

Mean Square Error Analysis of Stochastic Runge–Kutta Integrators for Langevin Dynamics

Xuda Ye (Purdue)

The sharp Wasserstein-1 bound is joint work with Yuliang Wang (Duke)

September 20, 2026

SIAM Great Lakes Section Meeting

Langevin Dynamics and Mean Square Error

- Overdamped Langevin dynamics, target $\pi \propto e^{-U}$:

$$dX_t = -\nabla U(X_t) dt + \sqrt{2} dB_t.$$

- Integrator: Markov chain $Z_k \approx X_{kh}$; time averages estimate $\pi(f) := \int f d\pi$.
- **Mean square error** (MSE) of one trajectory:

$$\text{MSE}(N, h) := \mathbb{E} \left[\left| \frac{1}{N} \sum_{k=0}^{N-1} f(Z_k) - \pi(f) \right|^2 \right].$$

- Missing: **second order bound** on $\text{MSE}(N, h)$ without global convexity.

- Euler–Maruyama: $Z_{k+1} = Z_k - h \nabla U(Z_k) + \sqrt{2} (B_{(k+1)h} - B_{kh})$, strong order 1.
- Higher order with ∇U alone: **stochastic Runge–Kutta (SRK)** integrators SRK-I and SRK-II [Yang & Wang 2026], strong order $\frac{3}{2}$.
- Weak and Wasserstein bounds compare **laws**: local weak error propagated by the exact semigroup, or a coupling of numerical and exact solutions. They control $\mathbb{E}f(Z_k)$ step by step.
- The MSE is the error of **one trajectory**: it needs the variance of the time average, hence the decay of correlations along the chain. Step-by-step control of $\mathbb{E}f(Z_k)$ cannot give it.

Two SRK Integrators

Input per step: Brownian increment and **time integral** $\int_{kh}^{(k+1)h} (B_s - B_{kh}) ds$, drawn jointly.

$$(I) : \quad \begin{aligned} H_k &= Z_k - \frac{3h}{4} \nabla U(Z_k) + \frac{3\sqrt{2}}{2h} \int_{kh}^{(k+1)h} (B_s - B_{kh}) ds, \\ Z_{k+1} &= Z_k - \frac{h}{3} \nabla U(Z_k) - \frac{2h}{3} \nabla U(H_k) + \sqrt{2} (B_{(k+1)h} - B_{kh}). \end{aligned}$$

$$(II) : \quad \begin{aligned} H_k^\pm &= Z_k - \frac{h}{2} \nabla U(Z_k) + \frac{\sqrt{2}}{h} \left(1 \pm \frac{1}{\sqrt{2}}\right) \int_{kh}^{(k+1)h} (B_s - B_{kh}) ds, \\ Z_{k+1} &= Z_k - \frac{h}{2} (\nabla U(H_k^+) + \nabla U(H_k^-)) + \sqrt{2} (B_{(k+1)h} - B_{kh}). \end{aligned}$$

Both integrators have strong order $\frac{3}{2}$ and weak order 2.

Assumptions

We assume sufficient regularity but convexity only outside a bounded region.

- **Assumption 1** (potential): $U \in C^5(\mathbb{R}^d)$ with

$$|\nabla U(0)| \leq M, \quad \max_{2 \leq i \leq 5} \|\nabla^i U(x)\| \leq M, \quad \nabla^2 U(x) \succeq m I_d \text{ for } |x| \geq R.$$

- **Assumption 2** (test function): $f \in C^3(\mathbb{R}^d)$ with

$$|\nabla f(x)| \leq L, \quad \|\nabla^2 f(x)\| \leq L, \quad \|\nabla^3 f(x)\| \leq L.$$

Main Results: MSE Estimate

For the SRK-I and SRK-II [Yang & Wang 2026] integrators, we prove

MSE estimate (Ye, 2026)

$$\text{MSE}(N, h) \leqslant CL^2(1 + |Z_0|)^{12} \left(\frac{1}{Nh} + h^4 \right), \quad h \in (0, h_0], N \geqslant 1.$$

C depends on m, M, R, d only; h_0 a step-size threshold. Both terms optimal: $\frac{1}{Nh}$ sampling variance, h^4 squared second order bias.

- **No global convexity**: convex outside a bounded region, multiple wells allowed.
- **MSE at a finite N** : explicit in N and h , no ergodic limit.
- **Second order**: discretization at h^4 , gradient evaluations only.

Main Results: Wasserstein-1 Bound

For the same integrators, the law of the chain against the law of the exact solution:

Wasserstein-1 estimate (Wang & Ye, 2026)

$$\mathcal{W}_1(\text{Law}(Z_k), \text{Law}(X_{kh})) \leq C(1 + |Z_0|)^6 h^2, \quad h \in (0, h_0], \quad k \geq 1.$$

- **Uniform in time:** holds at every k ; the same against π , up to $e^{-\lambda kh}$.
- **Sharp:** h^2 attained on a Gaussian target, and h^4 in the MSE with it.
- Finite horizon: weak order 2 against smooth test functions.

Where This Sits

Gradient-only integrators, orders proved without global convexity. μ_k, ν_k : laws of chain and exact solution; π_h : invariant measure of the chain; Langevin Monte Carlo (LMC).

integrator	strong	weak	long-time convergence	MSE
Euler–Maruyama	1	1	$\mathcal{W}_2(\mu_k, \pi) \lesssim e^{-\lambda kh} + h$	h^2
Leimkuhler–Matthews	1	1	$ \mathbb{E}f(Z_k) - \pi(f) \lesssim e^{-\lambda kh} + h^2$	—
stochastic Heun	1	2	$ \pi_h(f) - \pi(f) \lesssim h^2$	—
randomized midpoint	1	1	$\mathcal{W}_2(\mu_k, \pi) \lesssim e^{-\lambda kh} + h$	—
random splitting LMC	1	2	$\mathcal{W}_1(\mu_k, \nu_k) \lesssim h^2$	—
SRK-I/II [Yang & Wang 2026]	3/2	3/2	$\mathcal{W}_2(\mu_k, \pi) \lesssim e^{-\lambda kh} + h^{3/2}$	—
SRK-I/II, our works	3/2	2	$\mathcal{W}_1(\mu_k, \nu_k) \lesssim h^2$	h^4

Route here: the [discrete Poisson equation](#) [Lu, Ye & Zhou 2025], carried to the nonconvex regime. SRK-I/II have strong order $\frac{3}{2}$; the [second order](#) MSE requires a refined analysis of the Taylor remainder.

Discrete Poisson Equation

Generator $\mathcal{L}\psi = \Delta\psi - \nabla U \cdot \nabla\psi$; semigroup $e^{t\mathcal{L}}g(x) = \mathbb{E}^x[g(X_t)]$, $\mathbb{E}^x$ the expectation from $X_0 = x$; $\mathcal{I}$ the identity.

- Kolmogorov solution: $u(x, t) := e^{t\mathcal{L}}f(x) - \pi(f)$.
- Discrete Poisson solution:

$$\phi_h(x) := h \sum_{k=0}^{\infty} u(x, kh) = h (\mathcal{I} - e^{h\mathcal{L}})^{-1} (f(x) - \pi(f)).$$

- Discrete Poisson equation:

$$\frac{\mathcal{I} - e^{h\mathcal{L}}}{h} \phi_h = f - \pi(f).$$

- One exact step turns ϕ_h into a martingale increment plus $h(f - \pi(f))$: time average $\rightarrow$ one-step quantities.

Time-Averaged Error Formula

Let $Z_k(h)$ be the exact solution of the Langevin dynamics after time h from Z_k , on the same Brownian path as Z_{k+1} . One step, split at $Z_k(h)$, read through ϕ_h :

$$\frac{\phi_h(Z_{k+1}) - \phi_h(Z_k)}{h} + f(Z_k) - \pi(f) = \Upsilon_k + \Psi_k,$$

$$\Upsilon_k := \frac{\phi_h(Z_{k+1}) - \phi_h(Z_k(h))}{h}, \quad \Psi_k := \frac{\phi_h(Z_k(h)) - \phi_h(Z_k)}{h} + f(Z_k) - \pi(f).$$

Υ_k : **discretization**, the local error through ϕ_h . Ψ_k : **martingale difference**, $\mathbb{E}_k \Psi_k = 0$ with $\mathbb{E}_k$ conditional on Z_k . Telescoping:

$$\frac{1}{N} \sum_{k=0}^{N-1} (f(Z_k) - \pi(f)) = \underbrace{\frac{\phi_h(Z_0) - \phi_h(Z_N)}{Nh}}_{\text{boundary}} + \underbrace{\frac{1}{N} \sum_{k=0}^{N-1} \Upsilon_k}_{\text{discretization}} + \underbrace{\frac{1}{N} \sum_{k=0}^{N-1} \Psi_k}_{\text{martingale}}.$$

Heuristic Taylor Expansion

Interpolate between the exact and the numerical step, $Y_\theta := \theta Z_{k+1} + (1 - \theta)Z_k(h)$:

$$\phi_h(Z_{k+1}) - \phi_h(Z_k(h)) = \nabla \phi_h(Z_k)^\top \Delta Z_k + \int_0^1 (\nabla \phi_h(Y_\theta) - \nabla \phi_h(Z_k))^\top \Delta Z_k d\theta.$$

- First term: mean-zero part of ΔZ_k drops out in conditional mean, remainder $O(h^3)$.
- Integral bounded by $\|\nabla^2 \phi_h\|_\infty |\Delta Z_k| |Y_\theta - Z_k|$: only $O(h^{3/2})$, the **strong order**.
- One more expansion:

$$\int_0^1 (\nabla \phi_h(Y_\theta) - \nabla \phi_h(Z_k))^\top \Delta Z_k d\theta \approx \frac{1}{2} \nabla^2 \phi_h(Z_k) : \underbrace{\Delta Z_k (Z_{k+1} + Z_k(h) - 2Z_k)^\top}_{=: \Delta Q_k}.$$

- Needed from the integrator: ΔZ_k and ΔQ_k split into a mean-zero part and an $O(h^3)$ remainder, plus a fourth moment of ΔZ_k .

Elliptic Estimates of ϕ_h

Elliptic estimates of ϕ_h (Ye, 2026)

Under Assumptions 1–2, there exist $\lambda > 0$ and C depending only on m, M, R, d , such that

$$|u(x, t)| \leq CL(1 + |x|) e^{-\lambda t}, \quad \max\{|\nabla u|, \|\nabla^2 u\|, \|\nabla^3 u\|\}(x, t) \leq CL e^{-\lambda t}.$$

Hence for $h \leq \lambda^{-1}$,

$$\max\{|\nabla \phi_h|, \|\nabla^2 \phi_h\|, \|\nabla^3 \phi_h\|\} \leq CL \quad \text{uniformly in } h.$$

∇u : reflection coupling. $\nabla^2 u, \nabla^3 u$: derivative formulae of Elworthy–Li type.

Local Error Decomposition

The key step is to split the local errors ΔZ_0 and ΔQ_0 into a **mean-zero part** and a **remainder**.

Local error decomposition (Ye, 2026)

For SRK-I/II, mean-zero parts $\mathcal{M}$ and remainders $\mathcal{R}$, $\mathbb{E}_0[\mathcal{M}(\Delta Z_0)] = \mathbb{E}_0[\mathcal{M}(\Delta Q_0)] = 0$:

$$\begin{aligned}\Delta Z_0 &= \mathcal{M}(\Delta Z_0) + \mathcal{R}(\Delta Z_0), & \Delta Q_0 &= \mathcal{M}(\Delta Q_0) + \mathcal{R}(\Delta Q_0), \\ \mathbb{E}_0[|\mathcal{M}(\Delta Z_0)|^4]^{1/4} &\leq Ch^2(1 + |Z_0|), & \mathbb{E}_0[|\mathcal{R}(\Delta Z_0)|^4]^{1/4} &\leq Ch^3(1 + |Z_0|)^4, \\ \mathbb{E}_0[\|\mathcal{M}(\Delta Q_0)\|^2]^{1/2} &\leq Ch^{5/2}(1 + |Z_0|), & \mathbb{E}_0[\|\mathcal{R}(\Delta Q_0)\|^2]^{1/2} &\leq Ch^3(1 + |Z_0|)^5.\end{aligned}$$

Expand $\phi_h(Z_{k+1})$ and $\phi_h(Z_k(h))$ about Z_k :

$$h \Upsilon_k = \nabla \phi_h(Z_k)^\top \Delta Z_k + \frac{1}{2} \nabla^2 \phi_h(Z_k) : \Delta Q_k + \varrho_k.$$

- Local error decomposition inserted: $\Upsilon_k = \mathcal{M}(\Upsilon_k) + \mathcal{R}(\Upsilon_k)$, $\mathbb{E}_k \mathcal{M}(\Upsilon_k) = 0$.
- Martingale part $\mathcal{M}(\Upsilon_k) = \mathcal{O}(h)$: additive variance.
- Remainder $\mathcal{R}(\Upsilon_k) = \mathcal{O}(h^2)$: accumulates linearly in N ; squared, h^4 .
- ϱ_k : third-order Taylor remainder, bounded by $\|\nabla^3 \phi_h\|_\infty$ (elliptic estimates).

Mean Square Error Bound

MSE estimate (Ye, 2026)

Under Assumptions 1–2, for $h \in (0, h_0]$ and $N \geq 1$,

$$\text{MSE}(N, h) \leq CL^2(1 + |Z_0|)^{12} \left(\frac{1}{Nh} + h^4 \right).$$

- $\frac{1}{Nh}$: variance of a time average over length Nh , from Ψ_k ; present for the exact dynamics too.
- h^4 : square of the $O(h^2)$ bias. Weak order 1 would leave h^2 .
- Optimal $h \asymp N^{-1/5}$: $\text{MSE} = O(N^{-4/5})$, against $O(N^{-2/3})$ at weak order 1.
- Stability: moments of the chain that do not grow with k .
- First **second order** MSE bound at finite N : gradient-only integrator, no global convexity.

Uniform-in-Time Wasserstein-1 Bound

Wasserstein-1 estimate (Wang & Ye, 2026)

Under Assumption 1, for $h \in (0, h_0]$ and $N \geq 1$,

$$\mathcal{W}_1(\text{Law}(Z_N), \text{Law}(X_{Nh})) \leq C(1 + |Z_0|)^6 h^2,$$

$$\mathcal{W}_1(\text{Law}(Z_N), \pi) \leq C(1 + |Z_0|) e^{-\lambda N h} + C(1 + |Z_0|)^6 h^2.$$

- Lipschitz f only; the MSE estimate needs 3 derivatives: the cost of a single trajectory.
- **Uniform in time**: within $O(h^2)$ of the exact law at every N , of π up to $e^{-\lambda N h}$.

Second Order Is Sharp

Gaussian example (Wang & Ye, 2026)

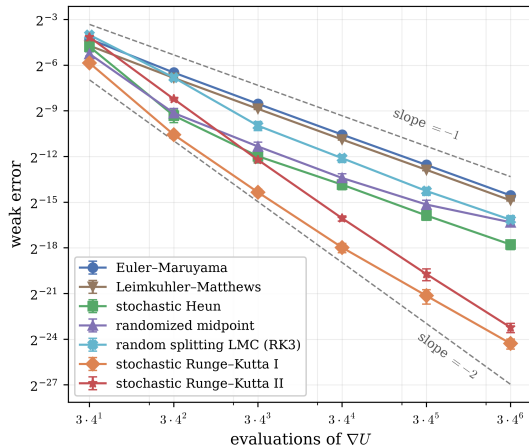
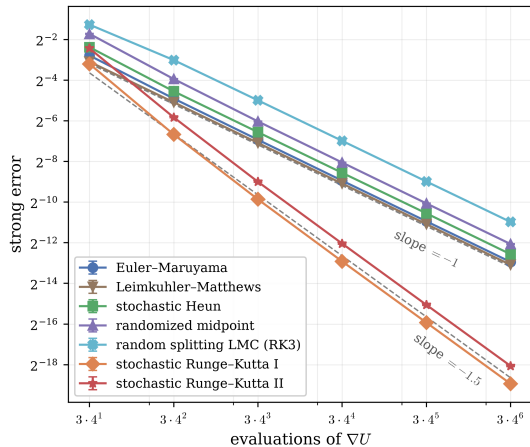
$U(x) = \frac{1}{2}x^2$, $\pi = \mathcal{N}(0, 1)$. SRK-I and SRK-II coincide, invariant law $\pi_h = \mathcal{N}(0, \sigma_h^2)$, and as $h \rightarrow 0$

$$\mathcal{W}_1(\pi_h, \pi) = \frac{1}{12} \sqrt{\frac{2}{\pi}} h^2 + O(h^3),$$

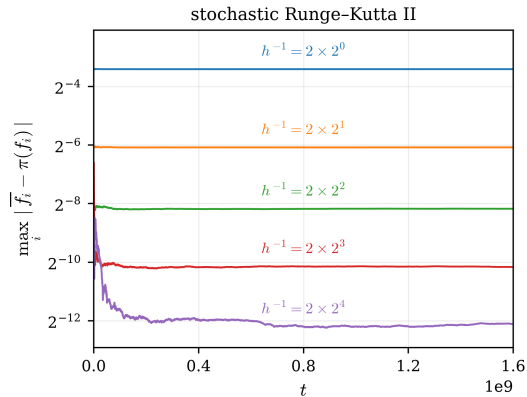
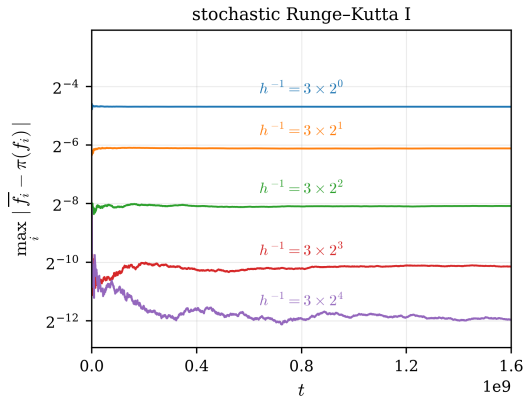
$$\liminf_{N \rightarrow \infty} \text{MSE}(N, h) \geq \frac{e^{-1}}{144} h^4 + O(h^5) \quad \text{for } f(x) = \cos x.$$

- Both SRK-I/II reduce to $Z_{k+1} = (1 - h + \frac{h^2}{2})Z_k + \text{noise}$, $\sigma_h^2 = 1 - \frac{h^2}{6} + O(h^3)$.
- h^2 of the Wasserstein-1 estimate and h^4 of the MSE estimate are **attained**.

Numerics I: Finite-Time Strong and Weak Error



Numerics II: Uniform-in-Time Sampling Error



Numerics II: Sampling Error at the Horizon

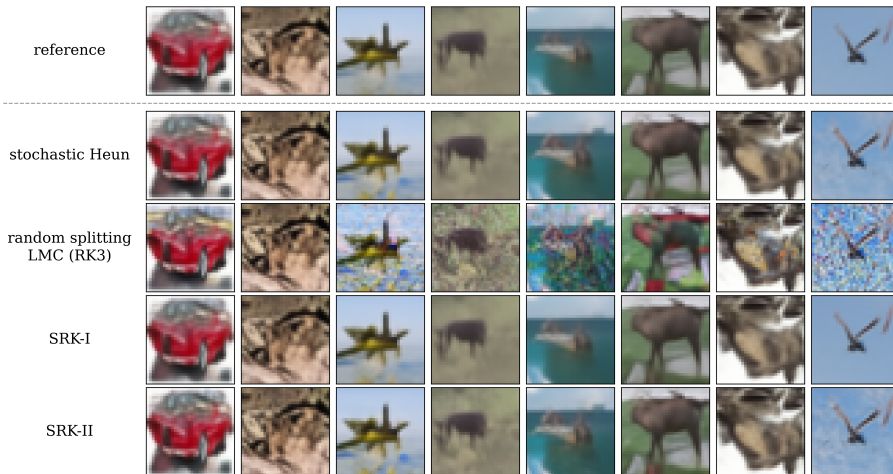
A scheme at e evaluations per step runs at $h = e/C$.

C	EM	LM	Heun	RS-LMC	SRK-I	SRK-II
6	$3.7 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$	$4.7 \cdot 10^{-2}$	$1.5 \cdot 10^{-1}$	$3.9 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$
12	$1.8 \cdot 10^{-2}$	$7.6 \cdot 10^{-4}$	$1.9 \cdot 10^{-2}$	$5.6 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$
24	$8.7 \cdot 10^{-3}$	$2.0 \cdot 10^{-4}$	$5.1 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$
48	$4.3 \cdot 10^{-3}$	$8.3 \cdot 10^{-5}$	$1.2 \cdot 10^{-3}$	$6.5 \cdot 10^{-3}$	$8.8 \cdot 10^{-4}$	$8.7 \cdot 10^{-4}$
96	$2.1 \cdot 10^{-3}$	$9.4 \cdot 10^{-5}$	$3.2 \cdot 10^{-4}$	$1.8 \cdot 10^{-3}$	$2.5 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$
fitted order	1.02	1.68	1.97	1.67	1.96	2.01

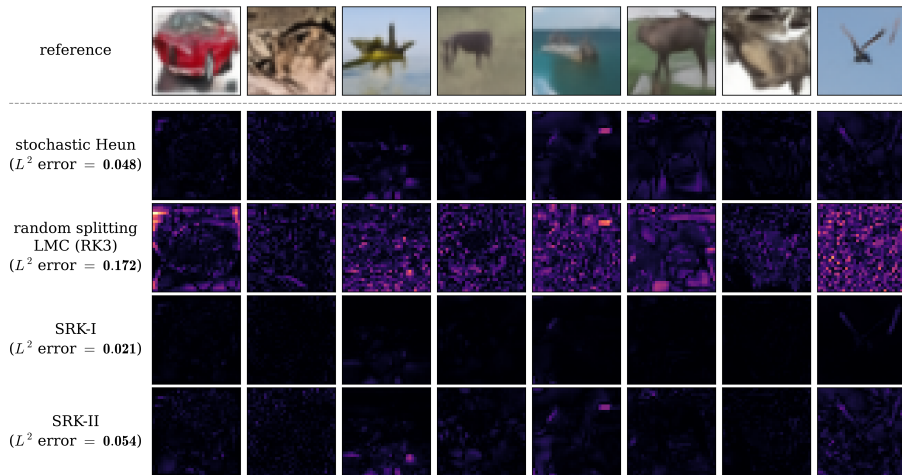
SRK-I, SRK-II: orders 1.96, 2.01, the h^2 bias of the Wasserstein-1 estimate, entering the MSE estimate as h^4 . Leimkuhler–Matthews: smallest sampling error at every cost.

Numerics III: Diffusion Model of CIFAR-10

Score-based diffusion model, $N = 60$ steps, a shared Brownian path, reference by fine SRK-LD.



Numerics III: Error to the Reference



- **Local error** $\rightarrow$ **MSE of a time average**: split the local error, convert through the discrete Poisson equation, derivatives bounded uniformly in h without global convexity (elliptic estimates).
- SRK-I, SRK-II: $\text{MSE} \leq C(\frac{1}{Nh} + h^4)$ at finite N ; weak order 2; $\mathcal{W}_1 \lesssim h^2$ uniformly in time.
- Second order **sharp** on the Gaussian: invariant law and time average.
- Numerics: best gradient-only schemes over a finite horizon and in the diffusion model; second order long-time bias. Leimkuhler–Matthews: smallest sampling error at every cost.
- Open: higher weak order by more stages; stochastic gradients.

- X. Ye. Mean square error analysis of stochastic Runge–Kutta integrators. arXiv:2609.11528, 2026.
- Y. Wang, X. Ye. A sharp uniform-in-time Wasserstein-1 bound for stochastic Runge–Kutta integrators of Langevin dynamics. To appear, 2026.
- B. Yang, X. Wang. Accelerating Langevin Monte Carlo via efficient stochastic Runge–Kutta methods beyond log-concavity. arXiv:2605.07939, 2026.
- J. Lu, X. Ye, Z. Zhou. Mean square error analysis of stochastic gradient and variance-reduced sampling algorithms. arXiv:2511.04413, 2025.
- A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* 166, 851–886, 2016.
- L. Li, C. Wang, M. Wang. Ergodicity and error estimate of laws for a random splitting Langevin Monte Carlo. arXiv:2510.07676, 2025.

Publications

<https://xuda-ye.com/publications/>

Code, data and the companion note

<https://github.com/xuda-ye-math/Poisson-Overdamped>