Skew-symmetric schemes for stochastic differential equations with non-Lipschitz drift: an unadjusted Barker algorithm

samuel.livingstone@ucl.ac.uk

‘Skew26’, Padova, January 2026

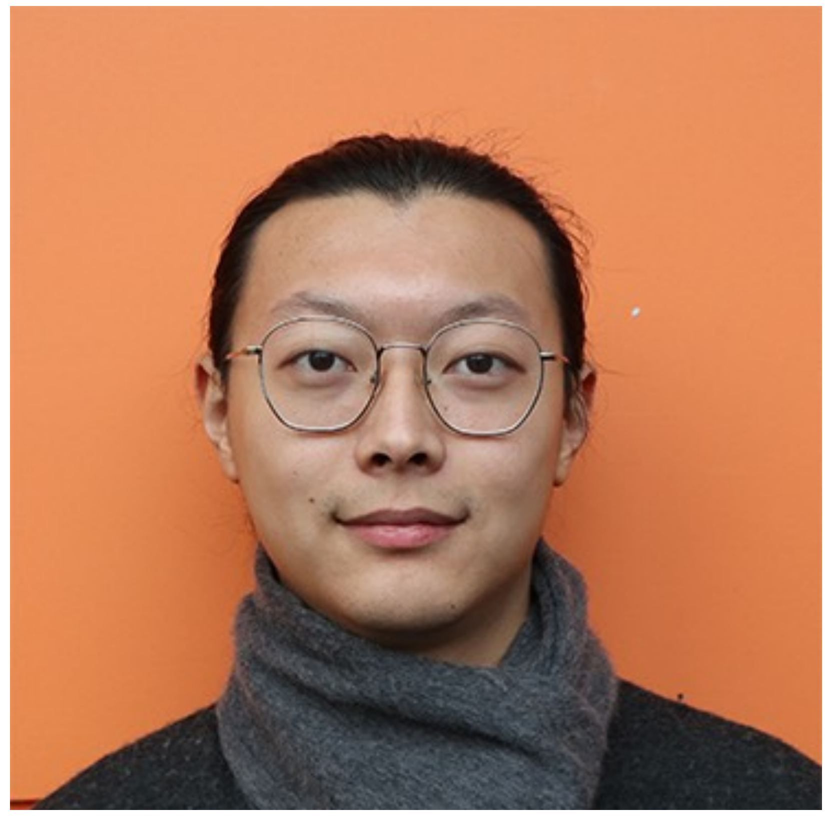
UCL

Collaborators

Yuga Iguchi
Lancaster University
(Former UCL PhD student)

Nikolas Nüsken
King's College London,

Giorgos Vasdekis
Newcastle University
(Former UCL postdoc)



Ruiyang Zhang
Lancaster University
(Former UCL undergraduate)

(earlier work with Giacomo Zanella, Jure Vogrinc & Max Hird).

Setup and motivation

Setup and motivation

- Consider a (uniformly elliptic) diffusion process $(Y_t)_{t\geq 0}$ on $\mathbb{R}^d$ with dynamics

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t$$

Setup and motivation

- Consider a (uniformly elliptic) diffusion process $(Y_t)_{t\geq 0}$ on $\mathbb{R}^d$ with dynamics

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t$$

- Questions:
  - How can we generate an approximate skeleton sample path $\{X_0, X_h, X_{2h}, \ldots\}$?
  - How can we approximate the transition density $p(x_0)f(x_h \mid x_0)f(x_{2h} \mid x_h)\ldots$ associated with this skeleton?

Setup and motivation

- Consider a (uniformly elliptic) diffusion process $(Y_t)_{t\geq 0}$ on $\mathbb{R}^d$ with dynamics

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t$$

- Questions:
  - How can we generate an approximate skeleton sample path $\{X_0, X_h, X_{2h}, \ldots\}$?
  - How can we approximate the transition density $p(x_0)f(x_h \mid x_0)f(x_{2h} \mid x_h)\ldots$ associated with this skeleton?
- Statistical applications: sampling algorithms (e.g. Bayesian inference), parameter inference for diffusion processes, diffusion priors for Bayesian modelling

Motivating application: sampling algorithms

Motivating application: sampling algorithms

- To sample from a distribution $\pi(dy) := e^{-U(y)}dy$ on $\mathbb{R}^d$, we can simulate the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

Motivating application: sampling algorithms

- To sample from a distribution $\pi(dy) := e^{-U(y)}dy$ on $\mathbb{R}^d$, we can simulate the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Under mild conditions on U the distribution $\mathscr{L}(Y_t)$ of the process at time t satisfies

$$\lim_{t\to\infty} \|\mathscr{L}(Y_t) - \pi\|_{TV} = 0$$

Motivating application: sampling algorithms

- To sample from a distribution $\pi(dy) := e^{-U(y)}dy$ on $\mathbb{R}^d$, we can simulate the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Under mild conditions on U the distribution $\mathscr{L}(Y_t)$ of the process at time t satisfies

$$\lim_{t\to\infty} \|\mathscr{L}(Y_t) - \pi\|_{TV} = 0$$

- To generate approximate samples/integrals from π, we can therefore simulate the process numerically over a long time period and take the end point/ergodic averages

Motivating application: sampling algorithms

- To sample from a distribution $\pi(dy) := e^{-U(y)}dy$ on $\mathbb{R}^d$, we can simulate the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Under mild conditions on U the distribution $\mathscr{L}(Y_t)$ of the process at time t satisfies

$$\lim_{t \to \infty} \|\mathscr{L}(Y_t) - \pi\|_{TV} = 0$$

- To generate approximate samples/integrals from π, we can therefore simulate the process numerically over a long time period and take the end point/ergodic averages
- If desired, after each time step a Metropolis—Hastings filter can also be applied

Motivating application: sampling algorithms

- To sample from a distribution $\pi(dy) := e^{-U(y)}dy$ on $\mathbb{R}^d$, we can simulate the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Under mild conditions on U the distribution $\mathscr{L}(Y_t)$ of the process at time t satisfies

$$\lim_{t \to \infty} \|\mathscr{L}(Y_t) - \pi\|_{TV} = 0$$

- To generate approximate samples/integrals from π, we can therefore simulate the process numerically over a long time period and take the end point/ergodic averages
- If desired, after each time step a Metropolis—Hastings filter can also be applied
- In Bayesian inference applications we can take π as the posterior.
    - NOTE! Only $\nabla U(x)$ needed, i.e. only unnormalised π needed (well-suited to Bayes)

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Classical approach: **Euler—Maruyama**

$$X_{(n+1)h} = X_{nh} + h\mu(X_{nh}) + \xi_{nh}$$

- Step-size $h > 0$, $\xi_{nh} \sim N(0, h\sigma(X_{nh})\sigma(X_{nh})^T)$

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Classical approach: **Euler—Maruyama**

$$X_{(n+1)h} = X_{nh} + h\mu(X_{nh}) + \xi_{nh}$$

- Step-size $h > 0$, $\xi_{nh} \sim N(0, h\sigma(X_{nh})\sigma(X_{nh})^T)$
- This leads to a Gaussian transition density, i.e. $X_{(n+1)h} = X_{nh} + \zeta_{nh}$ where

$$\zeta_{nh} \sim N\left(h\mu(X_{nh}), h\sigma(X_{nh})\sigma(X_{nh})^T\right)$$

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Classical approach: **Euler—Maruyama**

$$X_{(n+1)h} = X_{nh} + h\mu(X_{nh}) + \xi_{nh}$$

- Step-size $h > 0$, $\xi_{nh} \sim N(0,h\sigma(X_{nh})\sigma(X_{nh})^T)$
- This leads to a Gaussian transition density, i.e. $X_{(n+1)h} = X_{nh} + \zeta_{nh}$ where

$$\zeta_{nh} \sim N\left(h\mu(X_{nh}), h\sigma(X_{nh})\sigma(X_{nh})^T\right)$$

- When μ is not globally Lipschitz (e.g. super-linear) the scheme can be numerically unstable

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Classical approach: **Euler—Maruyama**

$$X_{(n+1)h} = X_{nh} + h\mu(X_{nh}) + \xi_{nh}$$

- Step-size $h > 0$, $\xi_{nh} \sim N(0, h\sigma(X_{nh})\sigma(X_{nh})^T)$
- This leads to a Gaussian transition density, i.e. $X_{(n+1)h} = X_{nh} + \zeta_{nh}$ where

$$\zeta_{nh} \sim N\left(h\mu(X_{nh}), h\sigma(X_{nh})\sigma(X_{nh})^T\right)$$

- When μ is not globally Lipschitz (e.g. super-linear) the scheme can be numerically unstable
    - **Sometimes fine** if x_{nh}, h small enough. But even then leads to inefficiencies

Numerical simulation (of skeleton paths)

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Classical approach: **Euler—Maruyama**

$$X_{(n+1)h} = X_{nh} + h\mu(X_{nh}) + \xi_{nh}$$

- Step-size $h > 0$, $\xi_{nh} \sim N(0, h\sigma(X_{nh})\sigma(X_{nh})^T)$
- This leads to a Gaussian transition density, i.e. $X_{(n+1)h} = X_{nh} + \zeta_{nh}$ where

$$\zeta_{nh} \sim N\left(h\mu(X_{nh}), h\sigma(X_{nh})\sigma(X_{nh})^T\right)$$

- When μ is not globally Lipschitz (e.g. super-linear) the scheme can be numerically unstable
    - **Sometimes fine** if x_{nh}, h small enough. But even then leads to inefficiencies
    - **Alternatives:** implicit schemes (expensive), tamed Euler, truncated Euler...

Instability of E-M when $\mu(x)$ grows super-linearly

$$\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$$

Instability of E-M when $\mu(x)$ grows super-linearly

$$\lim_{\|x\|\to\infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$$

- **Finite time simulation:** For finite T the mean squared error diverges as $h \downarrow 0$, meaning ($X^{(h)}$ is numerical scheme with time step h) (Hutzenthaler, Jentzen & Kloeden, 2011)

$$\lim_{h\to 0} \mathbb{E}\left[| Y_T - X_T^{(h)} |^2 \right] = \infty$$

Instability of E-M when $\mu(x)$ grows super-linearly

$$\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$$

- **Finite time simulation:** For finite T the mean squared error diverges as $h \downarrow 0$, meaning ($X^{(h)}$ is numerical scheme with time step h) (Hutzenthaler, Jentzen & Kloeden, 2011)

$$\lim_{h \to 0} \mathbb{E}\left[|Y_T - X_T^{(h)}|^2 \right] = \infty$$

- **Equilibrium sampling:** As $T \to \infty$ for fixed $h > 0$ the numerical process is typically a transient Markov chain (Roberts & Tweedie, 1996), i.e. $\forall$ probability distributions π_h, as $n \to \infty$

$$\|P_h^n(x, \cdot) - \pi_h\|_{TV} \nrightarrow 0$$

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\|\to\infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

- Consider the stochastic differential equation

$$dY_t = -Y_t^3 dt + dW_t$$

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\|\to\infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

- Consider the stochastic differential equation

$$dY_t = -Y_t^3 dt + dW_t$$

- The resulting Euler—Maruyama scheme with step-size h is

$$X_{(n+1)h} = X_{nh} - hX_{nh}^3 + \sqrt{2h}\xi_{nh}$$

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

- Consider the stochastic differential equation

$$dY_t = -Y_t^3 dt + dW_t$$

- The resulting Euler—Maruyama scheme with step-size h is

$$X_{(n+1)h} = X_{nh} - hX_{nh}^3 + \sqrt{2h}\xi_{nh}$$

- Take $h \ll 1$, and let $X_{nh} \sim h^{-1}$. Then

$$\begin{aligned} X_{(n+1)h} &\sim h^{-1} - h \cdot h^{-3} + \sqrt{2} \cdot h^{1/2} \\ &\sim h^{-2} \end{aligned}$$

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

- Consider the stochastic differential equation

$$dY_t = -Y_t^3 dt + dW_t$$

- The resulting Euler—Maruyama scheme with step-size h is

$$X_{(n+1)h} = X_{nh} - hX_{nh}^3 + \sqrt{2h}\xi_{nh}$$

- Take $h \ll 1$, and let $X_{nh} \sim h^{-1}$. Then

$$\begin{aligned} X_{(n+1)h} &\sim h^{-1} - h \cdot h^{-3} + \sqrt{2} \cdot h^{1/2} \\ &\sim h^{-2} \end{aligned}$$

- Using this intuition, it can be shown that as $N \to \infty$, $|X_{Nh}| \to \infty$ w.p. 1 from any starting point

Example instability of Euler—Maruyama

Super-linear drift $\lim_{\|x\| \to \infty} \frac{\|\mu(x)\|}{\|x\|} = +\infty$

- Consider the stochastic differential equation

$$dY_t = -Y_t^3 dt + dW_t$$

- The resulting Euler—Maruyama scheme with step-size h is

$$X_{(n+1)h} = X_{nh} - hX_{nh}^3 + \sqrt{2h}\xi_{nh}$$

- Take $h \ll 1$, and let $X_{nh} \sim h^{-1}$. Then

$$\begin{aligned} X_{(n+1)h} &\sim h^{-1} - h \cdot h^{-3} + \sqrt{2} \cdot h^{1/2} \\ &\sim h^{-2} \end{aligned}$$

- Using this intuition, it can be shown that as $N \to \infty$, $|X_{Nh}| \to \infty$ w.p. 1 from any starting point
- Taking $\pi(dy) \propto e^{-y^4/4}dy$, then this SDE is the corresponding overdamped Langevin diffusion

Super-linear ∇U in statistical modelling

A more realistic example

Super-linear ∇U in statistical modelling

A more realistic example

- In fact, 'light-tailed' Bayesian posterior distributions (∇U is super-linear) are easy to find

Super-linear ∇U in statistical modelling

A more realistic example

- In fact, 'light-tailed' Bayesian posterior distributions (∇U is super-linear) are easy to find
- **Poisson Generalised Linear Model:** data $\{(z_i, y_i)\}_{i=1}^n$, $z_i \in \mathbb{R}^d$, $y_i \in \mathbb{Z}_+$, model

$$Y_i \,|\, (Z_i = z_i) \sim \text{Poisson}(e^{z_i^T \beta})$$

Super-linear ∇U in statistical modelling

A more realistic example

- In fact, 'light-tailed' Bayesian posterior distributions (∇U is super-linear) are easy to find
- **Poisson Generalised Linear Model:** data $\{(z_i, y_i)\}_{i=1}^n$, $z_i \in \mathbb{R}^d$, $y_i \in \mathbb{Z}_+$, model

$$Y_i \,|\, (Z_i = z_i) \sim \text{Poisson}(e^{z_i^T \beta})$$

- Setting $\pi_0(\beta) \propto e^{-\frac{1}{2}\beta^T A \beta}$ gives a posterior $\pi(\beta) \propto e^{-U(\beta)}$ with

$$\nabla U(\beta) = \sum_i (e^{z_i^T \beta} - y_i) z_i + A\beta$$

i.e. **exponential** in β

A skew-symmetric numerical scheme

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

A skew-symmetric numerical scheme

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Consider instead the numerical scheme (in one dimension to begin)

$$X_{(n+1)h} = X_{nh} + \zeta_{nh}$$

A skew-symmetric numerical scheme

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Consider instead the numerical scheme (in one dimension to begin)

$$X_{(n+1)h} = X_{nh} + \zeta_{nh}$$

- Here $\zeta_{nh} \sim \nu_{X_{nh},h}$ with $\nu_{x,h}$ a **skew-symmetric probability distribution**, having density

$$\nu_{x,h}(\zeta) = 2 \cdot p_h(x,\zeta) \cdot \phi_{x,h}(\zeta)$$

  - $\phi_{x,h}$ is a $N(0,h\sigma^2(x))$ density
  - $p_h(x,\zeta) \in [0,1]$ satisfies $p_h(x,\zeta) + p(x,-\zeta) = 1$ (e.g. a CDF of symmetric r.v.)

A skew-symmetric numerical scheme

for the general diffusion process $dY_t = \mu(Y_t)dy + \sigma(Y_t)dW_t$

- Consider instead the numerical scheme (in one dimension to begin)

$$X_{(n+1)h} = X_{nh} + \zeta_{nh}$$

- Here $\zeta_{nh} \sim \nu_{X_{nh},h}$ with $\nu_{x,h}$ a **skew-symmetric probability distribution**, having density

$$\nu_{x,h}(\zeta) = 2 \cdot p_h(x,\zeta) \cdot \phi_{x,h}(\zeta)$$

  - $\phi_{x,h}$ is a $N(0,h\sigma^2(x))$ density
  - $p_h(x,\zeta) \in [0,1]$ satisfies $p_h(x,\zeta) + p(x,-\zeta) = 1$ (e.g. a CDF of symmetric r.v.)

- We stipulate the additional requirement that

$$\partial_\zeta p_h(x,\zeta)\,|_{\zeta=0} = \frac{\mu(x)}{2\sigma^2(x)}$$

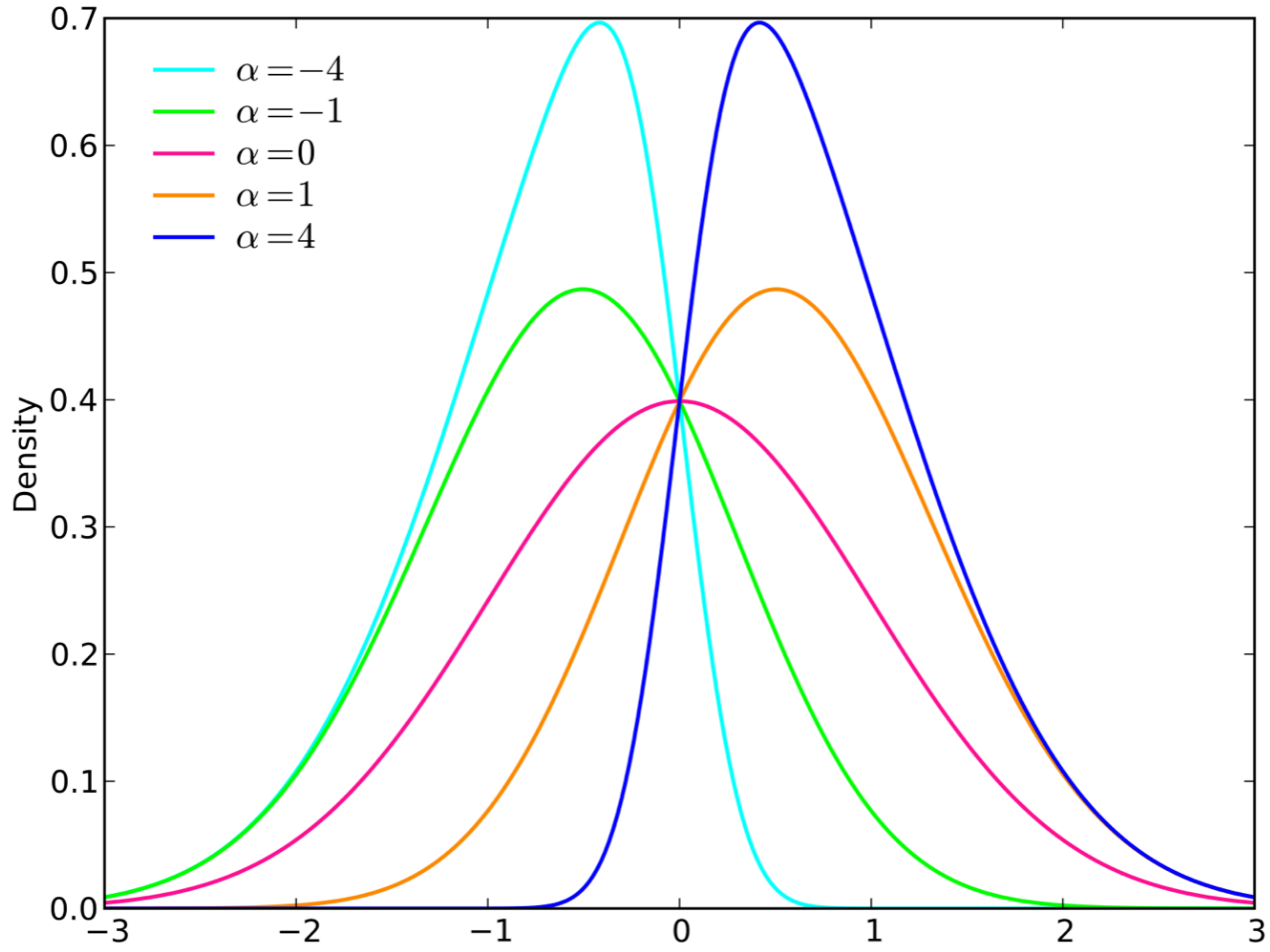
Example: skew-Normal distribution

(picture: 'Skew-Normal distribution' wikipedia page)

e.g. Skew-Normal
$s(z) = 2\Phi(\alpha z)\phi(z)$

The skew-Normal distribution is an example with explicit mean

$$\sqrt{\frac{2}{\pi}} \cdot \frac{\alpha}{\sqrt{1+\alpha^2}}$$



As $\alpha \to \infty$ density converges to **half-Normal**

Simulating the skew-symmetric scheme

Simulating the skew-symmetric scheme

- Simulation can be achieved by the re-parametrisation

$$X_{(n+1)h} = X_{nh} + b_{nh} \cdot \xi_{nh}$$

where $\xi_{nh} \sim N(0, h\sigma^2(X_{nh}))$ and

$$b_{nh} = \begin{cases} +1 & \text{with probability } p(X_{nh}, \xi_{nh}) \\ -1 & \text{otherwise} \end{cases}$$

Simulating the skew-symmetric scheme

- Simulation can be achieved by the re-parametrisation

$$X_{(n+1)h} = X_{nh} + b_{nh} \cdot \xi_{nh}$$

where $\xi_{nh} \sim N(0, h\sigma^2(X_{nh}))$ and

$$b_{nh} = \begin{cases} +1 & \text{with probability } p(X_{nh}, \xi_{nh}) \\ -1 & \text{otherwise} \end{cases}$$

- Note that $\mathbb{E}[b_{nh} \cdot \xi_{nh}] \neq 0$ since b_{nh} depends on ξ_{nh}

Simulating the skew-symmetric scheme

- Simulation can be achieved by the re-parametrisation

$$X_{(n+1)h} = X_{nh} + b_{nh} \cdot \xi_{nh}$$

where $\xi_{nh} \sim N(0, h\sigma^2(X_{nh}))$ and

$$b_{nh} = \begin{cases} +1 & \text{with probability } p(X_{nh}, \xi_{nh}) \\ -1 & \text{otherwise} \end{cases}$$

- Note that $\mathbb{E}[b_{nh} \cdot \xi_{nh}] \neq 0$ since b_{nh} depends on ξ_{nh}
- In words:

1. Simulate a mean-zero **Gaussian increment** ξ_{nh}
2. **Skew in the direction of the drift** $\mu(X_n)$ via multiplication by b_{nh}

Implementation details in higher dimensions

Implementation details in higher dimensions

- In multiple dimensions we apply the skewing operation **coordinate-wise**

Implementation details in higher dimensions

- In multiple dimensions we apply the skewing operation **coordinate-wise**
- Introduce a $b_{i,nh} \in \{-1, +1\}$ for each dimension $i \in \{1,\ldots,d\}$

Implementation details in higher dimensions

- In multiple dimensions we apply the skewing operation **coordinate-wise**
- Introduce a $b_{i,nh} \in \{-1, +1\}$ for each dimension $i \in \{1,\ldots,d\}$
- Set $\mathbb{P}[b_{i,nh} = +1] = p_i(X_{nh}, \xi_{nh})$

Implementation details in higher dimensions

- In multiple dimensions we apply the skewing operation **coordinate-wise**
- Introduce a $b_{i,nh} \in \{-1, +1\}$ for each dimension $i \in \{1, \ldots, d\}$
- Set $\mathbb{P}[b_{i,nh} = +1] = p_i(X_{nh}, \xi_{nh})$
- A simple choice of p_i is

$$p_i(x, \xi) = F(\alpha_{i,x} \cdot \xi_i)$$

where F is CDF of a symmetric r.v. and (e.g. for σ diagonal)

$$\alpha_{i,x} = \frac{\mu_i(x)}{2F'(0)\sigma_{ii}(x)}$$

Result 1: Weak convergence

Result 1: Weak convergence

Result 1 (Weak Convergence): Under appropriate assumptions the skew-symmetric scheme is **weak order 1**, meaning as $h \to 0$, for $k \leq \lfloor T/h \rfloor$ and polynomially growing K and $f \in C^4(\mathbb{R}^d)$

$$\left| \mathbb{E}_x \left[f(X_{kh}) - f(Y_{kh}) \right] \right| \leq K(x)h$$

Result 1: Weak convergence

Result 1 (Weak Convergence): Under appropriate assumptions the skew-symmetric scheme is **weak order 1**, meaning as $h \to 0$, for $k \leq \lfloor T/h \rfloor$ and polynomially growing K and $f \in C^4(\mathbb{R}^d)$

$$\left| \mathbb{E}_x \left[f(X_{kh}) - f(Y_{kh}) \right] \right| \leq K(x)h$$

Remarks on assumptions & proof

- Drift μ is C^4 with polynomial growth, locally Lipschitz + satisfies *one-sided* Lipschitz condition

$$(\mu(x) - \mu(y)) \cdot (x - y) \leq C(1 + \|x - y\|^2)$$

- Volatility σ is C^4, Lipschitz
- Proof extends 'General convergence theorem' of Milstein & Tretyakov to non-Lipschitz case

Result 2: Convergence to equilibrium

Result 2: Convergence to equilibrium

Result 2 (equilibrium measure and geometric convergence): Under appropriate assumptions, for any fixed $h \leq 1$ the skew-symmetric scheme has an invariant probability measure π_h with finite moments of all orders, and if $X_{nh} \sim P_h^n(x, \cdot\,)$

$$\|P_h^n(x, \cdot\,) - \pi_h\|_{TV} \leq M(x)e^{-\lambda n}$$

Result 2: Convergence to equilibrium

Result 2 (equilibrium measure and geometric convergence): Under appropriate assumptions, for any fixed $h \leq 1$ the skew-symmetric scheme has an invariant probability measure π_h with finite moments of all orders, and if $X_{nh} \sim P_h^n(x, \cdot\,)$

$$\|P_h^n(x, \cdot\,) - \pi_h\|_{TV} \leq M(x)e^{-\lambda n}$$

Remarks on assumptions & proof

- Additional regularity on μ to allow contraction, e.g.

$$\lim_{\|x\| \to \infty, x_i < -\|x_i\|_\infty/2} \mu_i(x) = \infty, \qquad \lim_{\|x\| \to \infty, x_i > \|x_i\|_\infty/2} \mu_i(x) = -\infty$$

- Bounded + diagonal volatility, $N \leq \sigma_{ii}(x) \leq M$, some (weak) regularity of F
- Proof is based on Foster—Lyapunov drift condition (e.g. Meyn & Tweedie)

Result 3: Bias at equilibrium

Result 3: Bias at equilibrium

Result 3 (Bias at equilibrium): Under appropriate assumptions, as the step-size $h \to 0$ the bias between π_h and π is $O(h)$, i.e. for polynomially growing $f \in C^\infty(\mathbb{R}^d)$

$$\left| \int f(x)\pi_h(dx) - \int f(x)\pi(dx) \right| \leq Ch$$

Result 3: Bias at equilibrium

Result 3 (Bias at equilibrium): Under appropriate assumptions, as the step-size $h \to 0$ the bias between π_h and π is $O(h)$, i.e. for polynomially growing $f \in C^\infty(\mathbb{R}^d)$

$$\left| \int f(x)\pi_h(dx) - \int f(x)\pi(dx) \right| \leq Ch$$

Remarks on assumptions & proof

- True diffusion process has equilibrium π, which has moments of all orders
- Slightly different type of contraction on drift

$$\langle x, \mu(x) \rangle \leq -C\|x\|^a$$

- Proof based on theorem of Lelièvre & Stoltz (expansion of numerical scheme generator)

Remarks on skew-symmetry and strong convergence

Remarks on skew-symmetry and strong convergence

- The stability of the scheme in large part comes from the fact that

$$|X_{(n+1)h} - X_{nh}| = |\zeta_{nh}| = |\xi_{nh}|$$

which is independent of the drift $\mu(X_{nh})$ (a nice property of skew-symmetric distributions)

Remarks on skew-symmetry and strong convergence

- The stability of the scheme in large part comes from the fact that

$$|X_{(n+1)h} - X_{nh}| = |\zeta_{nh}| = |\xi_{nh}|$$

  which is independent of the drift $\mu(X_{nh})$ (a nice property of skew-symmetric distributions)

- A strong convergence result requires a coupling between the numerical process X and the true diffusion Y

Remarks on skew-symmetry and strong convergence

- The stability of the scheme in large part comes from the fact that

$$|X_{(n+1)h} - X_{nh}| = |\zeta_{nh}| = |\xi_{nh}|$$

which is independent of the drift $\mu(X_{nh})$ (a nice property of skew-symmetric distributions)

- A strong convergence result requires a coupling between the numerical process X and the true diffusion Y
  - For Euler—Maruyama typically achieved via a **synchronous coupling** (conditional on knowing X_{nh})

$$\sigma(X_{nh})^{-1/2}\xi_{nh} := W_{(n+1)h} - W_{nh}$$

Remarks on skew-symmetry and strong convergence

- The stability of the scheme in large part comes from the fact that

$$|X_{(n+1)h} - X_{nh}| = |\zeta_{nh}| = |\xi_{nh}|$$

  which is independent of the drift $\mu(X_{nh})$ (a nice property of skew-symmetric distributions)

- A strong convergence result requires a coupling between the numerical process X and the true diffusion Y
  - For Euler—Maruyama typically achieved via a **synchronous coupling** (conditional on knowing X_{nh})

$$\sigma(X_{nh})^{-1/2}\xi_{nh} := W_{(n+1)h} - W_{nh}$$

  - We have now established strong convergence for the scheme when ζ_{nh} is skew-Normal, using representation (e.g. Azzalini, 2005) ($\delta(\mu,\sigma) := \alpha(\mu,\sigma)/\sqrt{1+\alpha(\mu,\sigma)^2}$)

$$\zeta_{nh} := \sqrt{1-\delta(\mu,\sigma)^2}\xi_{nh} + \delta(\mu,\sigma))|\nu_{nh}|$$

Remarks on skew-symmetry and strong convergence

- The stability of the scheme in large part comes from the fact that

$$|X_{(n+1)h} - X_{nh}| = |\zeta_{nh}| = |\xi_{nh}|$$

  which is independent of the drift $\mu(X_{nh})$ (a nice property of skew-symmetric distributions)

- A strong convergence result requires a coupling between the numerical process X and the true diffusion Y
  - For Euler—Maruyama typically achieved via a **synchronous coupling** (conditional on knowing X_{nh})

$$\sigma(X_{nh})^{-1/2}\xi_{nh} := W_{(n+1)h} - W_{nh}$$

  - We have now established strong convergence for the scheme when ζ_{nh} is skew-Normal, using representation (e.g. Azzalini, 2005) ($\delta(\mu,\sigma) := \alpha(\mu,\sigma)/\sqrt{1+\alpha(\mu,\sigma)^2}$)

$$\zeta_{nh} := \sqrt{1-\delta(\mu,\sigma)^2}\xi_{nh} + \delta(\mu,\sigma))|\nu_{nh}|$$

  - **Skew-synchronous coupling:** couple ξ_{nh} synchronously to the Brownian motion W, but leave ν_{nh} independent

Application to Langevin sampling

Unadjusted *Barker* algorithm

Application to Langevin sampling

Unadjusted *Barker* algorithm

- Recall the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

Application to Langevin sampling

Unadjusted *Barker* algorithm

- Recall the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Numerical simulation via Euler—Maruyama scheme is known as the **Unadjusted Langevin algorithm** (ULA)

$$X_{(n+1)h} = X_{nh} - h\nabla U(X_{nh}) + \sqrt{2h}\xi_{nh}$$

Application to Langevin sampling

Unadjusted *Barker* algorithm

- Recall the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Numerical simulation via Euler—Maruyama scheme is known as the **Unadjusted Langevin algorithm** (ULA)

$$X_{(n+1)h} = X_{nh} - h\nabla U(X_{nh}) + \sqrt{2h}\xi_{nh}$$

- Transient when ∇U super-linear (shown previously)

Application to Langevin sampling

Unadjusted *Barker* algorithm

- Recall the overdamped Langevin diffusion

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2}dW_t$$

- Numerical simulation via Euler—Maruyama scheme is known as the **Unadjusted Langevin algorithm** (ULA)

$$X_{(n+1)h} = X_{nh} - h\nabla U(X_{nh}) + \sqrt{2h}\xi_{nh}$$

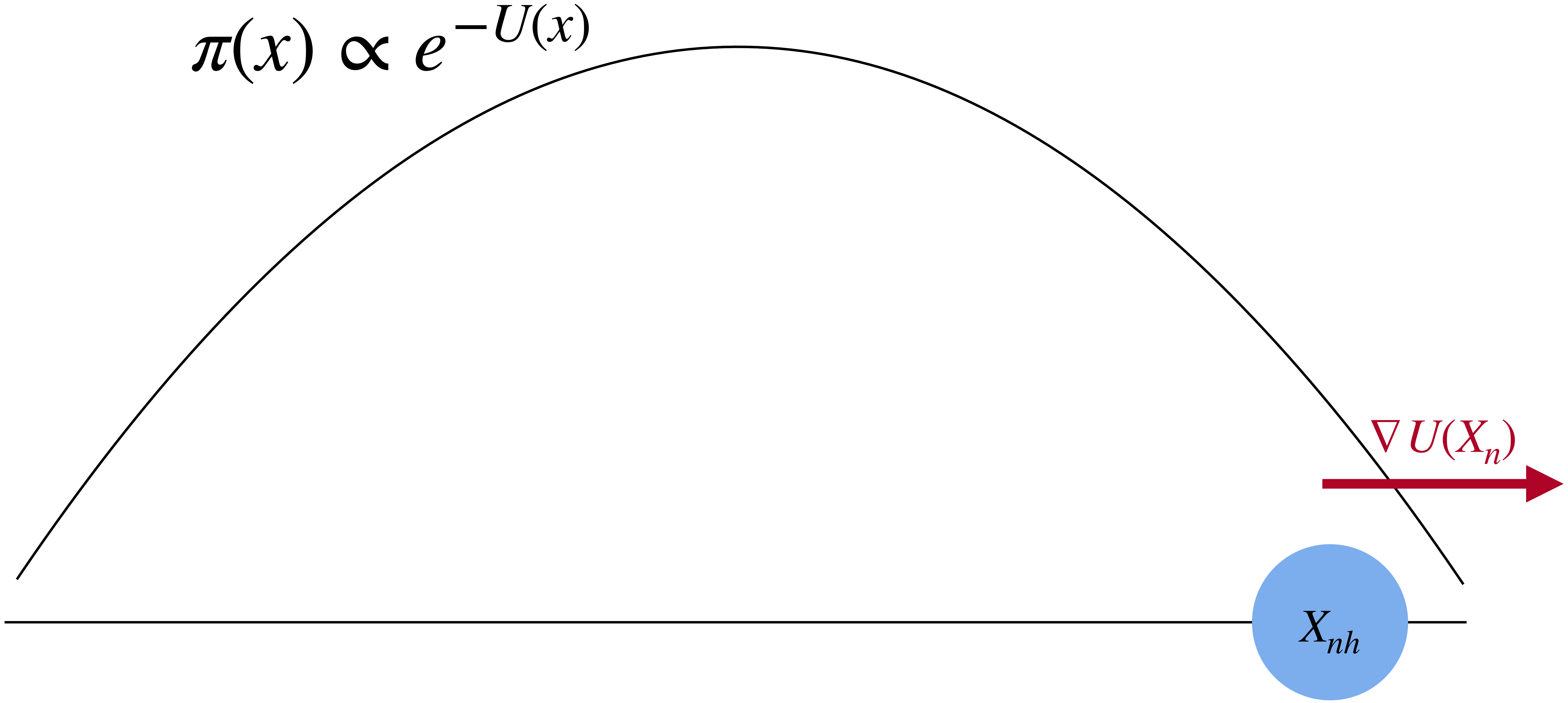
- Transient when ∇U super-linear (shown previously)

- Skew-symmetric scheme (using logistic CDF to skew) becomes $X_{i,(n+1)h} = X_{i,nh} + b_{i,nh} \cdot \xi_{i,nh}$ with

$$\mathbb{P}[b_{i,nh} = +1] = \frac{1}{1 + \exp\{\partial_i U(X_{nh}) \cdot \xi_{i,nh}\}}$$

Unadjusted Barker algorithm

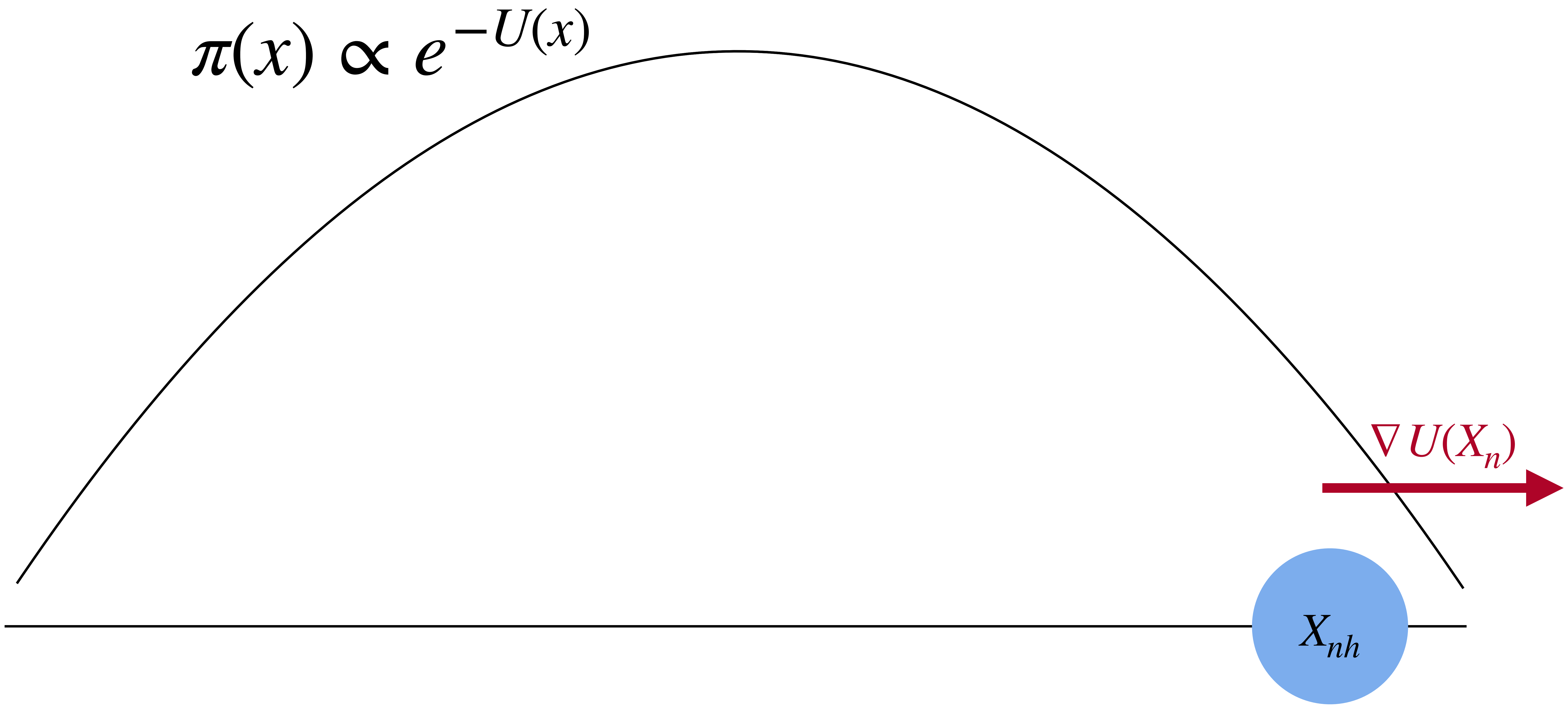
Example on $\mathbb{R}$



Unadjusted Barker algorithm

Example on $\mathbb{R}$

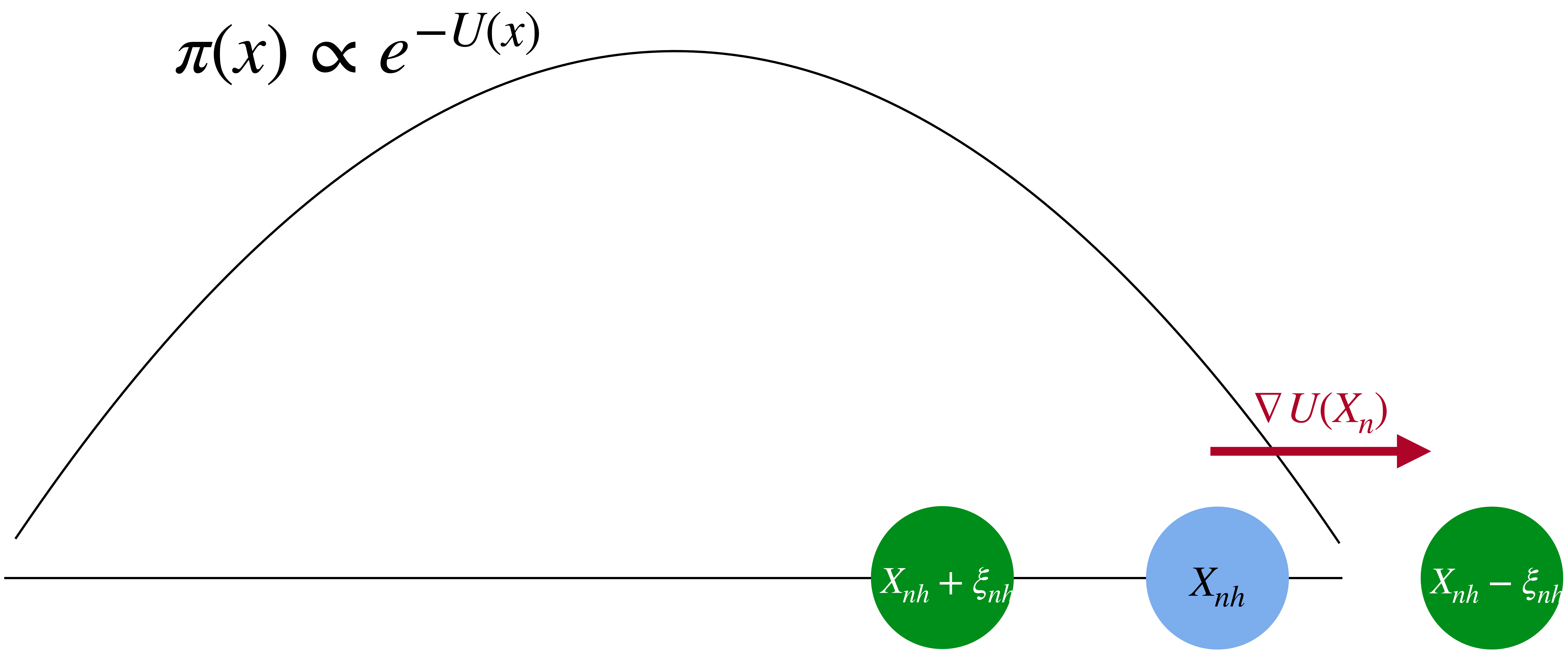
1. Draw $\xi_{nh} \sim N(0,2h)$



Unadjusted Barker algorithm

Example on $\mathbb{R}$

1. Draw $\xi_{nh} \sim N(0,2h)$

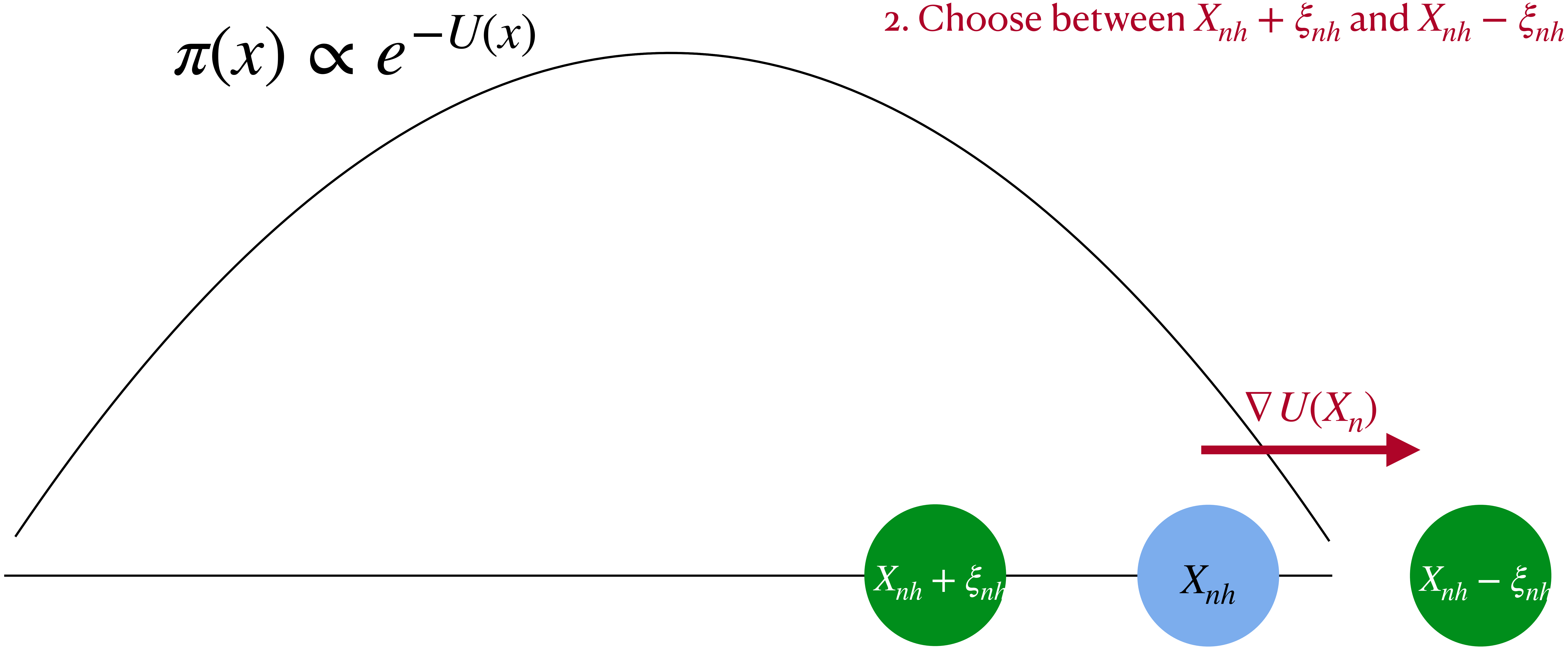


Unadjusted Barker algorithm

Example on $\mathbb{R}$

1. Draw $\xi_{nh} \sim N(0,2h)$

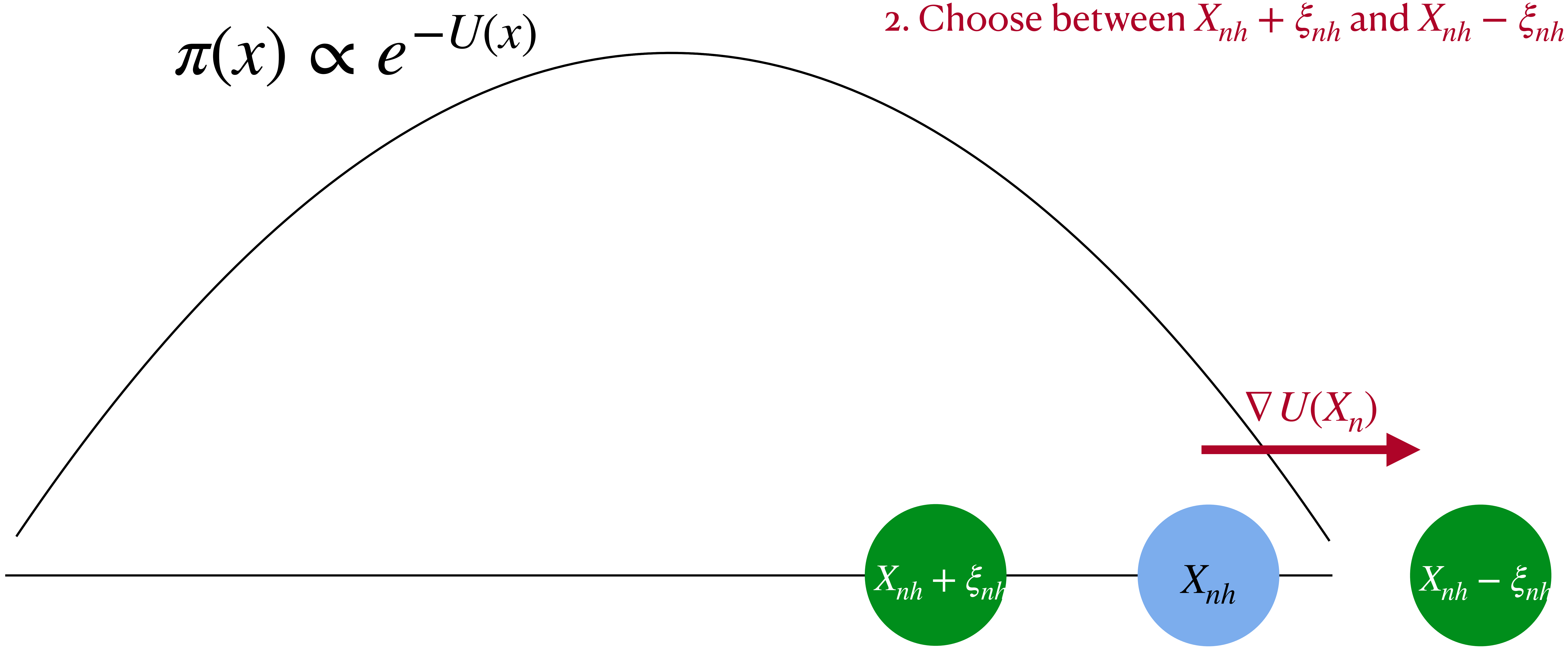
2. Choose between $X_{nh} + \xi_{nh}$ and $X_{nh} - \xi_{nh}$



Unadjusted Barker algorithm

Example on $\mathbb{R}$

1. Draw $\xi_{nh} \sim N(0,2h)$
2. Choose between $X_{nh} + \xi_{nh}$ and $X_{nh} - \xi_{nh}$



$p(X_{nh}, \xi_{nh}) > \frac{1}{2} \implies X_{nh} + \xi_{nh}$ the more likely step

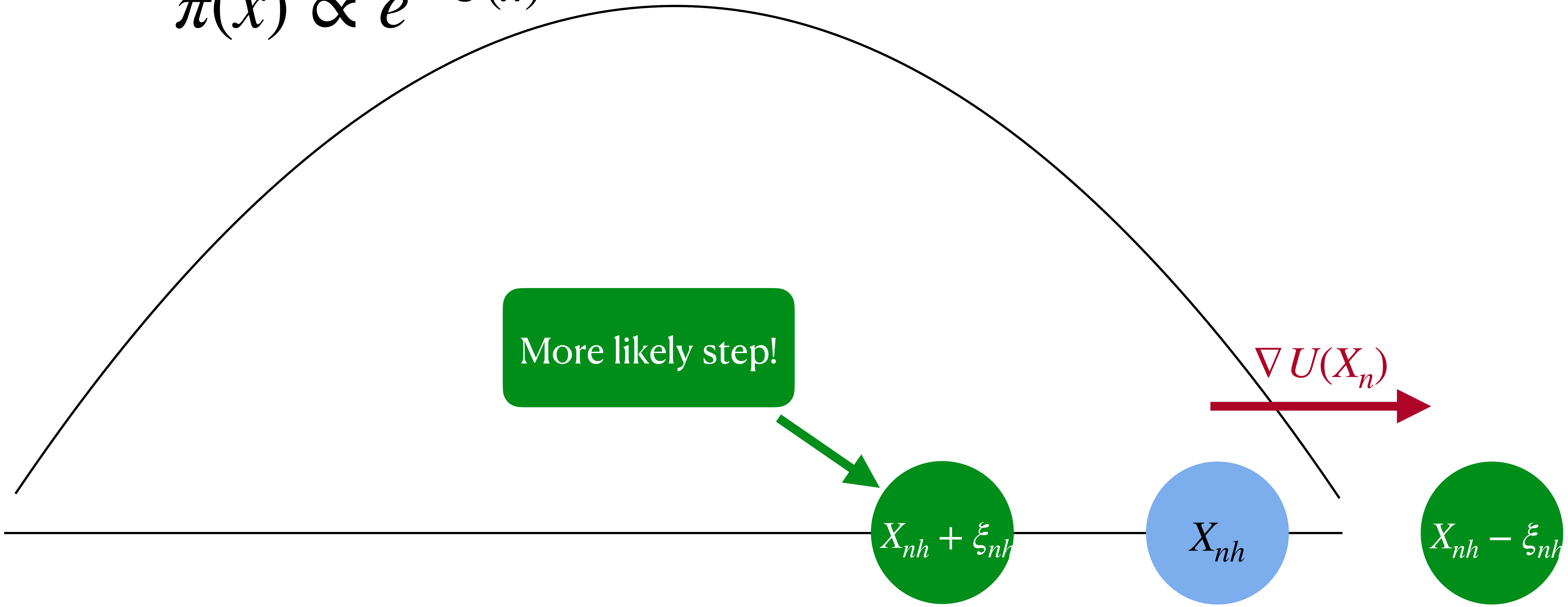
Unadjusted Barker algorithm

Example on $\mathbb{R}$

1. Draw $\xi_{nh} \sim N(0,2h)$

2. Choose between $X_{nh} + \xi_{nh}$ and $X_{nh} - \xi_{nh}$

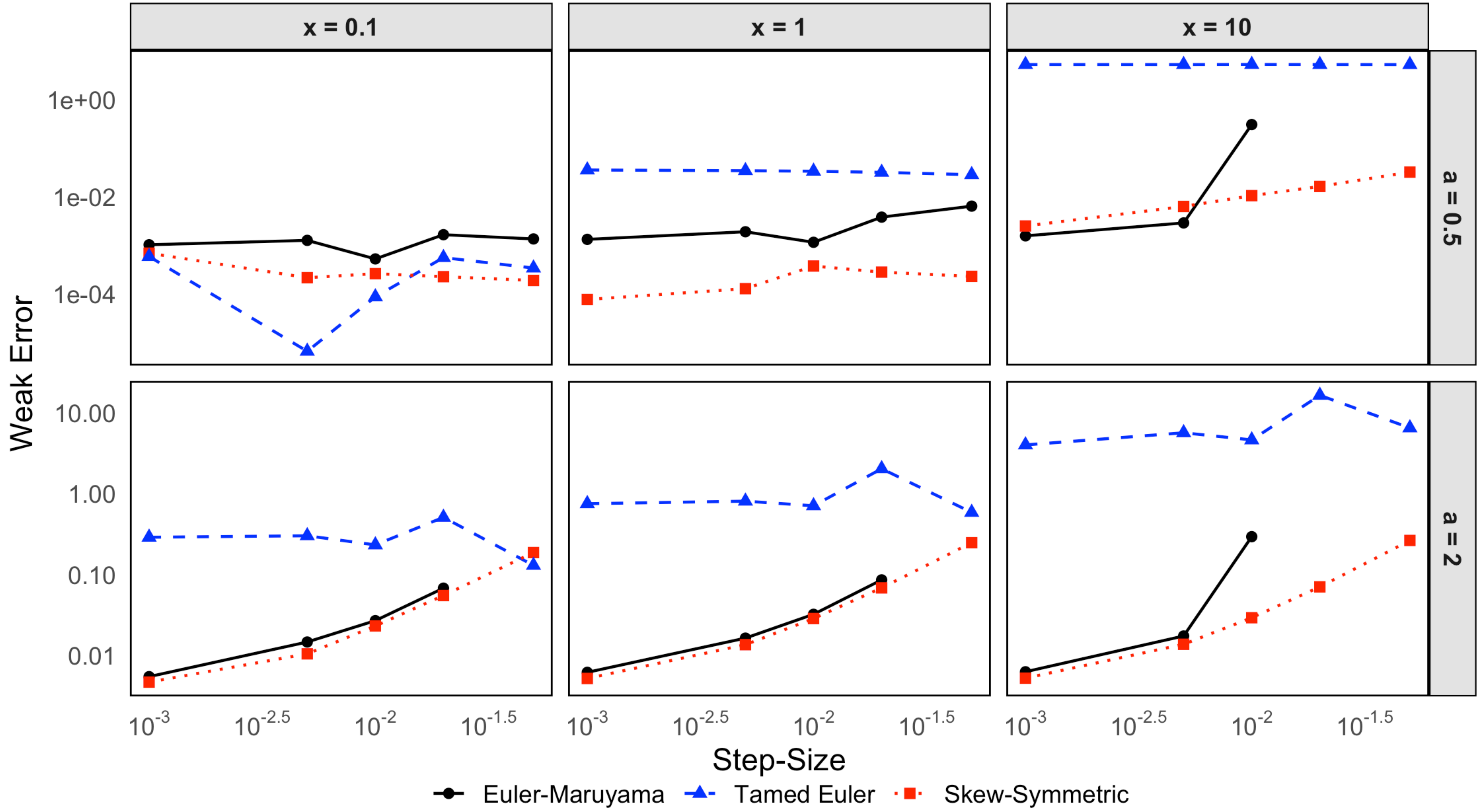
$$\pi(x) \propto e^{-U(x)}$$



$$p(X_{nh}, \xi_{nh}) > \frac{1}{2} \implies X_{nh} + \xi_{nh} \text{ the more likely step}$$

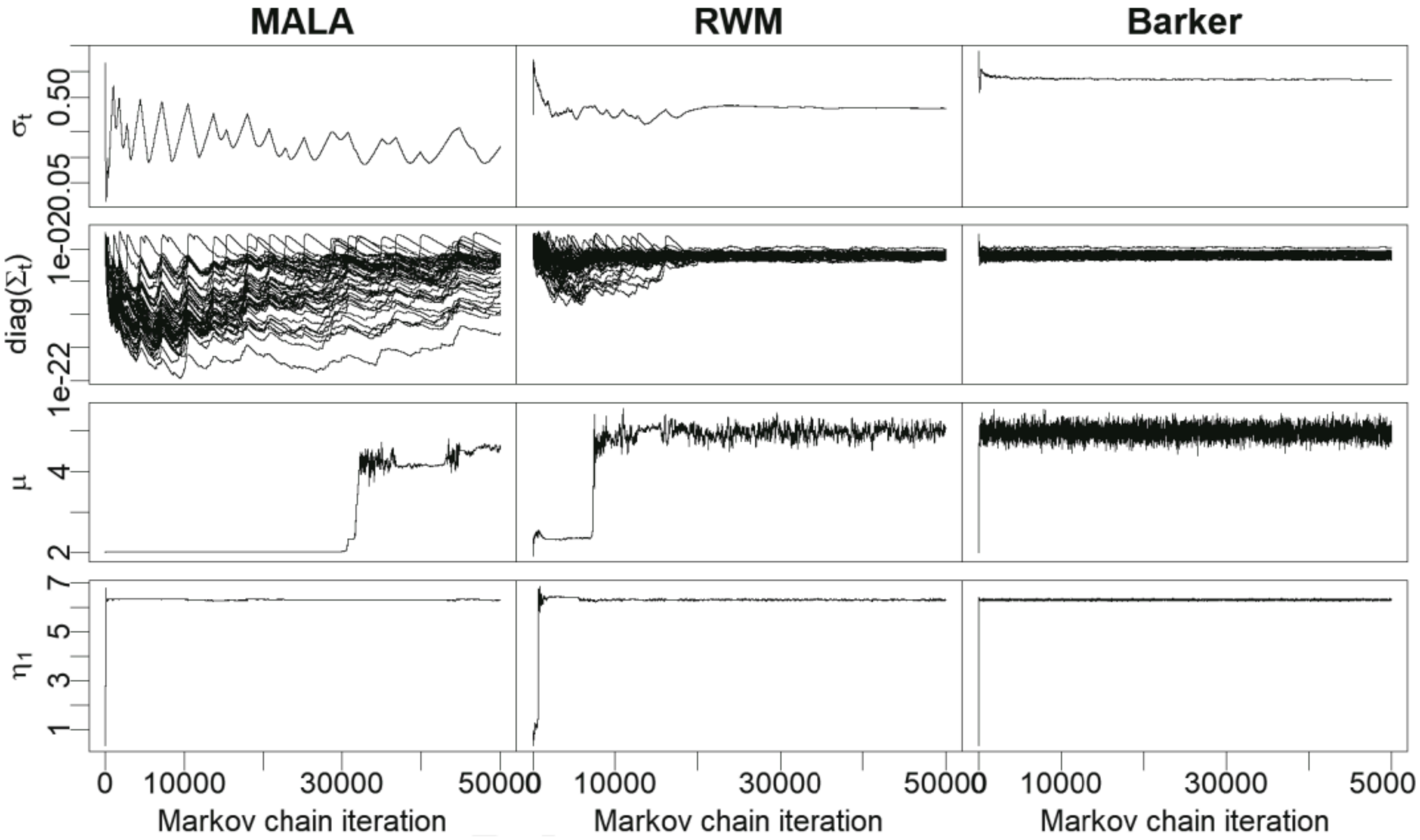
Example: finite time simulation (stochastic Ginzburg–Landau model)

$$dY_t = ((a^2/2)Y_t - Y_t^3)dt + aY_t dW_t, \quad Y_0 = x$$



Example: long time simulation (Metropolis-adjusted)

Poisson random effects model, adaptive step-size σ_t and diagonal preconditioned Σ_t



Discussion

- Current ongoing work (preprint coming in ~3 months)
  - Finalising strong mean-squared convergence/path-wise accuracy…multi-level Monte Carlo implementation
  - Hypo-elliptic scheme (e.g. kinetic Langevin sampling) + hybrid scheme (combined with Euler—Maruyama)
- Also check out the **rmcmc** R package for MCMC sampling with skew-symmetric distributions (via unadjusted/Metropolis-adjusted Barker algorithm) and general comparison of many different sampling algorithms (on CRAN)

- Further reading:

  - Iguchi, Y., Livingstone, S., Nüsken, N., Vasdekis, G., & Zhang, R. Y. (2026). Skew-symmetric schemes for stochastic differential equations with non-Lipschitz drift: an unadjusted Barker algorithm. *IMA Journal of Numerical Analysis (forthcoming)*. **(Unadjusted scheme)**
  - Livingstone, S., & Zanella, G. (2022). The Barker proposal: Combining robustness and efficiency in gradient-based MCMC. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 84(2), 496-523*. **(Metropolis-adjusted scheme)**
  - Vogrinc, J., Livingstone, S., & Zanella, G. (2023). Optimal design of the Barker proposal and other locally balanced Metropolis–Hastings algorithms. *Biometrika, 110(3), 579-595*. **(Metropolis-adjusted scheme)**
  - Hardcastle, L., Livingstone, S., & Baio, G. (2025). Diffusion piecewise exponential models for survival extrapolation using Piecewise Deterministic Monte Carlo. arXiv preprint arXiv:2505.05932. **(Diffusion priors for Bayesian models)**

Appendices

Example 1: Weak order and equilibrium bias

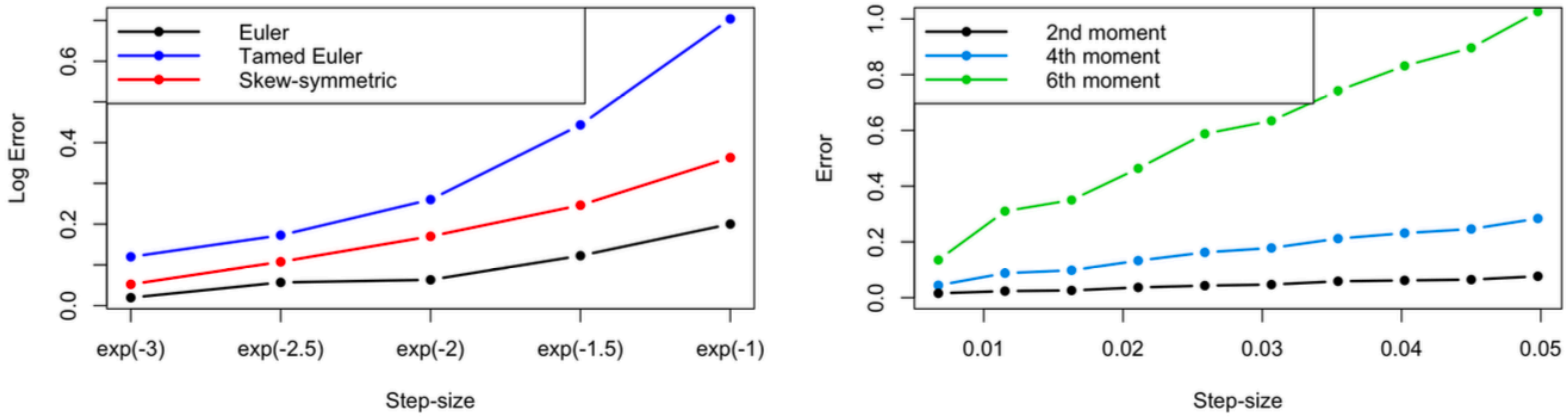


Figure 1: Figure on the left-hand side shows logarithm of the absolute error when computing $\mathbb{E}_x[Y_T^2]$ for T fixed for the Ornstein–Uhlenbeck process. Figure on the right-hand side shows absolute error when computing moments of the density $\pi(x) \propto e^{-x^4/4}$ by simulating the overdamped Langevin diffusion over long time scales.

Example 2: Poisson random effects model

$$
\begin{aligned}
y_{ij}|\eta_i &\sim \text{Poi}(e^{\eta_i}), && j = 1, \ldots, J, \\
\eta_i|\mu &\sim N(\mu, 1), && i = 1, \ldots, I, \\
\mu &\sim N(0, \sigma_\mu^2).
\end{aligned}
$$

$$\pi(x) \propto \exp\{-U(x)\}$$

$$U(x) = J\sum_i e^{\eta_i} - \sum_{i,j} y_{ij}\eta_i + \frac{1}{2}\sum_i(\eta_i - \mu)^2 + \frac{\mu^2}{2\sigma_\mu^2}.$$

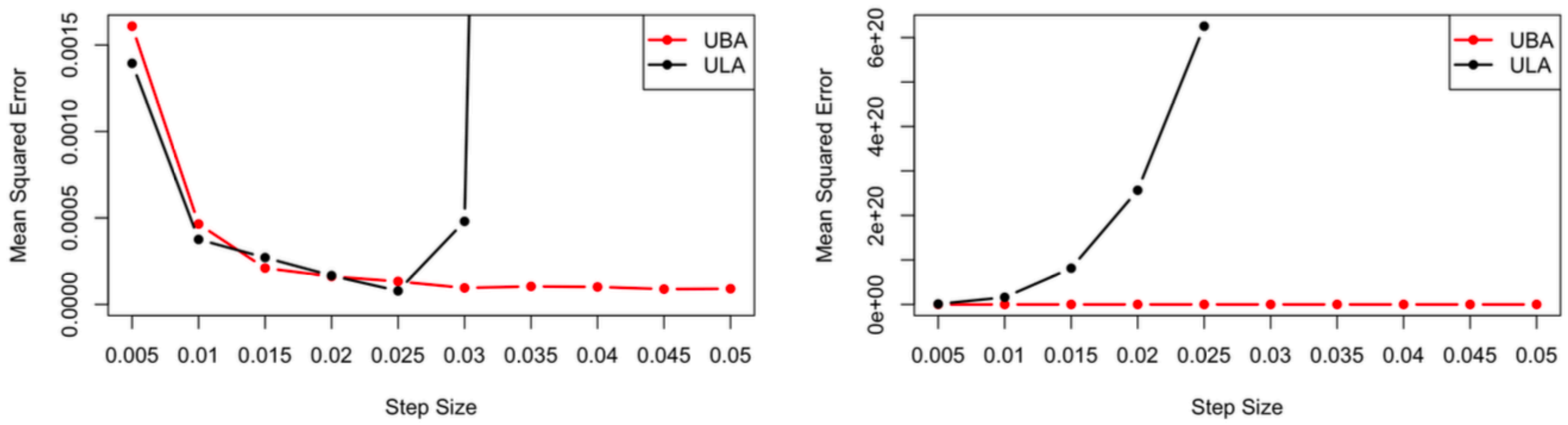


Figure 2: Mean squared error comparisons for the Poisson random effects example. Figure 2(a) (left-hand side) shows means squared error when initialised at the true value μ^*. Figure 2(b) (right-hand side) is initialised from a $N(5, 10^2)$ sample.

Example 3: Soft-spheres in an anharmonic trap

Example 3: Soft-spheres in an anharmonic trap

- Particle locations $X_t^{(i)} \in \mathbb{R}^2$ for $i \in \{1,2,\ldots,N\}$ follow dynamics (for $B, A, D, r > 0$)

$$dX_t^{(i)} = \underbrace{-4BX_t^{(i)}\|X_t^{(i)}\|^2 dt}_{\text{anharmonic trap}} + \underbrace{\frac{A}{Nr^2}\sum_{j=1}^{N}(X_t^{(i)} - X_t^{(j)})e^{-\|X_t^{(i)}-X_t^{(j)}\|^2/2r^2}dt}_{\text{Soft-spheres interaction}} + \sqrt{2D}\,dW_t$$

Example 3: Soft-spheres in an anharmonic trap

- Particle locations $X_t^{(i)} \in \mathbb{R}^2$ for $i \in \{1,2,\ldots,N\}$ follow dynamics (for $B, A, D, r > 0$)

$$dX_t^{(i)} = \underbrace{-4BX_t^{(i)}\|X_t^{(i)}\|^2 dt}_{\text{anharmonic trap}} + \underbrace{\frac{A}{Nr^2}\sum_{j=1}^{N}(X_t^{(i)} - X_t^{(j)})e^{-\|X_t^{(i)}-X_t^{(j)}\|^2/2r^2}dt}_{\text{Soft-spheres interaction}} + \sqrt{2D}dW_t$$

- We vary B and step-size h and compare **numerical explosion** probabilities for Euler—Maruyama and the skew-symmetric scheme

Example 3: Soft-spheres in an anharmonic trap

- Particle locations $X_t^{(i)} \in \mathbb{R}^2$ for $i \in \{1,2,\ldots,N\}$ follow dynamics (for $B, A, D, r > 0$)

$$dX_t^{(i)} = \underbrace{-4BX_t^{(i)}\|X_t^{(i)}\|^2 dt}_{\text{anharmonic trap}} + \underbrace{\frac{A}{Nr^2}\sum_{j=1}^{N}(X_t^{(i)} - X_t^{(j)})e^{-\|X_t^{(i)}-X_t^{(j)}\|^2/2r^2}dt}_{\text{Soft-spheres interaction}} + \sqrt{2D}\,dW_t$$

- We vary B and step-size h and compare **numerical explosion** probabilities for Euler—Maruyama and the skew-symmetric scheme

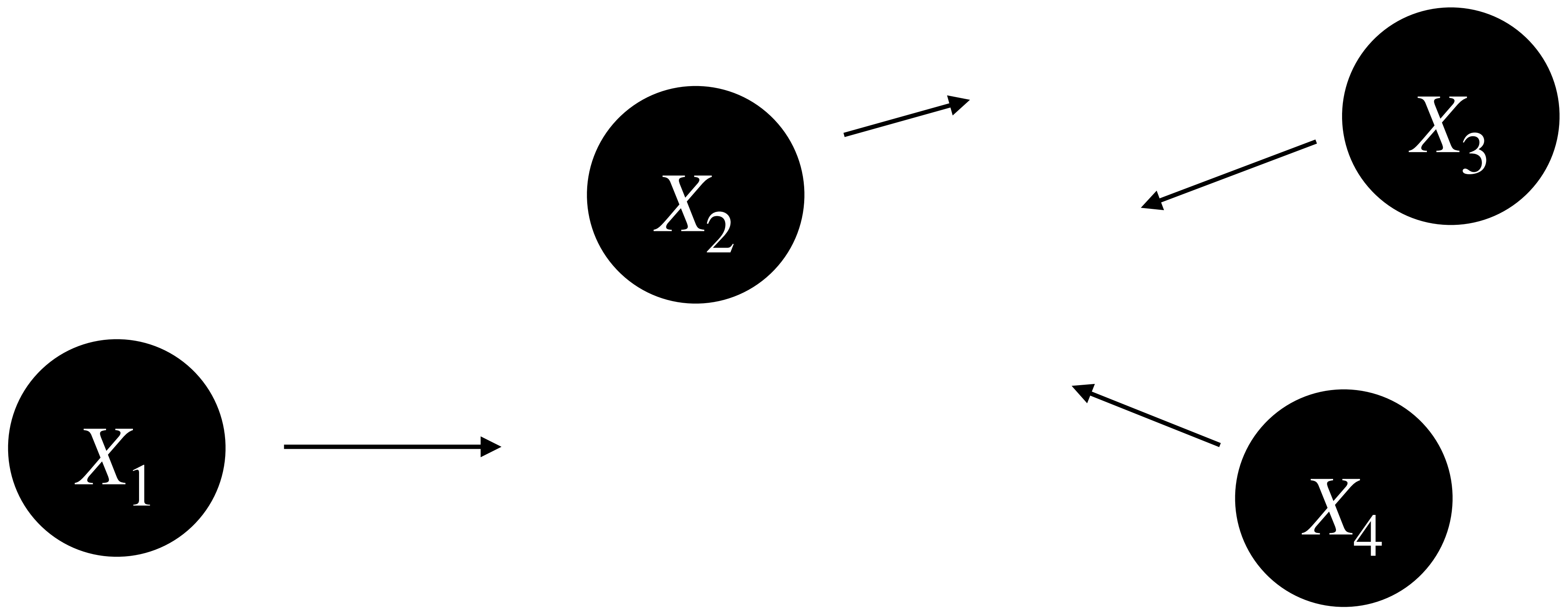
i.e. particle reaches a position too large for computer to store

Example 3: Soft-spheres in an anharmonic trap

- Particle locations $X_t^{(i)} \in \mathbb{R}^2$ for $i \in \{1,2,\ldots,N\}$ follow dynamics (for $B, A, D, r > 0$)

$$dX_t^{(i)} = \underbrace{-4BX_t^{(i)}\|X_t^{(i)}\|^2 dt}_{\text{anharmonic trap}} + \frac{A}{Nr^2} \underbrace{\sum_{j=1}^{N} (X_t^{(i)} - X_t^{(j)}) e^{-\|X_t^{(i)} - X_t^{(j)}\|^2/2r^2} dt}_{\text{Soft-spheres interaction}} + \sqrt{2D}\, dW_t$$

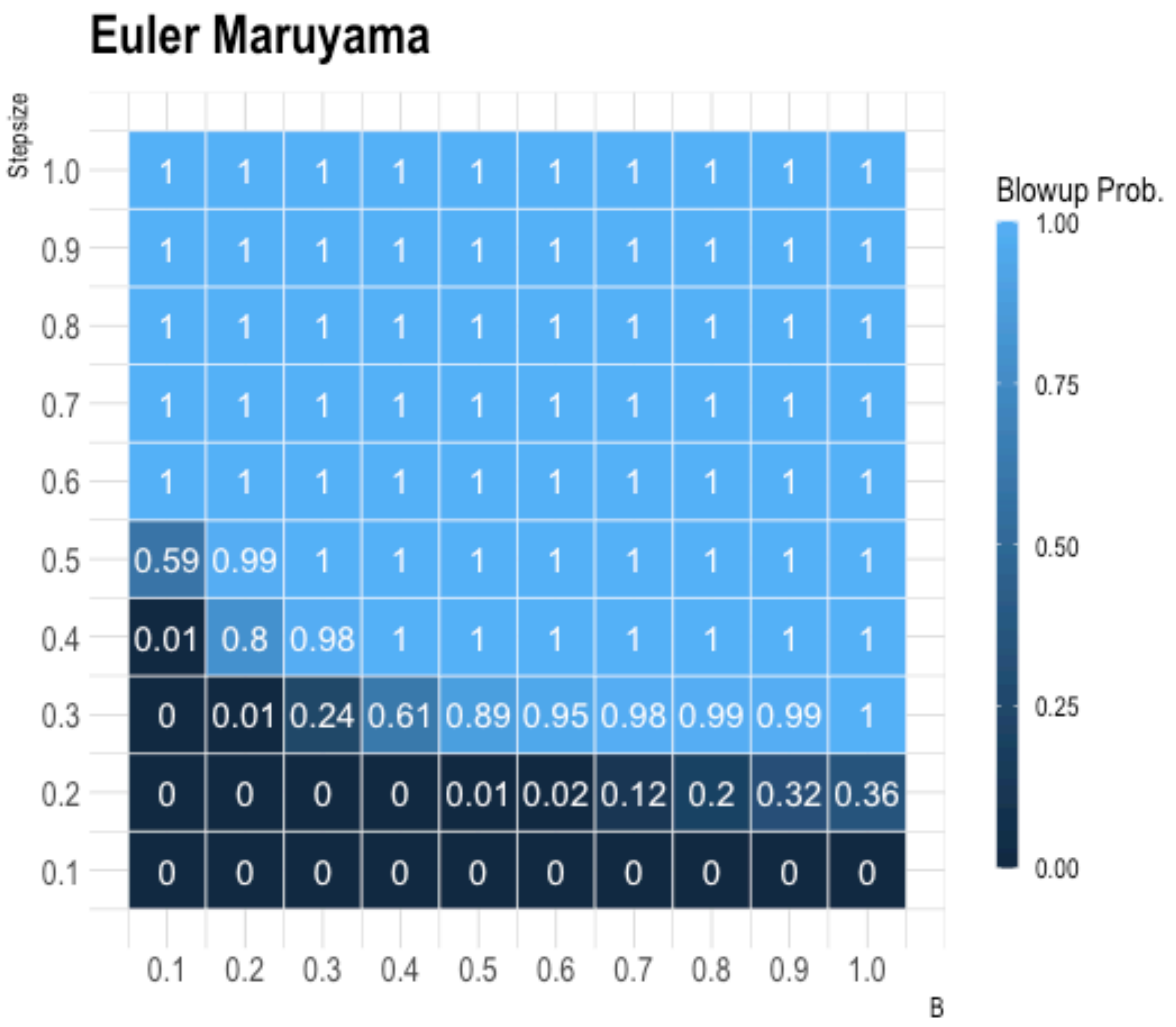
- We vary B and step-size h and compare **numerical explosion** probabilities for Euler—Maruyama and the skew-symmetric scheme

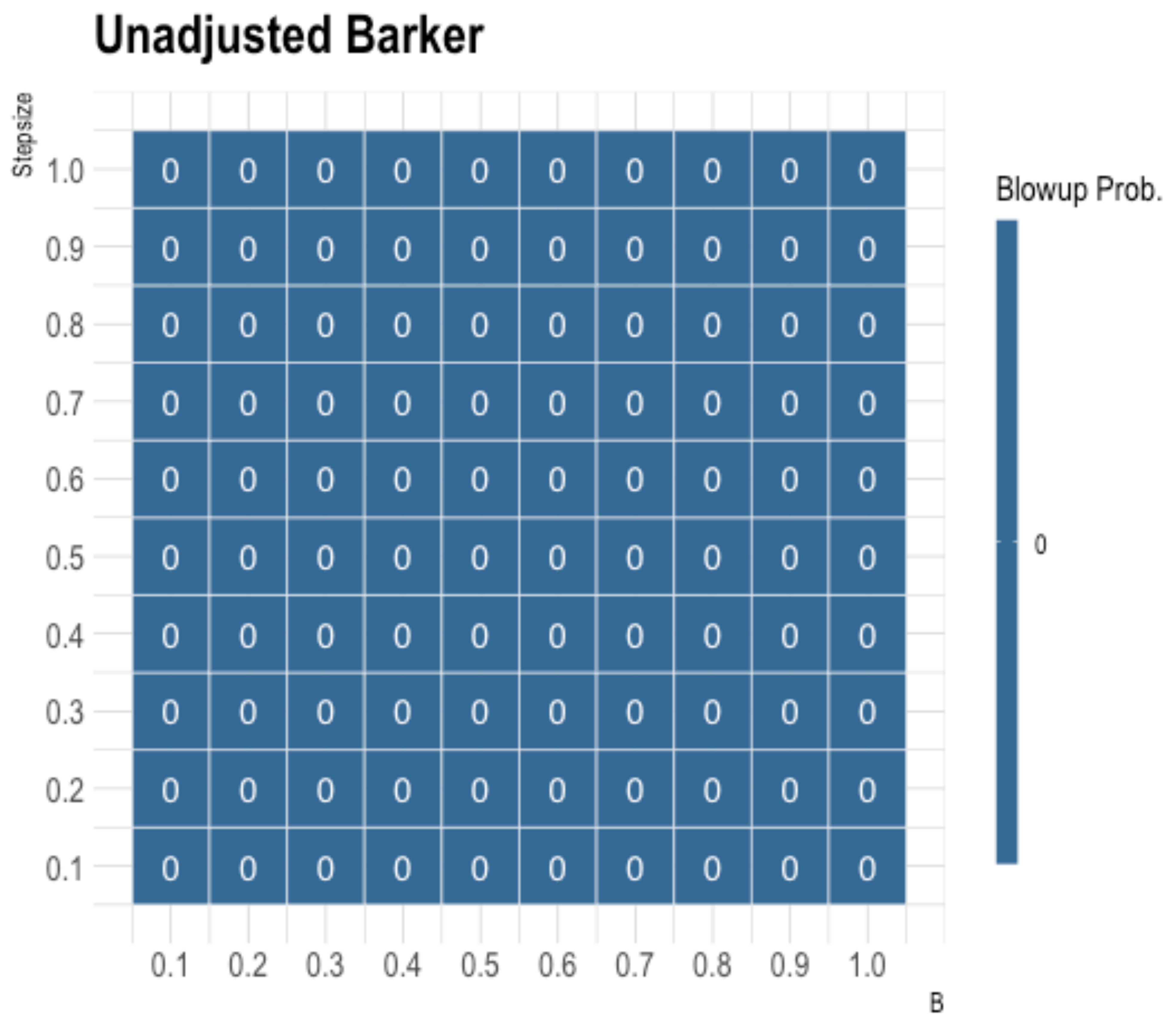


i.e. particle reaches a position too large for computer to store

Example 3: Soft-spheres in an anharmonic trap

Numerical explosion probability after 10 steps, 100 runs, N=50 particles





Worked examples

Worked examples

- In the skew-Normal case taking $\sigma(x) = 1$ gives

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = h \cdot \frac{\mu(X_n)}{\sqrt{1 + h \cdot \frac{\pi}{4} \cdot \mu(X_n)^2}}$$

Worked examples

- In the skew-Normal case taking $\sigma(x) = 1$ gives

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = h \cdot \frac{\mu(X_n)}{\sqrt{1 + h \cdot \frac{\pi}{4} \cdot \mu(X_n)^2}}$$

- Contrast with Euler—Maruyama

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = \mu(X_n)h$$

Worked examples

- In the skew-Normal case taking $\sigma(x) = 1$ gives

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = h \cdot \frac{\mu(X_n)}{\sqrt{1 + h \cdot \frac{\pi}{4} \cdot \mu(X_n)^2}}$$

- Contrast with Euler—Maruyama

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = \mu(X_n)h$$

- When μ grows **super linearly** (i.e. not Lipschitz) the latter is unstable

Worked examples

- In the skew-Normal case taking $\sigma(x) = 1$ gives

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = h \cdot \frac{\mu(X_n)}{\sqrt{1 + h \cdot \frac{\pi}{4} \cdot \mu(X_n)^2}}$$

- Contrast with Euler—Maruyama

$$\mathbb{E}[X_{n+h} - X_n \,|\, X_n] = \mu(X_n)h$$

- When μ grows **super linearly** (i.e. not Lipschitz) the latter is unstable
- To compare with Tamed Euler choose e.g. $\sigma(x) = \max(\epsilon, |x|)$ (see preprint)