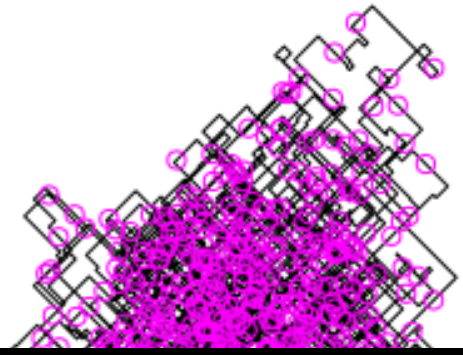
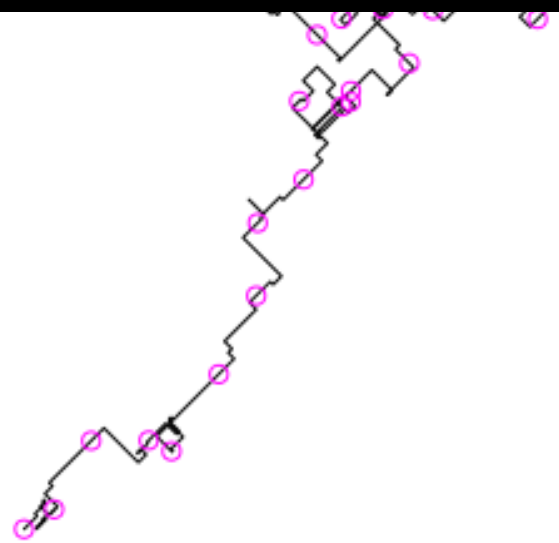


Piecewise Deterministic Monte Carlo

CIRM, Advances in Computational Statistical Physics

Joris Bierkens

21 September 2018



Acknowledgements

Collaborators

Kengo Kamatani

Gareth Roberts

Pierre-André Zitt

Financial support

EPSRC
Pioneering research and skills

Outline

Next Section

1. The Zig-Zag process and Bouncy Particle Sampler
2. Simulation and Subsampling
3. Convergence and efficiency

One-dimensional Zig-Zag process

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

The Zig-Zag process is a continuous time Markov process with states $(X(t), V(t)) \in \mathbb{R} \times \{-1, +1\}$.

One-dimensional Zig-Zag process

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

The Zig-Zag process is a continuous time Markov process with states $(X(t), V(t)) \in \mathbb{R} \times \{-1, +1\}$.

$X(t)$ moves in the direction $V(t)$, so $X(t) = X(0) + \int_0^t V(s)\, ds$.

One-dimensional Zig-Zag process

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

The Zig-Zag process is a continuous time Markov process with states $(X(t), V(t)) \in \mathbb{R} \times \{-1, +1\}$.

$X(t)$ moves in the direction $V(t)$, so $X(t) = X(0) + \int_0^t V(s)\, ds$.

$V(t)$ switches sign with switching intensity $\lambda(X(t), V(t))$, i.e. the first switching time T has distribution

$$\mathbb{P}(T \geq t) = \exp\left(- \int_0^t \lambda(X(s), V(s))\, ds \right).$$

One-dimensional Zig-Zag process

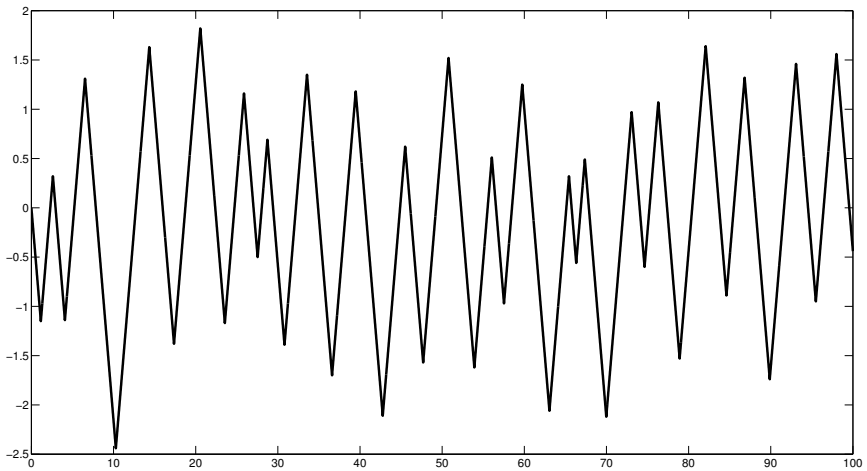
(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

The Zig-Zag process is a continuous time Markov process with states $(X(t), V(t)) \in \mathbb{R} \times \{-1, +1\}$.

$X(t)$ moves in the direction $V(t)$, so $X(t) = X(0) + \int_0^t V(s)\, ds$.

$V(t)$ switches sign with switching intensity $\lambda(X(t), V(t))$, i.e. the first switching time T has distribution

$$\mathbb{P}(T \geq t) = \exp\left(- \int_0^t \lambda(X(s), V(s))\, ds \right).$$



Relation between switching rate and potential

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

- Potential $U(x) = -\log \pi(x)$

Relation between switching rate and potential

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

- Potential $U(x) = -\log \pi(x)$
- π is stationary if and only if $\lambda(x, +1) - \lambda(x, -1) = U'(x)$ for all x.

Relation between switching rate and potential

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

- Potential $U(x) = -\log \pi(x)$
- π is stationary if and only if $\lambda(x, +1) - \lambda(x, -1) = U'(x)$ for all x.
- Equivalently,

$$\lambda(x, v) = \underbrace{\max(0, vU'(x))}_{\text{canonical switching rate}} + \underbrace{\gamma(x)}_{\text{excess switching rate}}, \quad \gamma(x) \geq 0.$$

Relation between switching rate and potential

(B., Roberts, *Ann. Appl. Prob. 2017*, https://arxiv.org/abs/1509.00302)

- Potential $U(x) = -\log \pi(x)$
- π is stationary if and only if $\lambda(x, +1) - \lambda(x, -1) = U'(x)$ for all x.
- Equivalently,

$$\lambda(x, v) = \underbrace{\max(0, vU'(x))}_{\text{canonical switching rate}} + \underbrace{\gamma(x)}_{\text{excess switching rate}}, \quad \gamma(x) \geq 0.$$

Example: Gaussian distribution $\mathcal{N}(0, \sigma^2)$

- Density $\pi(x) \propto \exp(-x^2/(2\sigma^2))$
- Potential $U(x) = x^2/(2\sigma^2)$
- Derivative $U'(x) = x/\sigma^2$
- Switching rates $\lambda(x, v) = (vx/\sigma^2)_+ + \gamma(x)$

Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, https://arxiv.org/abs/1607.03188**)**

- Target $\pi(x) = \exp(-U(x))$ on $\mathbb{R}^d$.
- Set of directions $v \in \{-1, +1\}^d$.
- Switching rates $\lambda_i(x, v) = (v_i \partial_i U(x))_+ + \gamma_i(x, v)$, for $i = 1, \ldots, d$.

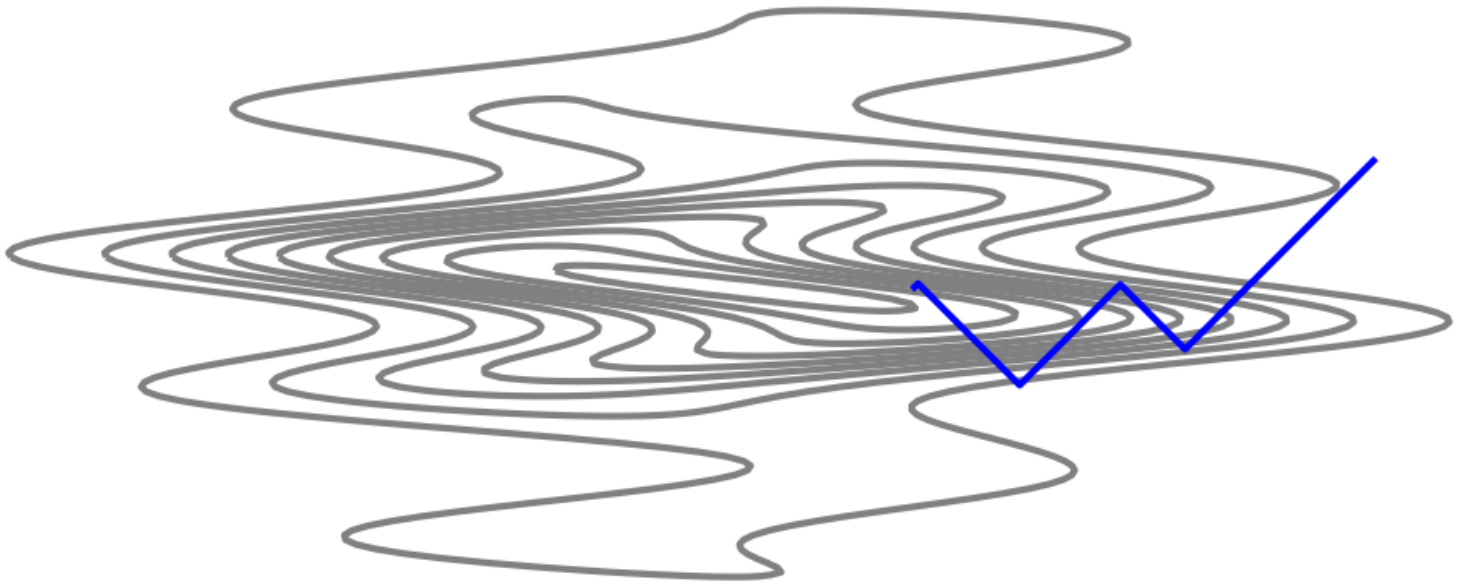
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, https://arxiv.org/abs/1607.03188**)**

- Target $\pi(x) = \exp(-U(x))$ on $\mathbb{R}^d$.
- Set of directions $v \in \{-1, +1\}^d$.
- Switching rates $\lambda_i(x, v) = (v_i \partial_i U(x))_+ + \gamma_i(x, v)$, for $i = 1, \ldots, d$.
- The excess switching rate $\gamma_i(x, v)$ should not depend on the i-th component of v.

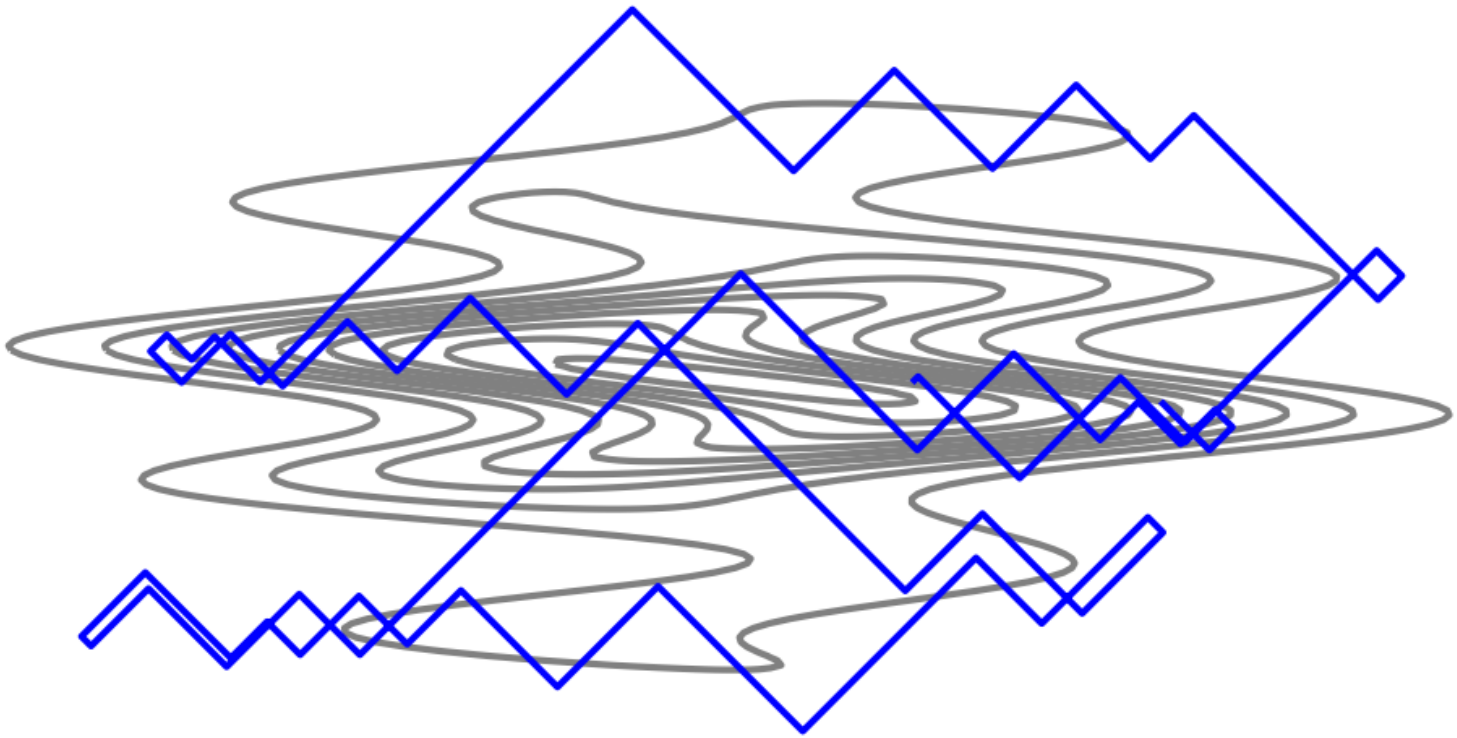
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



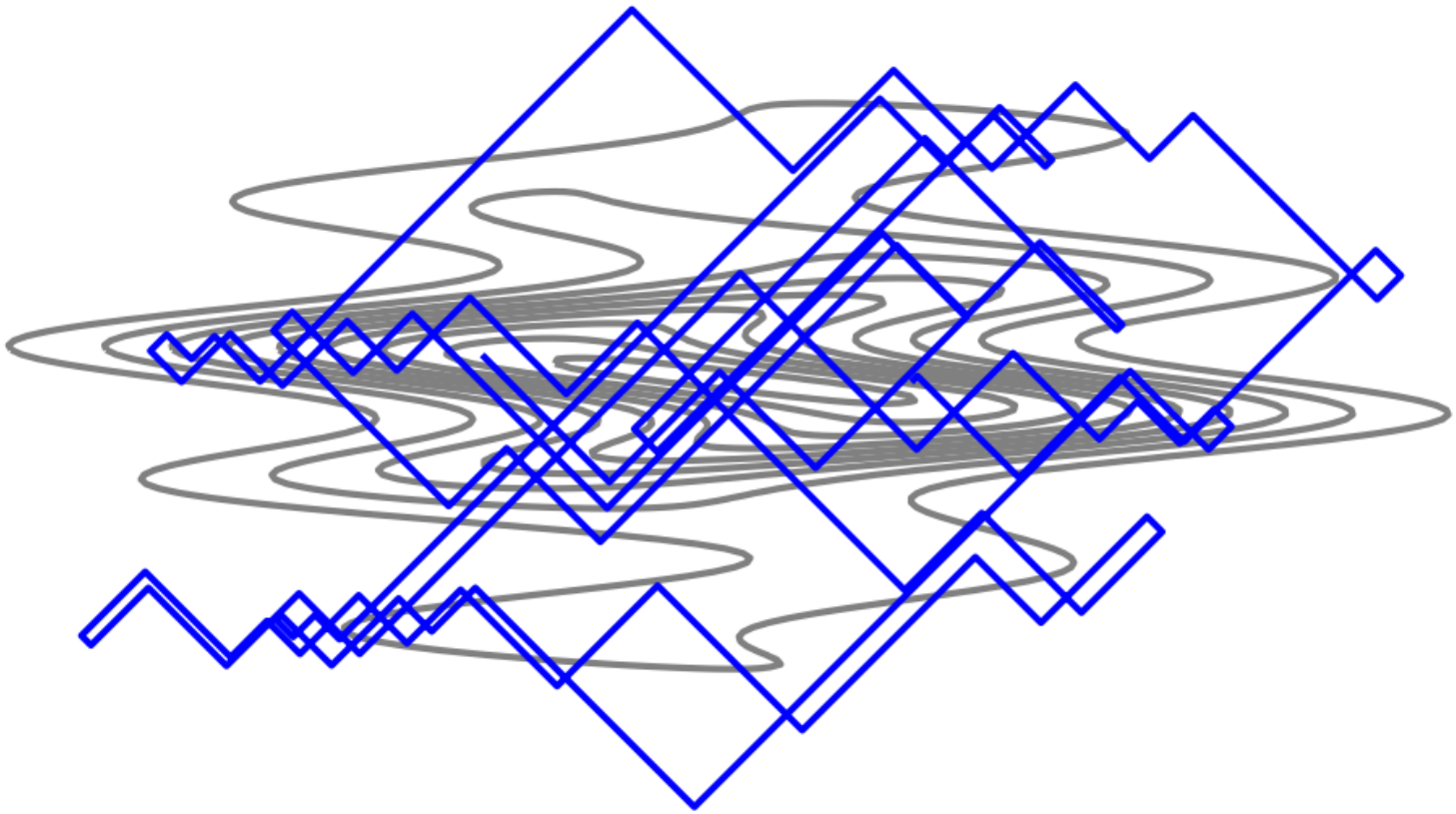
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



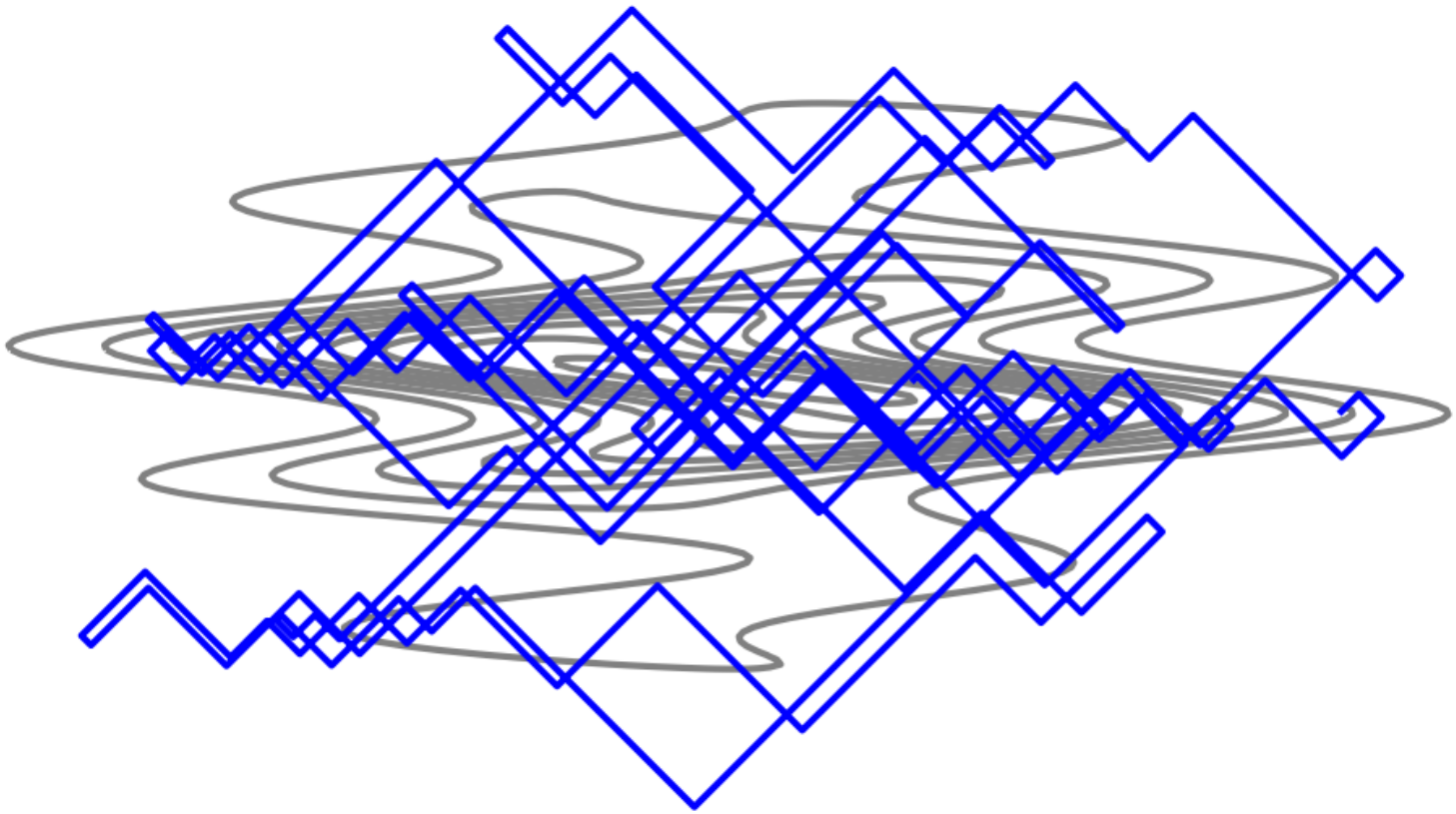
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



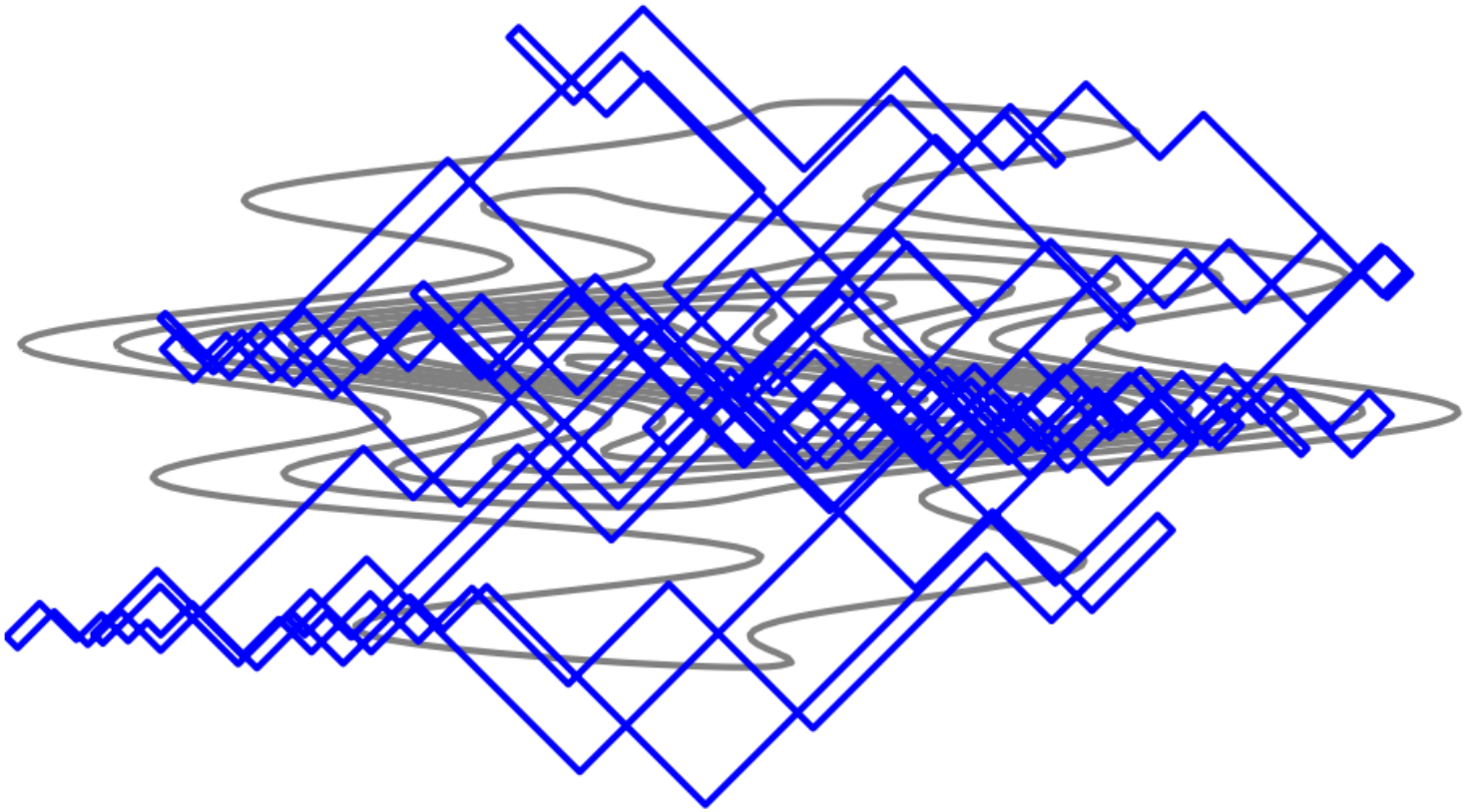
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



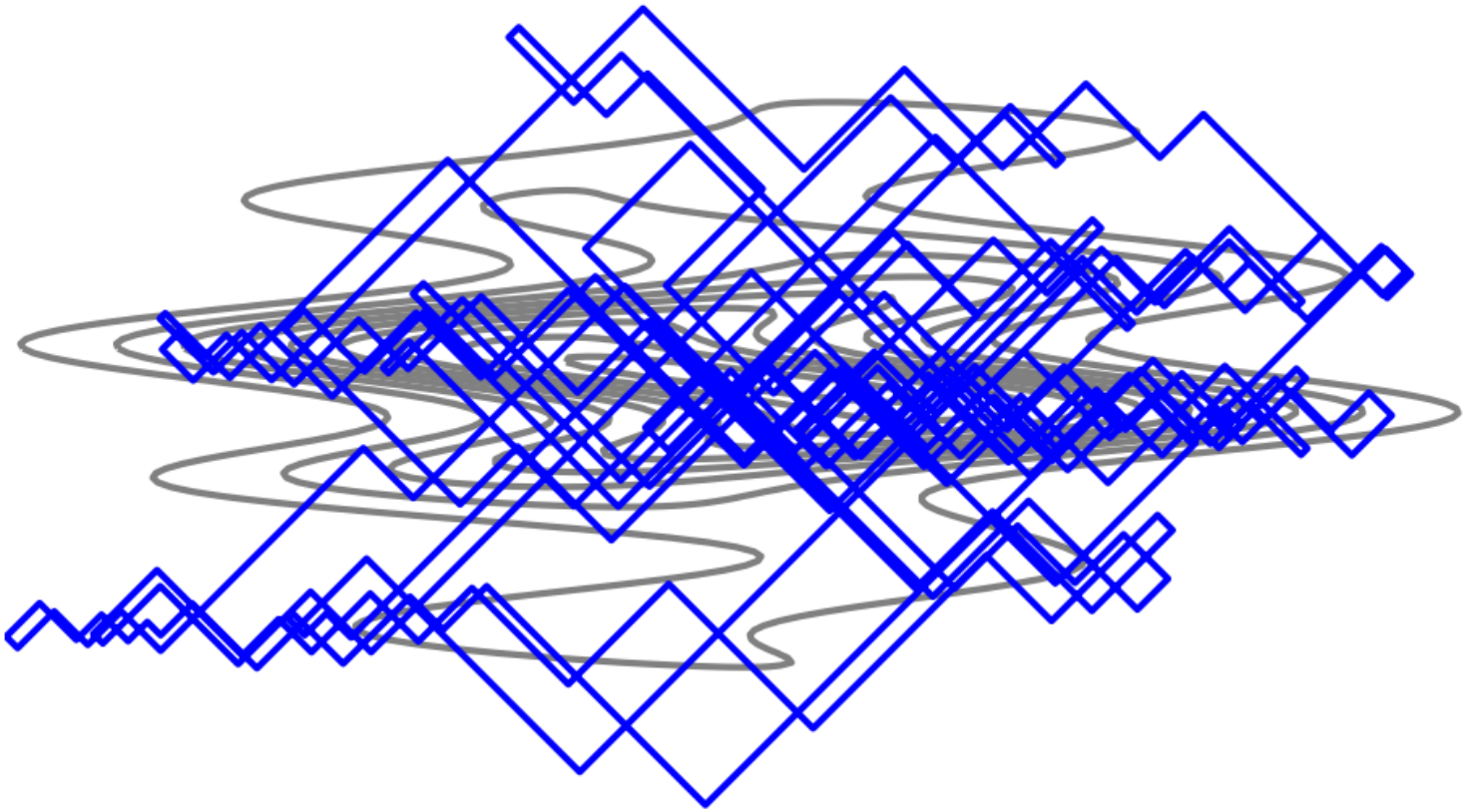
Multi-dimensional Zig-Zag process

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



Multi-dimensional Zig-Zag process

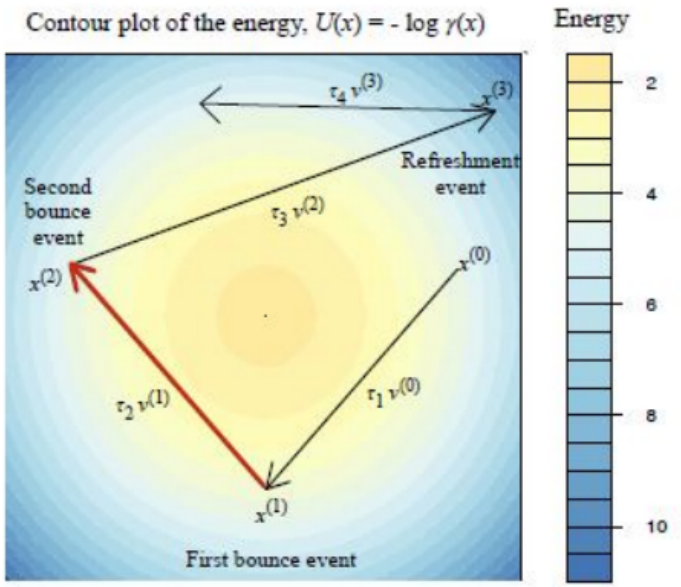
(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



The Bouncy Particle Sampler

(Bouchard-Côté, Vollmer, Doucet, JASA, 2017, https://arxiv.org/abs/1510.02451**)**

The *Bouncy Particle Sampler (BPS)* is a second canonical example of a PDMP which can be used for sampling.
State space is $\mathbb{R}^d \times \mathbb{R}^d$ with stationary distribution $\pi(x)\,dx \otimes \mathcal{N}(0, \sigma^2 I_n)$.
The BPS bounces off at random contours of π, through specular reflection.
Additionally, the momentum gets refreshed after at rate λ_{ref}.



Zig-Zag and BPS in terms of their generator

Zig-Zag process:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \sum_{i=1}^{d} [(v_i \partial_{x_i} U(x))^+ + \gamma_i(x,v)] \left(f(x, F_i v) - f(x,v)\right),$$

where $F_i v$ flips the i-th component of v

Zig-Zag and BPS in terms of their generator

Zig-Zag process:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \sum_{i=1}^{d} [(v_i \partial_{x_i} U(x))^+ + \gamma_i(x,v)] \left(f(x, F_i v) - f(x,v)\right),$$

where $F_i v$ flips the i-th component of v

Bouncy Particle Sampler:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \langle v, \nabla_x U(x)\rangle^+ (f(x, R(x)v) - f(x,v)) + \lambda_{\text{ref}} \int_V [f(x,u) - f(x,v)]\nu(dv),$$

Zig-Zag and BPS in terms of their generator

Zig-Zag process:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \sum_{i=1}^{d}[(v_i\partial_{x_i}U(x))^+ + \gamma_i(x,v)]\left(f(x,F_iv) - f(x,v)\right),$$

where F_iv flips the i-th component of v

Bouncy Particle Sampler:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \langle v, \nabla_x U(x)\rangle^+(f(x,R(x)v) - f(x,v)) + \lambda_{\text{ref}}\int_V [f(x,u) - f(x,v)]\nu(dv),$$

where

$$R(x)v = v - 2\frac{\langle \nabla_x U(x), v\rangle}{\|\nabla U(x)\|^2}\nabla U(x),$$

Zig-Zag and BPS in terms of their generator

Zig-Zag process:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \sum_{i=1}^{d}[(v_i\partial_{x_i}U(x))^+ + \gamma_i(x,v)]\left(f(x,F_iv) - f(x,v)\right),$$

where F_iv flips the i-th component of v

Bouncy Particle Sampler:

$$\mathcal{L}f(x,v) = \langle v, \nabla_x f(x)\rangle + \langle v, \nabla_x U(x)\rangle^+(f(x,R(x)v) - f(x,v))$$
$$+ \lambda_{\text{ref}}\int_V [f(x,u) - f(x,v)]\nu(dv),$$

where

$$R(x)v = v - 2\frac{\langle \nabla_x U(x), v\rangle}{\|\nabla U(x)\|^2}\nabla U(x),$$

and e.g

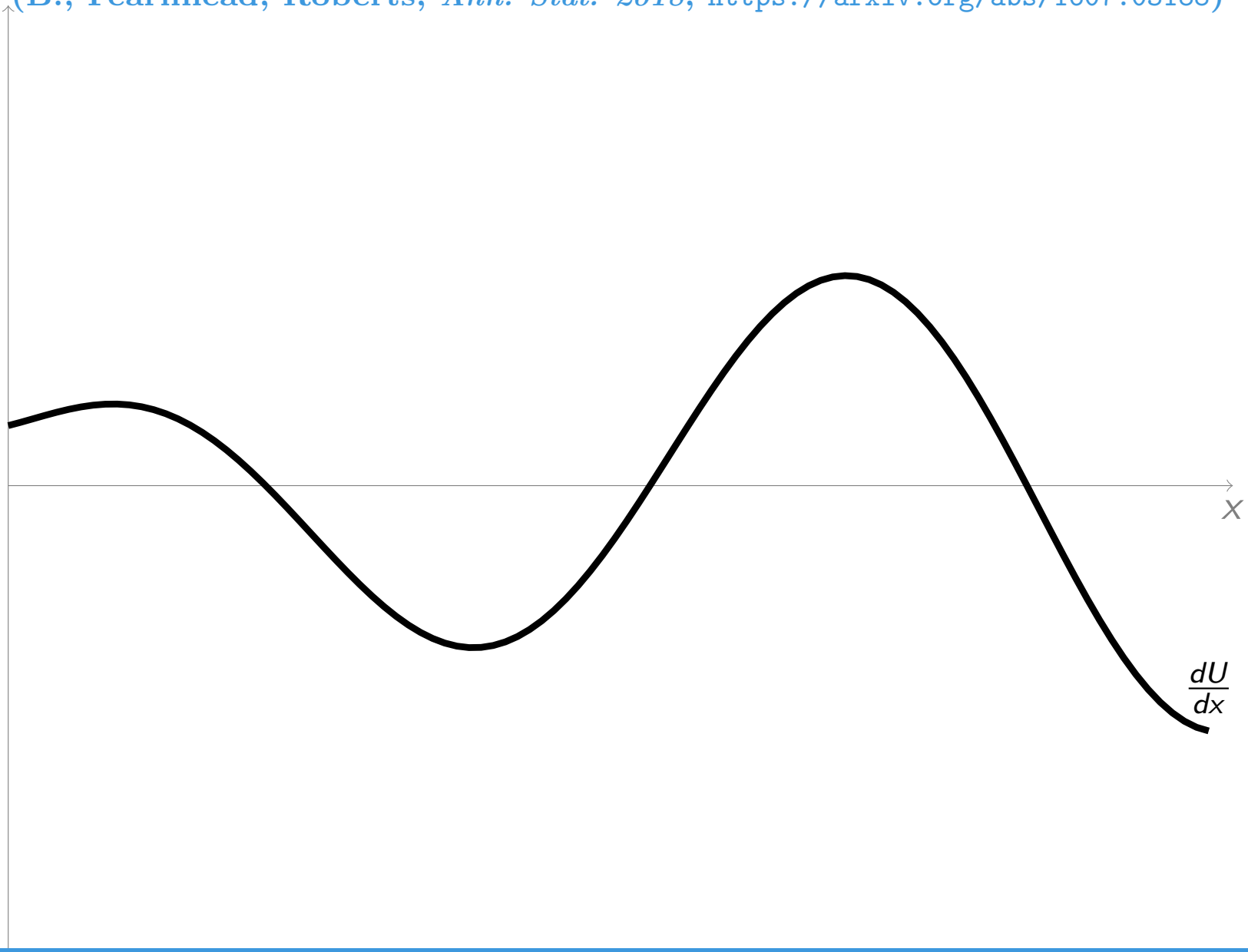
- $V = \mathbb{R}^d$, $\nu(dv) \sim \mathcal{N}(0, I_d)$, or
- $V = S^{d-1}$, $\nu(dv) \sim \mathrm{U}(S^{d-1})$.

Next Section

1. The Zig-Zag process and Bouncy Particle Sampler
2. Simulation and Subsampling
3. Convergence and efficiency

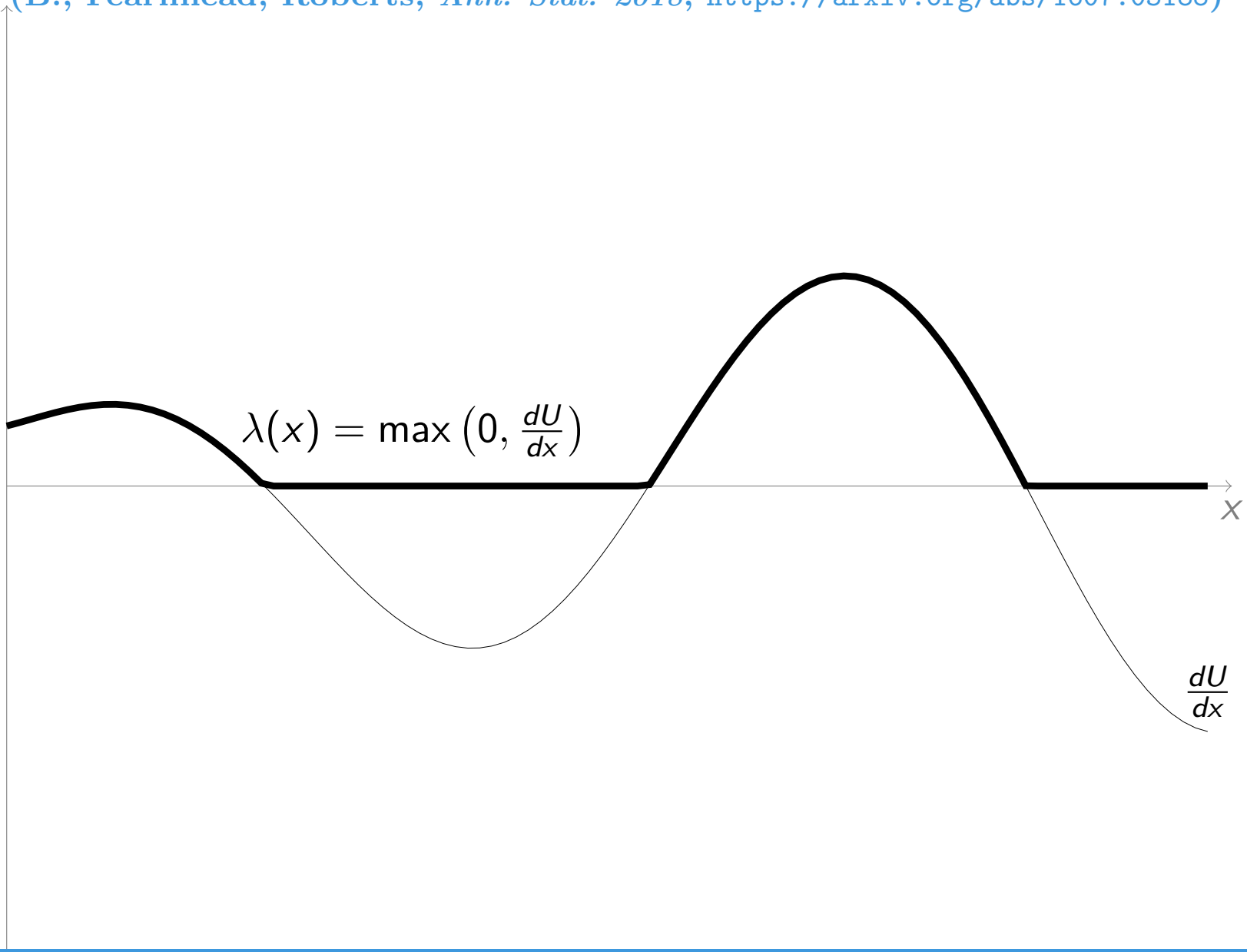
Sampling

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



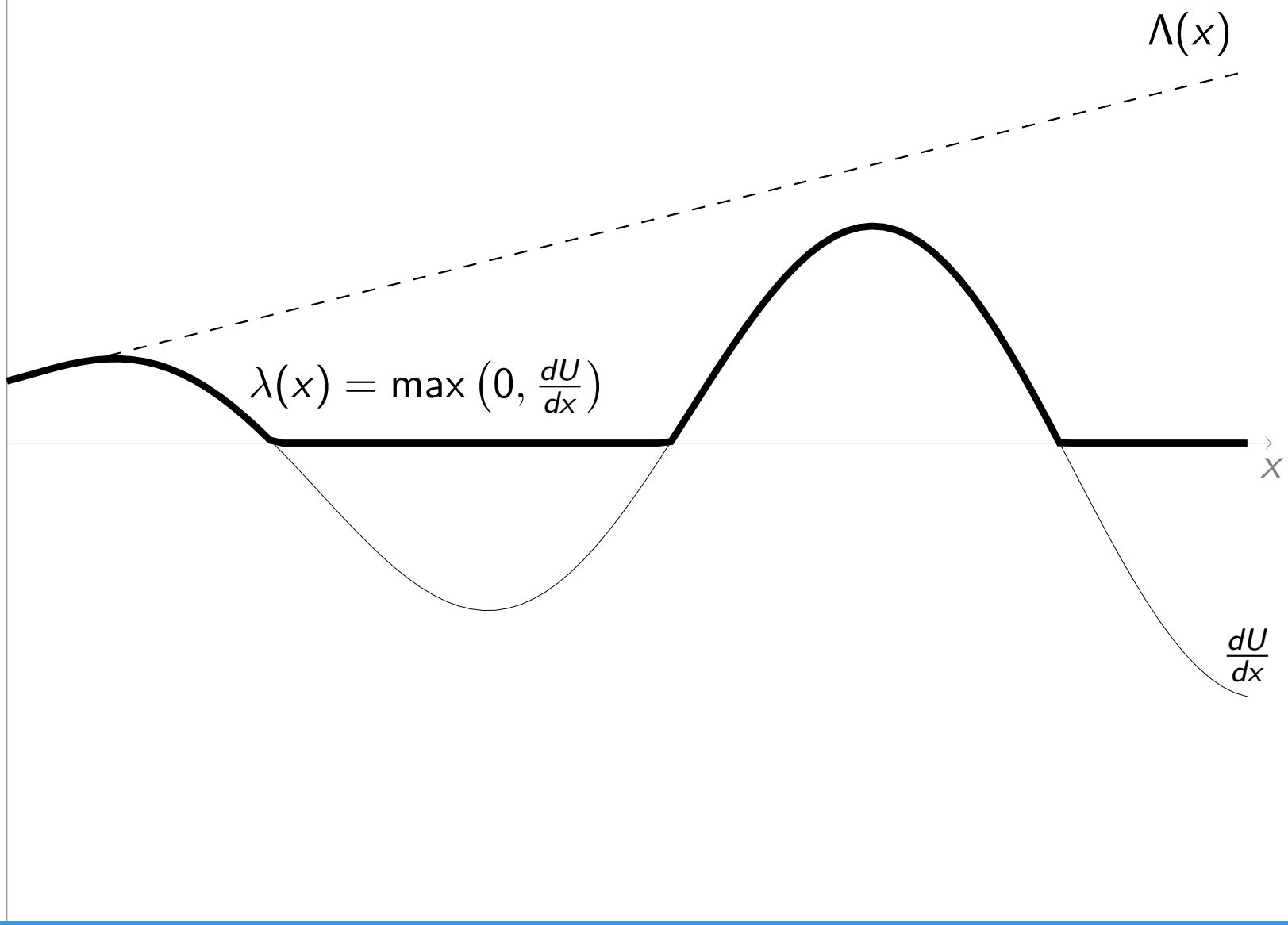
Sampling

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



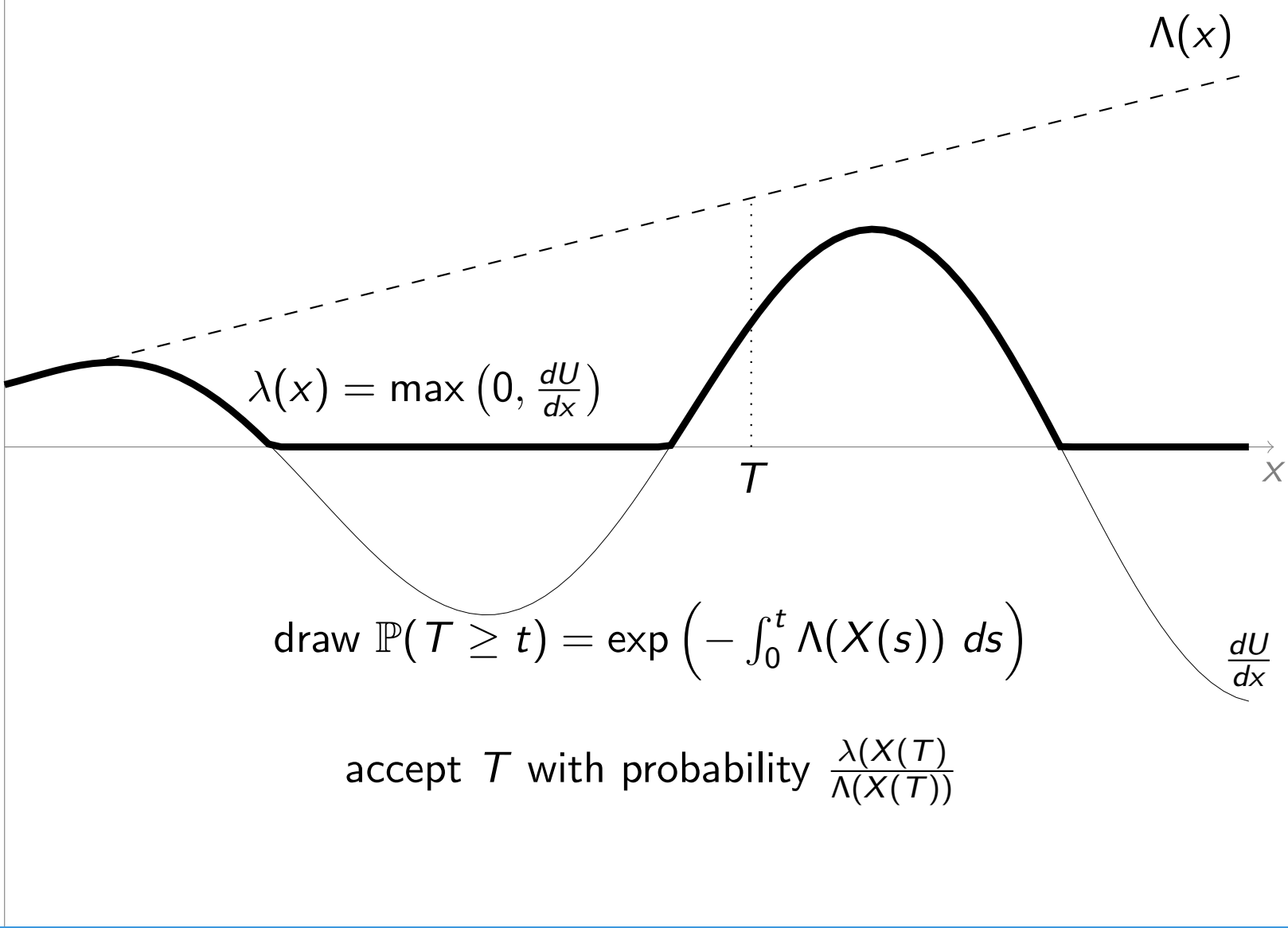
Sampling

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)



Sampling

(B., Fearnhead, Roberts, *Ann. Stat. 2018*, https://arxiv.org/abs/1607.03188)

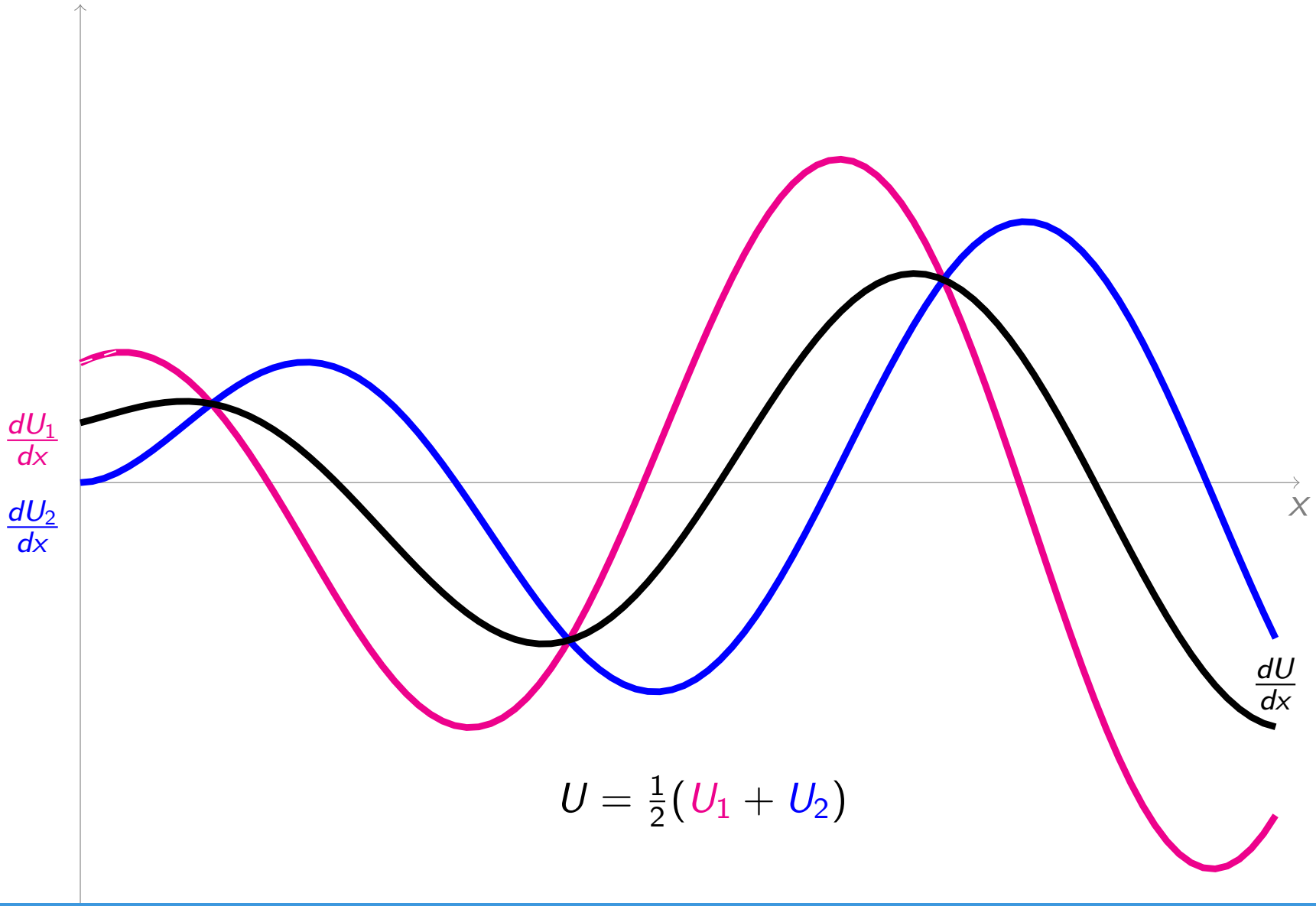


draw $\mathbb{P}(T \geq t) = \exp\left(-\int_0^t \Lambda(X(s))\, ds\right)$

accept T with probability $\frac{\lambda(X(T))}{\Lambda(X(T))}$

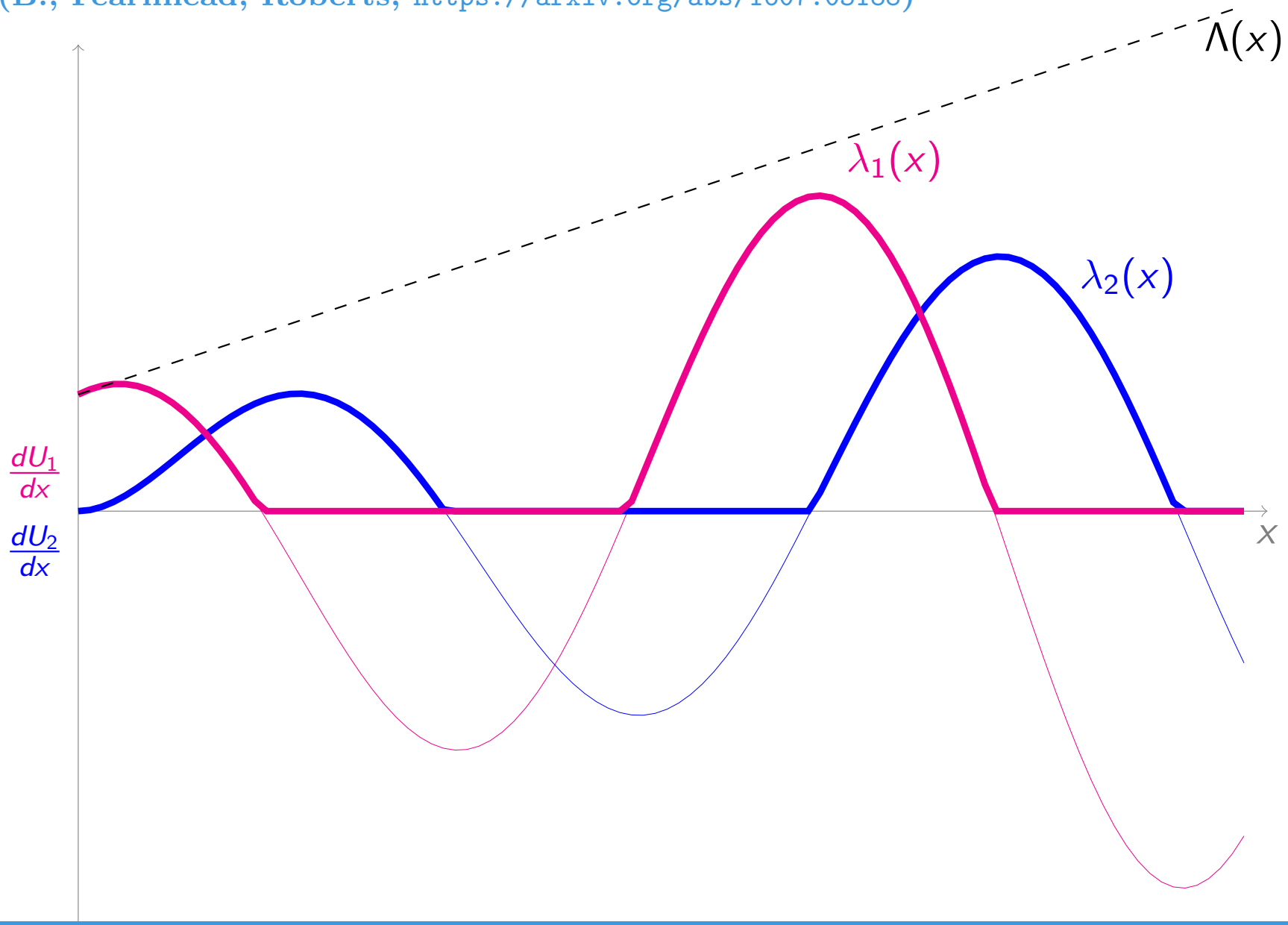
Subsampling

(B., Fearnhead, Roberts, https://arxiv.org/abs/1607.03188**)**



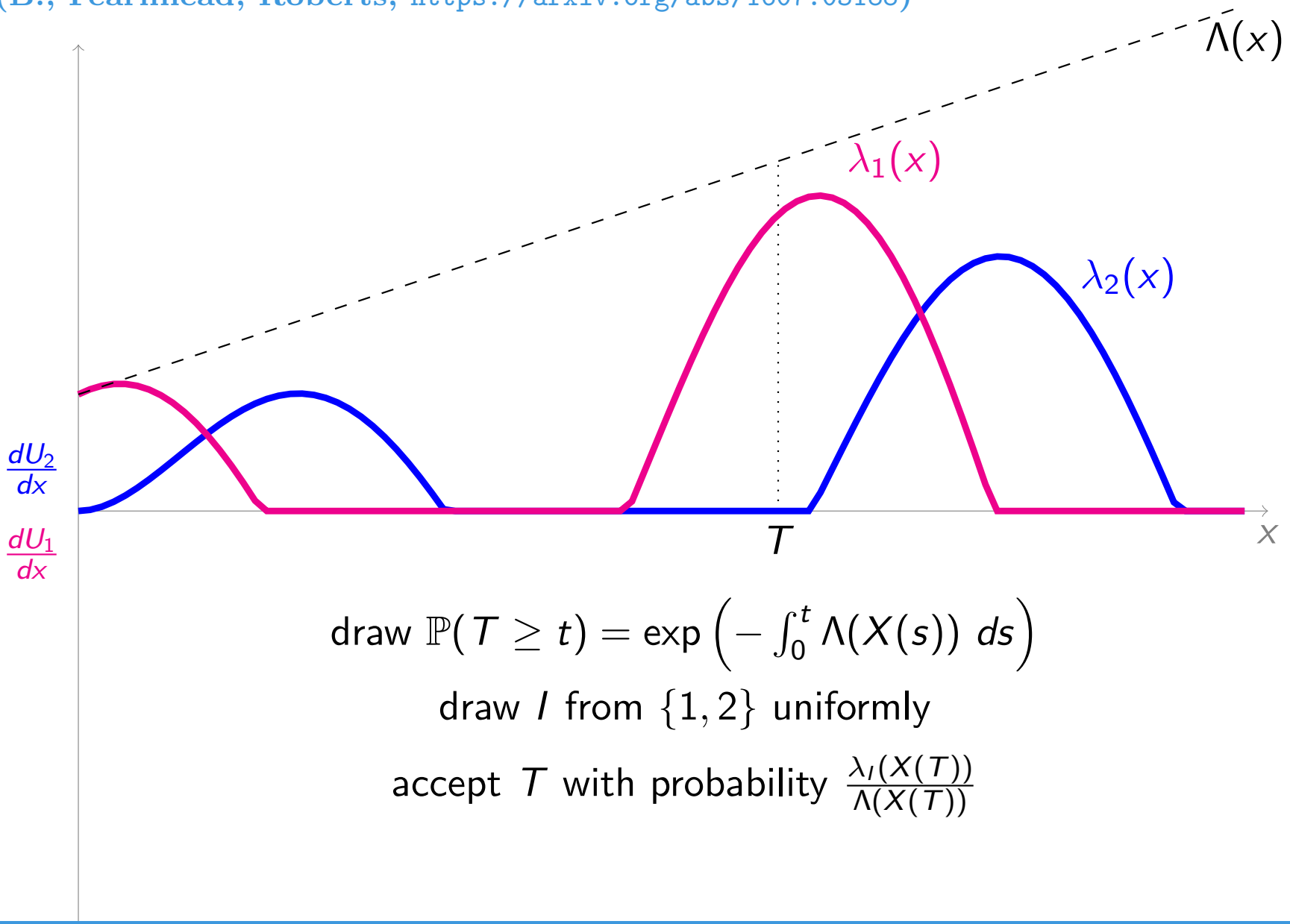
Subsampling

(B., Fearnhead, Roberts, https://arxiv.org/abs/1607.03188)



Subsampling

(B., Fearnhead, Roberts, https://arxiv.org/abs/1607.03188)



draw $\mathbb{P}(T \geq t) = \exp\left(-\int_0^t \Lambda(X(s))\, ds\right)$

draw I from $\{1, 2\}$ uniformly

accept T with probability $\frac{\lambda_I(X(T))}{\Lambda(X(T))}$

Next Section

Ergodicity of the multi-dimensional Zig-Zag process

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**

Ergodicity means that we should be able to construct *admissable paths* between any (x^1, v^1) and (x^2, v^2) in $\mathbb{R}^d \times \{-1, +1\}^d$.

Can we do this for the Zig-Zag process?

Ergodicity of the multi-dimensional Zig-Zag process

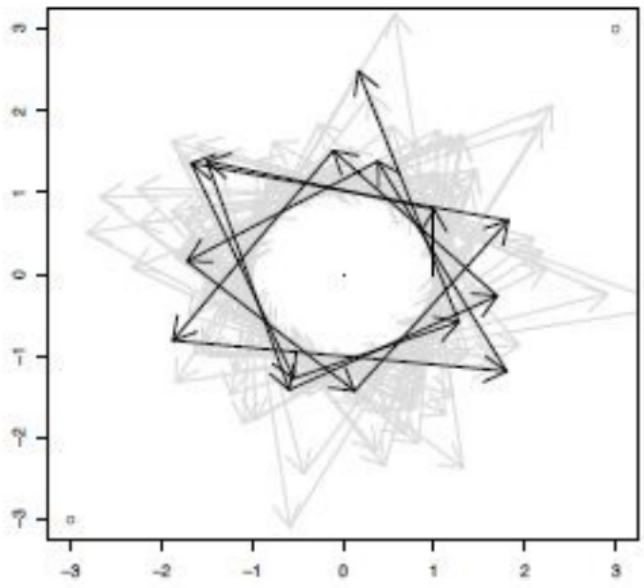
(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**

Ergodicity means that we should be able to construct *admissable paths* between any (x^1, v^1) and (x^2, v^2) in $\mathbb{R}^d \times \{-1, +1\}^d$.

Can we do this for the Zig-Zag process?

Remark: The Bouncy Particle Sampler is *not ergodic*, without refreshments.

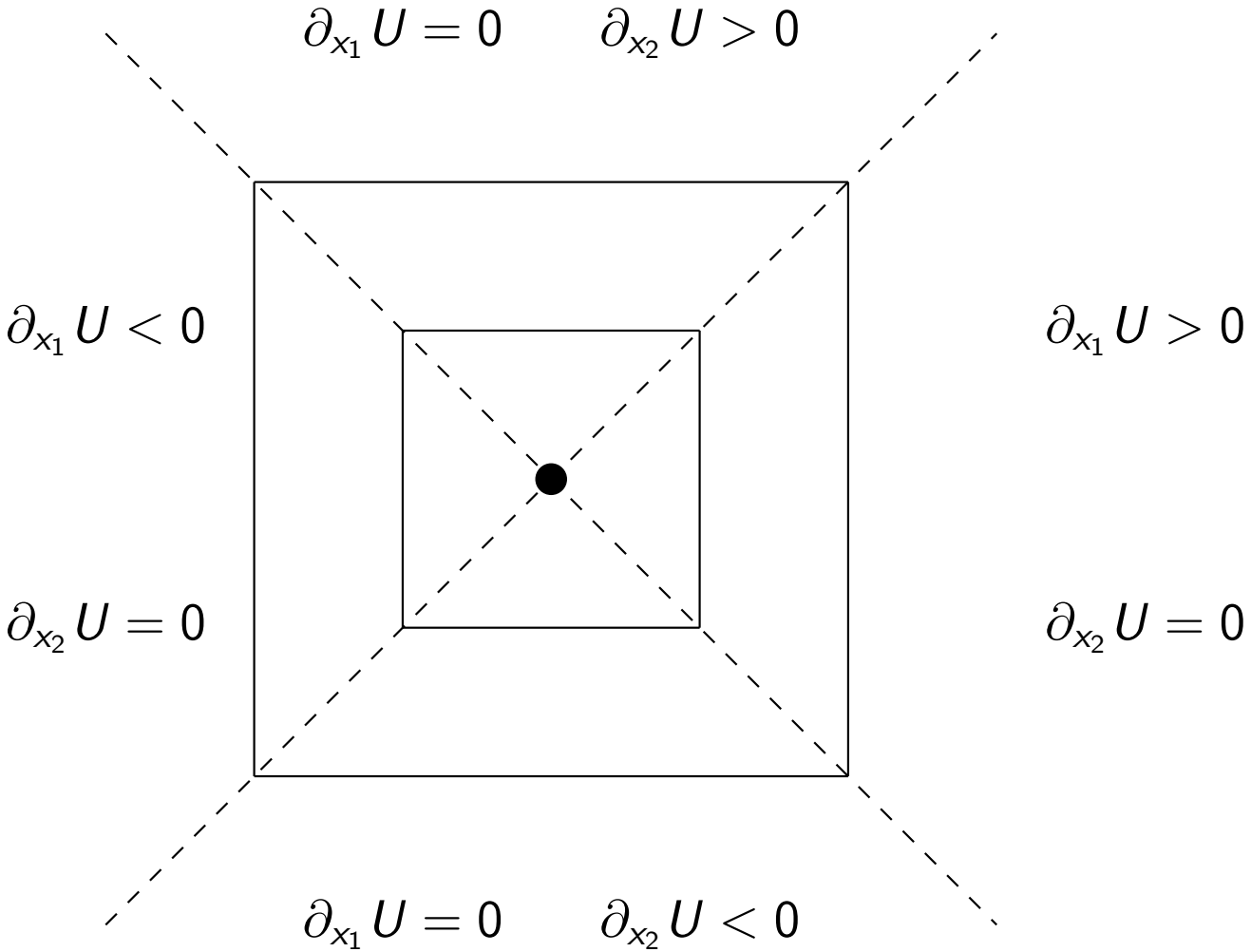
For BPS with refreshment, see (Deligiannidis, Bouchard-Côté, Doucet, *Annals of Statistics*, 2018).



(Bouchard-Côté, Vollmer, Doucet, 2017)

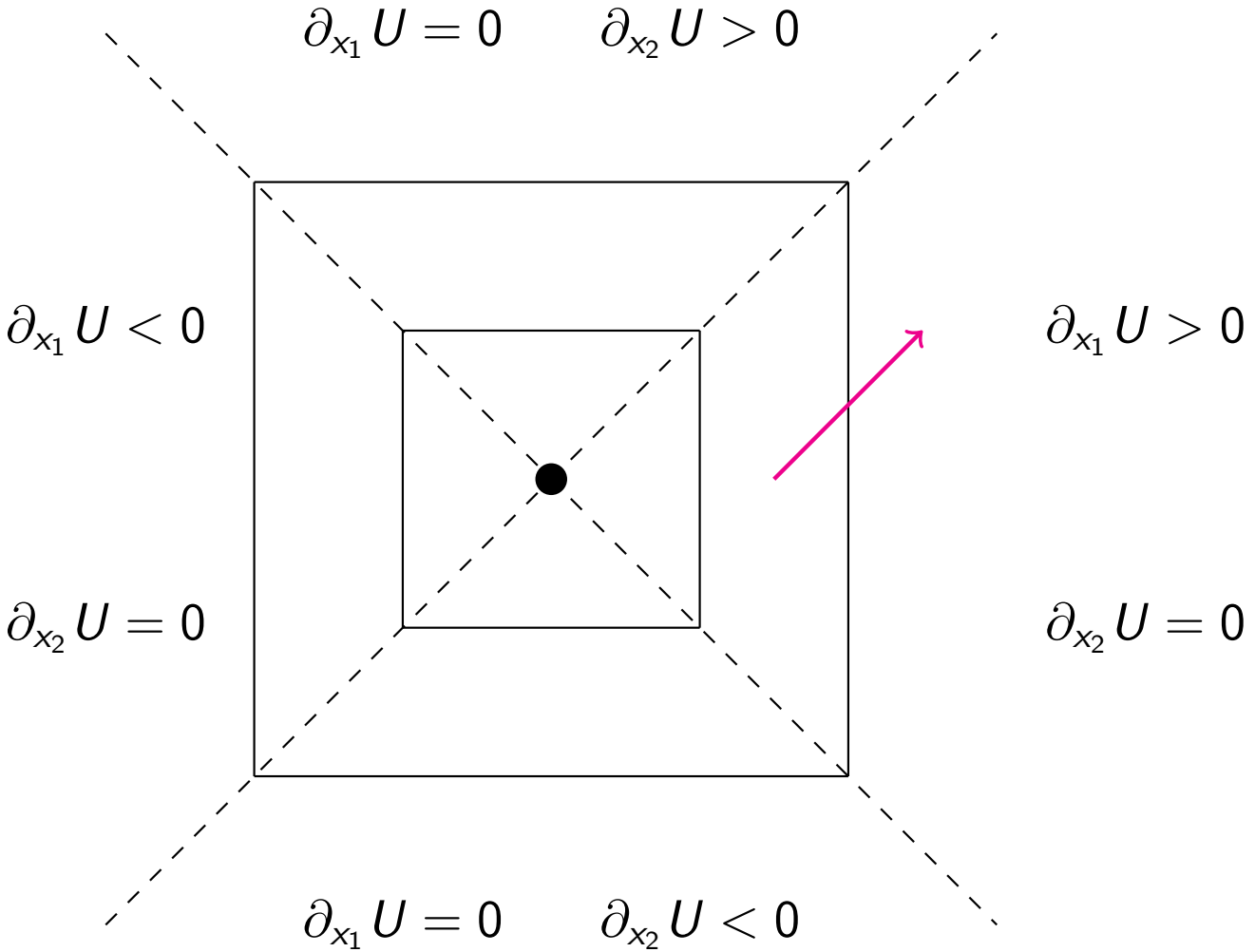
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



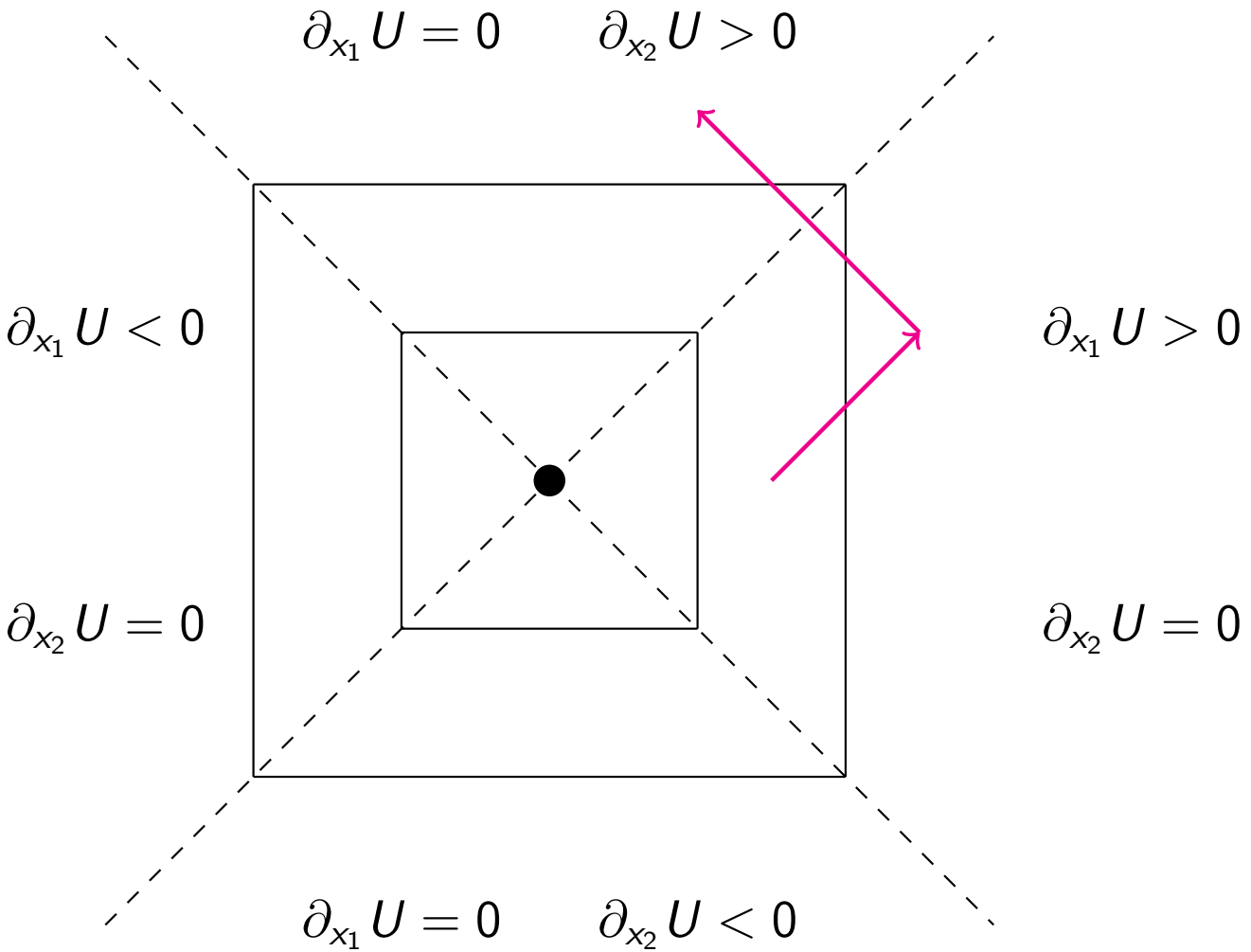
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



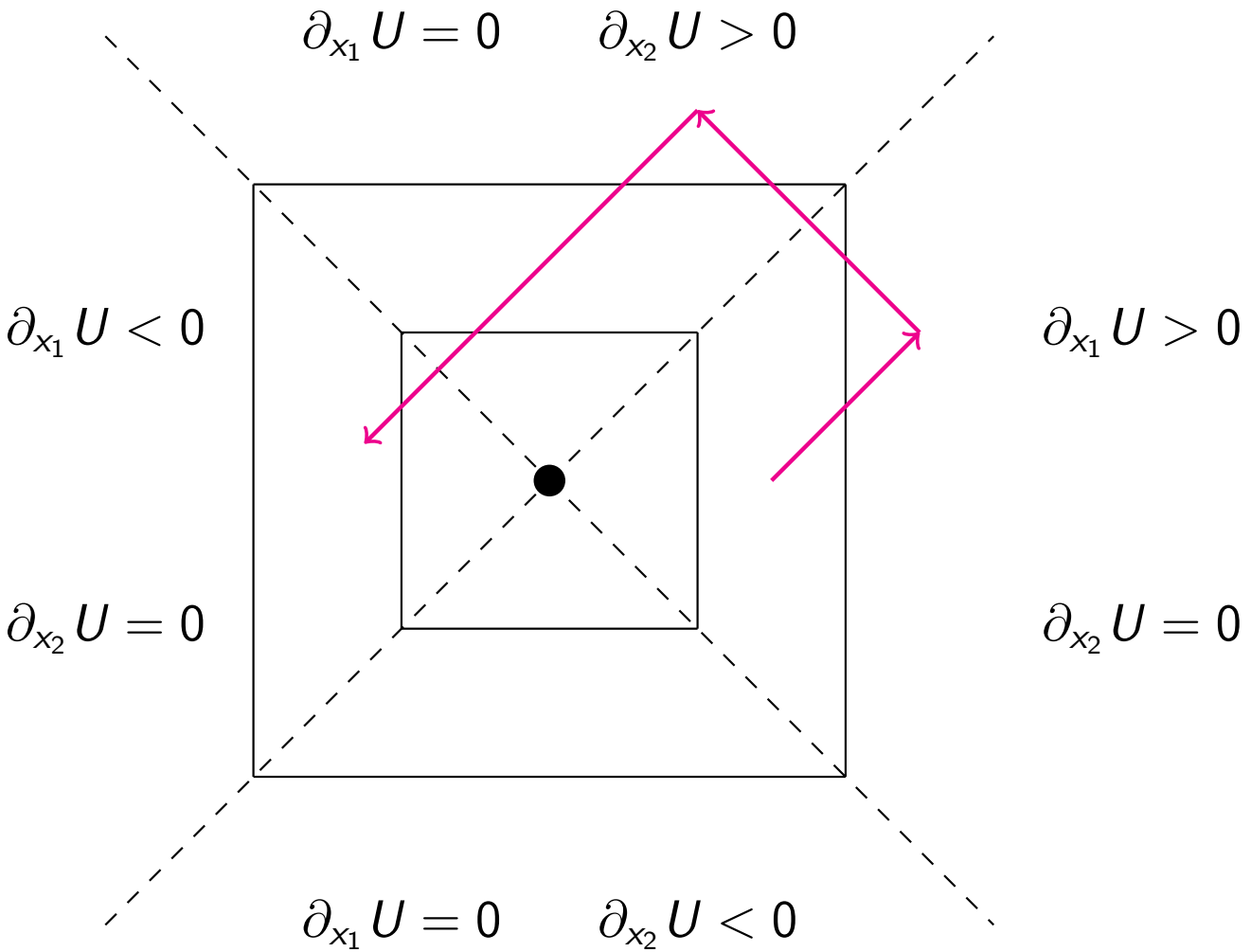
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



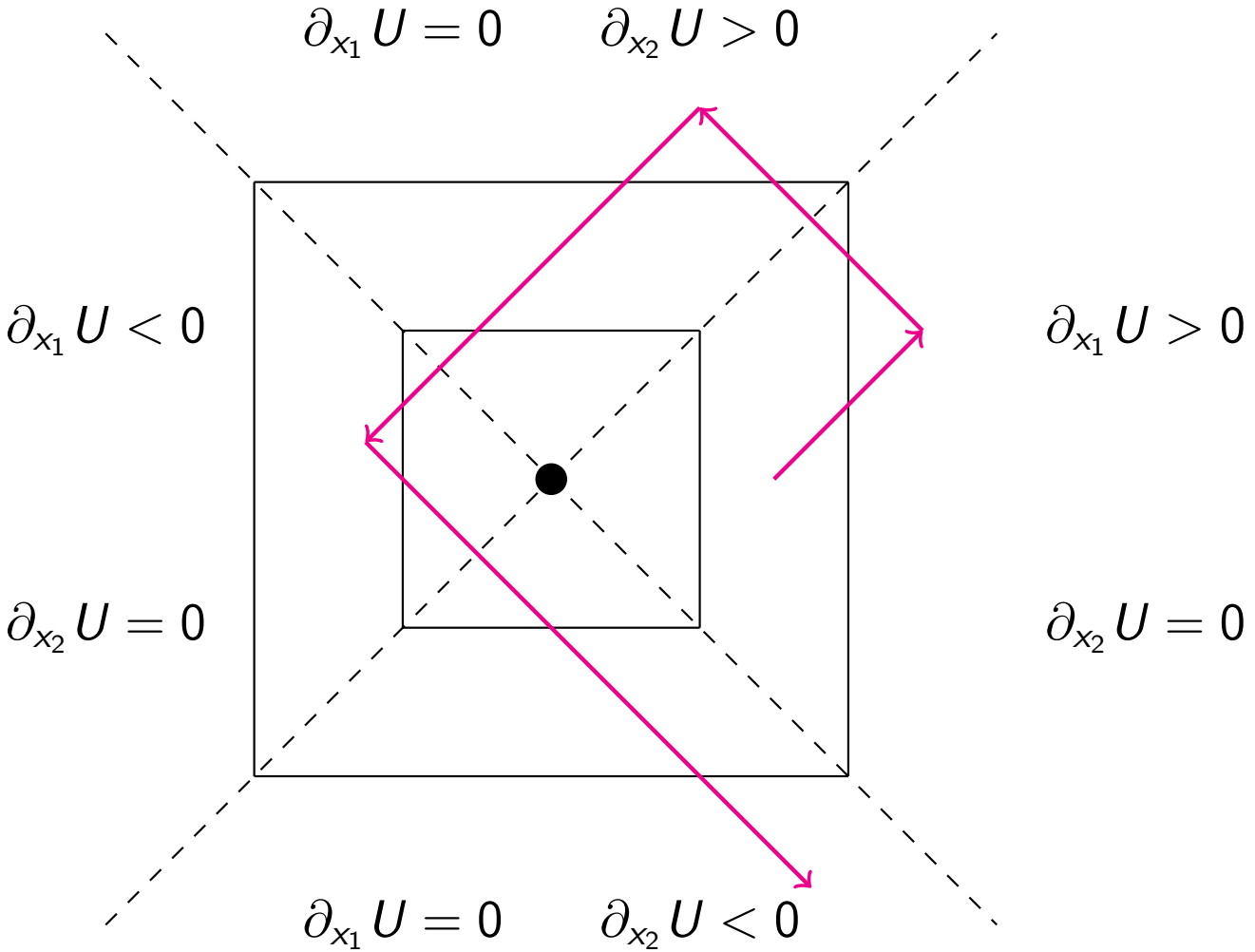
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



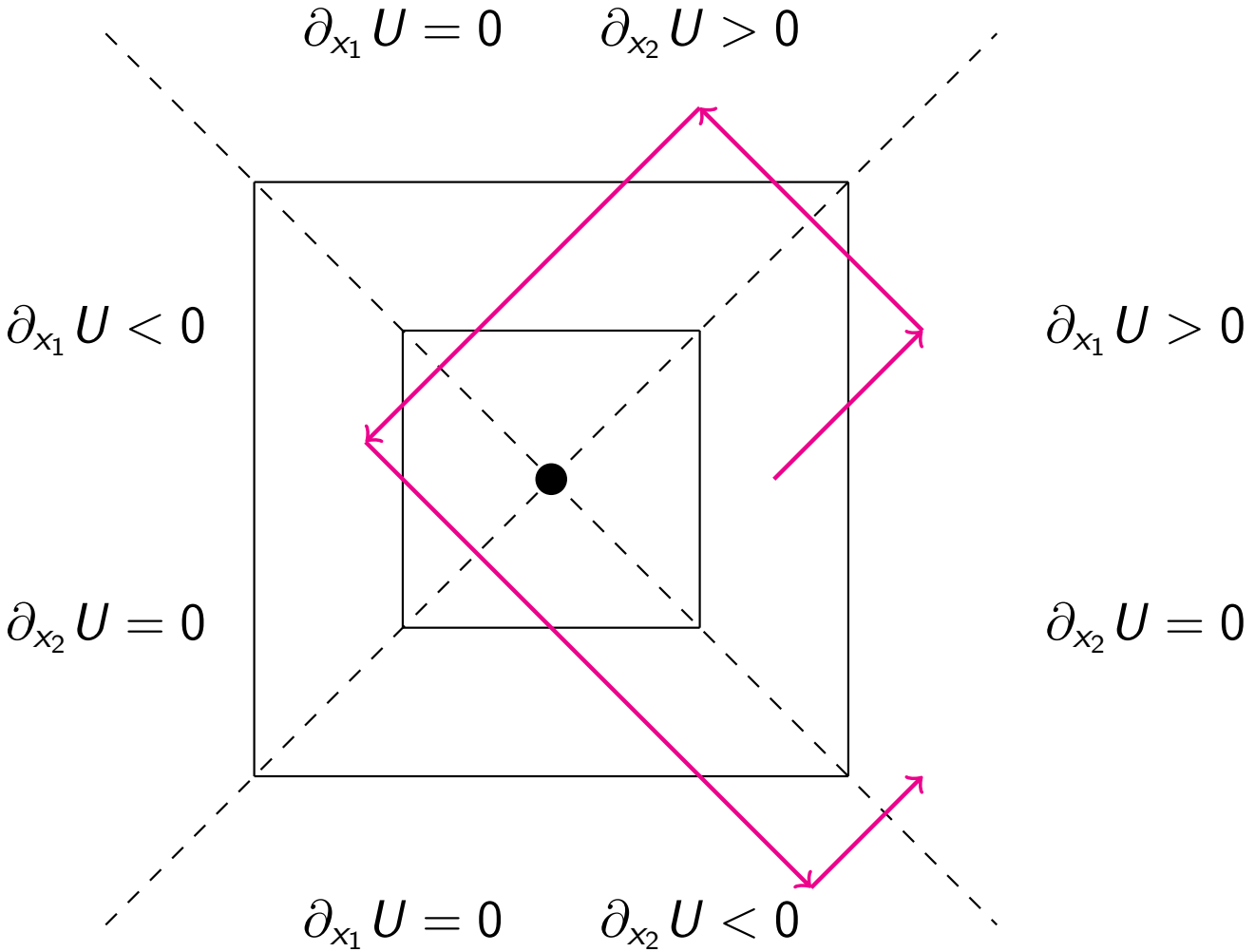
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



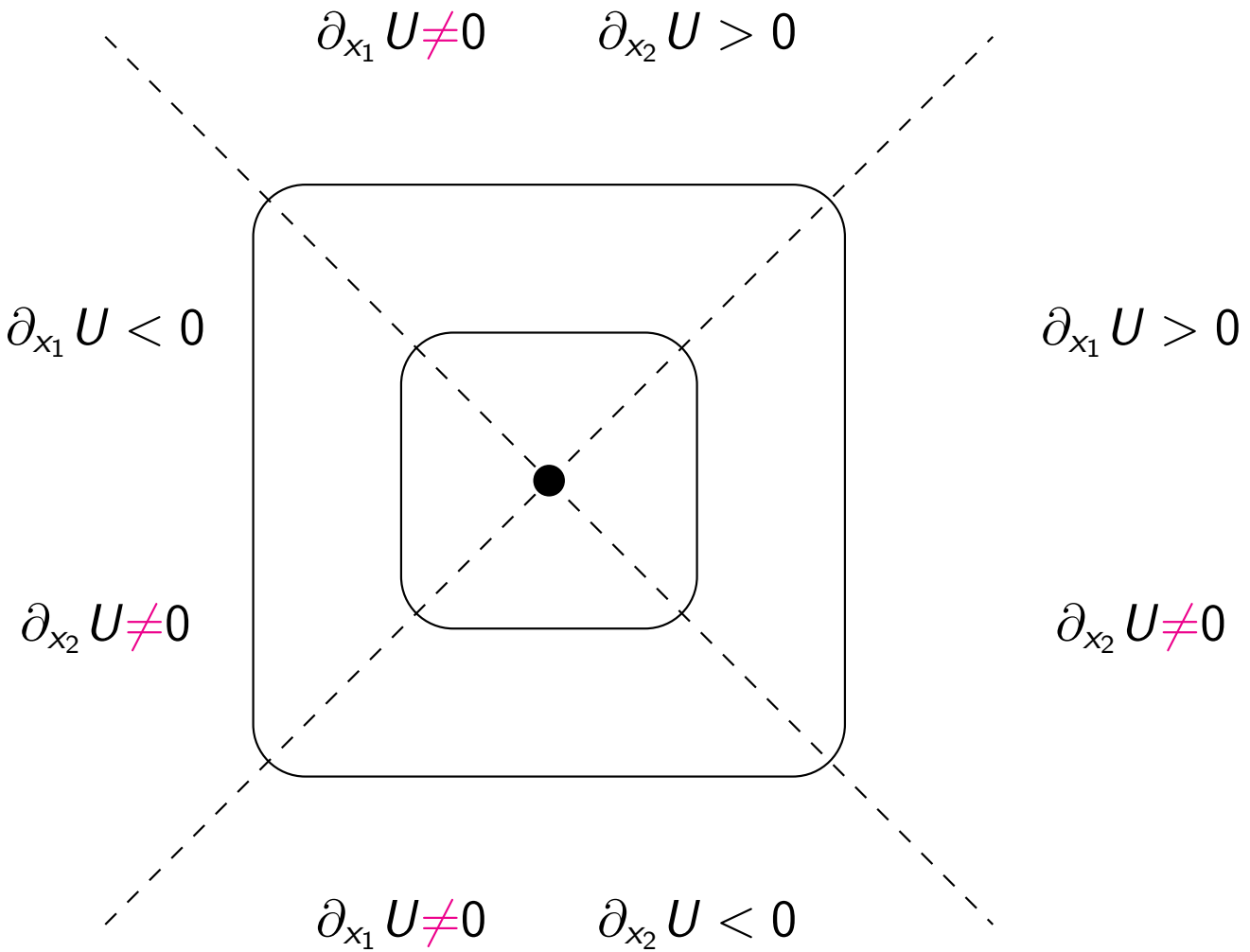
Counterexample of strict ergodicity, potential $U(x_1, x_2) = \max(|x_1|, |x_2|)$

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



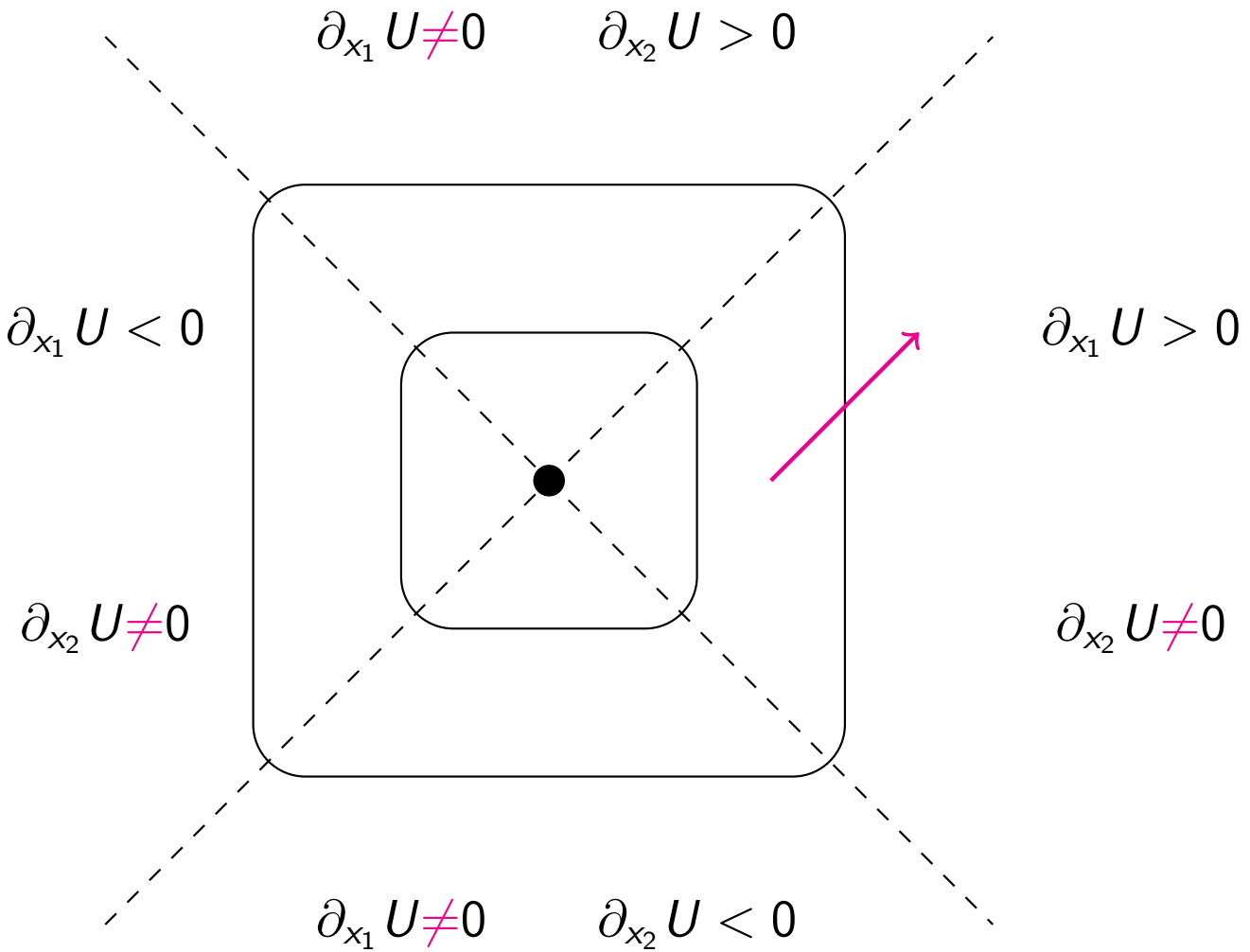
Resolution of counterexample

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



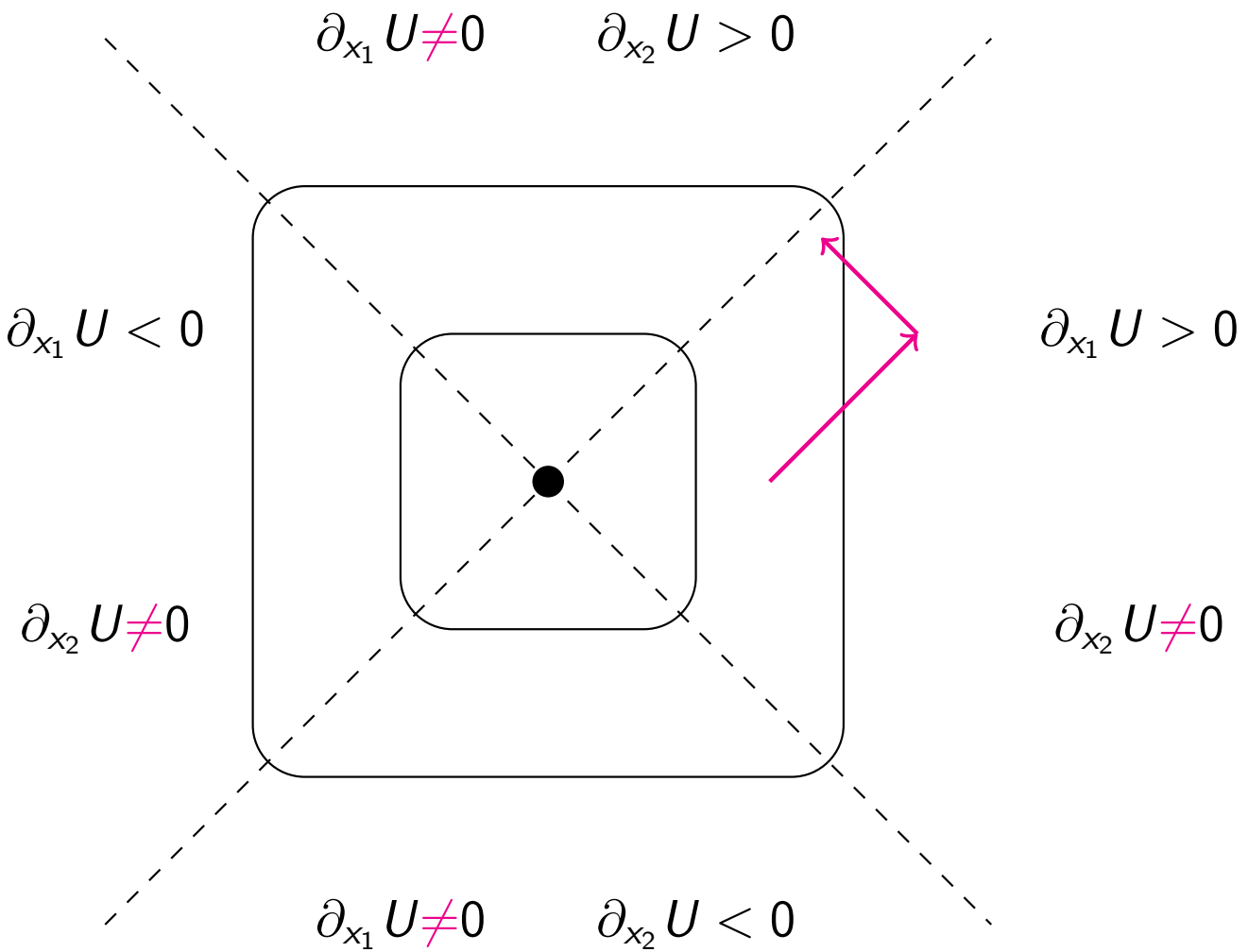
Resolution of counterexample

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



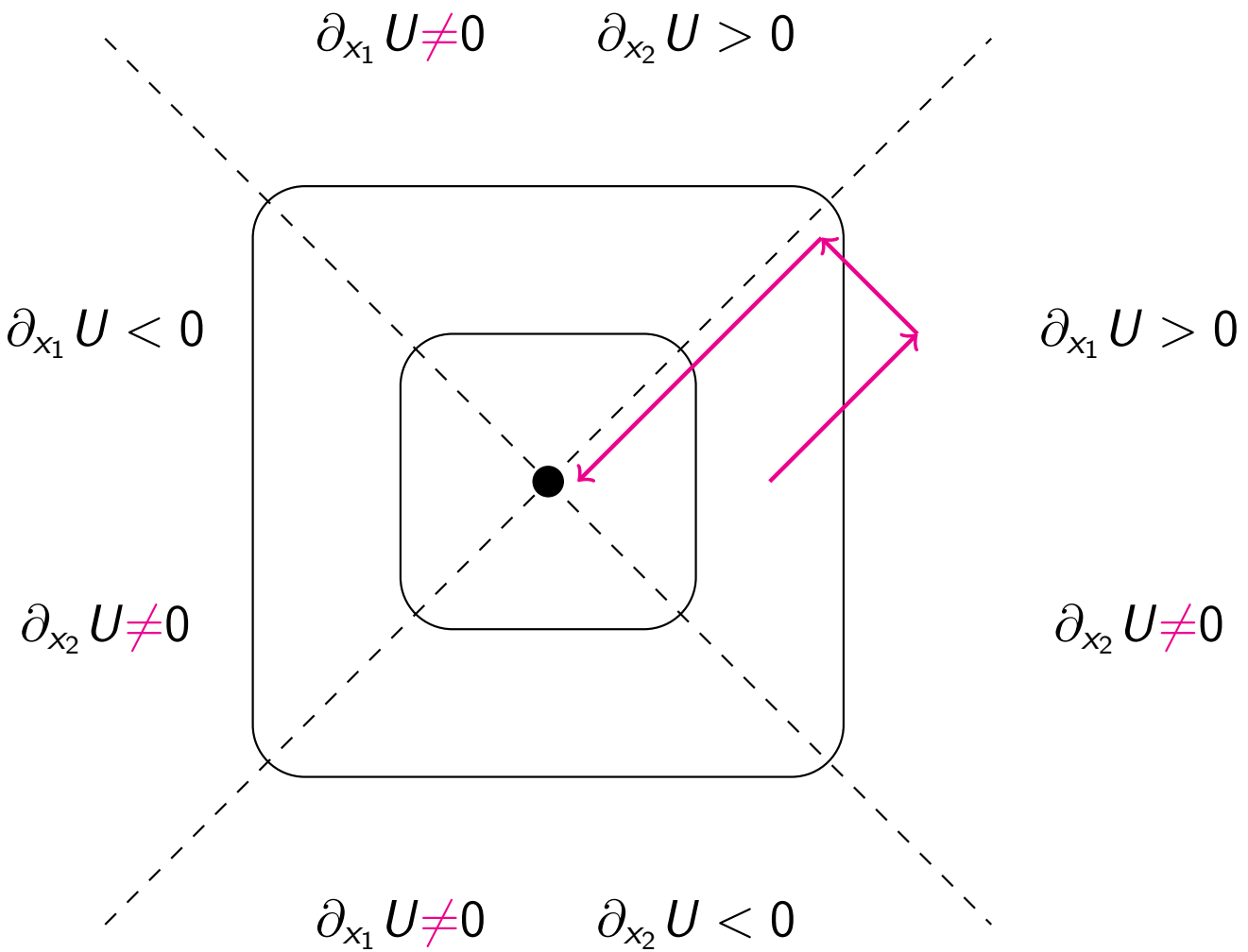
Resolution of counterexample

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



Resolution of counterexample

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875**)**



Ergodicity of the ZZP

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875)

Growth conditions

(GC1) $\lim_{|x|\to\infty} U(x) = \infty$

Theorem. Suppose $U \in \mathcal{C}^3(\mathbb{R}^d)$ has a non-degenerate local minimum.

1. If (GC1) is satisfied, then the Zig-Zag process is a ψ-irreducible aperiodic T-process.

Ergodicity of the ZZP

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875)

Growth conditions

(GC1) $\lim_{|x|\to\infty} U(x) = \infty$

(GC2) For some constants $c > d$, $c' \in \mathbb{R}$, $U(x) \geq c \ln(|x|) - c'$ for all $x \in \mathbb{R}^d$.

Theorem. Suppose $U \in \mathcal{C}^3(\mathbb{R}^d)$ has a non-degenerate local minimum.

1. If (GC1) is satisfied, then the Zig-Zag process is a ψ-irreducible aperiodic T-process.
2. If (GC2) is satisfied, then the law of $(X(t), V(t))$ converges to π in total variation for all initial conditions.

Ergodicity of the ZZP

(B., Roberts, Zitt, 2017, https://arxiv.org/abs/1712.09875)

Growth conditions

(GC1) $\lim_{|x|\to\infty} U(x) = \infty$

(GC2) For some constants $c > d$, $c' \in \mathbb{R}$, $U(x) \geq c \ln(|x|) - c'$ for all $x \in \mathbb{R}^d$.

(GC3)

$$\lim_{|x|\to\infty} \frac{\max(1, \|\operatorname{Hess} U(x)\|)}{|\nabla U(x)|} = 0 \quad \text{and} \quad \lim_{|x|\to\infty} \frac{|\nabla U(x)|}{U(x)} = 0.$$

Theorem. Suppose $U \in \mathcal{C}^3(\mathbb{R}^d)$ has a non-degenerate local minimum.

1. If (GC1) is satisfied, then the Zig-Zag process is a ψ-irreducible aperiodic T-process.
2. If (GC2) is satisfied, then the law of $(X(t), V(t))$ converges to π in total variation for all initial conditions.
3. If the excess switching rates $(\gamma_i)_{i=1}^d$ are bounded and (GC3) is satisfied, then the ZZP is exponentially ergodic.

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

Consider a factorized target distribution $\pi(x) = \prod_{i=1}^{d} \pi_i(x_i)$ with $\pi_i(y) = \exp(-U_i(y))$.

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

Consider a factorized target distribution $\pi(x) = \prod_{i=1}^{d} \pi_i(x_i)$ with $\pi_i(y) = \exp(-U_i(y))$.

Switching rates: $\lambda_i(x, v) = (v_i U_i'(x_i))_+$.

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

Consider a factorized target distribution $\pi(x) = \prod_{i=1}^{d} \pi_i(x_i)$ with $\pi_i(y) = \exp(-U_i(y))$.

Switching rates: $\lambda_i(x, v) = (v_i U_i'(x_i))_+$.

Every component of the Zig-Zag process mixes at $\mathcal{O}(1)$.

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

Consider a factorized target distribution $\pi(x) = \prod_{i=1}^{d} \pi_i(x_i)$ with $\pi_i(y) = \exp(-U_i(y))$.

Switching rates: $\lambda_i(x, v) = (v_i U_i'(x_i))_+$.

Every component of the Zig-Zag process mixes at $\mathcal{O}(1)$.

Compare to RWM $\mathcal{O}(d)$, MALA $\mathcal{O}\left(d^{1/3}\right)$, HMC $\mathcal{O}\left(d^{1/4}\right)$.

Scaling with dimension

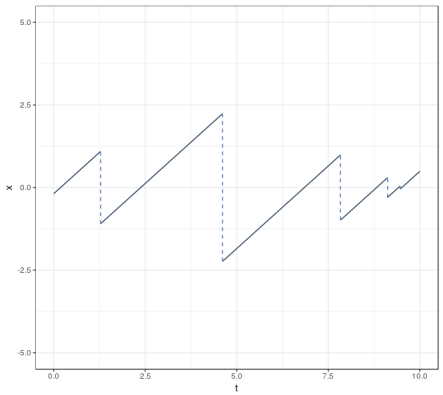
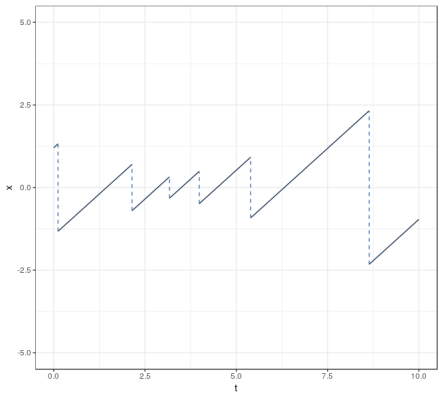
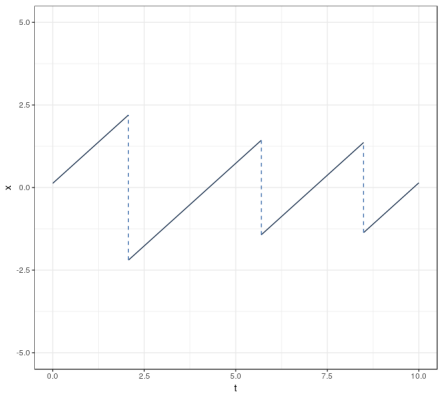
(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

Setup:

- Multivariate standard normal marginal distribution for (X_t) on $\mathbb{R}^d$.
- BPS with speeds $V_t \in d^{1/2} S^{d-1}$ (uniformly distributed)
- BPS with refreshment rate $\rho d^{1/2}$.
- Scaling limits for:
  - Angular momentum $\langle X_t, V_t \rangle$
  - Negative log target $d^{1/2}(d^{-1}\|X_t\|^2 - 1)$
  - First coordinate $X_{1,t}$

Scaling with dimension – angular momentum

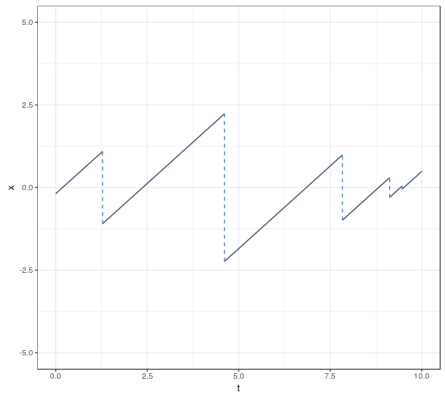
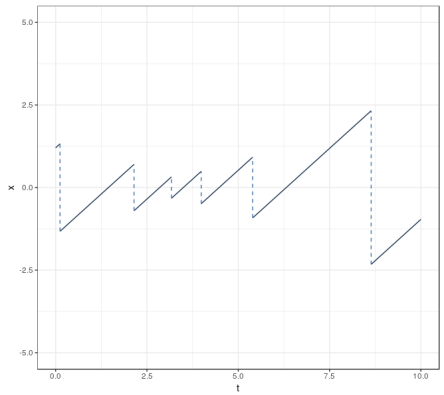
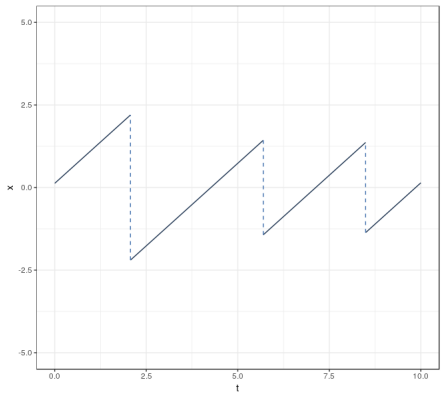
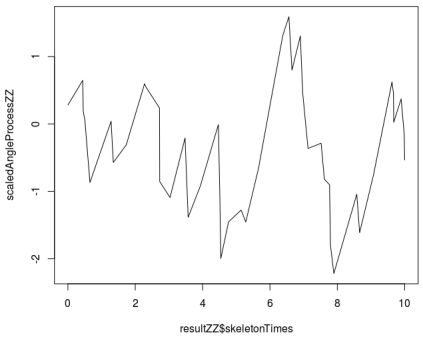
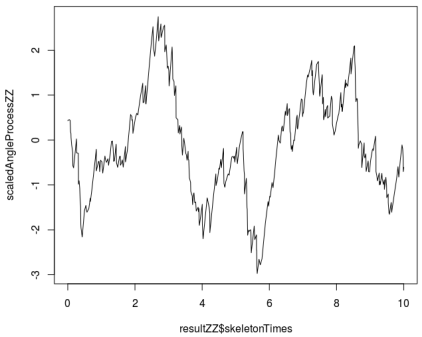
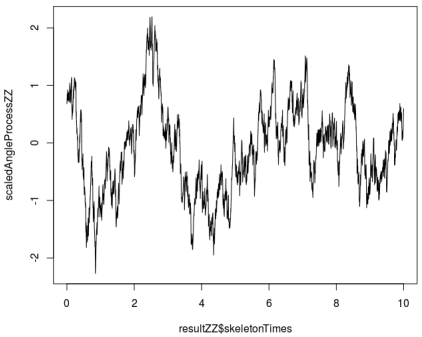
(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS, $\rho = 0$			

BPS limit: PDMP in $\mathbb{R}$ with generator $Lf(x) = f'(x) + (x)^+(f(-x) - f(x))$.

Scaling with dimension – angular momentum

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS, $\rho = 0$			
ZZ			

BPS limit: PDMP in $\mathbb{R}$ with generator $Lf(x) = f'(x) + (x)^+(f(-x) - f(x))$.

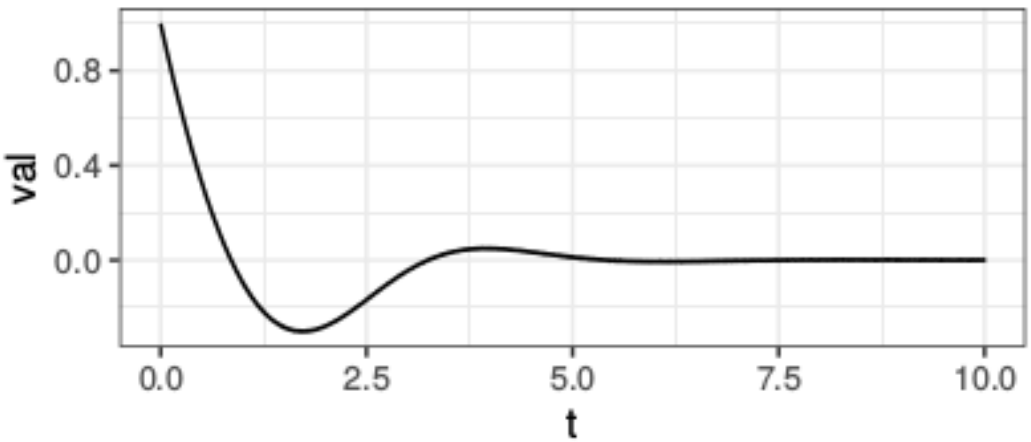
ZZ limit: non-Markovian Gaussian process with same covariance structure as the BPS limit.

Scaling with dimension – angular momentum

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

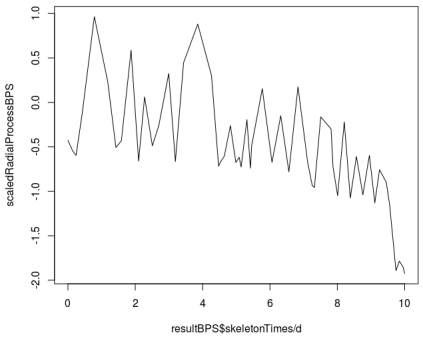
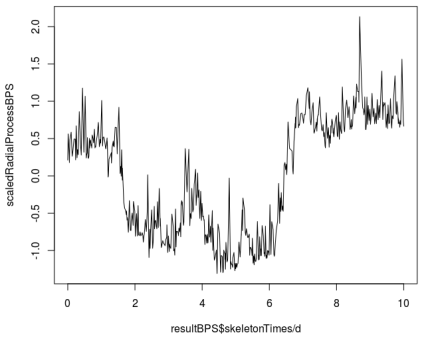
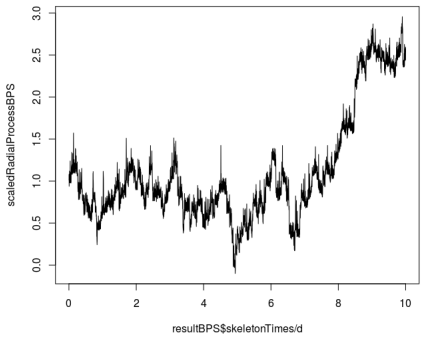
BPS limit: PDMP in $\mathbb{R}$ with generator $Lf(x) = f'(x) + (x)^+(f(-x) - f(x))$.

ZZ limit: non-Markovian Gaussian process with same covariance structure as the BPS limit.
Covariance kernel $K(t)$, satisfying $\int_0^\infty K(t)\,dt = 0$.



Scaling with dimension – log density

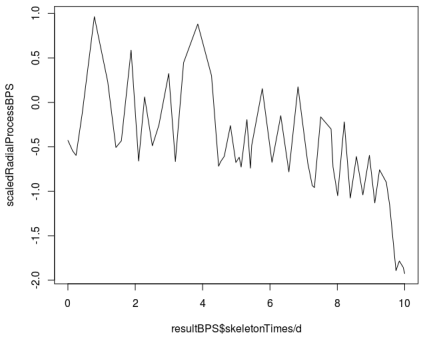
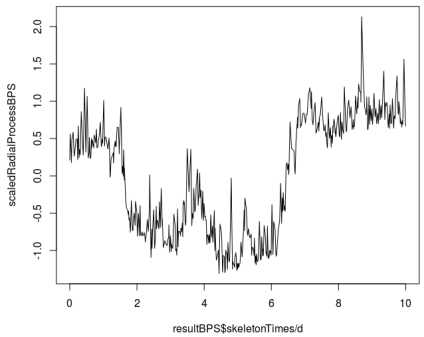
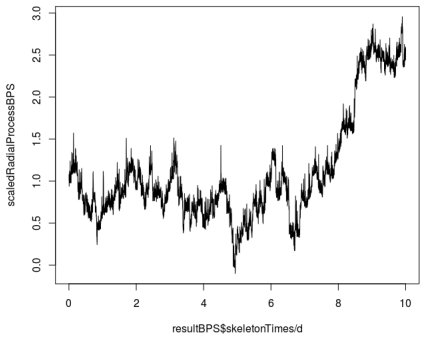
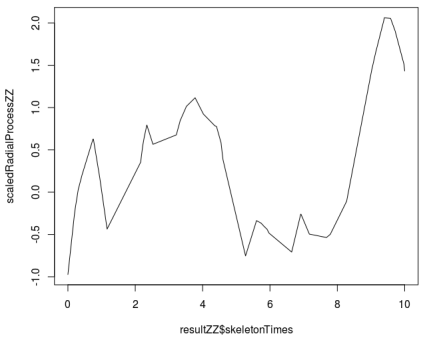
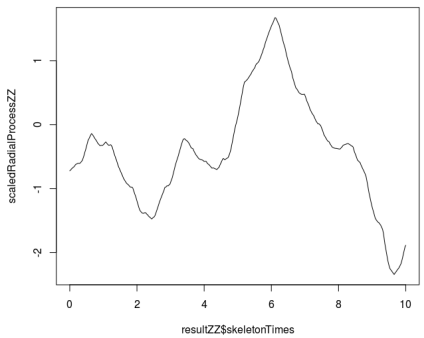
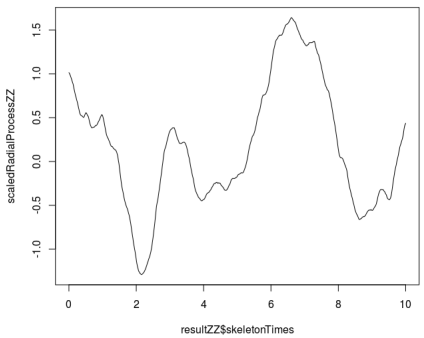
(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS			

BPS limit: Ornstein Uhlenbeck process $dY_t = -\frac{\sigma^2(\rho)}{4} Y_t \, dt + \sigma(\rho) \, dW_t$

Scaling with dimension – log density

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS			
ZZ			

BPS limit: Ornstein Uhlenbeck process $dY_t = -\frac{\sigma^2(\rho)}{4} Y_t \, dt + \sigma(\rho) \, dW_t$

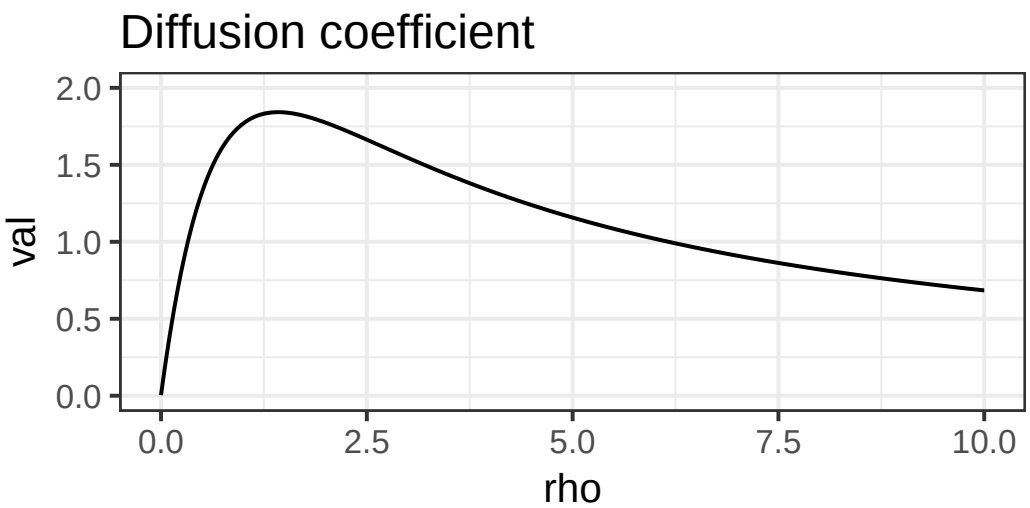
ZZ limit: "time integral" of angular momentum limit process

Scaling with dimension – log density

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

BPS limit for log density process: OU process

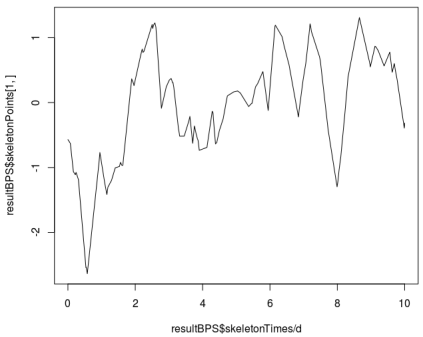
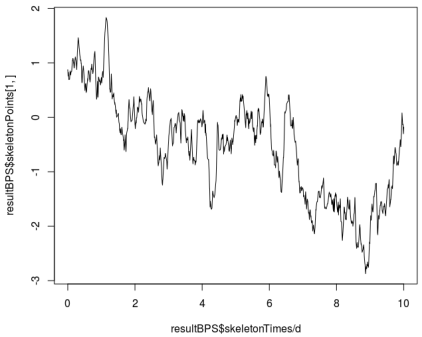
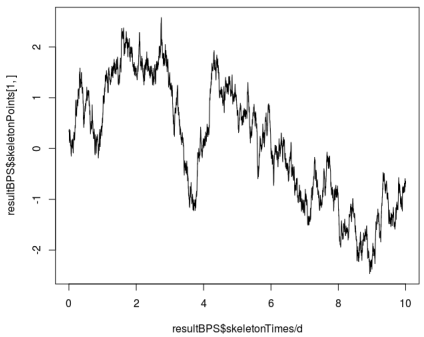
$$dY_t = -\frac{\sigma^2(\rho)}{4} Y_t \, dt + \sigma(\rho) \, dW_t$$



Note that $\sigma(0) = 0$ and $\lim_{\rho \to \infty} \sigma(\rho) = 0$, so a non-vanishing, finite diffusion coefficient is required!
Maximized at $\rho^\star \approx 1.424$

Scaling with dimension – first coordinate

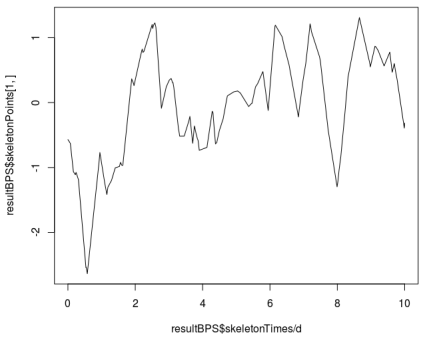
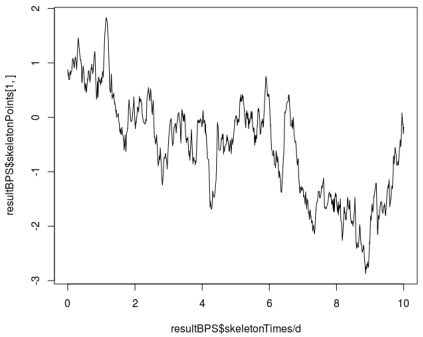
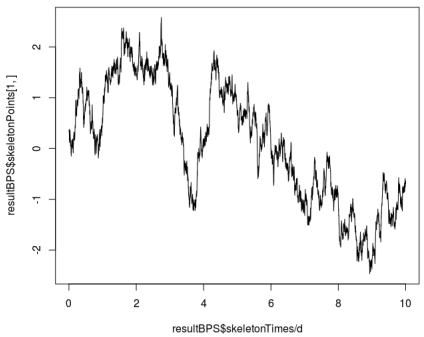
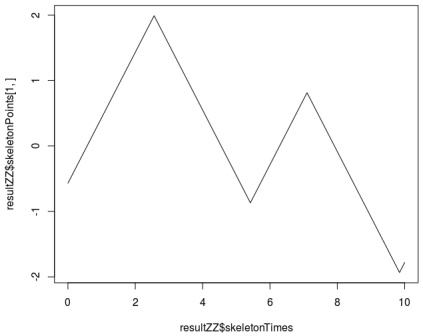
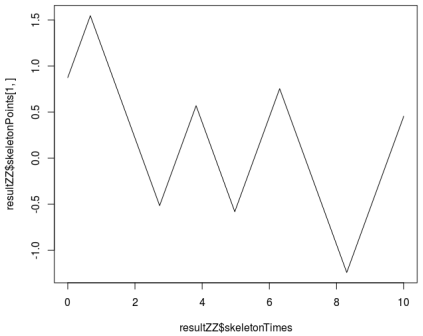
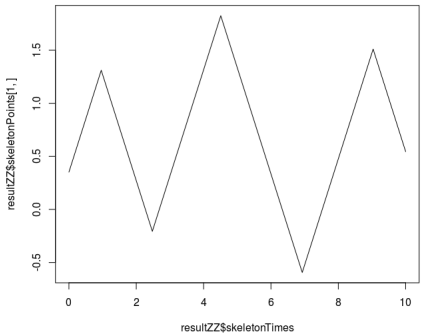
(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS			

BPS limit: Ornstein Uhlenbeck process $dY_t = -\rho^{-1} Y_t + \sqrt{2\rho^{-1}}\, dW_t$

Scaling with dimension – first coordinate

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

d	10	100	1000
BPS			
ZZ			

BPS limit: Ornstein Uhlenbeck process $dY_t = -\rho^{-1} Y_t + \sqrt{2\rho^{-1}}\, dW_t$

ZZ limit: 1-dimensional Zig-Zag process!

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

- Mixing times:

Method	Angular momentum	Radius	1st Coordinate
ZZ	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
BPS	$\mathcal{O}(d^{-1/2})$	$\mathcal{O}(d^{1/2})$	$\mathcal{O}(d^{1/2})$
MALA	N/A	$\mathcal{O}(d^{1/3})$	$\mathcal{O}(d^{1/3})$

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358**)**

- Mixing times:

Method	Angular momentum	Radius	1st Coordinate
ZZ	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
BPS	$\mathcal{O}(d^{-1/2})$	$\mathcal{O}(d^{1/2})$	$\mathcal{O}(d^{1/2})$
MALA	N/A	$\mathcal{O}(d^{1/3})$	$\mathcal{O}(d^{1/3})$

- Computational effort:

Method	Switches/time	Comp. effort/switch	Effort/time
ZZ	$\mathcal{O}(d)$	$\mathcal{O}(1)$	$\mathcal{O}(d)$
BPS	$\mathcal{O}(d^{1/2})$	$\mathcal{O}(d)$	$\mathcal{O}(d^{3/2})$
MALA	1	$\mathcal{O}(d)$	$\mathcal{O}(d)$

Scaling with dimension

(B., Kamatani, Roberts, 2018, https://arxiv.org/abs/1807.11358)

- Mixing times:

Method	Angular momentum	Radius	1st Coordinate
ZZ	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
BPS	$\mathcal{O}(d^{-1/2})$	$\mathcal{O}(d^{1/2})$	$\mathcal{O}(d^{1/2})$
MALA	N/A	$\mathcal{O}(d^{1/3})$	$\mathcal{O}(d^{1/3})$

- Computational effort:

Method	Switches/time	Comp. effort/switch	Effort/time
ZZ	$\mathcal{O}(d)$	$\mathcal{O}(1)$	$\mathcal{O}(d)$
BPS	$\mathcal{O}(d^{1/2})$	$\mathcal{O}(d)$	$\mathcal{O}(d^{3/2})$
MALA	1	$\mathcal{O}(d)$	$\mathcal{O}(d)$

- Full computational effort before mixing:

Method	Angular momentum	Radius	1st Coordinate
ZZ	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$
BPS	$\mathcal{O}(d)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$
MALA	N/A	$\mathcal{O}(d^{4/3})$	$\mathcal{O}(d^{4/3})$

Piecewise Deterministic Monte Carlo

- We can use piecewise deterministic Markov processes for sampling
- Unbiased estimate for the log density gradient results in correct stationary distribution.
- Theoretical scaling results for Zig-Zag very promising, but relying upon
  - Tight computational bounds
  - Product distribution (independence) assumption
- Scaling results comparable to: Andrieu, Durmus, Nüsken, Roussel, *Hypocoercivity of Piecewise Deterministic Markov Process-Monte Carlo*, http://arxiv.org/abs/1808.08592
  - Efficiency of Zig-zag is estimated to depend on the amount of target anisotropy
  - Positive refreshment for Zigzag as a technical condition.
- R package (so far: Gaussian, Logistic Regression) `RZigZag` available on CRAN

References

http://jbierkens.nl/pdmps

Thank you!