
Bellman Defect Correction

Inference Time Algorithms
for SciML and GenAI

Yiping Lu

Beijing International Center for Mathematical Research
Peking University

More thought can improve the answer

• Flash

Quick response

YOU

A bat and a ball cost \$1.10 together.
The bat costs \$1.00 more than the ball.
What does the ball cost?

AI

The ball costs **\$0.10**.
 $1.10 - 1.00 = 0.10$

✗ Fails the check: $\$0.10 + \$1.10 = \$1.20$

• Pro

Reason and verify

YOU

A bat and a ball cost \$1.10 together.
The bat costs \$1.00 more than the ball.
What does the ball cost?

AI

Let the ball cost x . The bat then costs $x + 1$.
 $x + (x + 1) = 1.10 \implies x = 0.05$

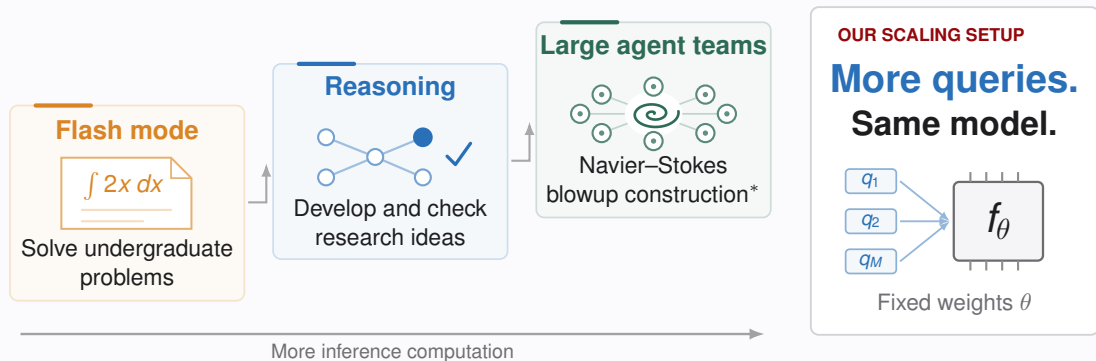
The ball costs **\$0.05**.

✓ Check: $\$0.05 + \$1.05 = \$1.10$

Same question. **More computation for reasoning and checks.**

Illustrative AI dialogues

From fast answers to scientific discovery



Which algorithm turns more queries into better answers?

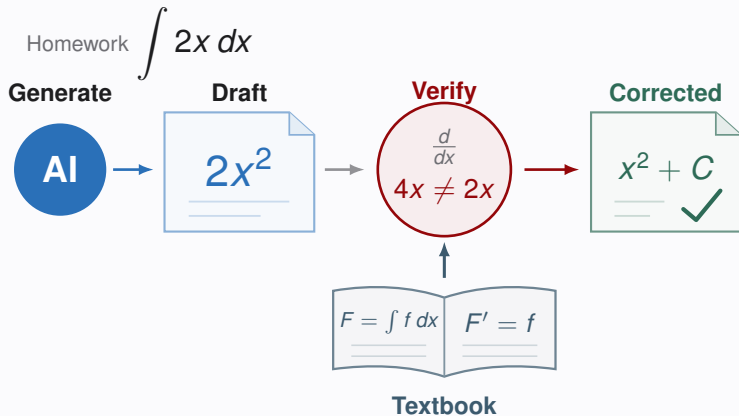
Illustrative task progression. * Reported NS case: $\sim 10,000$ agents, with model upgrades. OpenAI (2026).

Textbook knowledge checks an AI answer

MY AI PHILOSOPHY

**Generate first.
Verify with
known rules.**

The error tells us
what to correct.



Defect correction has exactly two steps

Write the Law of Defect

Derive the equation obeyed by $\check{u} := u - \hat{u}$.

1

known equation + autodiff

Simulate the Law of Defect

Spend new compute on stochastic paths for $\check{u}$.

2

linear: Feynman Kac + Monte Carlo
semilinear: multilevel Picard

$$u_{\text{corr}} = \hat{u} + \hat{\check{u}}_N$$

Write analytically. Spend inference compute stochastically.

No retraining: additional compute targets only what the model missed.

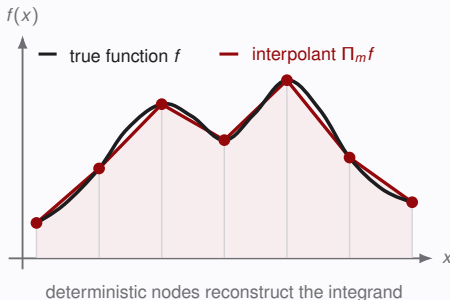
Source: SCaSML paper, structure preserving Law of Defect and stochastic correction.

Integration from random samples

Rare events, minimax rates, and optimal correction.

Linear case: function approximation

For intuition, take $q = 1$ and $I(f) = \int_{\Omega} f d\mu$.



APPROXIMATE, THEN INTEGRATE

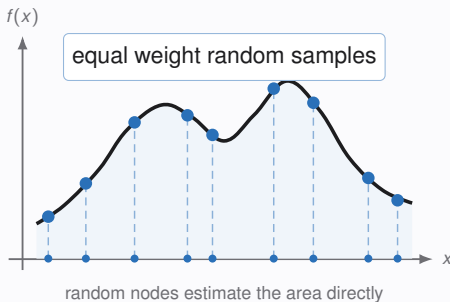
$$f \approx \Pi_m f = \sum_{j=1}^m f(x_j) \ell_j,$$

$$Q_m(f) = \int_{\Omega} \Pi_m f d\mu = \sum_{j=1}^m w_j f(x_j).$$

APPROXIMATION RATE

$$\begin{aligned} |I(f) - Q_m(f)| &\leq \|f - \Pi_m f\|_{L^1(\mu)} \\ &\lesssim m^{-s/d}. \end{aligned}$$

Linear case: Monte Carlo



SAMPLE, THEN AVERAGE

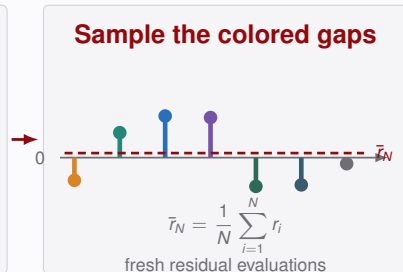
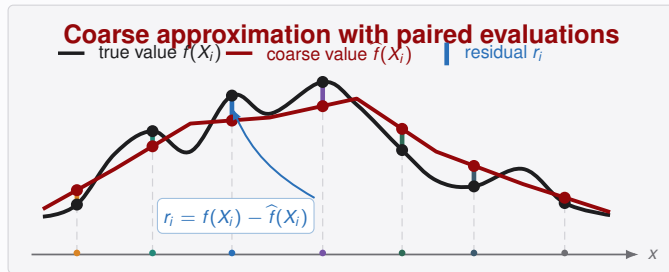
$$X_i \stackrel{\text{iid}}{\sim} \mu, \quad \hat{I}_N^{\text{MC}} = \frac{1}{N} \sum_{i=1}^N f(X_i).$$

$$\mathbb{E}[\hat{I}_N^{\text{MC}}] = I(f).$$

Sampling rate

$$\text{RMSE} = \frac{\sqrt{\text{Var}_{\mu}(f)}}{\sqrt{N}}.$$

Learn the bulk, simulate only the defect



$$I(f) = I(\hat{f}) + \mathbb{E}[r(X)], \quad \hat{I}_{\text{CV}} = I(\hat{f}) + \bar{r}_N.$$

Training learns the bulk; inference averages only what the surrogate missed.

Residual Monte Carlo multiplies the two rates

Condition on a surrogate trained with independent data.

$$\mathbb{E}[\hat{l}_{\text{CV}} \mid \hat{f}] = l(f), \quad \text{RMSE}(\hat{l}_{\text{CV}} \mid \hat{f}) \leq \frac{\|f - \hat{f}\|_{L^2(\mu)}}{\sqrt{N}}.$$

Two independent gains multiply

$$\underbrace{\|f - \hat{f}_m\|_2}_{O(m^{-\gamma})} \times \underbrace{N^{-1/2}}_{\text{residual samples}} = \underbrace{O(m^{-\gamma} N^{-1/2})}_{\text{corrected RMSE}}.$$

MATCHED BUDGETS: $N \asymp m$ $\text{RMSE} = O(m^{-\gamma-1/2})$

Source: Blanchet, Chen, Lu, and Ying, arXiv:2305.16527, Theorems 4 and 5 for the noisy linear functional.

Moment estimation with rare peaks

Estimate a moment from random function values

$$\mathcal{D}_n = \{(X_i, f(X_i))\}_{i=1}^n, \quad X_i \stackrel{\text{iid}}{\sim} \text{Unif}(\Omega)$$

$\Omega = [0, 1]^d$, n exact function values



$$T_q(f) = \int_{\Omega} f(x)^q dx = \mathbb{E}[f(X)^q]$$

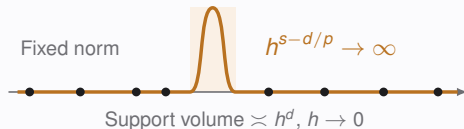
$q = 1$: integral $q \geq 2$: nonlinear moment

Sobolev regularity allows rare peaks

$$\mathcal{F} = \{f : \|f\|_{W^{s,p}(\Omega)} \leq 1\}, \quad \alpha = s/d$$

$$W^{s,p} \hookrightarrow L^{p^*}, \quad \frac{1}{p^*} = \frac{1}{p} - \alpha$$

$0 < \alpha < 1/p$: finite p^* , no uniform height bound



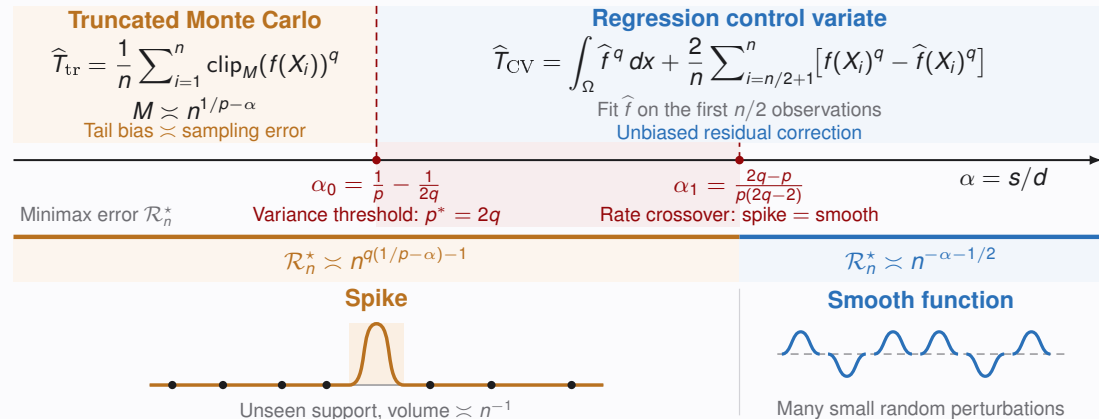
What is the optimal estimation rate?

$$\mathcal{R}_n^* = \inf_{\hat{T} = \hat{T}(\mathcal{D}_n)} \sup_{f \in \mathcal{F}} \mathbb{E} |\hat{T} - T_q(f)|$$

Moment theorem: integer $q \geq 2$, $p > 2$, $q < p < 2q$. Expected absolute error. Boundary cases omitted.

Source: Blanchet, Chen, Lu, and Ying, [arXiv:2305.16527](https://arxiv.org/abs/2305.16527), Section 2, Figure 1, Sections 3.1 and 3.2.

Optimal algorithms and minimax rates



Expected absolute error. CV: optimal regression oracle, exact surrogate integral. Strict regimes, schematic axis.

Source: Blanchet, Chen, Lu, and Ying, [arXiv:2305.16527](https://arxiv.org/abs/2305.16527), Theorems 1 to 3, Figure 2, Assumption 3.1.

Influence functions control nonlinear variance

Sample split estimation of $T_q(f) = \int_{\Omega} f^q$, $q \geq 2$

$$\hat{T}_{q,\text{CV}} = \int_{\Omega} \hat{f}^q + \frac{2}{n} \sum_{i=n/2+1}^n [f(X_i)^q - \hat{f}(X_i)^q]$$

Influence expansion

$$e = f - \hat{f}$$

$$f^q - \hat{f}^q = \underbrace{q f^{q-1}}_{\text{IF}_f(x)} e + R_q$$

$$\text{Var}(\hat{T}_{q,\text{CV}} \mid \hat{f}) \leq \frac{C_q}{n} \left(\underbrace{\int_{\Omega} |f|^{2q-2} |e|^2}_{\text{A: influence fluctuation}} + \underbrace{\int_{\Omega} |e|^{2q}}_{\text{B: nonlinear propagation}} \right)$$

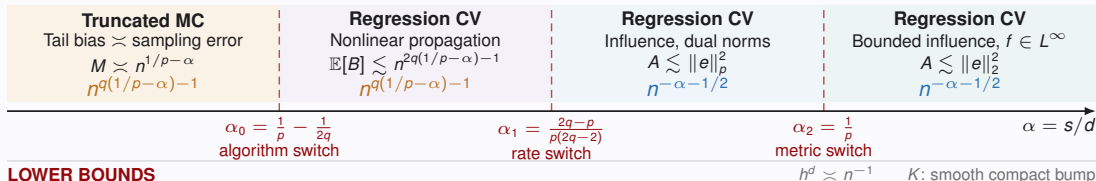
Source: Blanchet, Chen, Lu, and Ying, arXiv:2305.16527, equation 7 and Theorem 2.

Proof mechanisms behind the phase diagram

$$e = f - \hat{f}, \quad f^q - \hat{f}^q = \underbrace{q f^{q-1} e}_{\text{IF}_f} + R_q, \quad n \text{Var}(\hat{T}_{q,\text{CV}} | \hat{f}) \lesssim \underbrace{\|f^{q-1} e\|_2^2}_{\text{A: influence}} + \underbrace{\|e\|_{2q}^{2q}}_{\text{B: propagation}}.$$

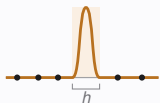
$$\mathcal{R}_n^* \asymp \max\{n^{q(1/p-\alpha)-1}, n^{-\alpha-1/2}\}, \quad \alpha = s/d, \quad p > 2, \quad q < p < 2q.$$

UPPER BOUNDS



LOWER BOUNDS

Spike

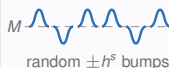


$$g(x) = h^{s-d/p} K\left(\frac{x-c}{h}\right)$$

Unseen support, $\Pr(\text{miss}) \asymp 1$

$$\int g^q \asymp n^{q(1/p-\alpha)-1}$$

Smooth function



$$f_\eta(x) = M + h^s \sum_j \eta_j K\left(\frac{x-c_j}{h}\right)$$

Sign bias $\asymp n^{-1/2}$, $\eta_j \in \{\pm 1\}$

$$\Delta T_q \asymp h^s n^{-1/2} = n^{-\alpha-1/2}$$

Expected absolute error, bounded Sobolev ball. Boundary cases omitted. CV upper bounds: regression oracle.

Source: Blanchet, Chen, Lu, and Ying, [arXiv:2305.16527](https://arxiv.org/abs/2305.16527), Theorems 1 to 3, Figure 2, Appendix B.2.

A two scale method

$$f(x) = \underbrace{\hat{f}(x)}_{\text{smooth, coarse scale}} + \underbrace{r(x)}_{\text{fine scale residual}}.$$

Smooth component



Function approximation

Small scale, high frequency residual



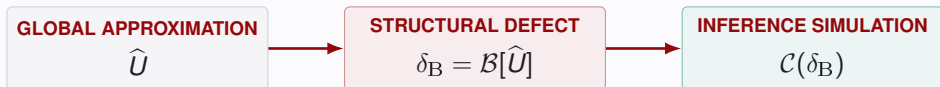
Monte Carlo

$$\hat{I} = \int_{\Omega} \hat{f} d\mu + \frac{1}{N} \sum_{i=1}^N r(X_i), \quad X_i \stackrel{\text{iid}}{\sim} \mu.$$

$$\text{RMSE}(\hat{I} | \hat{f}) = \frac{\sqrt{\text{Var}_{\mu}(r)}}{\sqrt{N}} \leq \frac{\|r\|_{L^2(\mu)}}{\sqrt{N}}.$$

Small residual energy controls the variance, not frequency alone.

Bellman residuals lift the same identity from integrals to paths



$$U_{\text{corr}} = \hat{U} + \mathcal{C}(\delta_B).$$

INTEGRATION

$\hat{f}$
 $f - \hat{f}$
 residual Monte Carlo

SCIENTIFIC ML

$\hat{u}$
 PDE Bellman residual
 Feynman Kac or MLP

GENERATIVE AI

$\hat{V}$ or H
 Bellman or Doob defect
 SMC, Girsanov, or two queries

The defect is computable because the governing structure is known.

Scientific machine learning

SCaSML: simulate the model's defect.

Linear case: subtraction gives the defect law

EQUATION 1 GOVERNING EQUATION: THE BELLMAN EQUATION

$$\partial_t u + \mathcal{L}u = f, \quad u(T, \cdot) = g,$$

EQUATION 2 ANOTHER BELLMAN EQUATION THAT THE SURROGATE SATISFIED

$$\partial_t \hat{u} + \mathcal{L}\hat{u} = \hat{f}, \quad \hat{u}(T, \cdot) = \hat{g}.$$

EQUATION 1 MINUS EQUATION 2

$$\underbrace{(\partial_t + \mathcal{L})(u - \hat{u})}_{(\partial_t + \mathcal{L})\check{u}} = \underbrace{f - \hat{f}}_{\text{Bellman error } \delta_B}$$

Law of Defect: Value evaluation with Bellman Error

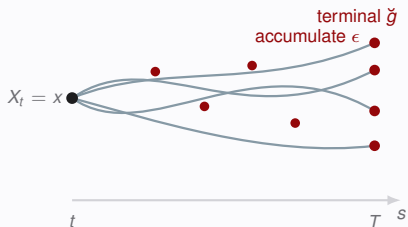
$$\check{u} := u - \hat{u}, \quad \check{g} := g - \hat{g}, \quad \delta_B := f - \hat{f} = -\epsilon,$$

$$(\partial_t + \mathcal{L})\check{u} = \delta_B, \quad \check{u}(T, \cdot) = \check{g}.$$

Feynman Kac simulates only the defect

SAMPLE THE GENERATOR

$$X_t = x, \quad dX_s = \mu(s, X_s) ds + \sigma(s, X_s) dW_s.$$



FEYNMAN KAC REPRESENTATION

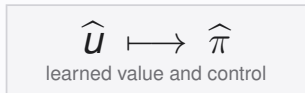
$$\begin{aligned} \Delta &:= \check{g}(X_T) - \int_t^T \delta_B(s, X_s) ds \\ &= \check{g}(X_T) + \int_t^T \epsilon(s, X_s) ds, \\ \check{u}(t, x) &= \mathbb{E}_{t,x}[\Delta]. \end{aligned}$$

MONTE CARLO REALIZATION

$$\hat{u}_{\text{CORR}}^{(N)}(t, x) = \hat{u}(t, x) + \frac{1}{N} \sum_{n=1}^N \Delta^{(n)}.$$

Simulator feedback for a learned controller

Learn a controller



current solution

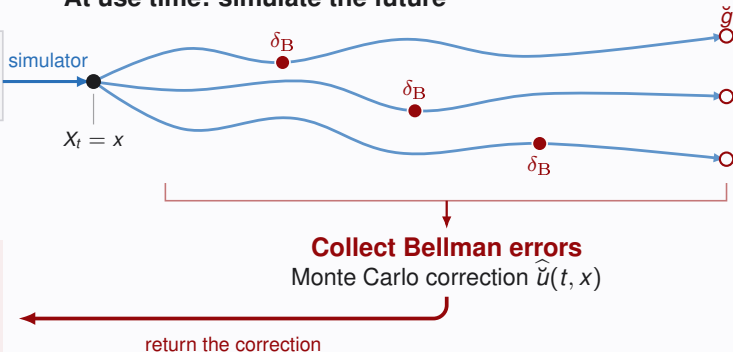
$$\hat{u}(t, x)$$

Correct the current value

$$\hat{U}_{\text{corr}} = \hat{U} + \hat{\tilde{U}}$$

update the control from $\hat{U}_{\text{corr}}$

At use time: simulate the future



Source: SCaSML, Law of Defect and stochastic correction.

Defect size controls the Monte Carlo variance

Monte Carlo Factor

$$\text{Var}\left(\hat{u}_{\text{corr}}^{(N)} \mid \hat{u}\right) = \frac{\text{Var}(\Delta \mid \hat{u})}{N}.$$

Defect Factor Condition on $\hat{u}$ and assume $\|\check{g}\|_{\infty}, \|\epsilon\|_{\infty} < \infty$.

$$|\Delta| \leq \|\check{g}\|_{\infty} + (T - t)\|\epsilon\|_{\infty},$$

The errors multiply

$$\|\check{g}\|_{\infty} \leq C_T e(\hat{u}), \|\epsilon\|_{\infty} \leq C_{\epsilon} e(\hat{u}) \implies \boxed{\text{RMSE}_{\hat{u}} \leq (C_T + (T - t)C_{\epsilon}) \frac{e(\hat{u})}{\sqrt{N}}}.$$

Use the surrogate as a control variate.

Better surrogate model $\rightarrow$ Lower correction variance

Nonlinear case: subtraction gives the defect law

Semi-Linear Equation

$$\mathcal{N}[v] := \partial_t v + \mathcal{L}v + F(v, \sigma^\top \nabla v).$$

EQUATION 1

$$\mathcal{N}[u] = f, \quad u(T) = g.$$

EQUATION 1 MINUS EQUATION 2

$$\begin{aligned} \partial_t \check{u} + \mathcal{L}\check{u} + F(\hat{u} + \check{u}, \hat{z} + \check{z}) - F(\hat{u}, \hat{z}) &= \underbrace{f - \hat{f}}_{\text{Bellman error } \delta_B} = -\epsilon, \\ \check{u}(T, \cdot) &= g - \hat{g}. \end{aligned}$$

EQUATION 2

$$\mathcal{N}[\hat{u}] = \hat{f}, \quad \hat{u}(T) = \hat{g}.$$

$$\text{Law of Defect: } \partial_t \check{u} + \mathcal{L}\check{u} + \check{F}(\check{u}, \check{z}) = 0.$$

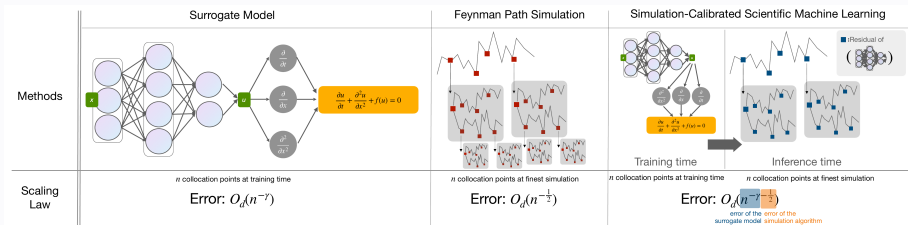
Key Observation: Closeness

$$\check{F}(v, z) := F(\hat{u} + v, \hat{z} + z) - F(\hat{u}, \hat{z}) + \epsilon$$

$$\check{u} := u - \hat{u}, \hat{z} := \sigma^\top \nabla \hat{u}, \check{z} := \sigma^\top \nabla \check{u}.$$

Training and inference exponents add

$$e(\hat{u}) = O(m^{-\gamma}), \quad N \asymp m \implies \text{Err}_{\text{SCaSML}} = O\left(m^{-\gamma} N^{-1/2+o(1)}\right) = O\left(m^{-\gamma-1/2+o(1)}\right).$$



TRAINING

 γ

+

INFERENCE

 $\frac{1}{2}$

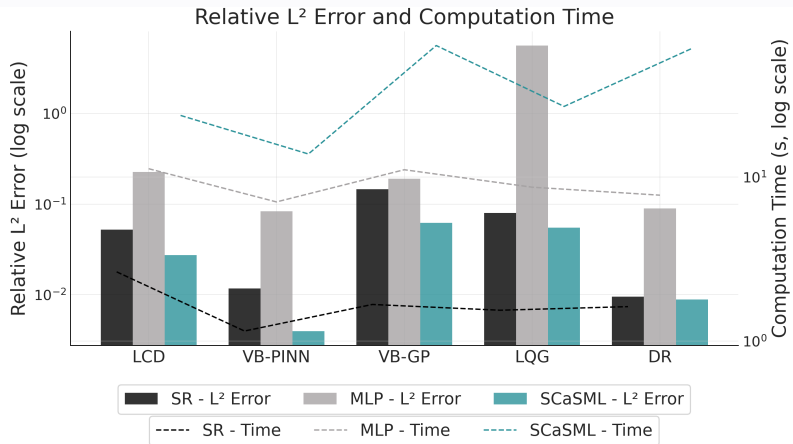
=

CORRECTED EXPONENT

 $\gamma + \frac{1}{2}$

Monte Carlo multiplies the surrogate error; it does not replace it.

SCaSML spends inference compute to buy accuracy



More inference compute lowers relative L^2 error across all five benchmarks.

Diffusion model

URGE: correct approximate guidance with path weights.

Guidance turns “at work” into a sampling direction

Suppose a diffusion model already knows how José looks.

BASE DISTRIBUTION $X_t \sim p_t$, public images of José



REWARD MODEL

Is José at work?

$G = \mathbb{P}(\text{José at work})$

extra drift follows predicted reward

$V(t)^2 \nabla G_t(X_t)$



GUIDED SAMPLE synthetic target



synthetic illustration • José at work

The base model supplies identity; the reward specifies the task; guidance changes the proposal.

Guidance follows the value, not the reward

BACKWARD VALUE

The terminal reward is propagated to time t through

$$h(x, t) = \mathbb{E}_{\mathbb{P}}[e^{r(X_T)} \mid X_t = x],$$

— Intractable in general

Ideal Doob guidance

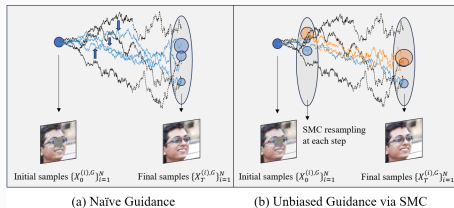
$$V(t)^2 \nabla \log h(x, t)$$

value gradient, **not** reward gradient $V(t)^2 \nabla_x r(x, t)$

PRACTICAL PROPOSAL

$$G \approx \log h \implies V(t)^2 \nabla \log h \longrightarrow V(t)^2 \nabla G.$$

$$G \neq \log h \implies \mathbb{P}^G \neq \mathbb{Q}: \text{correction is required.}$$



Reward scores the endpoint.

Value guides the path.

Girsanov gives the exact path correction

GUIDED PROPOSAL TO BASE DYNAMICS

$$\frac{d\mathbb{P}}{d\mathbb{P}^G}(X_{0:t}^G) = \exp\left(-\int_0^t V \nabla G^\top dW - \frac{1}{2} \int_0^t V^2 \|\nabla G\|_2^2 ds\right).$$

REWARD TILT

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X_{0:t}) \propto e^{r(X_t, t) - r(X_0, 0)}.$$

COMBINED PATH CORRECTION

$$\frac{d\mathbb{Q}}{d\mathbb{P}^G}(X_{0:t}^G) \propto \exp\left[r(X_t^G, t) - r(X_0^G, 0) - \int_0^t V \nabla G^\top dW - \frac{1}{2} \int_0^t V^2 \|\nabla G\|_2^2 ds\right].$$

Key advantage

The correction weight needs no reward gradient, reward Laplacian, or score.

URGE corrects diffusion paths sequentially

$$X_{k+1}^{i,G} = X_k^{i,G} + (v_k^i + V_k^2 \nabla G_k^i) \Delta t_k + V_k \sqrt{\Delta t_k} \xi_k^i, \quad \xi_k^i \sim \mathcal{N}(0, I).$$

$$\log \beta_k^i = r(X_{k+1}^{i,G}, t_{k+1}) - r(X_k^{i,G}, t_k) - V_k (\nabla G_k^i)^\top \sqrt{\Delta t_k} \xi_k^i - \frac{1}{2} V_k^2 \|\nabla G_k^i\|_2^2 \Delta t_k.$$

Sample particles

Update path weights

Normalize and resample

Particles at t_k

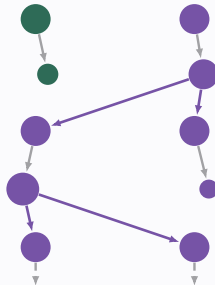
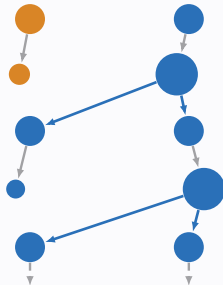
Sample and weight

Resample

Sample and weight

Resample

Color = ancestry



$N = 4$

Area = weight

Equal weights

Fresh noise per copy

Equal weights

Illustrative resampling outcomes

URGE is exact in an iterated limit

Iterated consistency

Under the paper's regularity assumptions and correct initial law,

$$\lim_{N \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \mathbb{E}_{\mathbb{P}}[M_T^{N, \Delta t}(\varphi)] = \frac{\mathbb{E}_{\mathbb{P}}[e^{r(X_T) - r(X_0, 0)} \varphi(X_T, T)]}{\mathbb{E}_{\mathbb{P}}[e^{r(X_T) - r(X_0, 0)}]}.$$

Scope of the theorem

Asymptotic consistency only: the result gives no finite N total variation or sample complexity bound.

Path and particle corrections agree only infinitesimally

Define the terminal conditioned instantaneous path weight rate

$$\lambda(x, t) = \lim_{s \uparrow t} \frac{\mathbb{E}_{\mathbb{P}^G}[w_{s,t}^{\text{URGE}} \mid X_t^G = x] - 1}{t - s}.$$

If the guided marginal density is sufficiently smooth, then $\lambda(x, t) = w_{\text{particle}}(x, t)$.

Previous Works:

$$w_{\text{particle}}(x, t) = \partial_t r - \frac{V^2}{2}(\Delta r + \|\nabla r\|^2) + \nabla r^\top (v + V^2 \nabla G) + V^2 \nabla_x \log p^\top \nabla_x (G - r) + V^2 \Delta G$$

PATH SPACE

One weight for the whole guided trajectory

$w_{s,t}^{\text{URGE}}$

Retains path history; no parent cloning.

$\lambda = w_{\text{particle}}$
only as $s \uparrow t$

$\neq$
finite $\Delta t, N$

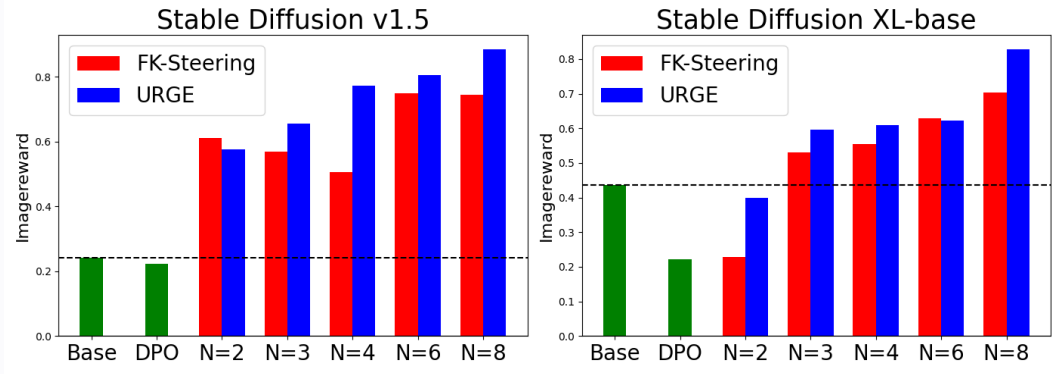
PARTICLE RESAMPLING

Local selection at each solver step

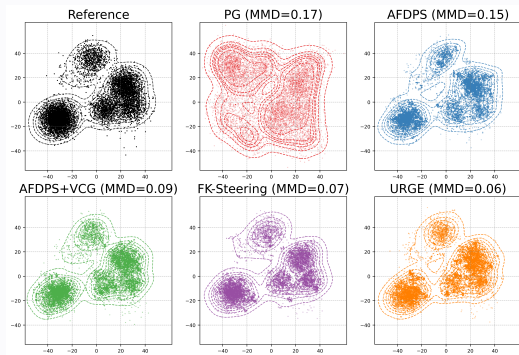
$w_{\text{particle}}(x, t)$

Kills or duplicates particles.

Inference scaling with more particles



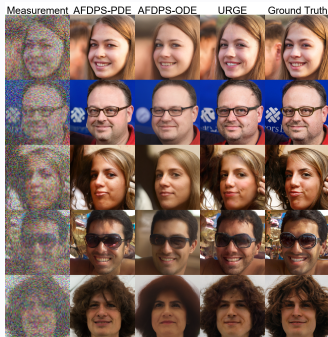
Pathwise correction recovers the target modes



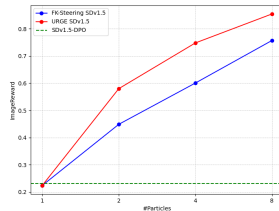
On the reported 30 dimensional Gaussian mixture, URGE gives the closest sample cloud to the analytical target, with **MMD 0.06**.

From inverse problems to neural rewards

FFHQ 256 SUPER RESOLUTION

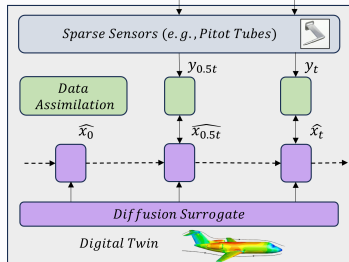
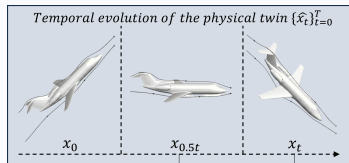


SD V1.5 PARTICLE SCALING



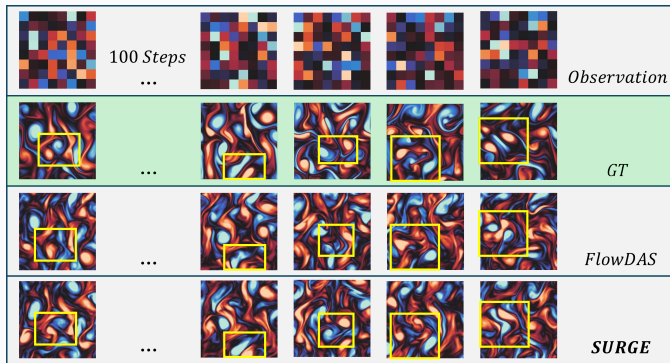
URGE matches derivative based restoration;
ImageReward exceeds FK Steering for
 $N = 2, 4, 8$.

Data Assimilation



REWARD

Reward measures **alignment with reality**.



Numerical analysis

Mesh uniform diffusion inference.

Micro SMC genealogy

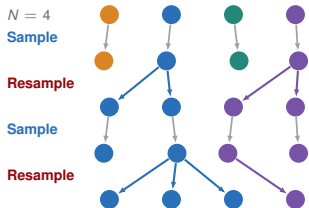
(a) Micro SMC genealogy

$$f_i = f(X_n^i) \text{ and } \bar{f}_N = N^{-1} \sum_i f_i$$

$$\text{Var}(\bar{f}_N) = N^{-1} \text{Var}(f_1) + (1 - N^{-1}) \text{Cov}(f_1, f_2)$$

marginal $\leq 1/(4N)$ $\text{Cov}(f_1, f_2) \asymp \min\{n/N, 1\}$

Shared ancestry induces covariance

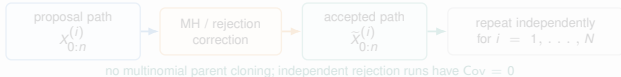


Equal weights. Color = ancestry.

$$\text{local ESS} = N$$

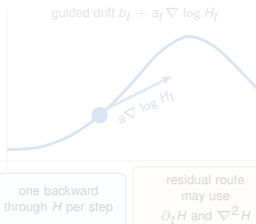
$$\text{MSE} \gtrsim \min\{n/N, 1\}$$

Independent paths: one-by-one generation + MH / rejection



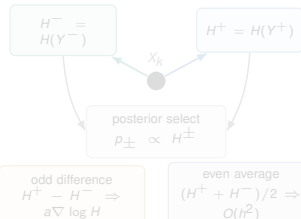
(b) Gradient Doob proposal

one value-model backward at every step



(c) 2Q-Doob proposal

two value forwards; zero derivatives



Perfect ESS can hide nonconvergence

NEUTRAL BROWNIAN CHAIN

$$X_p^{i,N} = X_{p-1}^{A_p^i,N} + n^{-1/2} \xi_p^i,$$

$$A_p^i \sim \text{Unif}\{1, \dots, N\}, \quad \xi_p^i \sim \mathcal{N}(0, 1).$$

Every local check passes

$\phi \equiv 1$, $\text{ESS}_p = N$, and

$X_n^{i,N} \sim \mathcal{N}(0, 1)$ exactly.

Every one particle marginal is exact.

Yet the empirical mean does not stabilize

For $g(x) = \sin(x)/2$,

$$\mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N g(X_n^{i,N}) \right)^2 \right] \geq \frac{1}{32e^2} \min \left\{ \frac{n}{N}, 1 \right\}.$$

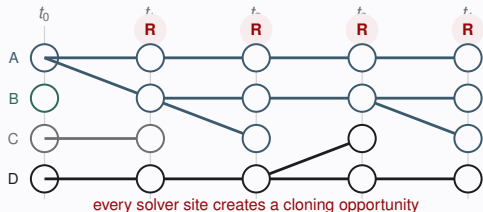
For fixed N , the error stays nonzero as $n \rightarrow \infty$.

RMSE $\leq \varepsilon$ forces $N = \Omega(n\varepsilon^{-2})$ and work $= \Omega(n^2\varepsilon^{-2})$.

Every microstep can become a statistical generation

FULL MICRO RESAMPLING

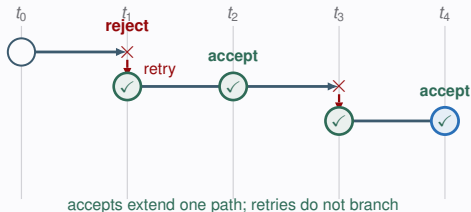
one parent pool



$$A_k^i \sim \text{Cat}(W_{k-1}^{1:N}), \quad X_k^i \sim \mathcal{M}_{k,n}(X_{k-1}^{A_k^i}, \cdot)$$

EXACT CHILD REJECTION

one sample; serial retries



$$Y \sim \mathcal{M}_{k,n}(x, \cdot), \quad U \sim \text{Unif}(0, 1)$$

$$\text{accept if } U \leq \frac{G_{k,n}(x, Y)}{\exp\{m_{k,n}(x) + a_{k,n}\}}$$

For N samples, repeat this one path sampler independently; parent normalizers require exact outer correction.

Resampling **couples lineages**. Rejection generates **one path at a time**.

Gradient Doob proposal

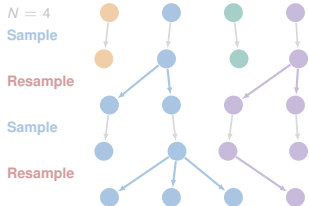
(a) Micro SMC genealogy

$$f_i = f(X_N^i) \text{ and } \bar{f}_N = N^{-1} \sum_i f_i$$

$$\text{Var}(\bar{f}_N) = N^{-1} \text{Var}(f_1) + (1 - N^{-1}) \text{Cov}(f_1, f_2)$$

marginal $\leq 1/(4N)$ $\text{Cov}(f_1, f_2) \asymp \min\{n/N, 1\}$

Shared ancestry induces covariance



Equal weights. Color = ancestry.

$$\text{local ESS} = N$$

$$\text{MSE} \gtrsim \min\{n/N, 1\}$$

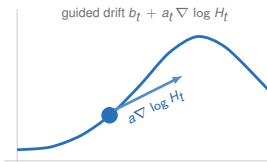
Independent paths: one-by-one generation + MH / rejection



no multinomial parent cloning; independent rejection runs have $\text{Cov} = 0$

(b) Gradient Doob proposal

one value-model backward at every step

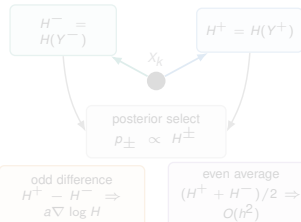


one backward
through H per step

residual route
may use
 $\partial_t H$ and $\nabla^2 H$

(c) 2Q-Doob proposal

two value forwards; zero derivatives



Direct guidance creates a Laplacian bottleneck

Gradient Doob proposal

$$dX_t^H = (b_t + a_t \nabla \log H_t)(X_t^H) dt + \sigma_t(X_t^H) dB_t$$

∇H_t at every solver step

Local residual repair

$$r_H(t, x) = \frac{(\partial_t + \mathcal{L}_t)H_t(x)}{H_t(x)}$$

Time derivative $\partial_t H_t$

$$\mathcal{L}_t H_t = b_t \cdot \nabla H_t + \frac{1}{2} a_t : \nabla^2 H_t$$

Hessian $\nabla^2 H_t$ Isotropic noise: ΔH_t

**How can we avoid the Laplacian
and still benefit from guidance?**

2Q Doob proposal

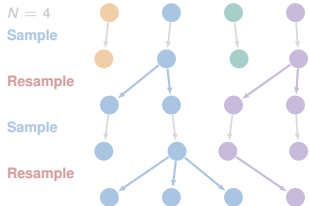
(a) Micro SMC genealogy

$$f_i = f(X_n^i) \text{ and } \bar{f}_N = N^{-1} \sum_i f_i$$

$$\text{Var}(\bar{f}_N) = N^{-1} \text{Var}(f_1) + (1 - N^{-1}) \text{Cov}(f_1, f_2)$$

marginal $\leq 1/(4N)$ $\text{Cov}(f_1, f_2) \asymp \min\{n/N, 1\}$

Shared ancestry induces covariance



Equal weights. Color = ancestry.

$$\text{local ESS} = N$$

$$\text{MSE} \gtrsim \min\{n/N, 1\}$$

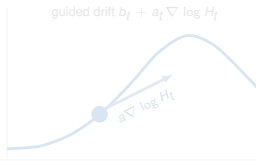
Independent paths: one-by-one generation + MH / rejection



no multinomial parent cloning; independent rejection runs have $\text{Cov} = 0$

(b) Gradient Doob proposal

one value-model backward at every step

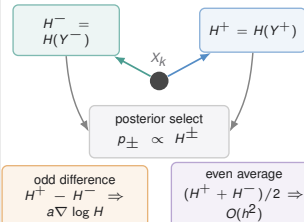


one backward through H per step

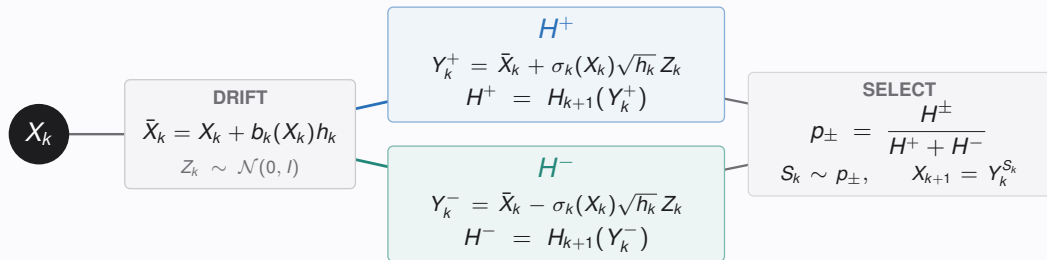
residual route may use $\partial_t H$ and $\nabla^2 H$

(c) 2Q-Doob proposal

two value forwards; zero derivatives



Two forward values recover the Doob drift



Odd part gives guidance

$H^+ - H^-$ biases the selected sign.

$$\mathbb{E}[Y_k^{S_k} - X_k] = h_k(b_k + a_k \nabla \log H_k) + O(h_k^{3/2})$$

Even part stabilizes correction

$$\text{RelVar}\left(\frac{H^+ + H^-}{2}\right) = O(h_k^2)$$

One child has relative variance $O(h_k)$.

Child selection and exact outer correction

A FINAL POOL CHOOSES THE CHILD

$$\mathcal{U}^{\text{final}} = \{Y_i\}_{i=1}^K,$$

$$S_{k,\text{final}} = \frac{1}{K} \sum_{i=1}^K H_{k+1}(Y_i),$$

$$p_i = \frac{H_{k+1}(Y_i)}{K S_{k,\text{final}}},$$

$$J \sim \text{Cat}(p_1, \dots, p_K), \quad \boxed{X_{k+1} = Y_J}.$$

B INITIAL POOL CORRECTS THE PATH WEIGHT

At step k , use the scalar saved **before** local MH:

$$\boxed{\Delta \log R_k = \log S_{k,0} - \log H_k(X_k)}$$

$$\log R = \log H_0(x_\star) + \sum_{k=0}^{n-1} \Delta \log R_k,$$

$$R = H_0(x_\star) \prod_{k=0}^{n-1} \frac{S_{k,0}}{H_k(X_k)}.$$

Final $\mathcal{U}^{\text{final}}$ selects $X_{k+1} \neq$

stored initial $S_{k,0}$ corrects R .

Width and local depth buy different guarantees

POOL WIDTH K CONTROLS CURVATURE NOISE

$$\text{Var}\left(\frac{S}{G}\right) \leq \frac{h^2 C_2^2}{2(K-1)}.$$

$$\Delta \leq 4A + \frac{C_2^2 Th}{2(K-1)}, \quad \text{MSE} \leq \frac{8e^\Delta}{N}.$$

More candidates reduce the variance constant as $1/(K-1)$.

Bounded A , fixed K, r : $N = O(\varepsilon^{-2})$, queries = $O(n\varepsilon^{-2})$.

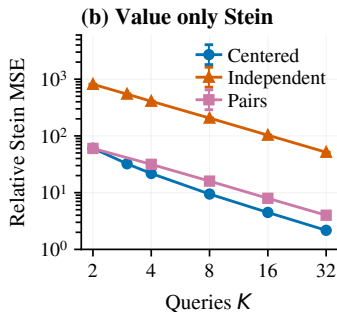
A measures accumulated dynamic inconsistency of H . At fixed raw query budget, width improves the curvature term by at most $2\times$ over $K=2$. Larger K supports parallel batching and odd pool widths.

LOCAL DEPTH r CONTROLS PROPOSAL ERROR

$$\|\mathcal{L}_{h,K,r} - \Pi_n\|_{\text{TV}} \leq \tanh\left(\frac{A}{2}\right) + C_r \sqrt{T} \frac{h^{r+1/2}}{(K-1)^{(r+1)/2}}.$$

Each local update gains one power of h in the correctable term.

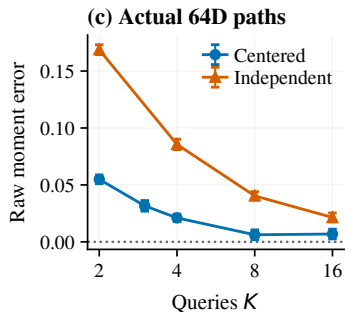
Centering sharpens value only guidance



21.8×

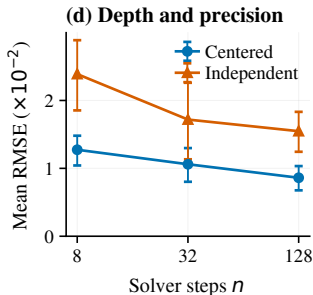
lower guidance MSE than independent candidates, $K = 8$.

$d = 64$. **Scalar value queries only. No gradient evaluations.**



0.6877 centered **0.7522 independent**
raw variance moment at $K = 4$, with target $2/3$.

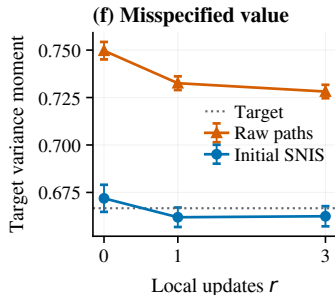
Depth refinement and correct path weights



FIXED $N = 128$, $K = 4$

RMSE: 0.0127 to 0.0086

ESS fraction: 0.390 to 0.929 as n goes from 8 to 128.



MISSPECIFIED VALUE, $r = 3$

Raw $q = 0.7282$ SNIS $q = 0.6624$

Exact target: $2/3$.

Local MH improves the proposal.

Initial pool factors provide the exact outer correction.

Reward guided LLM reasoning

Bellman error controls inference cost.

LLM reasoning is reward tilted path sampling

The reference model defines a trajectory law

$$\pi_{\text{ref}}(\mathbf{s}_{1:T} \mid \mathbf{s}_0) = \prod_{t=1}^T \pi_{\text{ref}}(s_t \mid \mathbf{s}_0, \mathbf{s}_{1:t-1}).$$

A terminal verifier or reward ϕ defines the target

$$\tilde{\pi}_T(\mathbf{s}_{1:T} \mid \mathbf{s}_0) \propto \pi_{\text{ref}}(\mathbf{s}_{1:T} \mid \mathbf{s}_0) \phi(\mathbf{s}_{1:T}).$$

Intermediate targets use a positive value approximation:

$$\tilde{\pi}_t(\mathbf{s}_{1:t} \mid \mathbf{s}_0) \propto \pi_{\text{ref}}(\mathbf{s}_{1:t} \mid \mathbf{s}_0) \hat{V}(\mathbf{s}_0, \mathbf{s}_{1:t}), \quad \hat{V}(\mathbf{s}_0, \mathbf{s}_{1:T}) = \phi(\mathbf{s}_{1:T}).$$

Inference compute allocation

The value model is not just a final ranker. It allocates particles among partial reasoning trajectories.

Source: Reward guided SMC for LLMs paper, target and intermediate Feynman Kac path measures.

Bellman consistency measures value quality

EXACT VALUE AND BELLMAN RECURSION

$$V_t^*(s_{1:t}) := \mathbb{E}_{\pi_{\text{ref}}}[\phi(S_{1:T}) \mid s_{1:t}] = \mathbb{E}_{\pi_{\text{ref}}}[V_{t+1}^*(S_{1:t+1}) \mid s_{1:t}].$$

LOCAL MULTIPLICATIVE DEFECT FOR $\hat{V} > 0$

$$\frac{1}{1 + \varepsilon} \leq \frac{\hat{V}(s_{1:t})}{\mathbb{E}_{\pi_{\text{ref}}}[\hat{V}(S_{1:t+1}) \mid s_{1:t}]} \leq 1 + \varepsilon,$$

Also assume the ratio bound $L^{-1} \leq \hat{V}(s_{1:t})/\hat{V}(s_{1:t-1}) \leq L$.

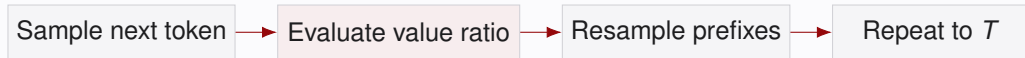
Source: Reward guided SMC for LLMs paper, optimal twist and local Bellman error assumption.

SMC turns a value approximation into local weights

For proposal π_p , use the local Feynman Kac potential

$$G_t(s_{1:t}) = \frac{\pi_{\text{ref}}(s_t \mid s_{1:t-1})}{\pi_p(s_t \mid s_{1:t-1})} \frac{\hat{V}(s_{1:t})}{\hat{V}(s_{1:t-1})}.$$

With $\pi_p = \pi_{\text{ref}}$, only the value ratio remains.



Output law

The theorem controls the averaged law of a uniformly selected final particle.

Source: Reward guided SMC for LLMs paper, naive value based SMC algorithm.

Small Bellman error makes particle correction polynomial

ONE PARTICLE

$$\|\tilde{\pi}_t - \hat{\pi}_t\|_{\text{TV}} \leq 2t\varepsilon.$$

A sufficient certificate, not a converse.

Naive particle SMC

$$N \geq \frac{L^6 T (1 + \varepsilon)^{6(T-1)}}{2\delta_{\text{TV}}} \implies \|\bar{\eta}_{T+1}^N - \tilde{\pi}_T\|_{\text{TV}} \leq \delta_{\text{TV}}.$$

In the clean regime $\varepsilon = O(1/T)$,

$$TC_{\text{naive SMC}} = O(L^6 T^2 / \delta_{\text{TV}}).$$

Source: Reward guided SMC for LLMs paper, single particle TV theorem and particle bias theorem.

Fixed Bellman error can make search exponential

NO GUESS ORACLE MODEL Every output prefix must have been visited through oracle queries.

Worst case lower bounds in the no guess oracle model

To achieve total variation error at most $1/3$ on every admissible instance,

$$TC = \Omega(L^{2T/3})$$

under the ratio bound alone. With $L > 1$ and $\varepsilon \in (0, L - 1]$ under the local Bellman bound,

$$TC = \Omega\left((1 + \varepsilon)^{2T/3}\right).$$

Interpretation. Fixed ε permits exponential worst case cost; $\varepsilon = O(1/T)$ gives a clean polynomial upper bound.

Source: Reward guided SMC for LLMs paper, lower bound theorem and Bellman error corollary.

One correction principle, four numerical languages

Setting	Learned object	Structural defect	Inference correction
Integration	$\hat{f}$	$f - \hat{f}$	residual Monte Carlo
Scientific ML	$\hat{u}$	PDE or Bellman residual	Feynman Kac or multilevel Picard
Diffusion	G or H	Doob or path law defect	Girsanov weights, two queries, or Stein pools
LLM reasoning	$\hat{V}$	local Bellman inconsistency	sequential Monte Carlo

Inference time scaling law

$$U_{\text{corr}} = \underbrace{\hat{U}}_{\text{learned bulk}} + \underbrace{\mathcal{C}(\delta_B)}_{\text{simulate the structural miss}}.$$

Learn the bulk once; spend inference compute only on the miss.

Thank you