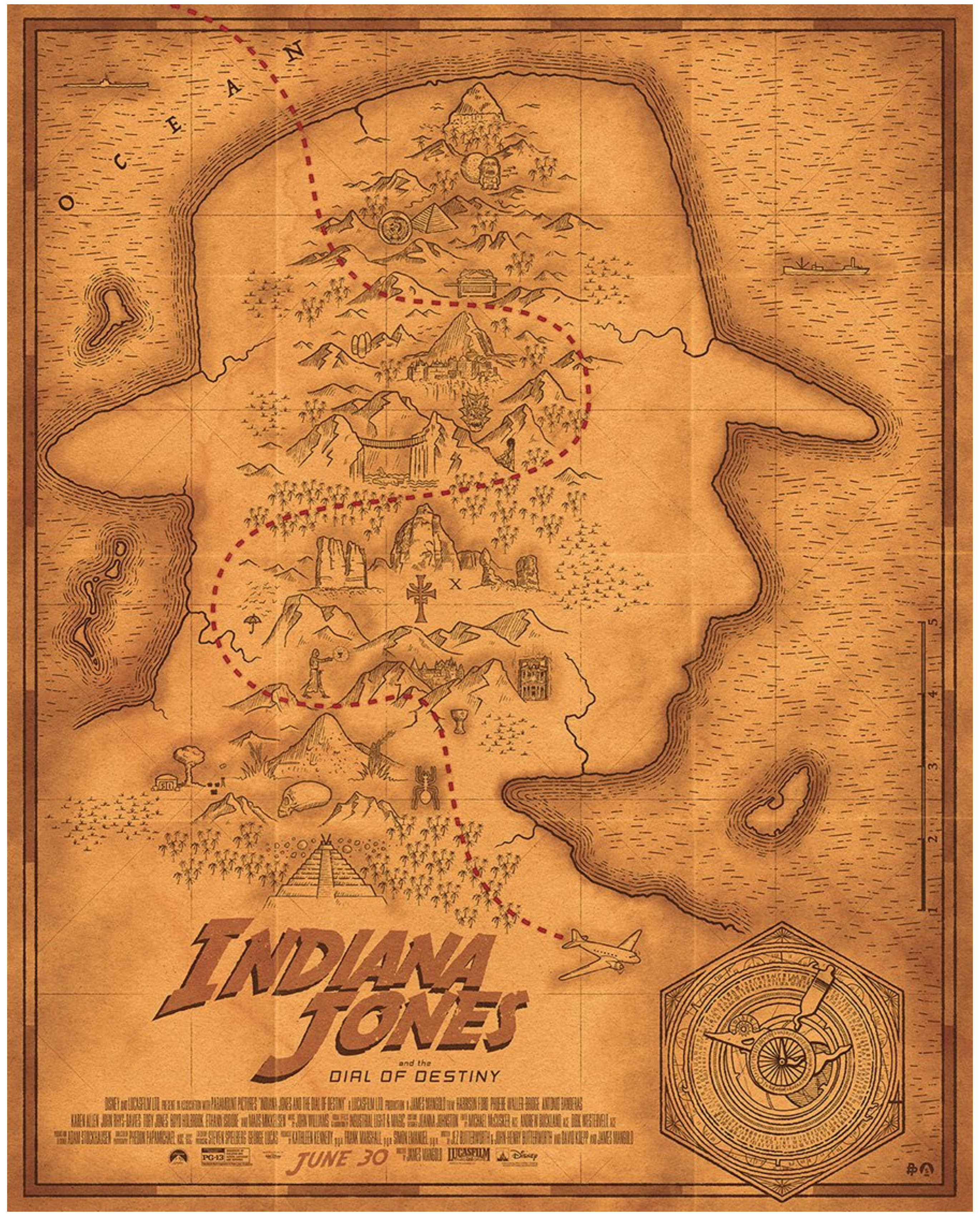
Markov chain Monte Carlo

Sampling methods & applications

Markov Chain Monte Carlo

Methods used for drawing samples that satisfy the strong law of large numbers and central limit theorem even if **not independent nor identically distributed**

One sample that explore the space moving around



What is a Markov Chain

Ingredients:

- Initial distribution: the initial configuration of the system
- Transition Kernel: probability of moving from a point to another

$$\pi(x_{0:n}) = \pi_0(x_0)\prod_{s=1}^{n} k(x_{s-1}, x_s)$$

How to use a Markov chain for sampling

We want to approximate a **distribution** using samples from a **Markov chain**

$$x_0 \sim \pi_0 \rightsquigarrow x_1 \sim \pi_1 \rightsquigarrow \dots$$

$$x_{1:I} \sim p$$

Problems:

- We are not sure that the chain converges to the target distribution *p*
- Even if it converges, there is no guarantee on Strong Law of Large numbers and central limit theorem

How to build a proper chain

Metropolis Hastings strikes back

Proposal density

$$k(x, x') = q(x' \mid x)\alpha(x, x') + (1 - \alpha(x, x'))\delta(x)$$

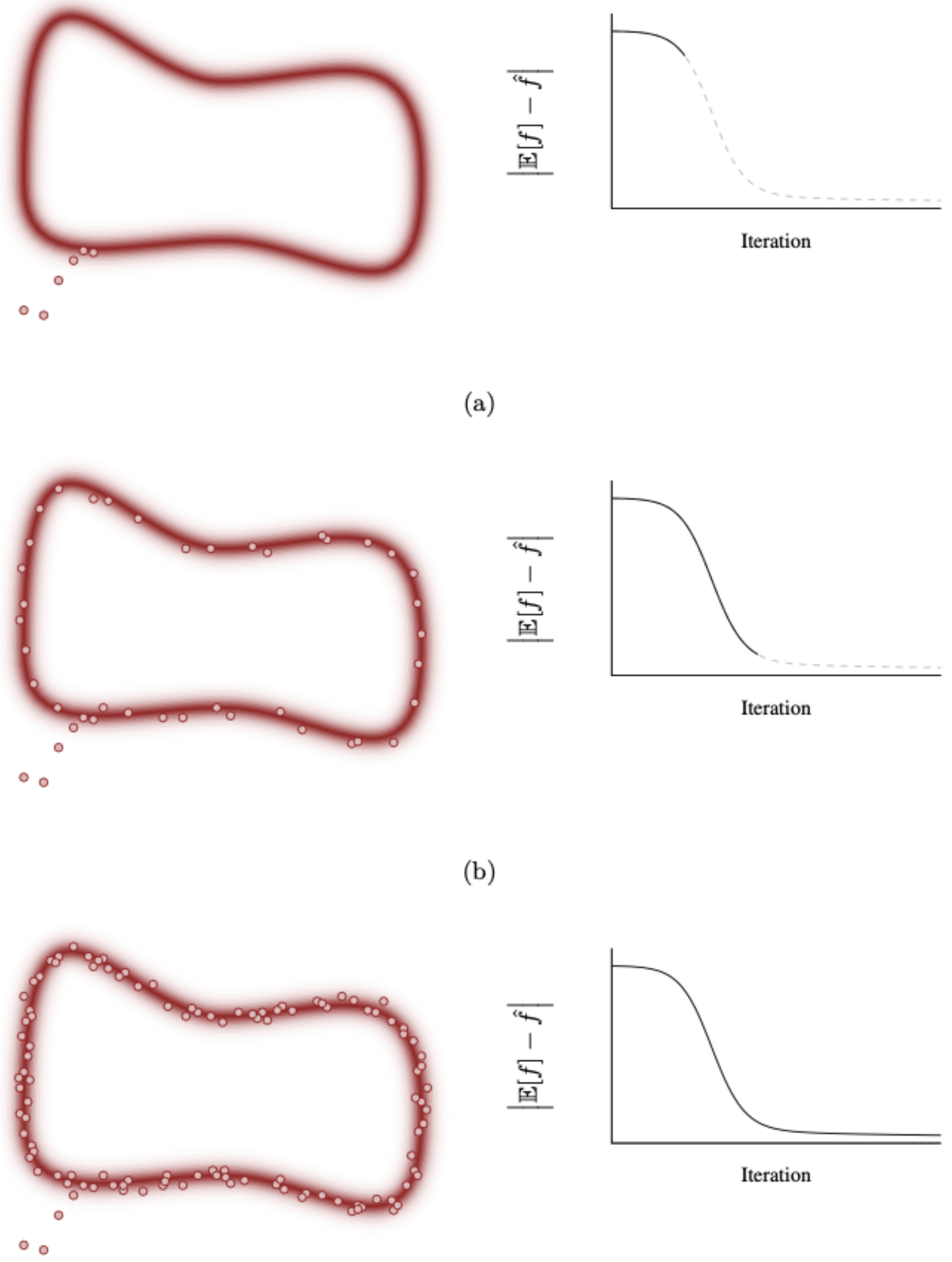
Acceptance Probability

- Propose a move from *x* to *x’* using distribution *q*
- Accept the move with probability α
- Reject the move with probability 1-α

Metropolis Hastings

One choice (of acceptance probability) to rule them all

$$\alpha(x, x') = 1 \wedge \frac{q(x' \mid x)p(x')}{q(x \mid x')p(x)}$$



This choice for the acceptance probability ensures:

- **Convergence** to *p* for the samples of the chain
- **Strong Law of large numbers** & **Central Limit theorem** for dependent samples
- **Maximization** for the acceptance probability

Pseudocode for Metropolis Hastings

$x_0 \sim \pi_0$

$for\ i\ in\ (0, I-1):$

$\quad \tilde{x} \sim q(\cdot \mid x_i)$

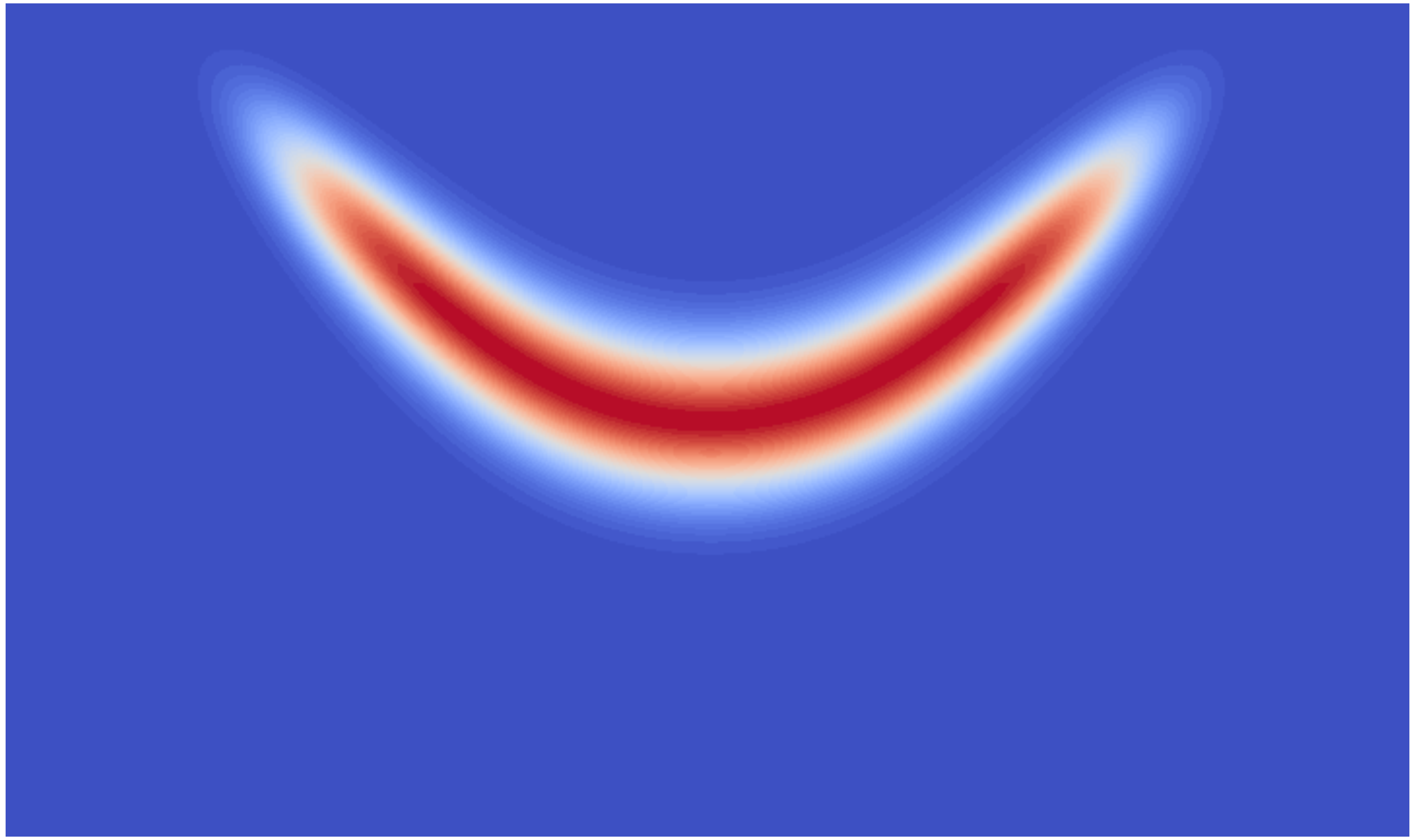
$\quad if\ \alpha(\tilde{x}, x_i) < rand(0,1) : x_{i+1} = \tilde{x}$

$\quad else : x_{i+1} = x_i$

Exercise

Apply Metropolis Hastings to obtain samples from

$$p(x, y) = \exp(-10(x^2 - y)^2 - (y - 0.25)^4)$$



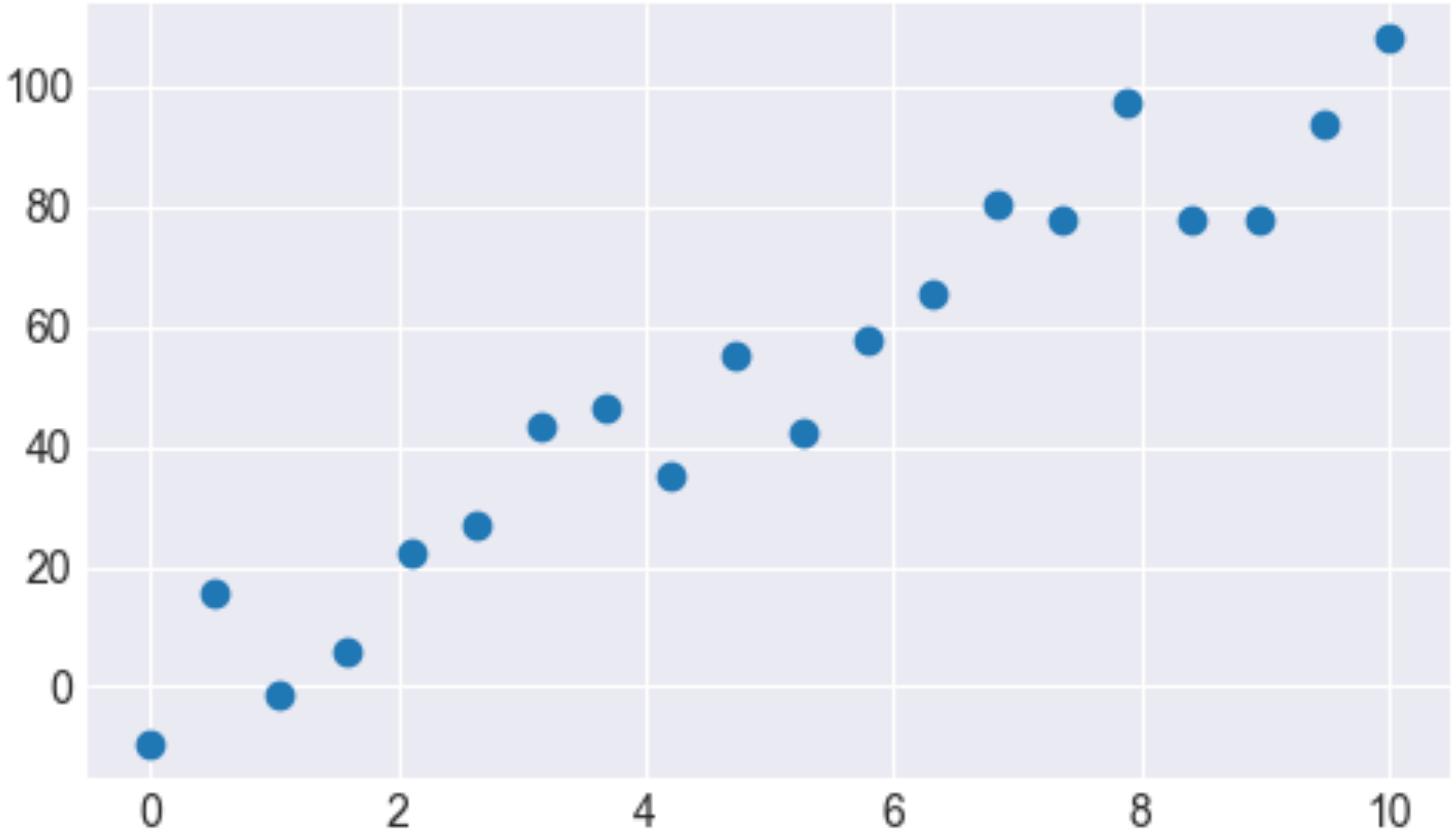
Exercise

Linear Regression

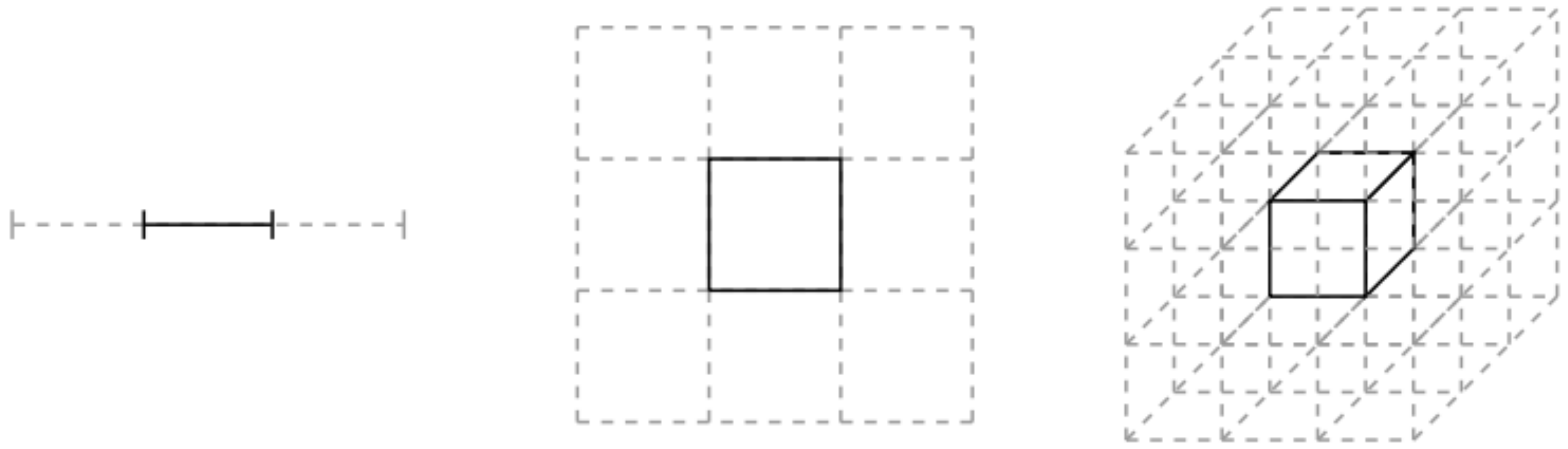
$$y = ax + b + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- 20 samples
- a=10
- b=1
- noise=10



Metropolis Hastings Limitations



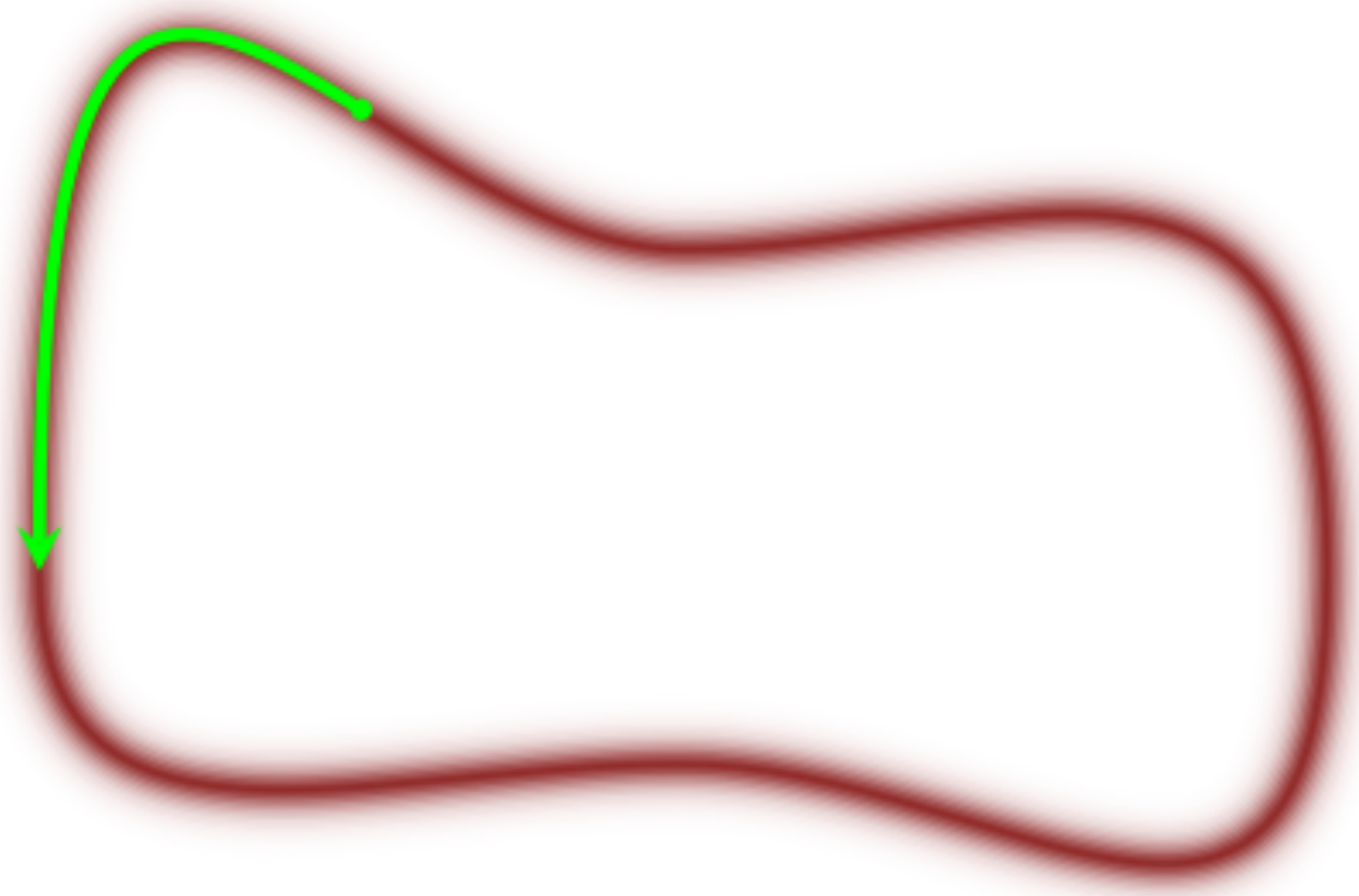
Metropolis Hastings limitations:

- Curse of dimensionality
- Possibility to get stuck in local maxima

Hamiltonian Monte Carlo

Introduce dynamics in MCMC

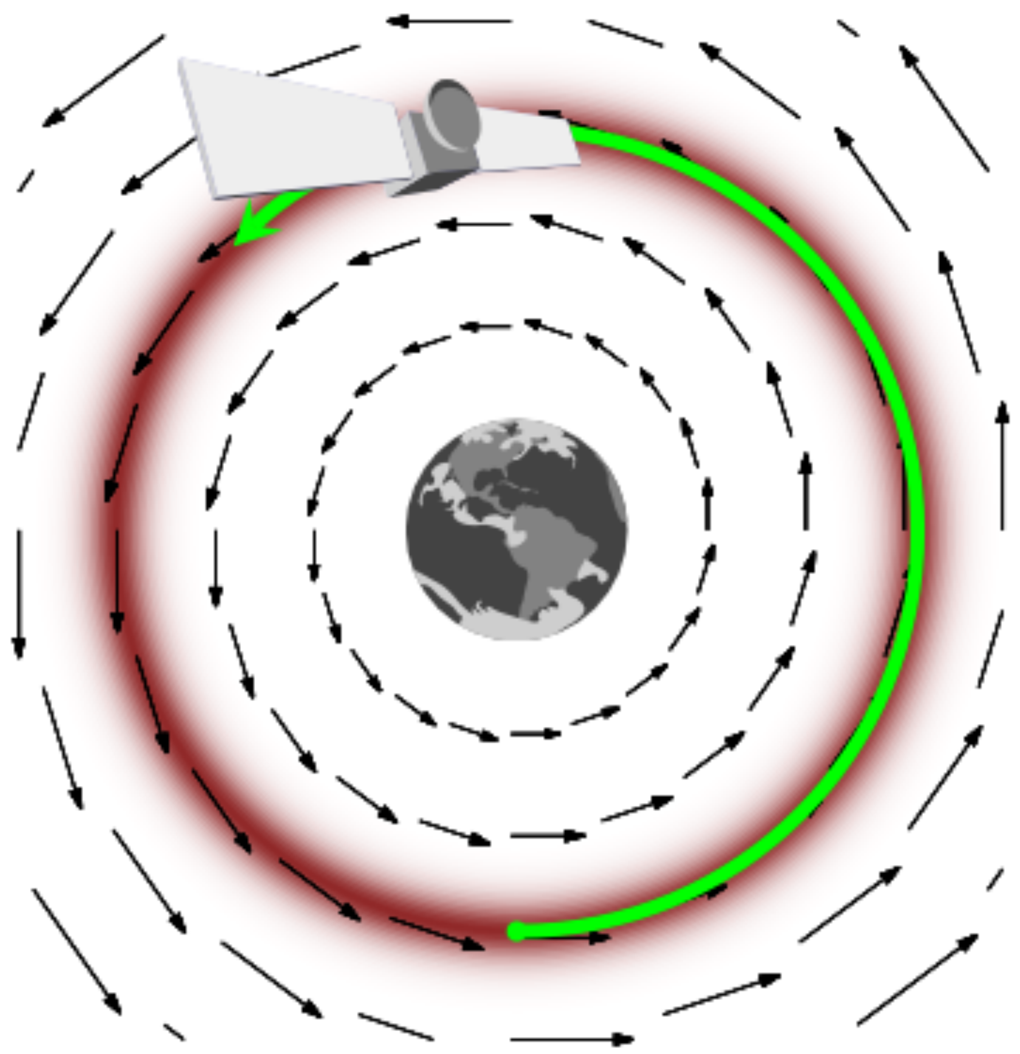
Idea: exploit geometrical information of the distribution to move across high probability regions

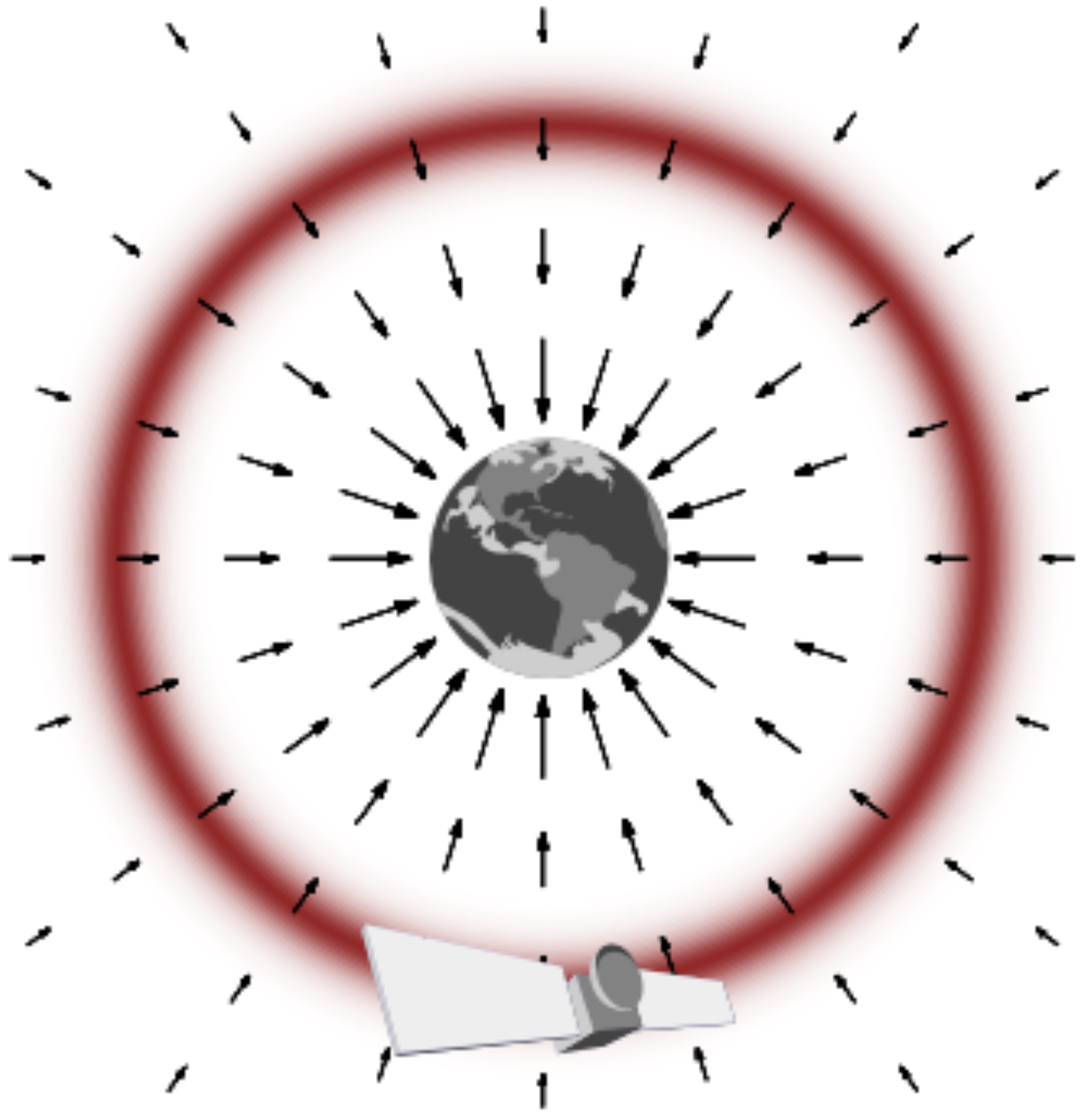


Hamiltonian Monte Carlo

Introduce dynamics in MCMC

Like in a dynamical system we have to **preserve the momentum** in order to maintain the system stable





Hamiltonian Monte Carlo

Introduce dynamics in MCMC

$$\pi(x) \rightsquigarrow \pi(x, m) = \pi(m \mid x)\pi(x)$$

We add the auxiliary variable of the MOMENTUM and thanks to the Hamilton equations, we get:

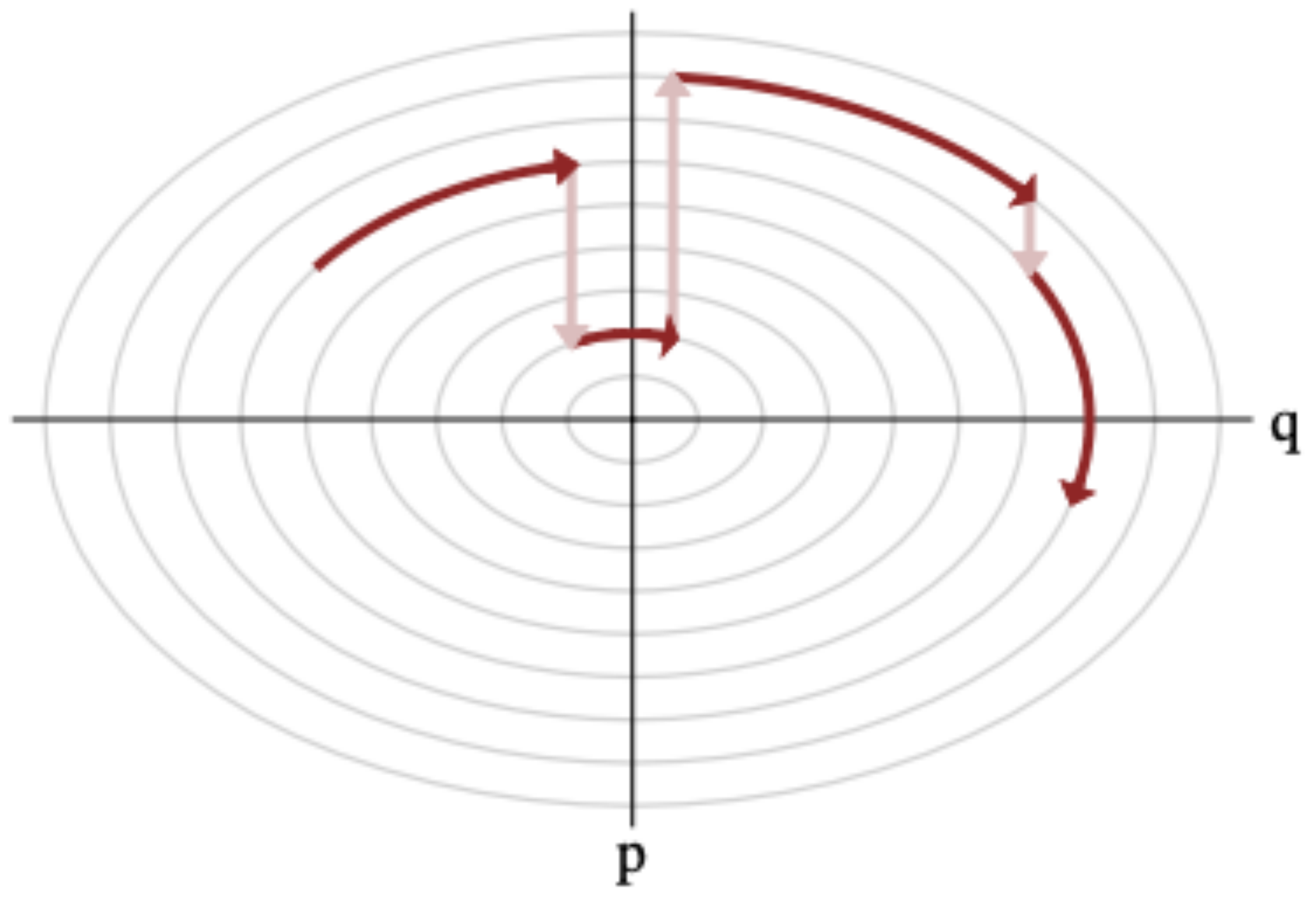
Kinetic energy

$$\partial_t m = \partial_x K$$

Potential energy

$$\partial_t x = - \partial_m K - \partial_m V$$

The idea is to **sample the momentum** and than look at the trajectory of the sample using the Hamilton's equations



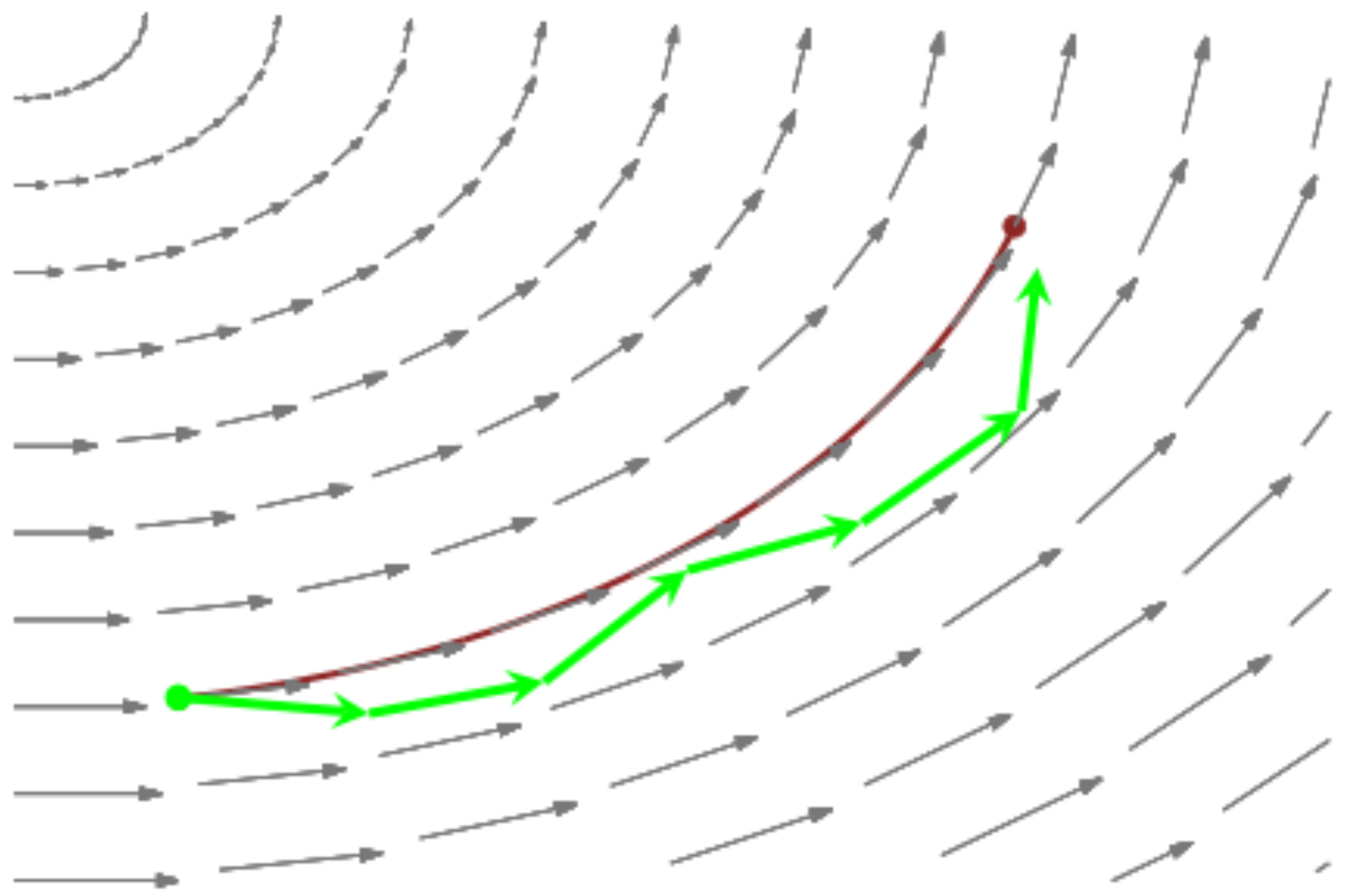
Hamiltonian Monte Carlo

Introduce dynamics in MCMC

- Select a kinetic energy function
- Sample a momentum *m*
- Select the time interval [0,T] and the discretization points
- Solve the Hamilton equations to find the new point *x*

$$\partial_t m = \partial_x K$$

$$\partial_t x = -\partial_m K - \partial_m V$$



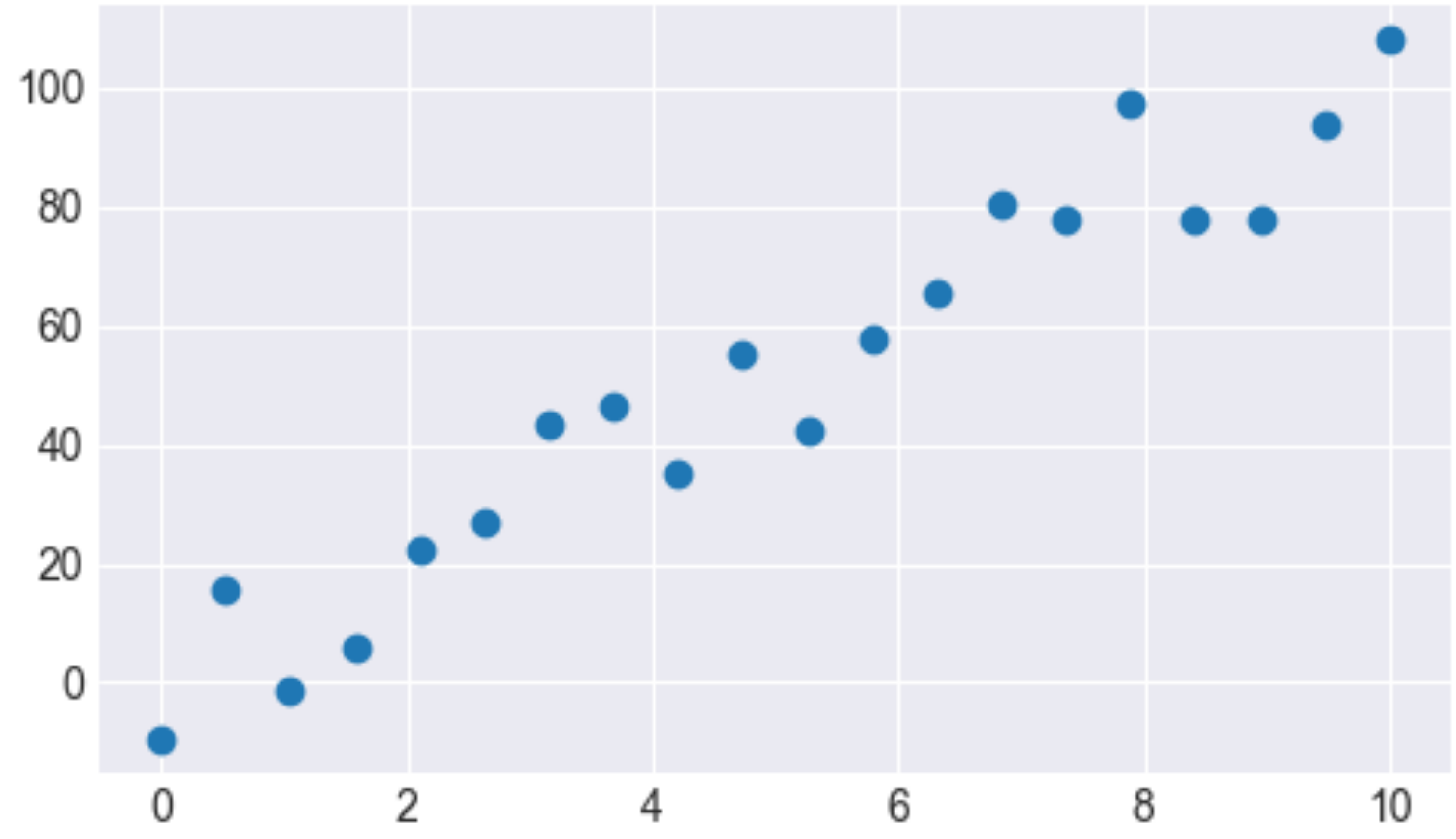
HMC WITH PYRO

https://chi-feng.github.io/mcmc-demo/app.html?algorithm=H2MC&target=banana

Solve the linear regression problem with HMC in pyro

$$y = ax + b + \varepsilon$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$



Approximate the mean and standard deviation of a gaussian distribution given some samples from it

$$y \sim \mathcal{N}(\mu, \sigma^2)$$