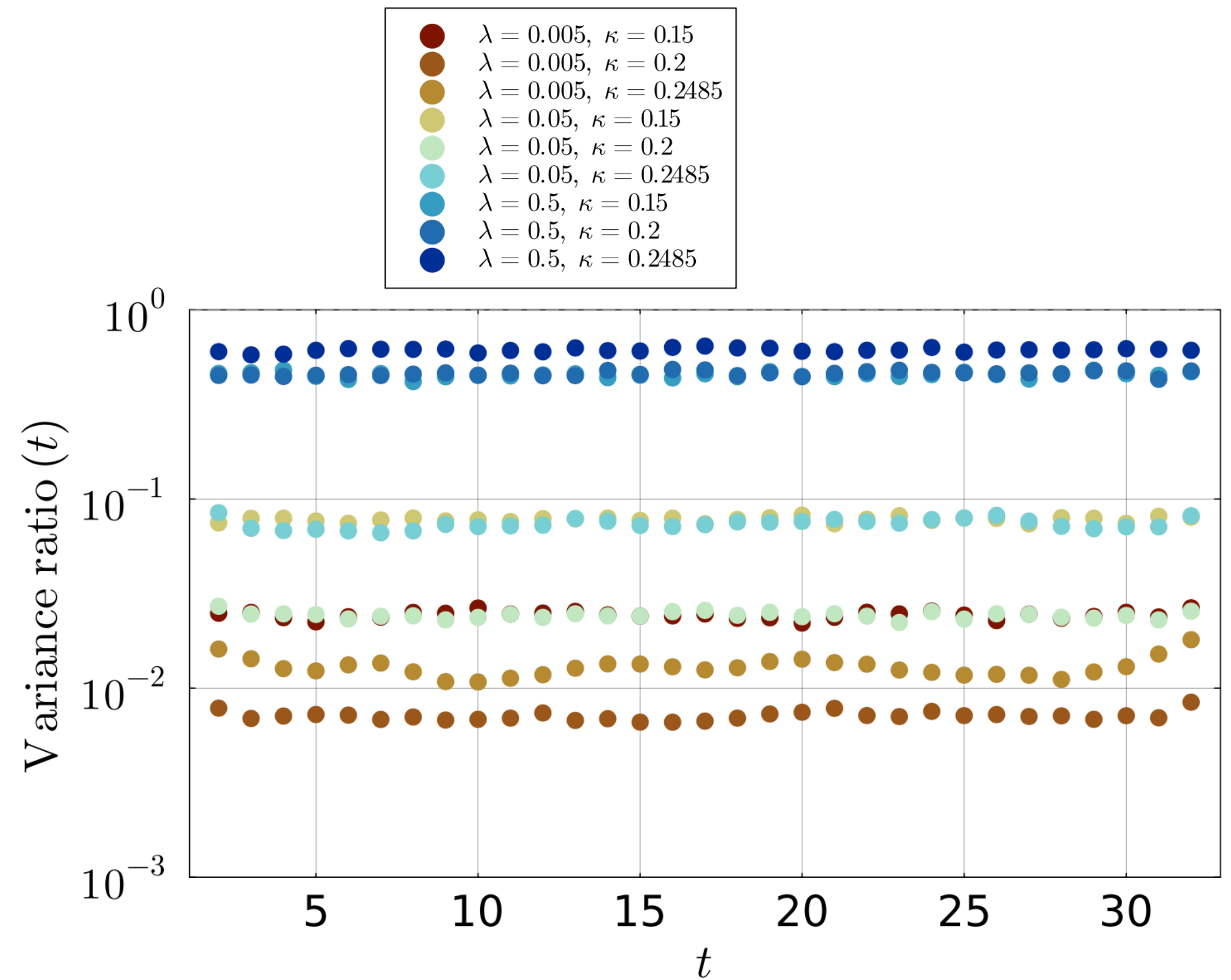
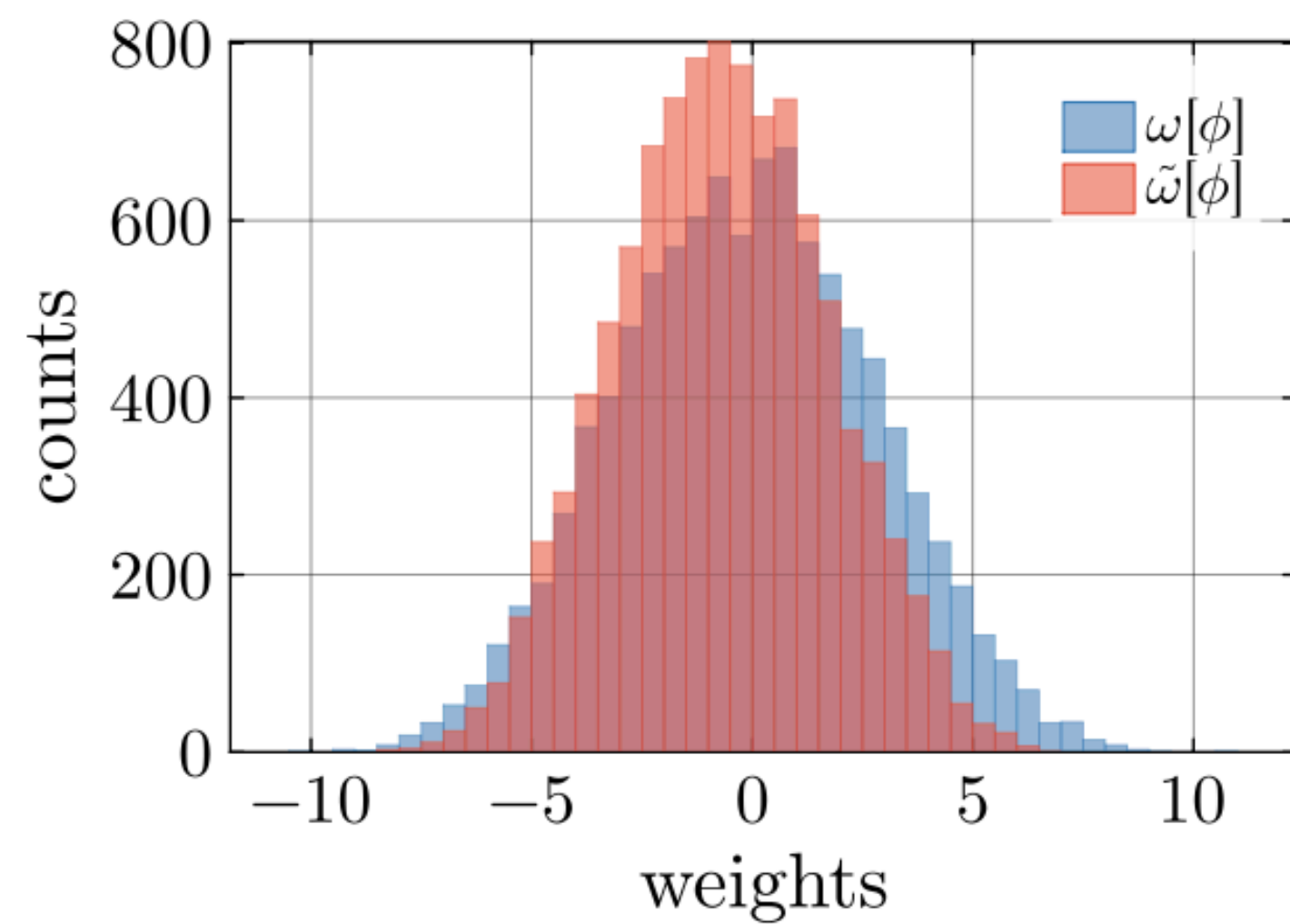
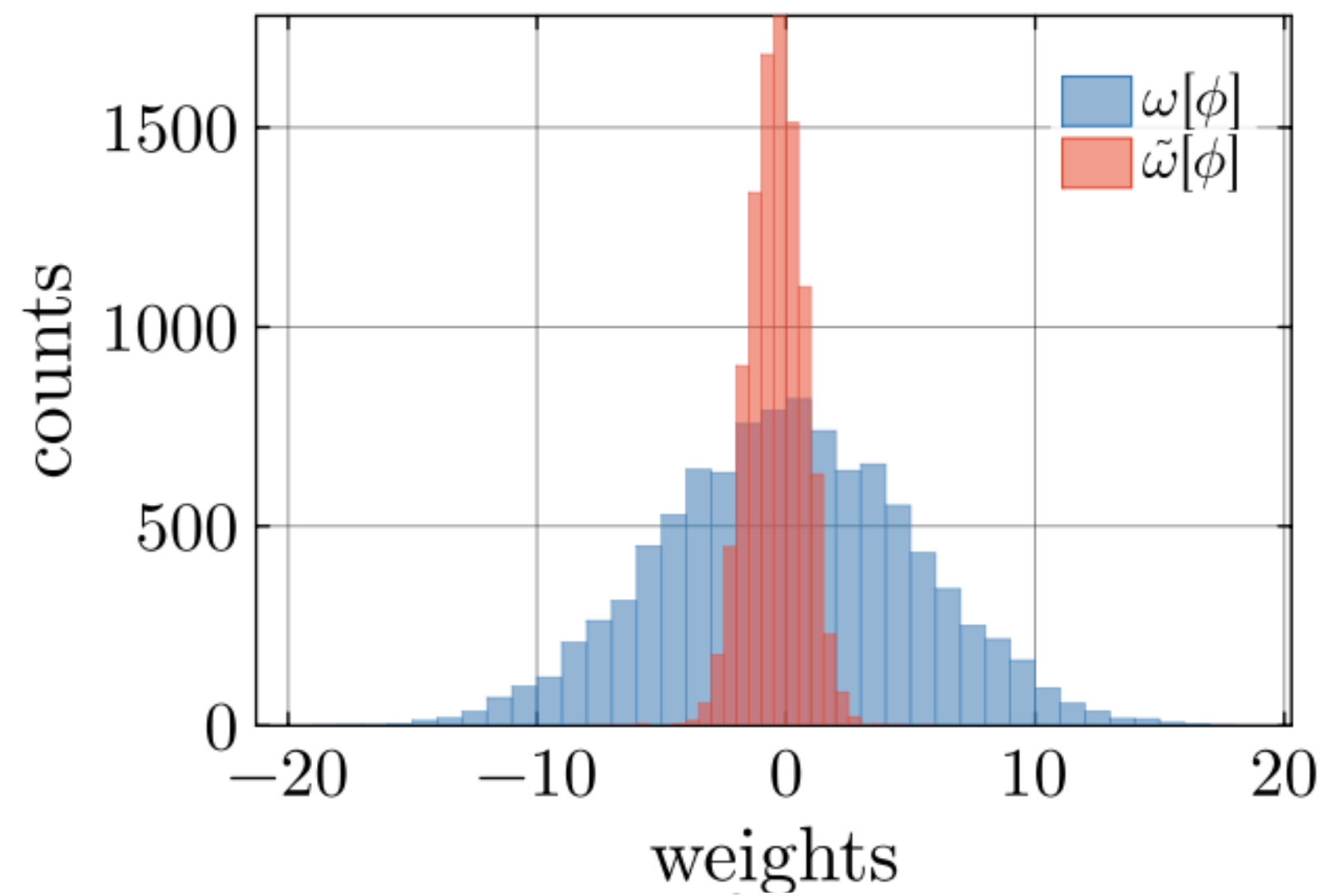
weights distribution before and after training ($\kappa = 0.2485$)



Direct ML minimisation



Direct ML minimisation



Stein operator is the generator of the action deformations and define the optimal path

$$\frac{dS_\varepsilon}{d\varepsilon} = -\mathcal{T}_r f \quad \mathcal{T}_r f \equiv \text{tr} J_f - f \cdot \nabla S$$

“Perfect flow” limit (aka zero variance) corresponds to the Stein-Poisson equation

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

Stochastic evolution provides a systematical way to reduce StN degradation

$$dZ_\tau = -\nabla S d\tau + \sqrt{2} dW_\tau$$

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

$$\hat{A}_t = \rho_T \mathcal{O}_t + f_T$$

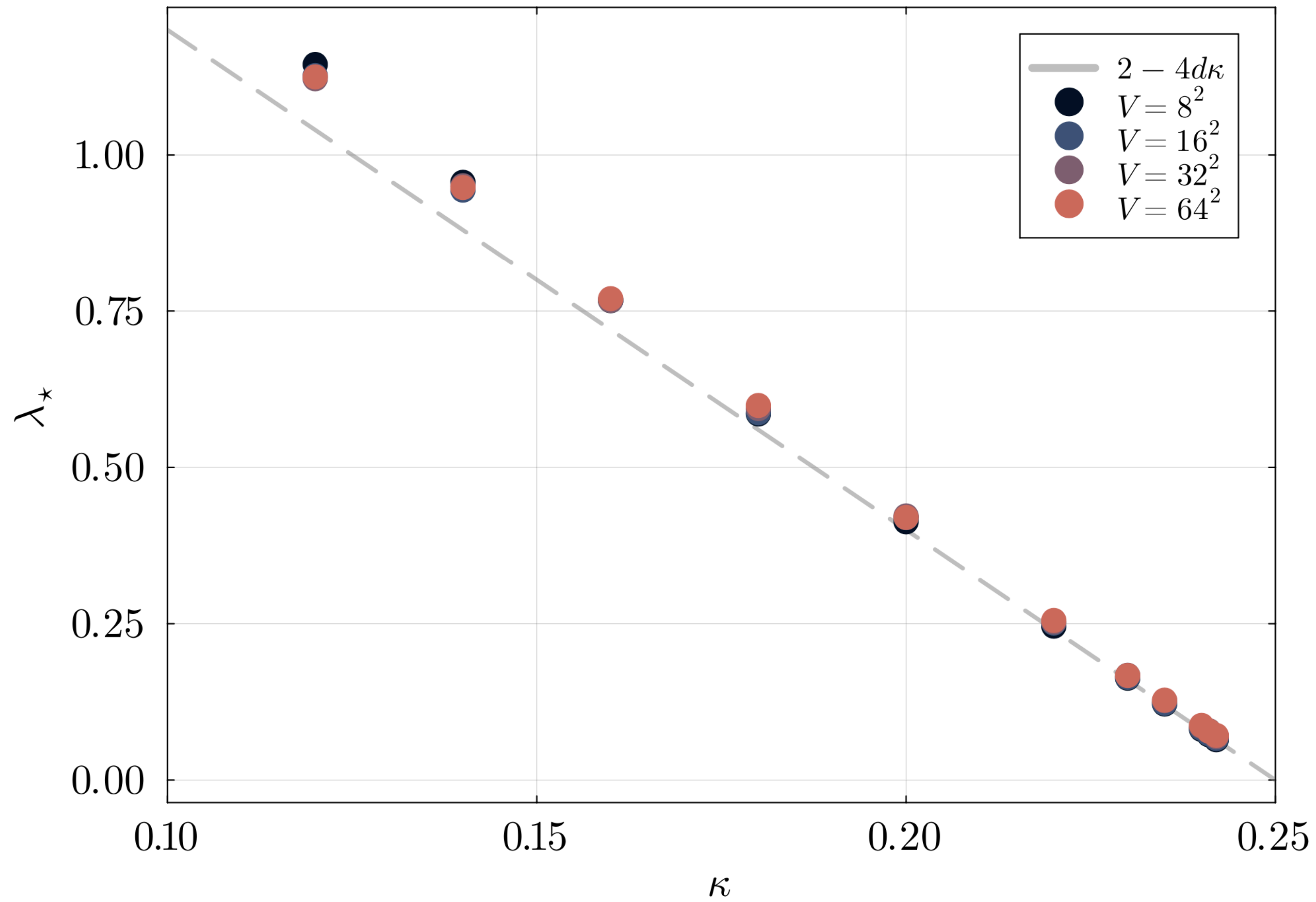
Stein operator is expensive, but needed for the bias => warm start NF!

$$\langle \rho_T \mathcal{O}_t \rangle \sim e^{-\lambda_* T}$$

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

BACKUP

The spectral gap



Minimal variance condition

Lattice / physics

transport / flow

Action

$$S$$

Action w/ a source

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

Base distribution

$$r = e^{-S} / Z_0$$

Target distribution

$$p = e^{-S_J} / Z_J$$

Transf. of configs induces effective action

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

Transport/flow map

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

Reweighting factors

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

Overlap factor / likelihood ratio

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

$$= -\text{KL}(q || p_\varepsilon)$$

$$\text{KL}(q_f || p_\varepsilon) = \frac{1}{2} \text{Var}[\tilde{w}] + \mathcal{O}(\langle (\tilde{w} - 1)^3 \rangle)$$

Beat 1: change of variable

Expectation values are invariant
under an invertible reparametrization

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int D\phi \mathcal{O}(\phi) e^{-S(\phi)} = \frac{1}{Z} \int D\tilde{\phi} \mathcal{O}(\tilde{\phi}) e^{-S(\tilde{\phi})} |\det J_T|^{-1}$$

The changes produce an effective action: the
Jacobian compensate for the change in confs

$$S_{\text{eff}}(\phi) = S(T(\phi)) - \log |\det J_T|$$

The volume element is our handle

A notable example: a flow $\phi_\varepsilon = T_\varepsilon(\phi)$

$$\frac{d\phi_\varepsilon}{d\varepsilon} = f(\phi_\varepsilon) \quad \phi_0 \equiv \phi$$

The Jacobian also flows accordingly (Liouville)

$$\frac{d}{d\varepsilon} \log |\det J_{T_\varepsilon}| = (\nabla \cdot f)(\phi_\varepsilon)$$

The effective action evolves as

$$\frac{dS_\varepsilon}{d\varepsilon} = f \cdot \nabla S - \nabla \cdot f$$

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 change of variable that minimises the KL divergence

$$\phi \rightarrow \phi + \varepsilon f(\phi) + \dots$$

Using flow-based models, only amounts to minimise directly

$$\mathcal{L}[f] = \left\langle (\text{tr} J_f - f \cdot \nabla S)^2 - 2f \cdot \nabla \mathcal{O}_0 \right\rangle$$

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

Signal-to-noise

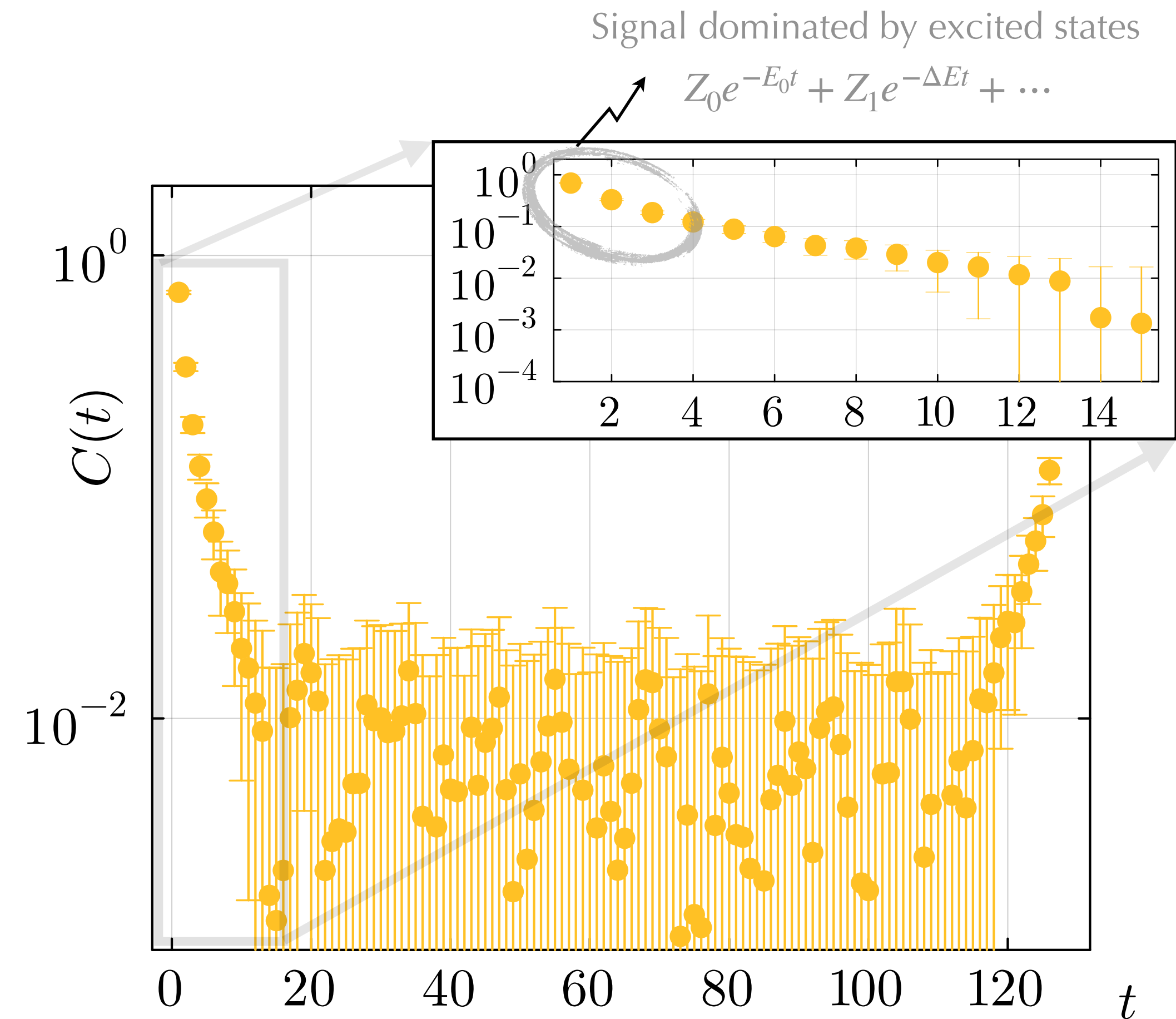
$$C(t) = \langle \mathcal{O}(t) \mathcal{O}^\dagger(0) \rangle \xrightarrow{t \gg (\Delta E)^{-1}} Z e^{-M_{\text{signal}} t}$$

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

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

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

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



Linearised flow maps

$$q_\varepsilon = (T_\varepsilon)_\# r$$

$$T_\varepsilon : \phi \mapsto \phi + \varepsilon f(\phi)$$

$$r(\phi)$$

$$\dot{r} = -\nabla \cdot (r f)$$

“curve” of prob. measures passing for r at $\varepsilon = 0$

$$\frac{dq_\varepsilon}{d\varepsilon} + \nabla \cdot (q_\varepsilon f) = 0$$

describes the evolution of “shifting” probability r along the vector field f

natural notion of
“tangent vector”

...what about length?

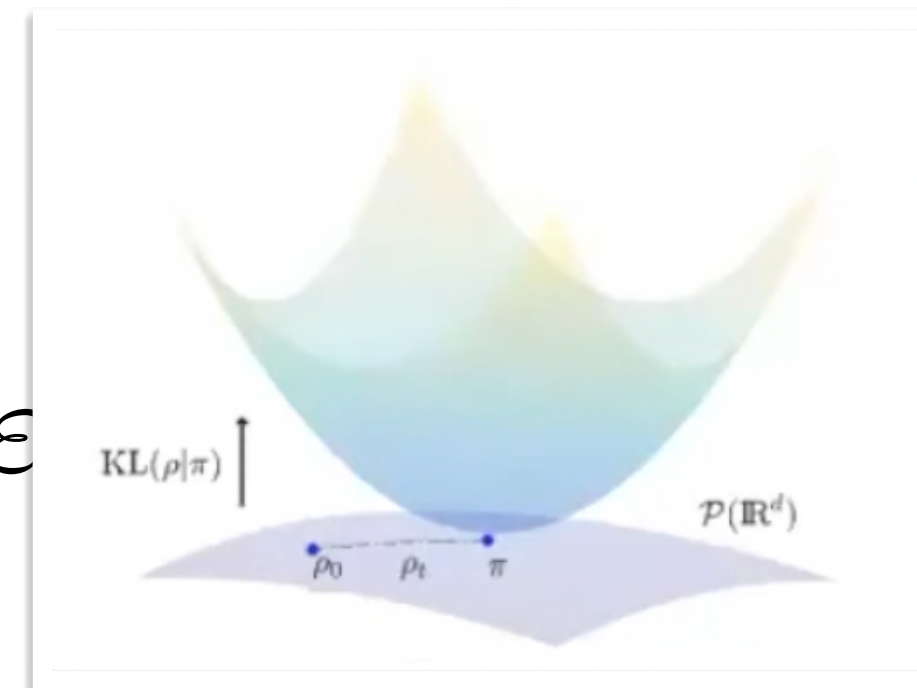
$$\|\dot{r}\|^2 \stackrel{?}{=} \langle f, f \rangle$$

...what about length?

Linearised flow maps

$$q_\varepsilon = (T_\varepsilon)_\# r$$

$$T_\varepsilon : \phi \mapsto \phi + \varepsilon$$



$$\langle f, f \rangle = \min_{f: \dot{r} = -\nabla \cdot (rf)} \|f\|^2$$

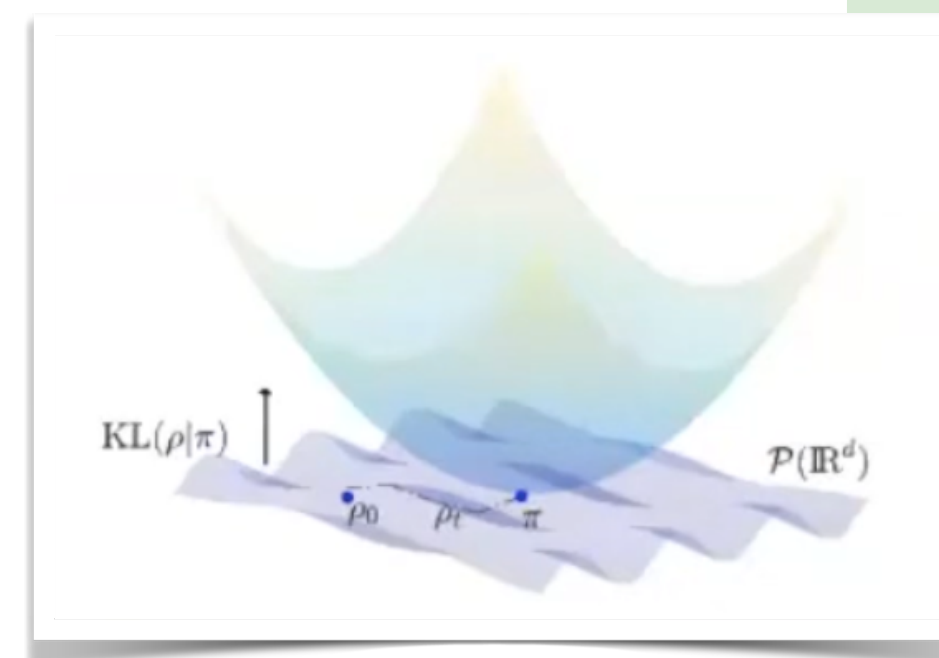
Wasserstein geometry à-la Benamou-Brenier

$$\|\cdot\|_{L^2}$$

“Minimum kinetic energy needed to transport r into p along a mass-conserving flow”

$$\dot{r} = -\nabla \cdot (rf)$$

$$r(\phi)$$

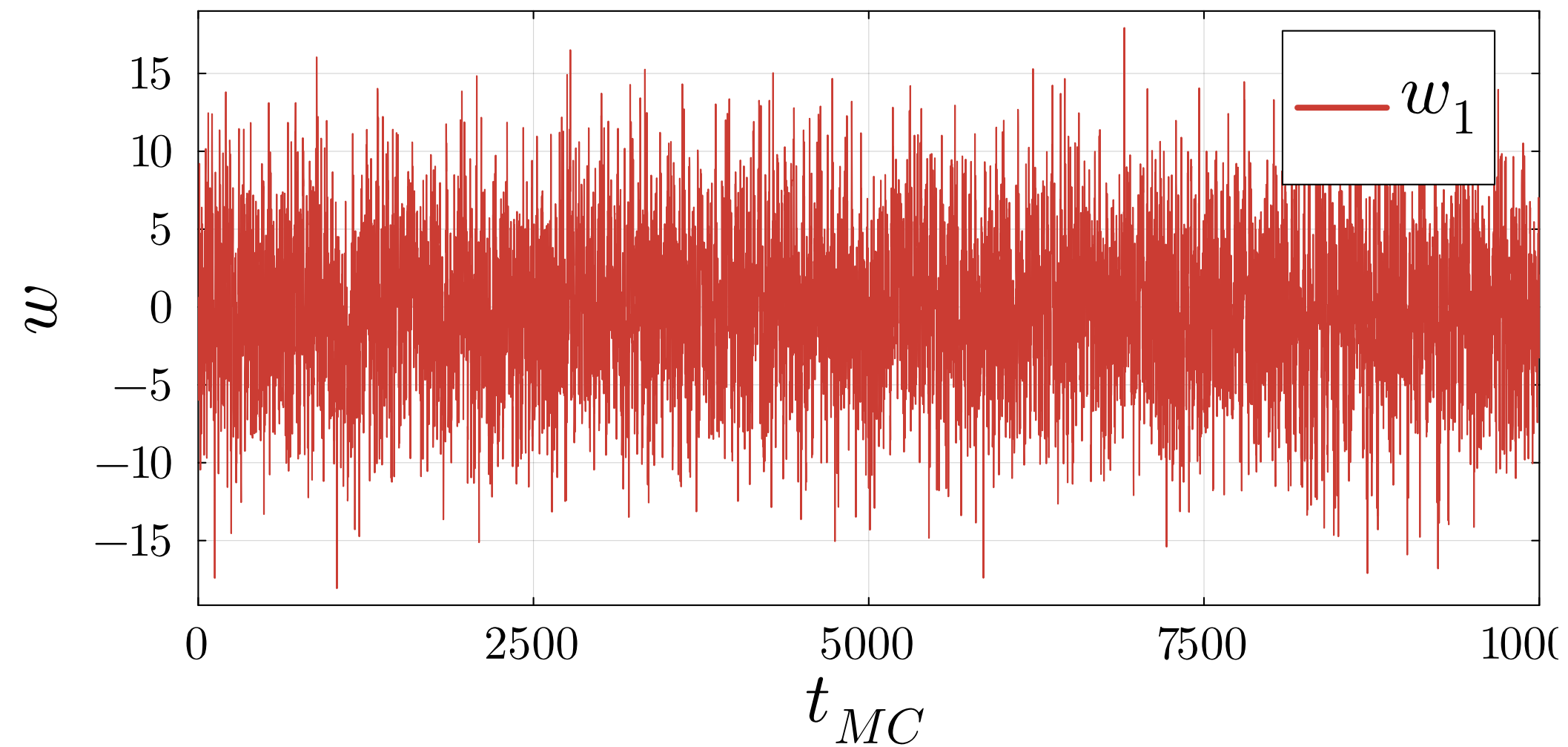


Stein-geometry

$$\|\cdot\|^2 = \langle (\mathcal{T}_r \cdot)^2 \rangle_r \quad \text{or} \quad \|\cdot\|_{\text{RKHS}}^2$$

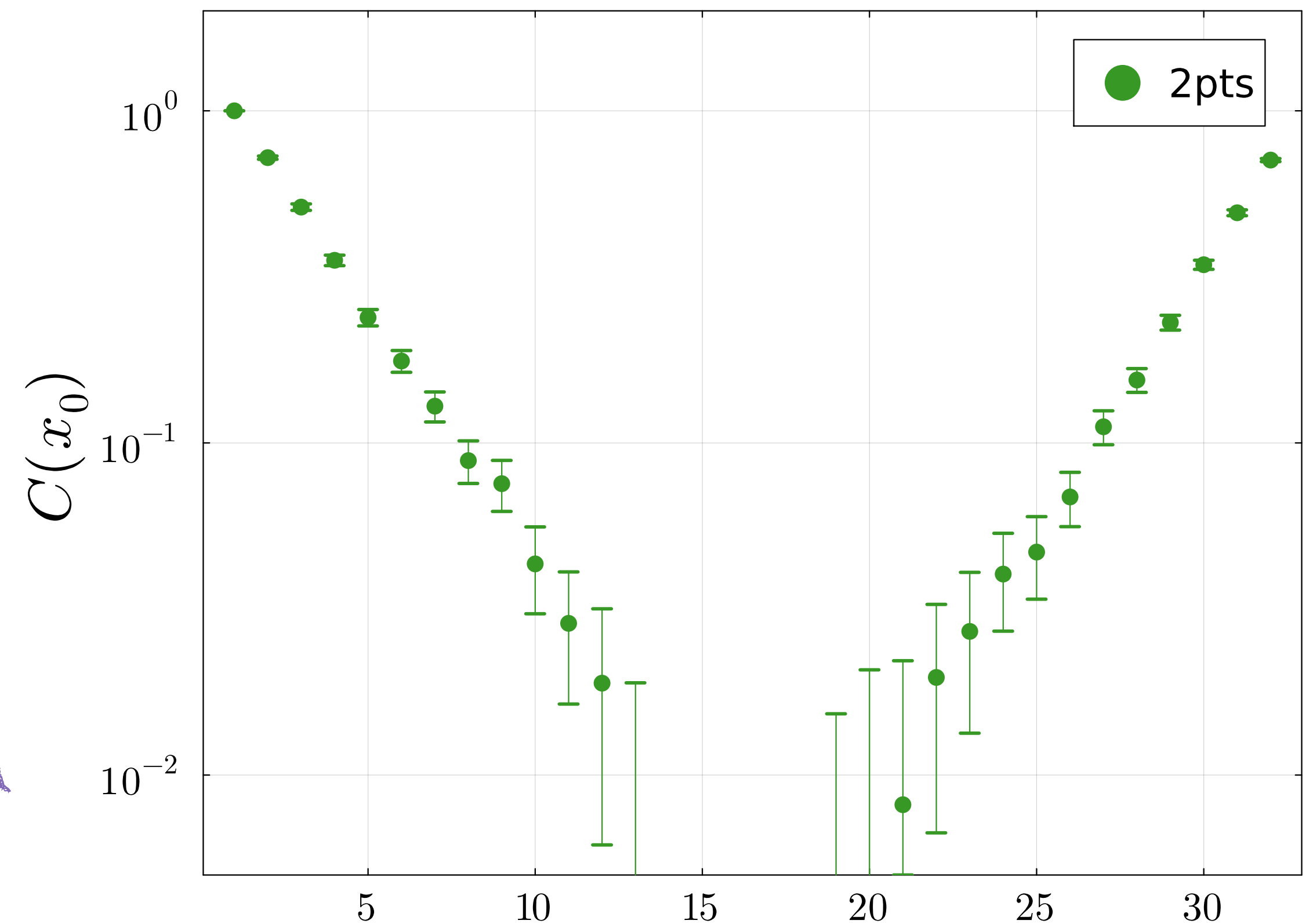
$$\langle w \mathcal{O}_t \rangle_S = \langle \mathcal{O}_t \rangle_S + \varepsilon C(t) + \dots$$

Correlators via reweighting



Solution: find sample transformation
that renders the w constant!

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



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

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

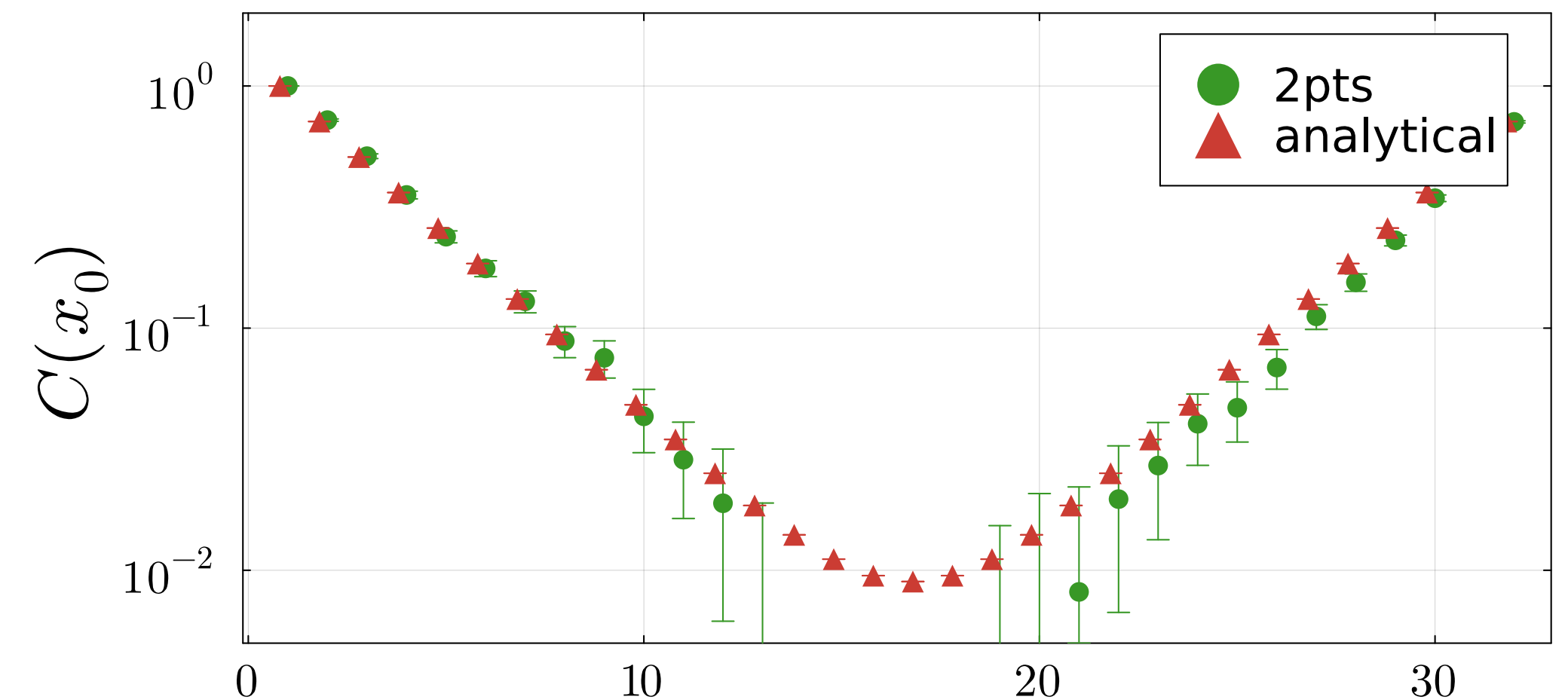
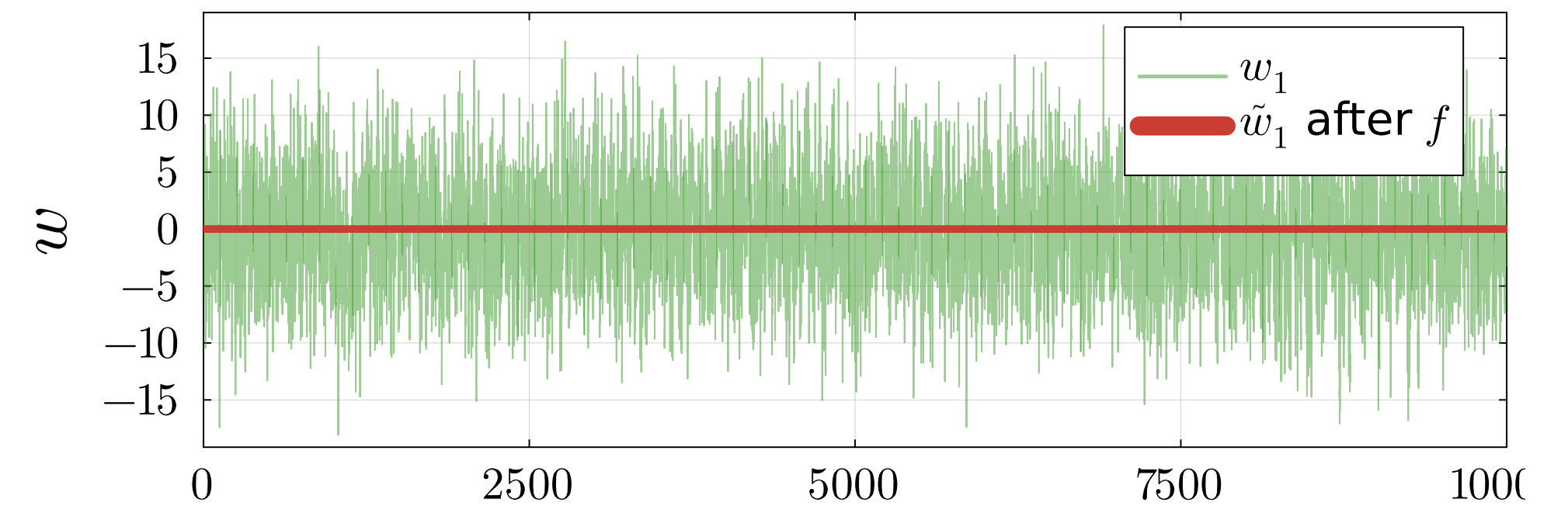
Improved reweighting

Transform samples before reweighting

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

Scalar free theory:

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



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

Configurations

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

Target distribution

Improved reweighting

Transform samples before reweighting

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

reweight with...

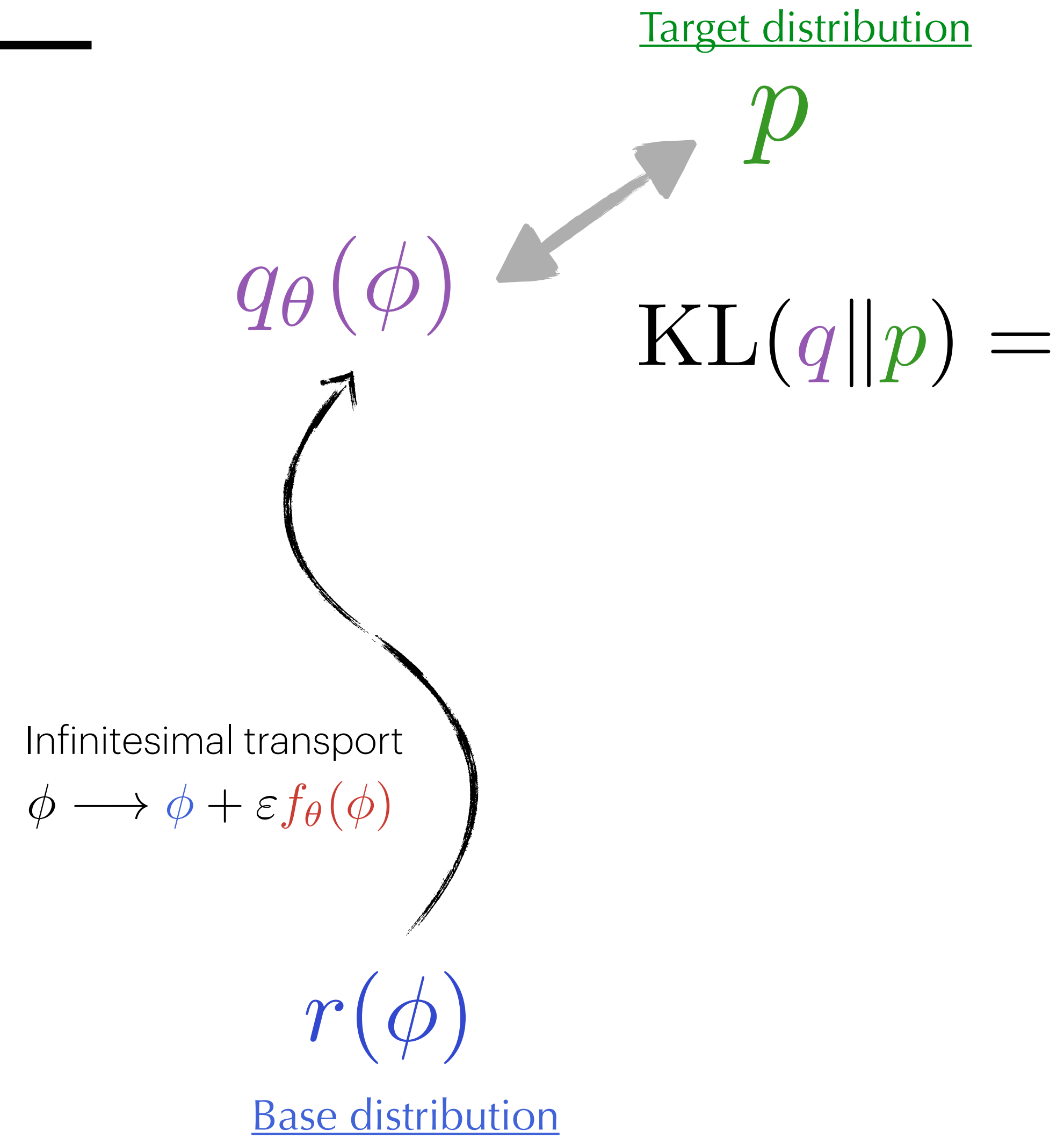
$$w = \exp \left[-S_J[\tilde{\phi}] + S[\phi] + \log |1 + \varepsilon J_f| \right]$$

Scalar ~~free~~ interacting theory:

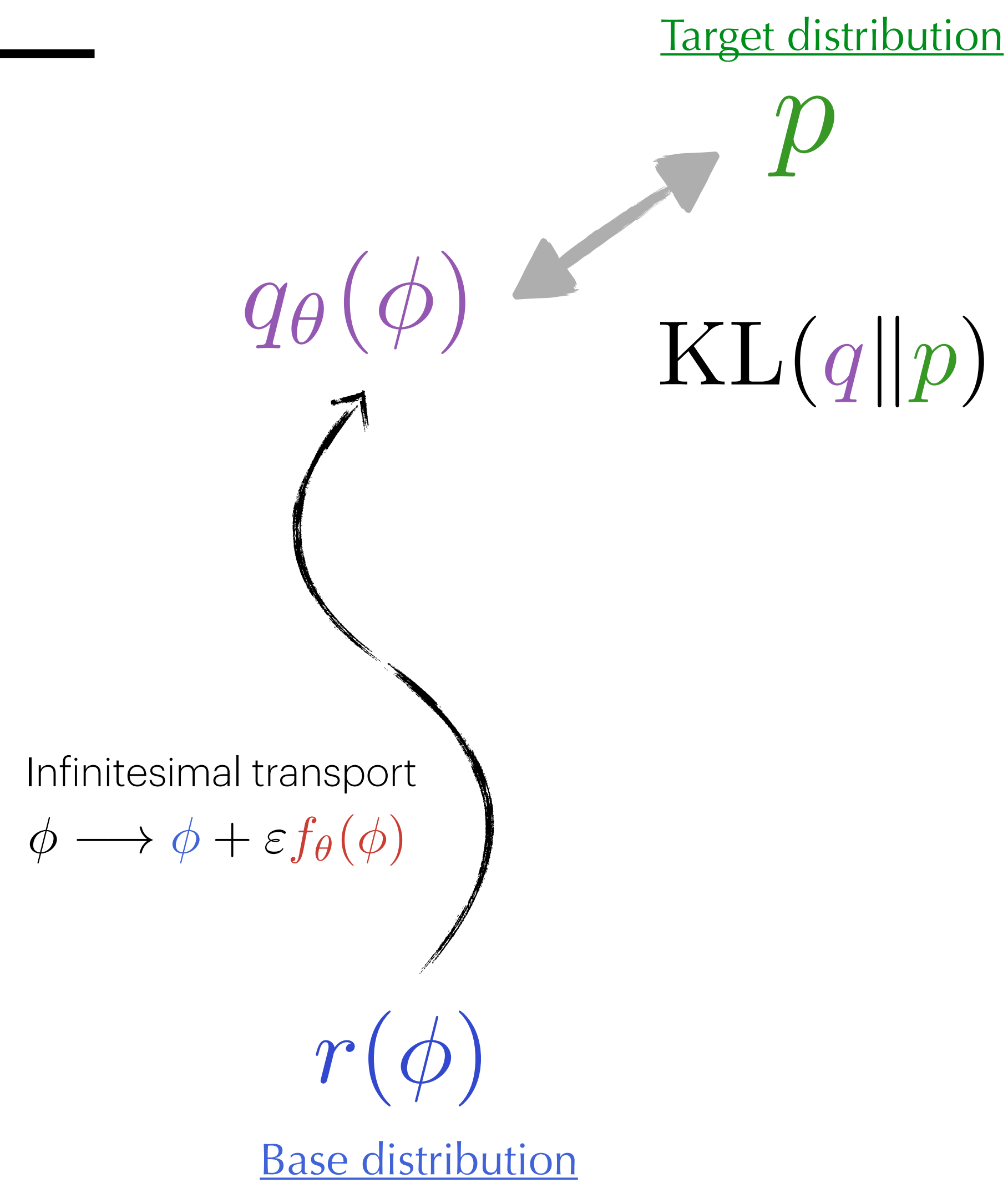
$$\phi_p \longrightarrow \tilde{\phi}_p = \phi_p + \frac{1}{2} \frac{\delta(\vec{p})}{\hat{p}^2 + \hat{m}_R^2(\phi)} \varepsilon$$

Renormalised mass

Linearised flow maps



Linearised flow maps



$$\left\langle \log \frac{r}{p} \right\rangle$$

base-target KL divergence (constant)

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

Score-mismatch or Stein's discrepancy

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

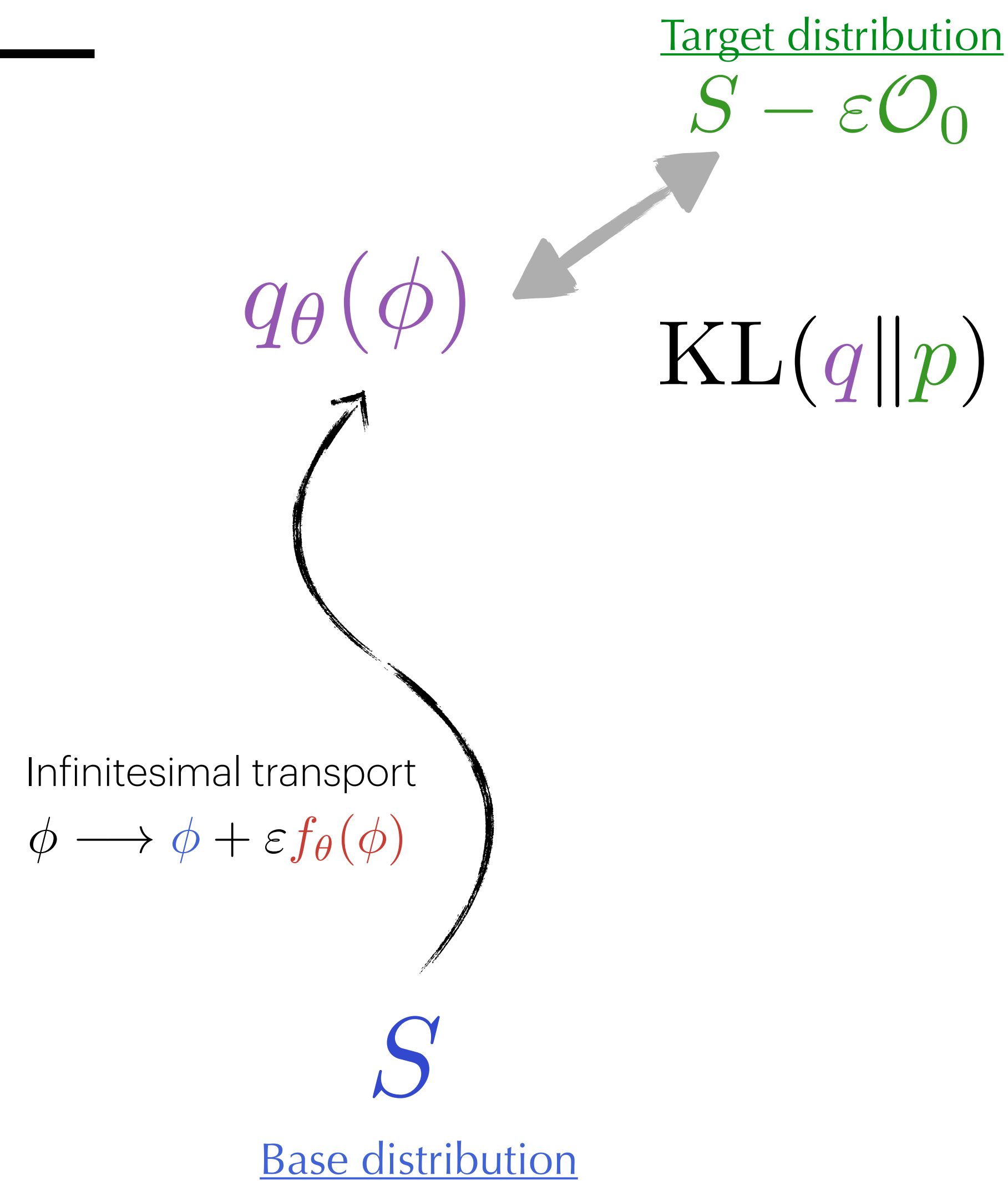
Riemannian "length" of f in Stein's geometry

Curvature instabilities (Hessian mismatch)

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

$\hbar$ QTC

Linearised flow maps



$$\text{KL}(q||p) =$$

$$\left\langle \log \frac{r}{p} \right\rangle$$

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

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

Curvature instabilities
(Hessian mismatch)

$\hbar$ QTC

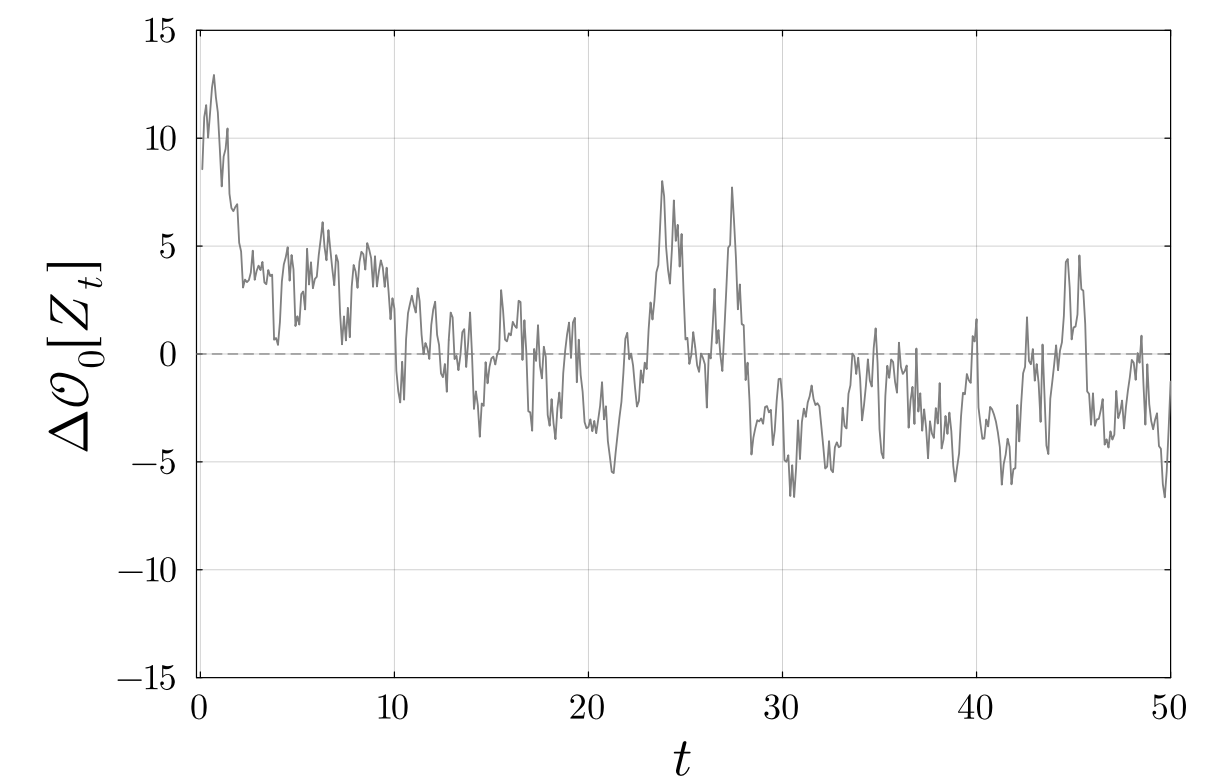
$$f(\phi) = \nabla_{\phi} \varphi$$

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

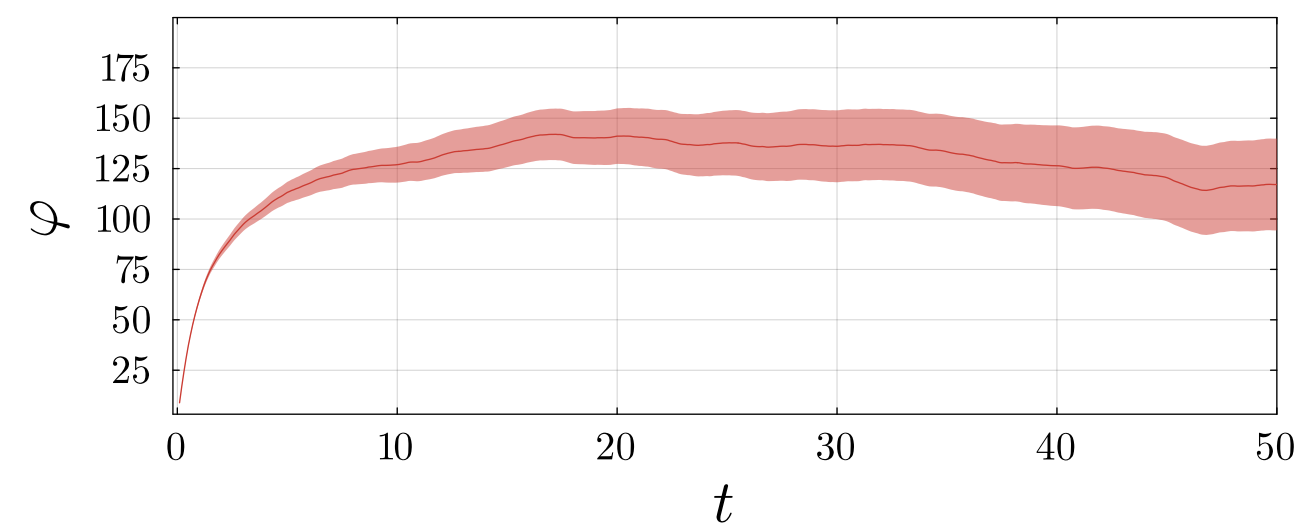
Feynman-Kac formula

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

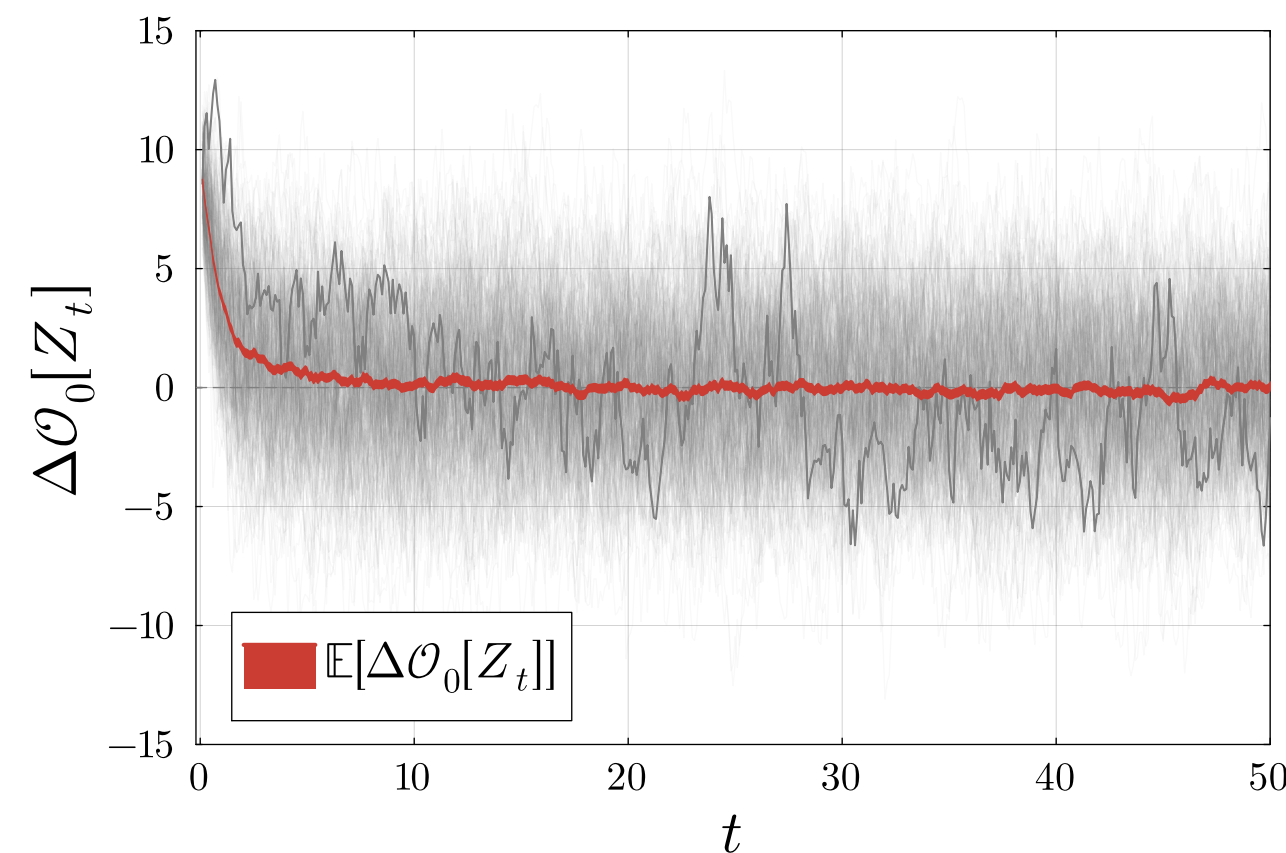
Integrate Langevin dynamics starting from $Z_0 = \phi$



Integrate in stochastic time



Expectation value over walkers (average over Brownian noise)



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

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

Feynman-Kac formula

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



Swap gradient and noise average

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

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

SDEsolve(...) forward to obtain

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

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

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

Works, but reproduce same hessian instabilities of HAD!

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

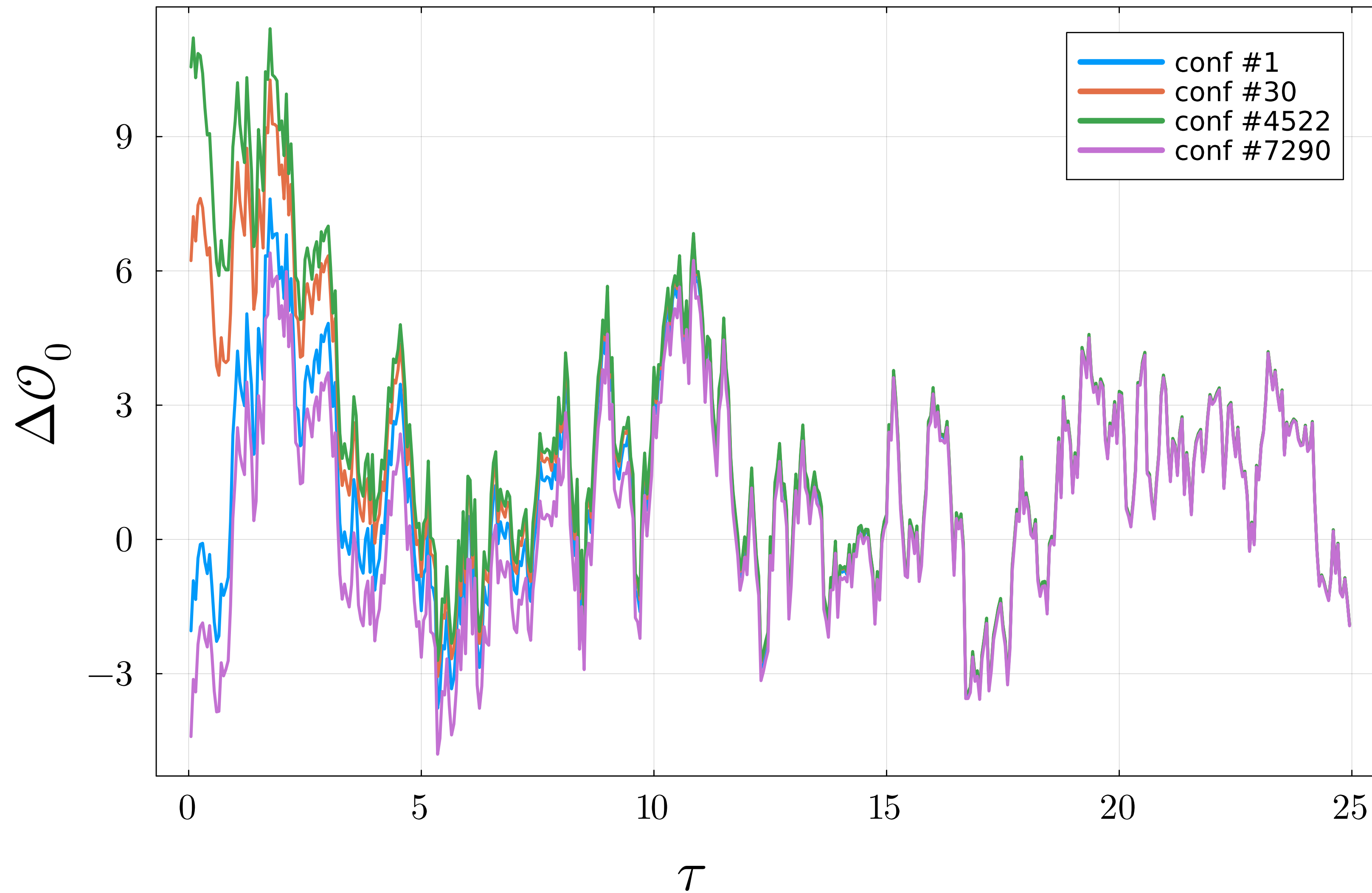


[Albergo, Eijnden] - *"NETS: A Non-Equilibrium Transport Sampler"*, e-Print: [2410.02711](https://arxiv.org/abs/2410.02711)



[Butti, Catumba, Nada, Spatscheck] - *in preparation*

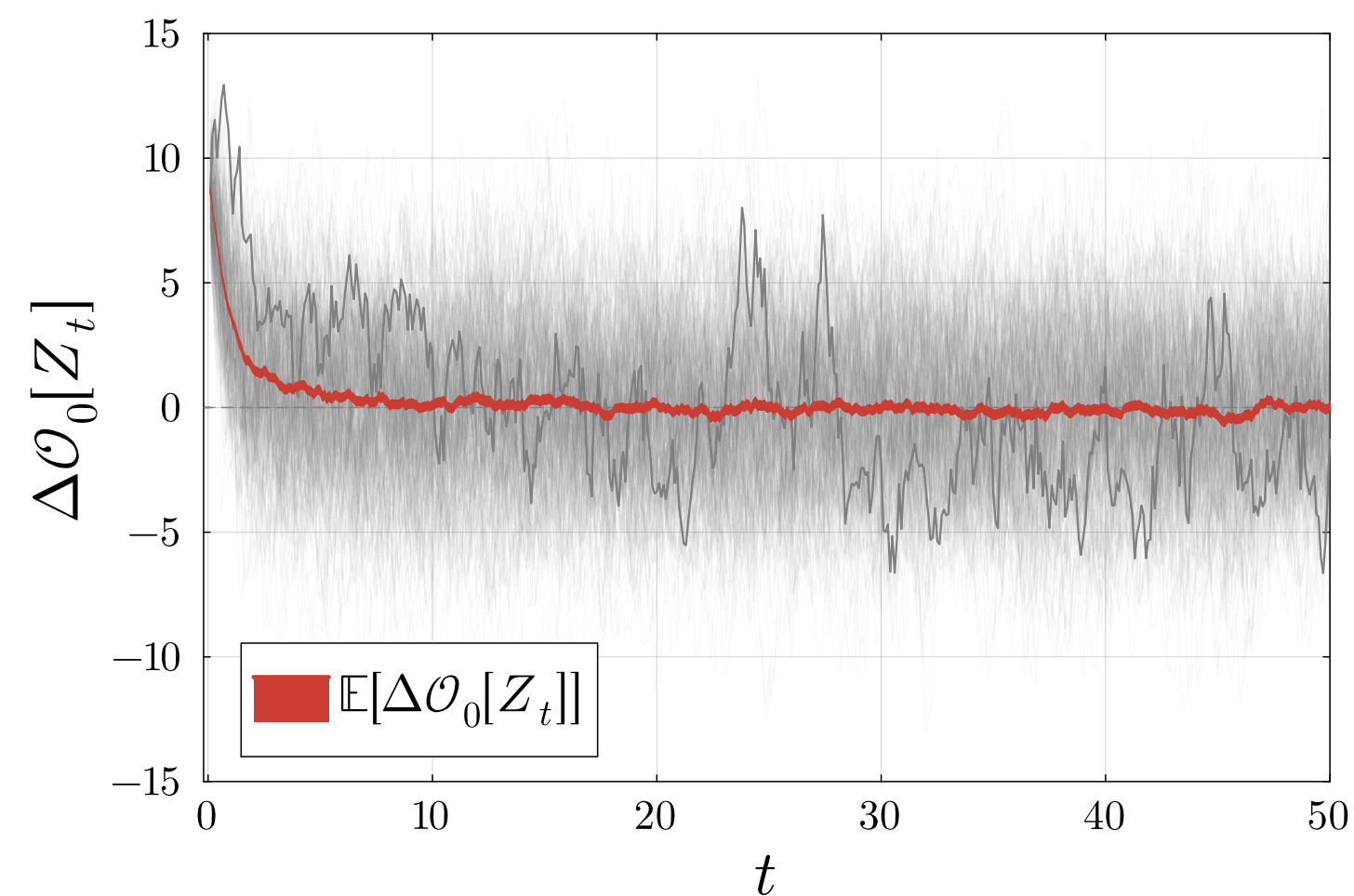
Maximally coupled noise



Feynman-Kac formula

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

Langevin evolution makes the integrand to converge to zero exponentially in τ



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 $\mathcal{O}$ -mom propagator

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

Define Langevin autocorrelation function

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

Green-Kubo relation:

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

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

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(phi))`

2.2 initial config: `Z[0] .= phi`

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

3. Repeat for N_w walkers

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

5. Compute the gradient wrt the original conf

6. Repeat for all configurations

τ

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

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(phi))`

2.2 initial config: `Z[0] .= phi`

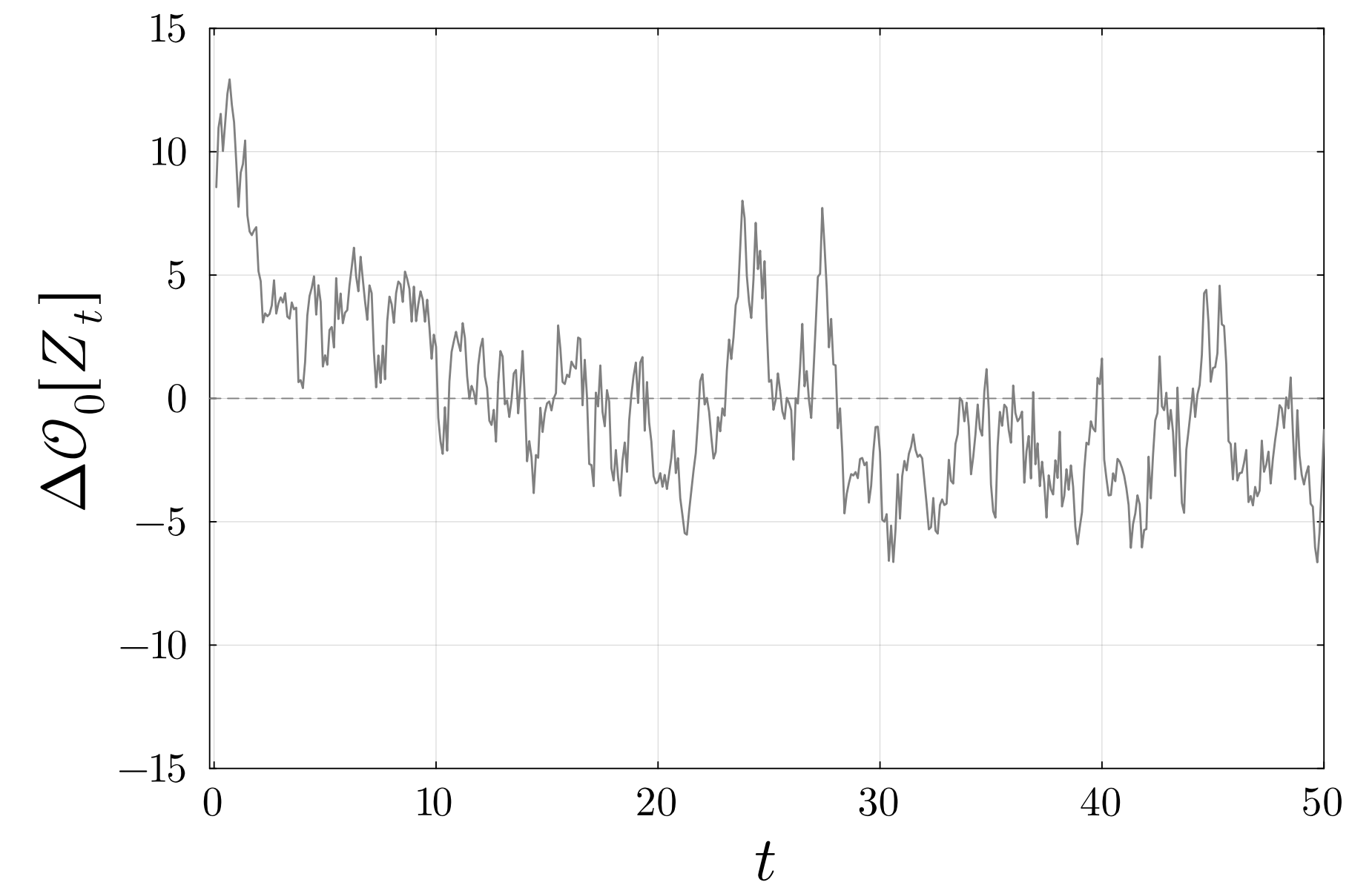
2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`

3. Repeat for N_w walkers

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

5. Compute the gradient wrt the original conf

6. Repeat for all configurations



τ

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

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(phi))`

2.2 initial config: `Z[0] .= phi`

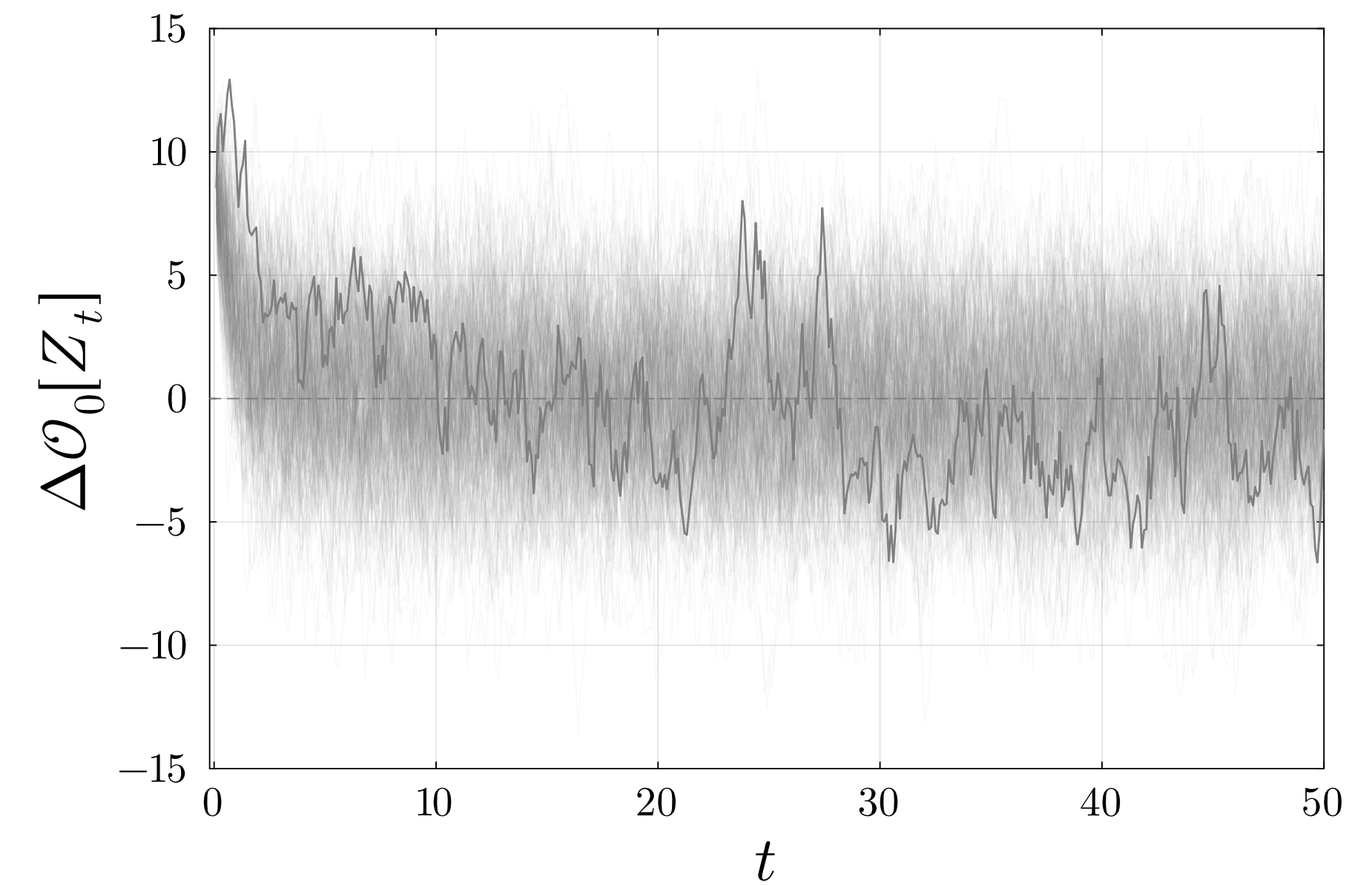
2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`

3. Repeat for N_w walkers

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

5. Compute the gradient wrt the original conf

6. Repeat for all configurations



τ

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

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(phi))`

2.2 initial config: `Z[0] .= phi`

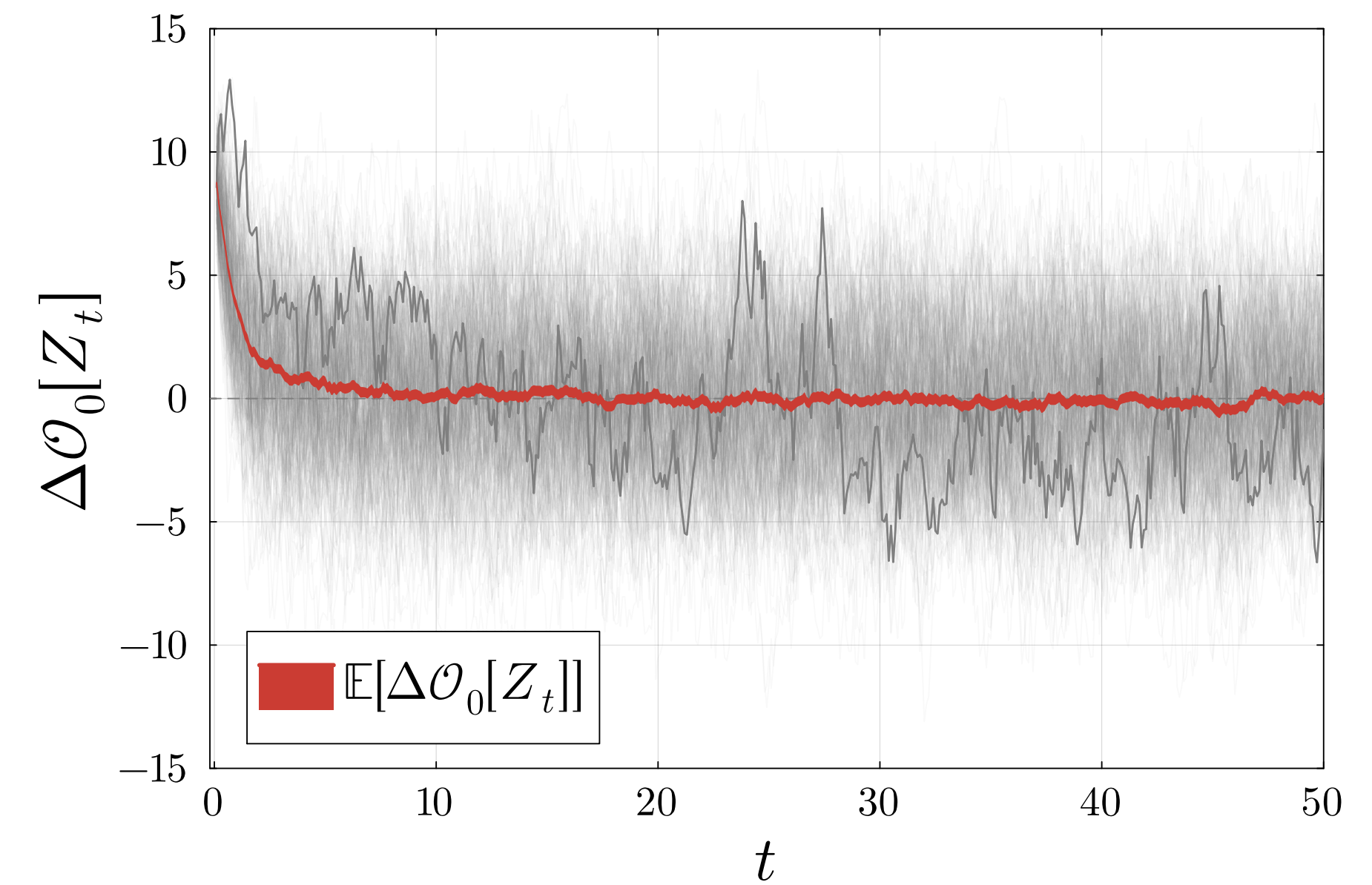
2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`

3. Repeat for N_w walkers

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

5. Compute the gradient wrt the original conf

6. Repeat for all configurations

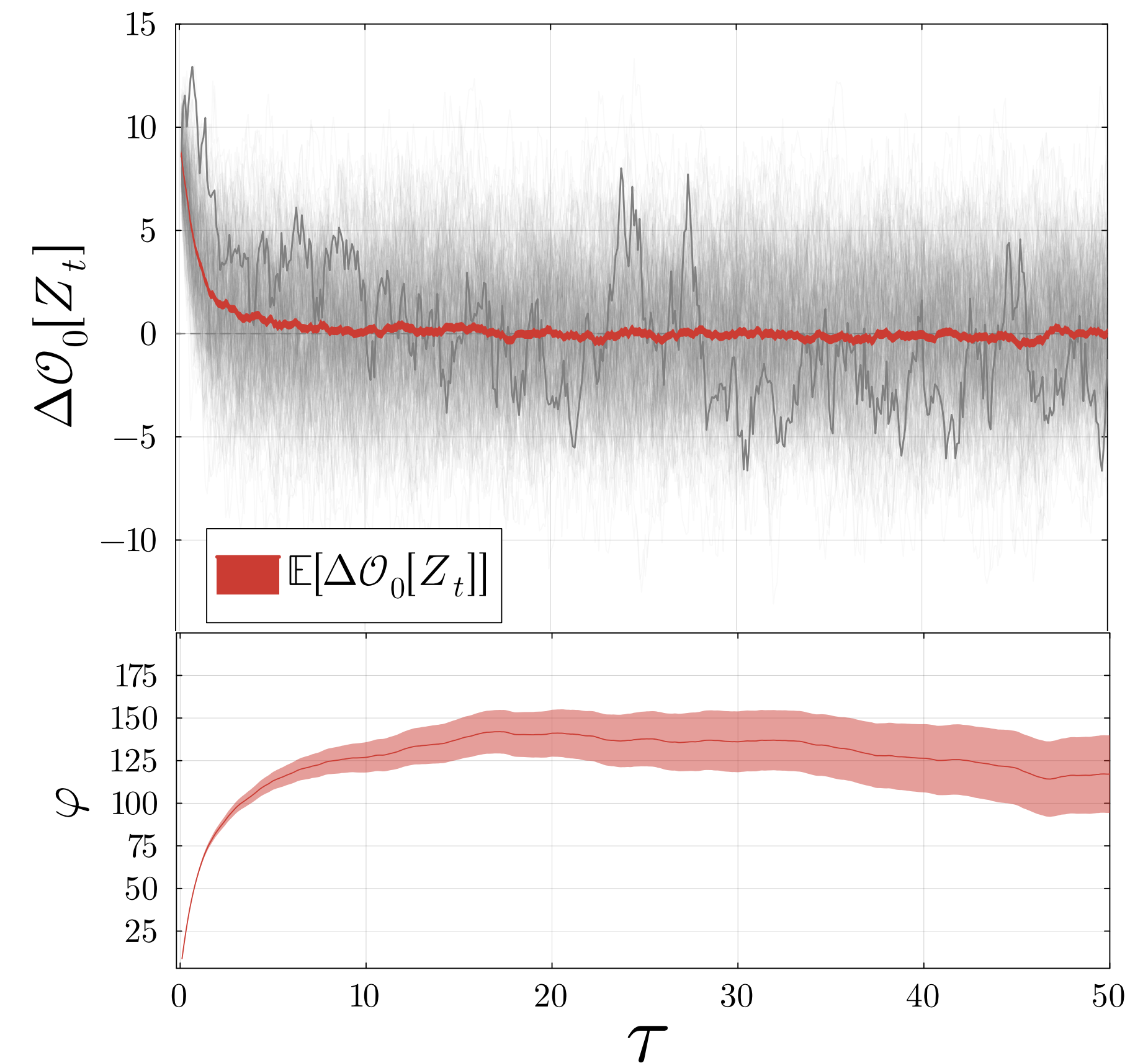


τ

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

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(phi))`
 - 2.2 initial config: `Z[0] .= phi`
 - 2.3 repeat: `Z[t+1] = Z[t] - ∇S*dt + sqrt(2*dt)*eps[t]`
3. Repeat for N_w walkers
4. Compute integral (correlated cumulative sum in $[0, T]$)
5. Compute the gradient wrt the original conf
6. Repeat for all configurations



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

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(phi))`

2.2 initial config: `Z[0] := phi, 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])`

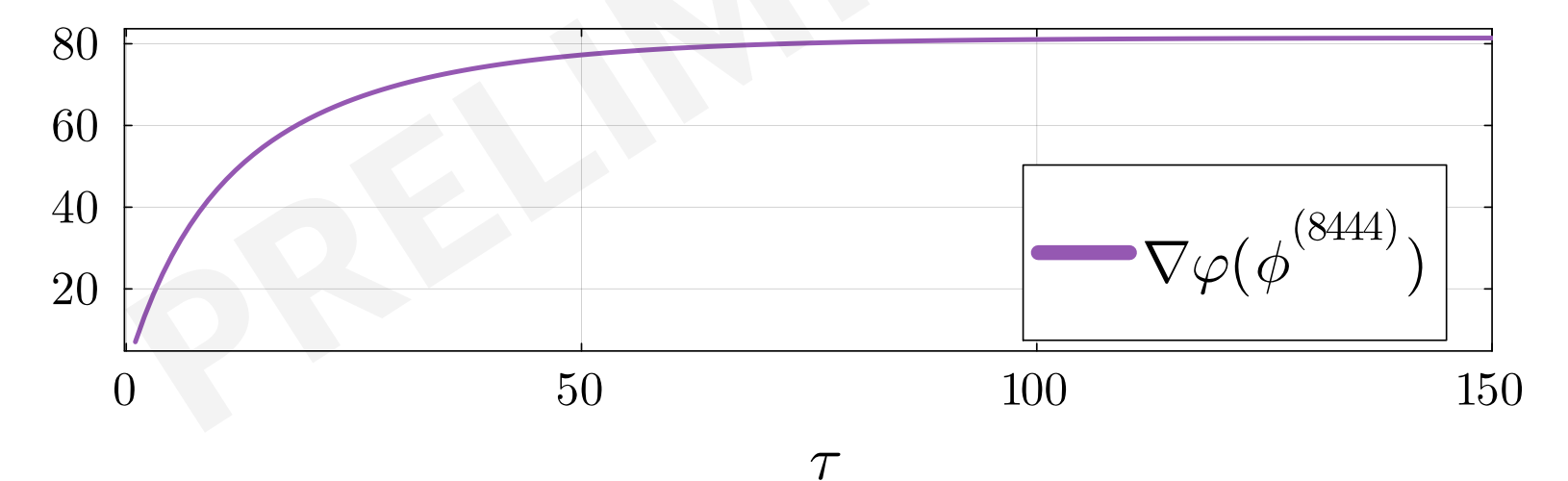
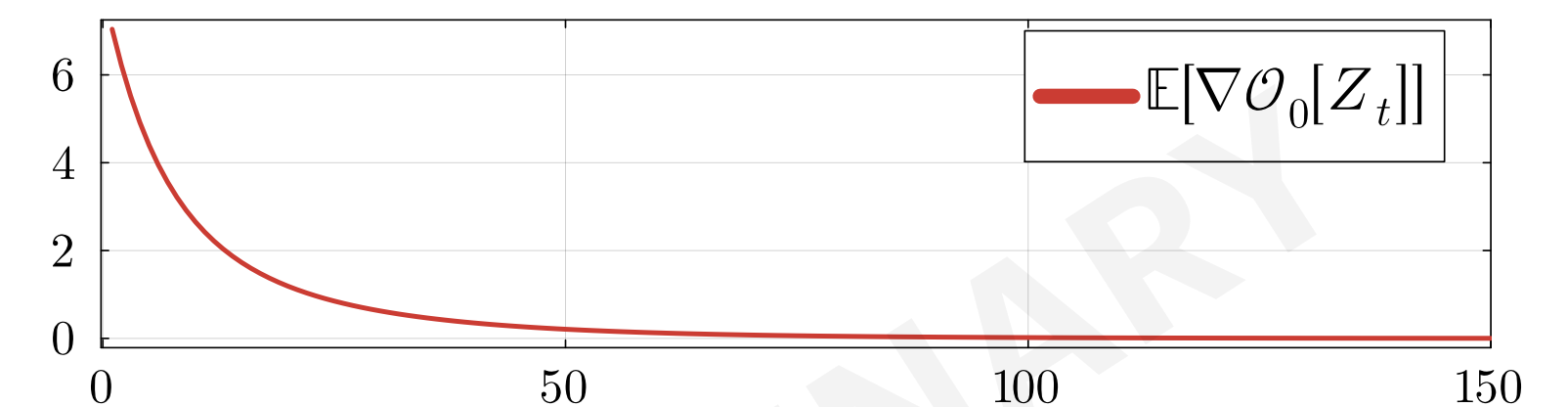
3. Compute $\hat{\phi}^{[0,T]}$ for single walker

4. Compute gradient $\nabla \hat{\phi}^{[0,T]}$ for a single walker *

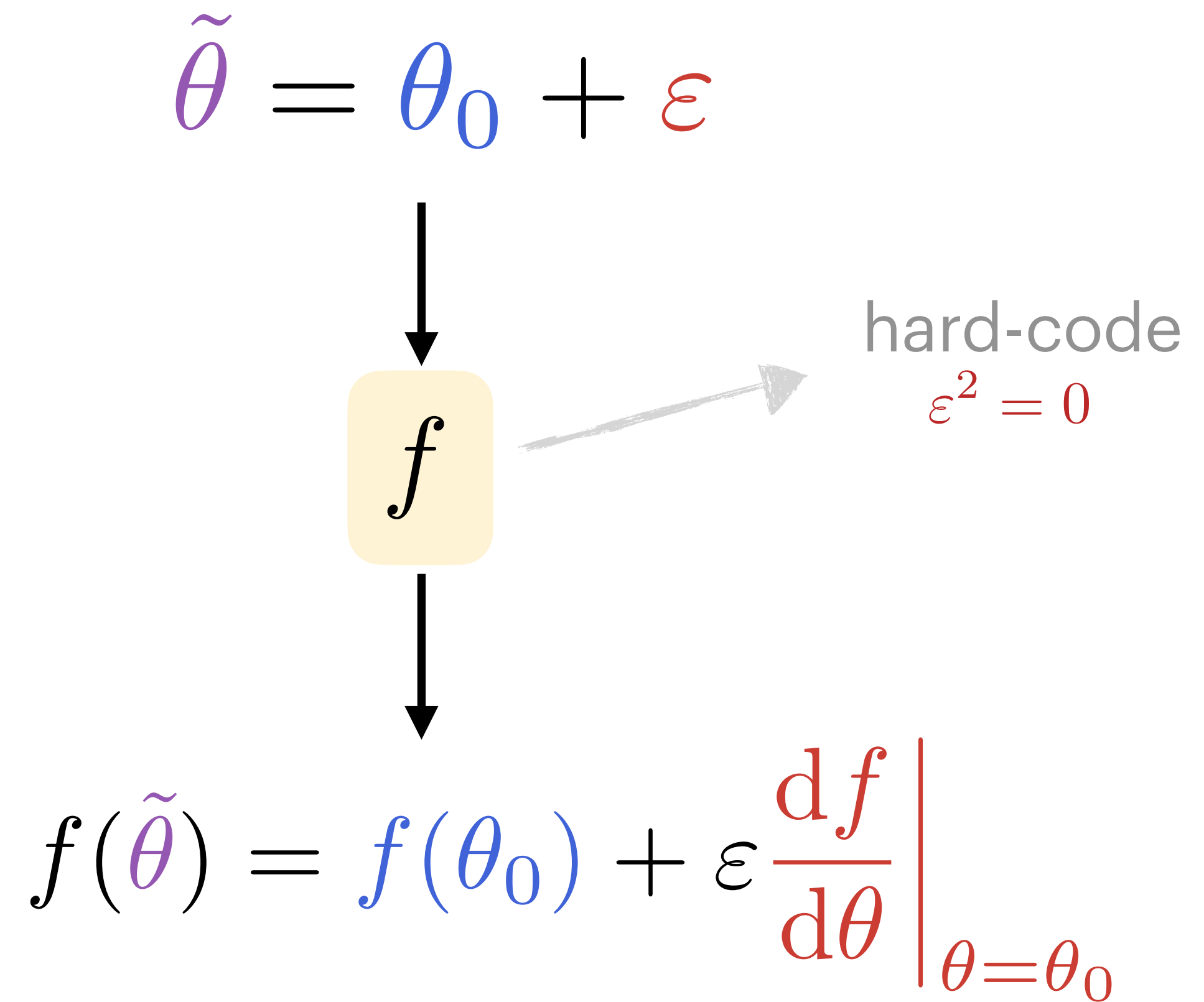
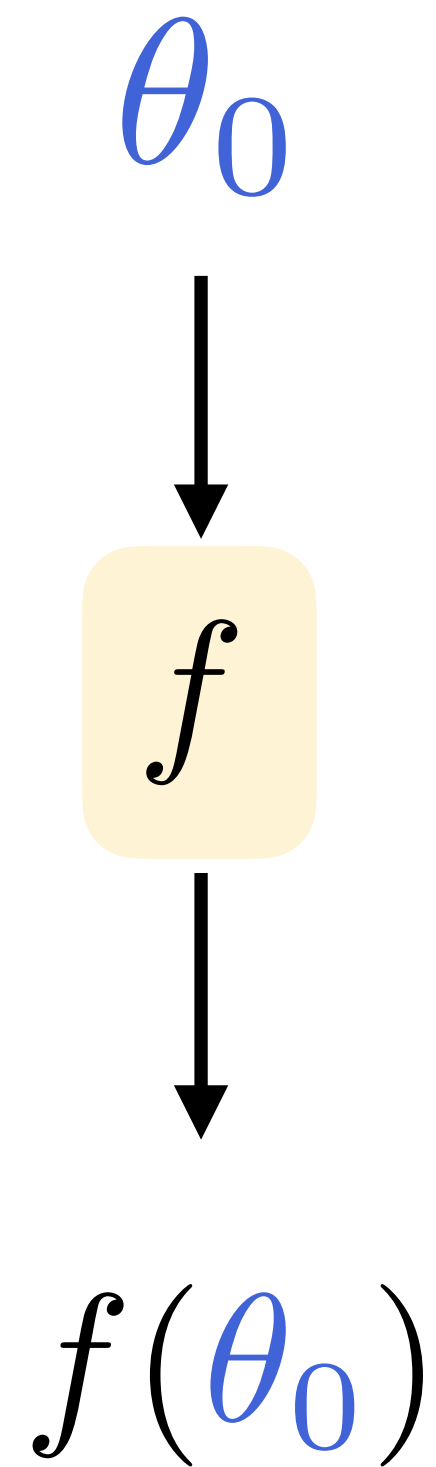
5. Repeat for N_w walkers with coupled noise

6. Repeat for all configurations

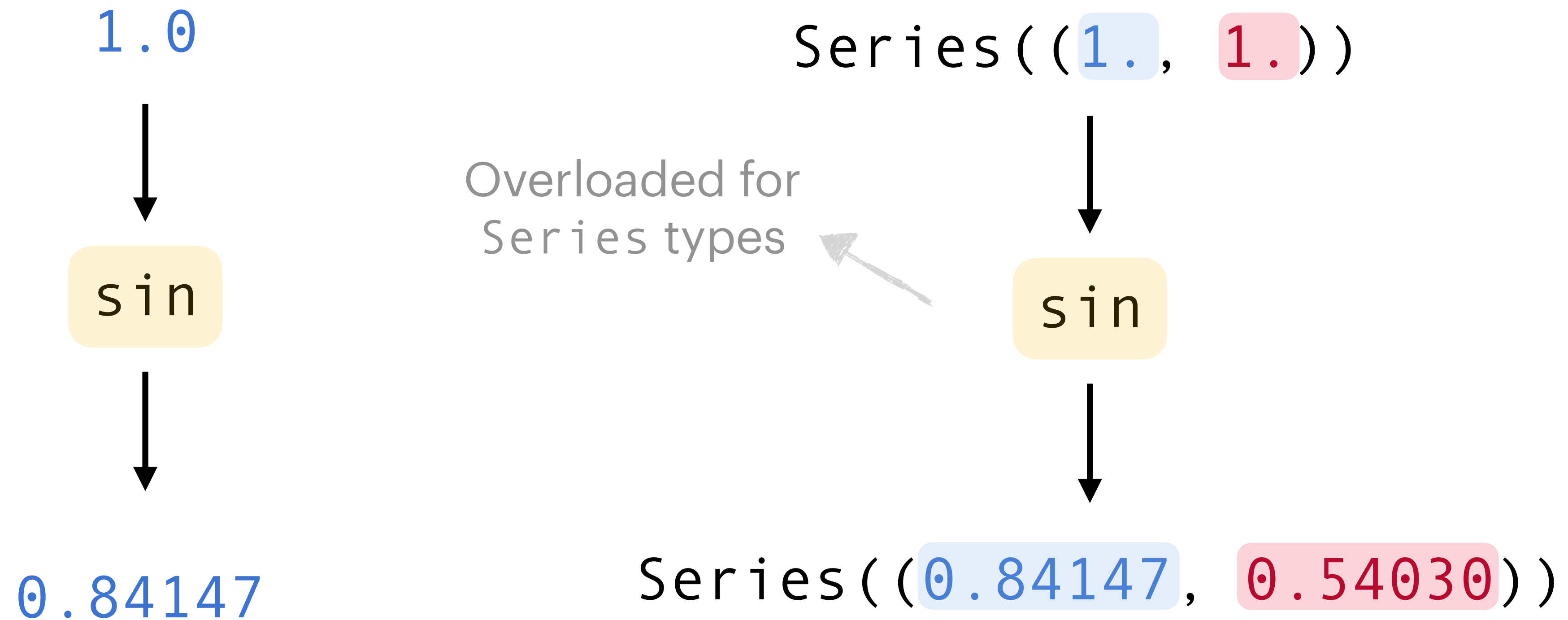
$$f(\phi) = \mathbb{E} \left[\nabla \underbrace{\int_0^\infty dt \mathcal{O}_0[Z_t]}_{\hat{\phi}^{[0,T]}} \right]$$



(forward) Automatic Differentiation



(forward) Automatic Differentiation



(forward) Automatic Differentiation

MCMC samples with S_β

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

$$\beta'$$

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

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

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

(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

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

$$\frac{\langle e^{S - S\tilde{J}} \mathcal{O} \rangle}{\langle e^{S - S\tilde{J}} \rangle} = \langle \mathcal{O} \rangle + \varepsilon \left. \frac{d}{dJ} \right|_{J=0} \langle \mathcal{O} \rangle$$

$$e^{-S}$$

Base distribution

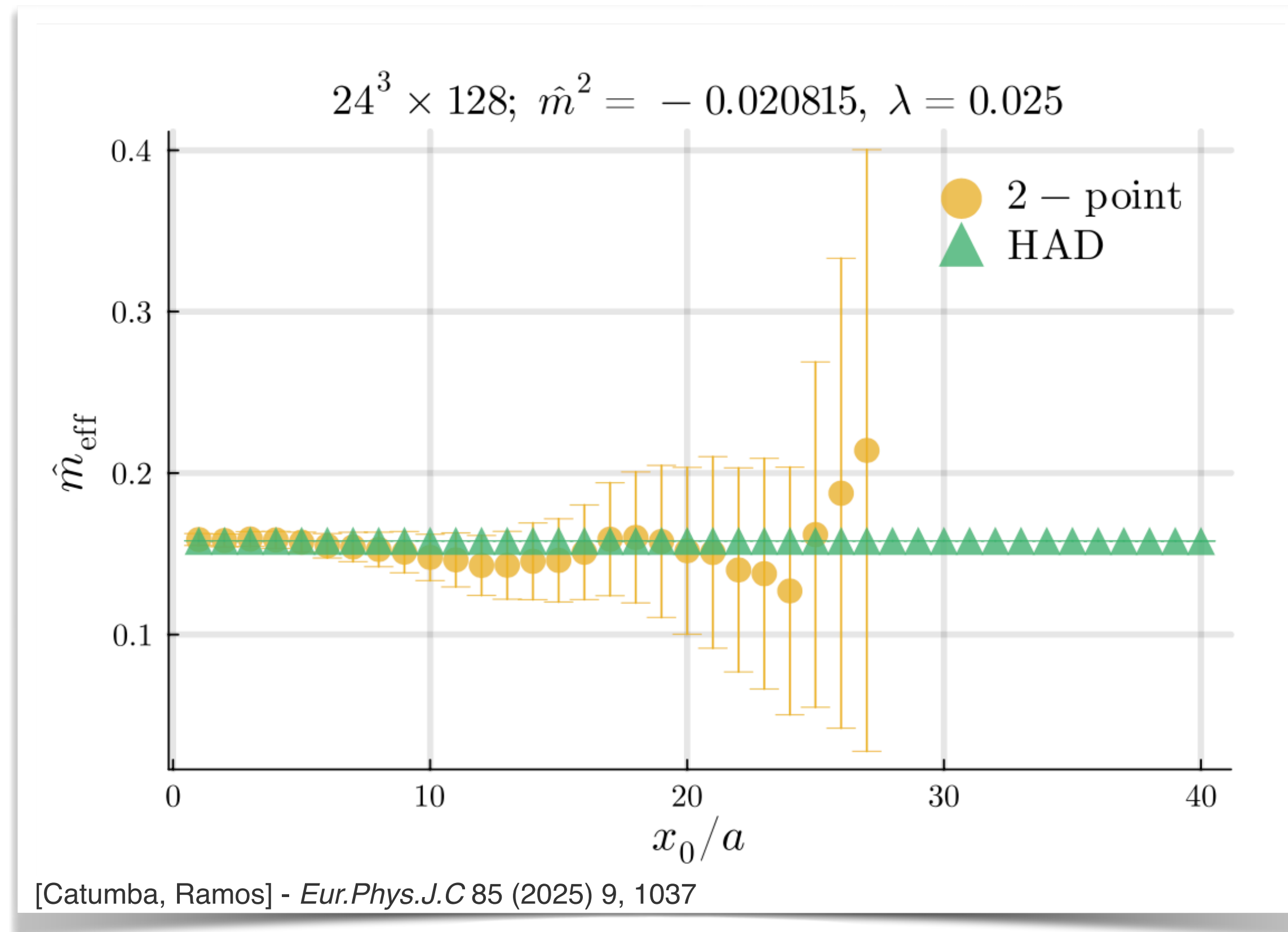
$$e^{-S+J\phi}$$

Target distribution

Stochastic (H)AD

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

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



- Promote also fields and momenta to formal series

$$\phi \rightarrow \phi + \varepsilon \quad \pi \rightarrow \pi + \varepsilon$$

- Integrate HMC equations order by order

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

- Samples now contains the dependence on the source

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

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

$$\tilde{x}\tilde{y} = x_0 y_0 + (x_0 y_1 + x_1 y_0) \varepsilon + (x_0 y_2 + x_1 y_1 + x_2 y_0) \varepsilon^2$$

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

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

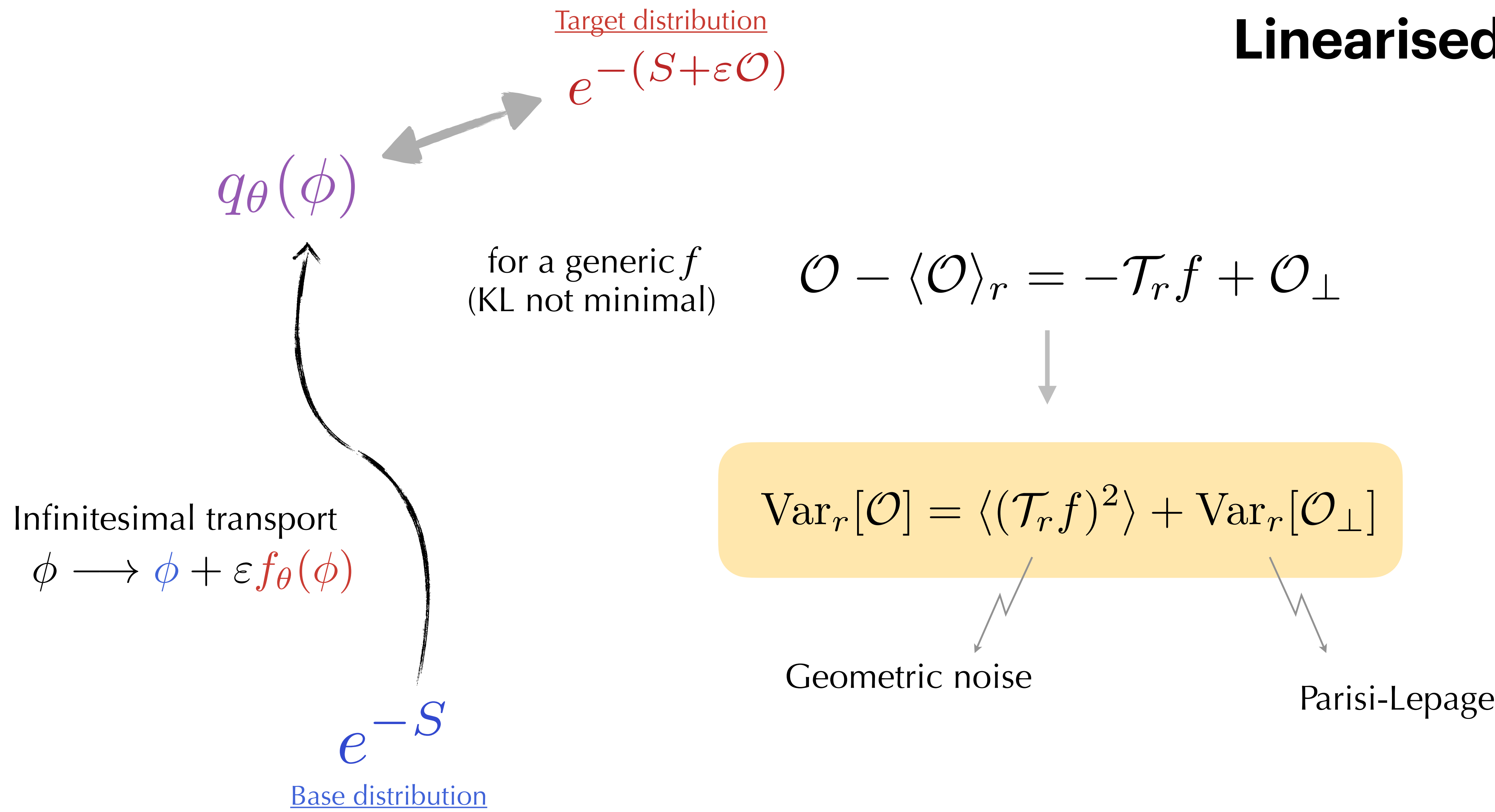
- Exact derivatives of any function f at x_0 are given by Taylor's theorem

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

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

```
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



Linearised flow

