Advanced statistical methods and Bayesian inference in scientific research

Lecture 10

W. Dębski

3.06.2024

Metropolisa-Hastings (MH) algorithm

$$x_{i+1} = x_i + \delta x_i$$

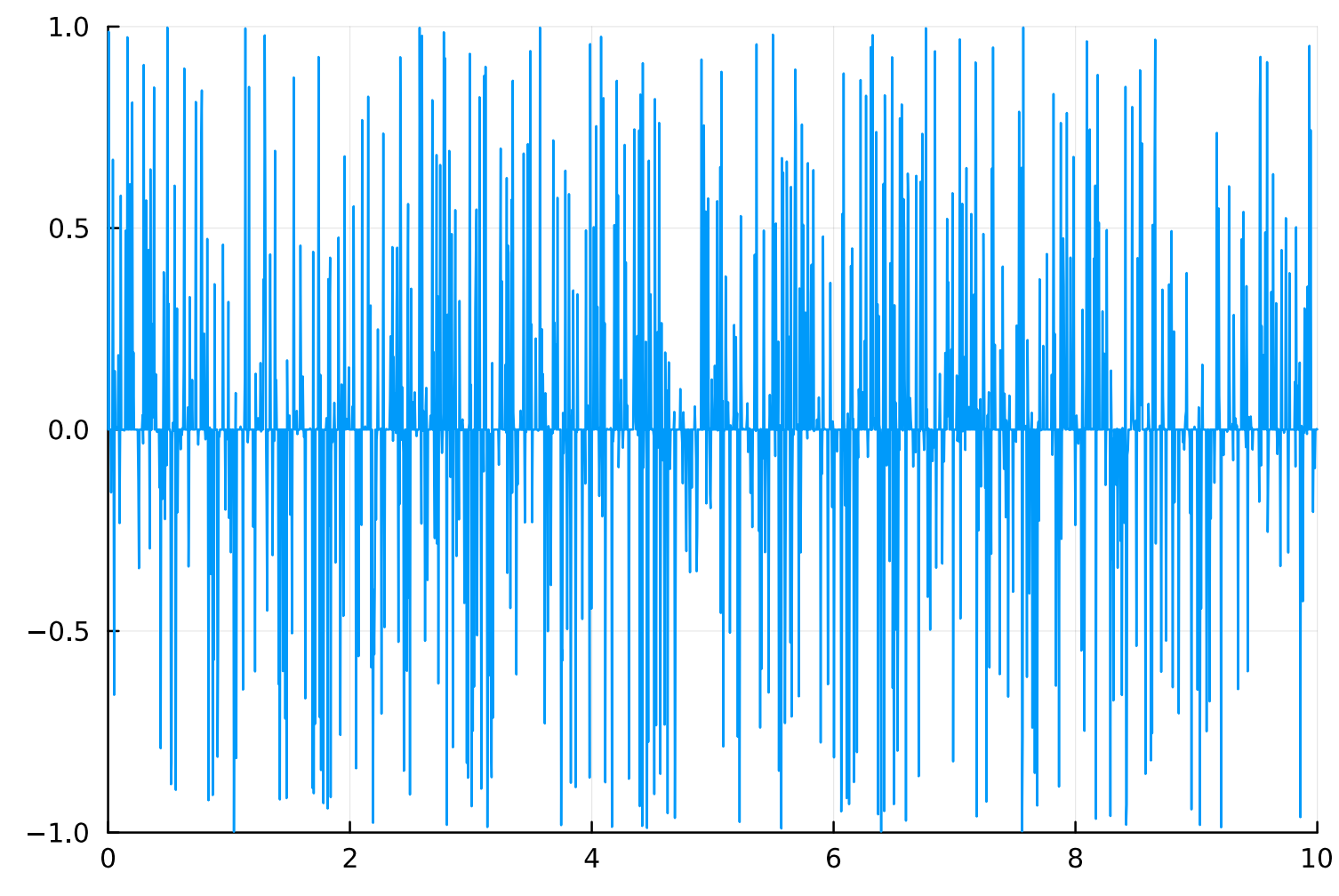
$$K(x_i; x_{i+1}) = \min\left\{1, \frac{p(x_{i+1})}{p(x_i)}\right\}$$

Metropolis algorithm

- Initialize $\mathbf{x}^0$
- Repeat
  - generate random test sample $\mathbf{x}^\beta = \mathbf{x}^\alpha + \delta * rand(0,1)$
  - accept $\mathbf{x}^\beta$ with probability $P(\mathbf{x}^\beta, \mathbf{x}^\alpha) = \min\left[1, \frac{\sigma(\mathbf{x}^\beta)}{\sigma(\mathbf{x}^\alpha)}\right]$

    - generate uniform random number $u \sim U(0,1)$
    - $u < P(\mathbf{x}^\beta, \mathbf{x}^\alpha)$
      $\mathbf{x}^{\alpha+1} = \mathbf{x}^\beta$
      otherwise
      $\mathbf{x}^{\alpha+1} = \mathbf{x}^\alpha$
- Continue until sufficient number of samples $\{\mathbf{x}^\alpha\}$ is generated

Metropolis algorithm



Metropolis algorithm - tips

$$\sigma(x) \;=\; ke^{-S(x)}$$

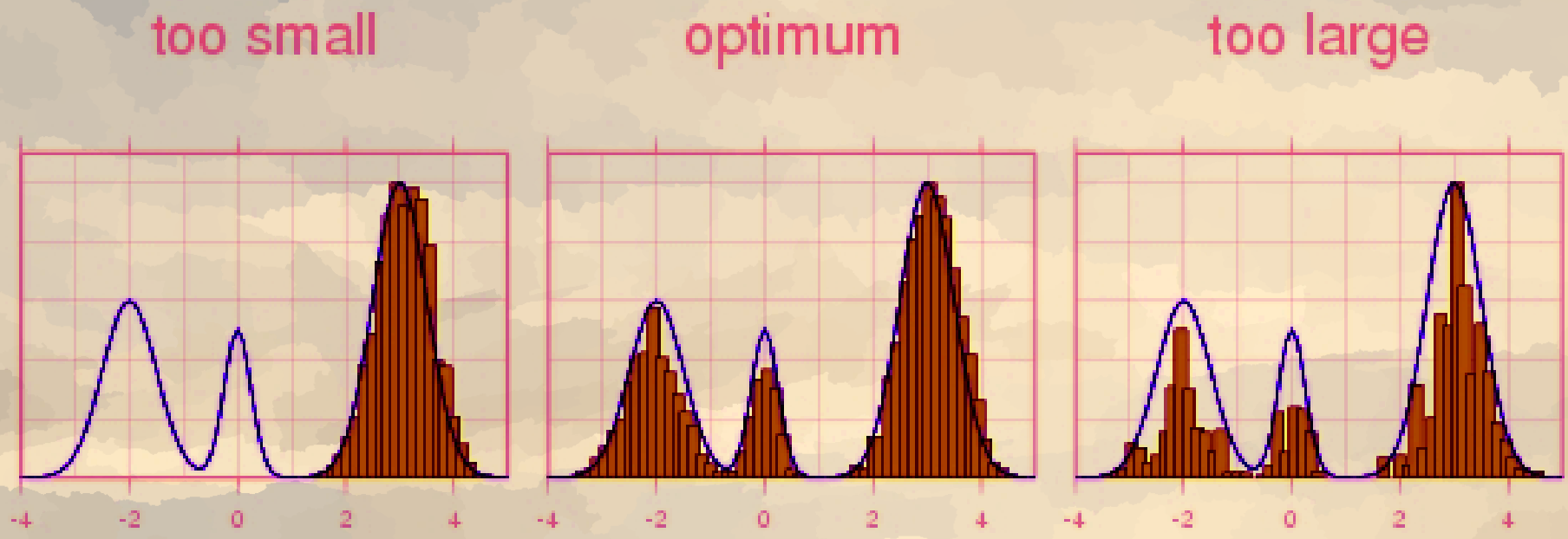
Then acceptance ratio

$$P(\mathbf{x}^{\beta}, \mathbf{x}^{\alpha}) \;=\; \min\left[1, e^{-\left(S(\mathbf{x}^{\beta}) - S(\mathbf{x}^{\alpha})\right)}\right]$$

is easier to calculate numerically (less prone to numerical accuracy)

Metropolis algorithm - generating proposal samples

The key point of the algorithm is **the proper** generating proposal samples (how large δ should be ?)



Metropolis Hasting algorithm - features

- after so called burn-in initial time generated samples follow $\sigma(\mathbf{x})$ probability distribution
- subsequent samples are strongly correlated - the chain must be run for a long time
- MH is optimum sampling algorithm if only $\sigma(\mathbf{x})$ is available.
- How many samples should be generated ?
- problem with generating "proper" test samples ($\mathbf{x}^{\beta}$)

Improving test sample generation

Runing MH chain for sufficiently long time we get some information about sampled σ distribution. Can we us this information to improve (make more robust, efficient, etc) sample generation process?

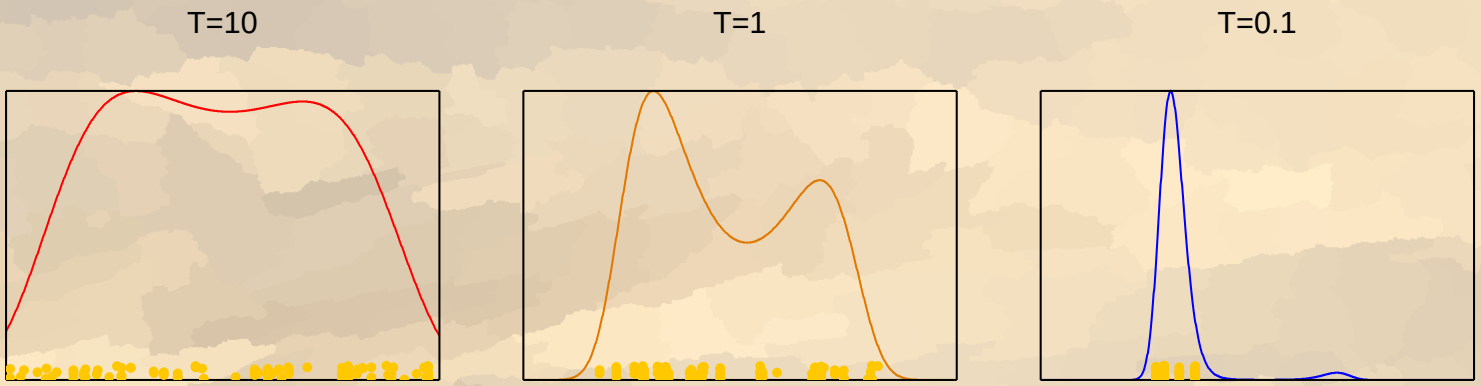
YES, but...

There is a **big risk** of biasing sampling procedure making the sampling procedure non-stationar

Non stationary MCMC - global optimization

Simulated annealing algorithm

$$p(\mathbf{x};t) \;=\; \exp\left(-\frac{S(\mathbf{x})}{kT}\right) \qquad T = T(t)$$



Simulated annealing - non-stationary sampler

The cooling process (decreasing T must be slow enought to allow sampler properly sample the whole space at "fixed T" If original MH is used one can prove that fastest cooling must acually be very slow

$$T(k) \sim \frac{1}{\ln(k)}$$

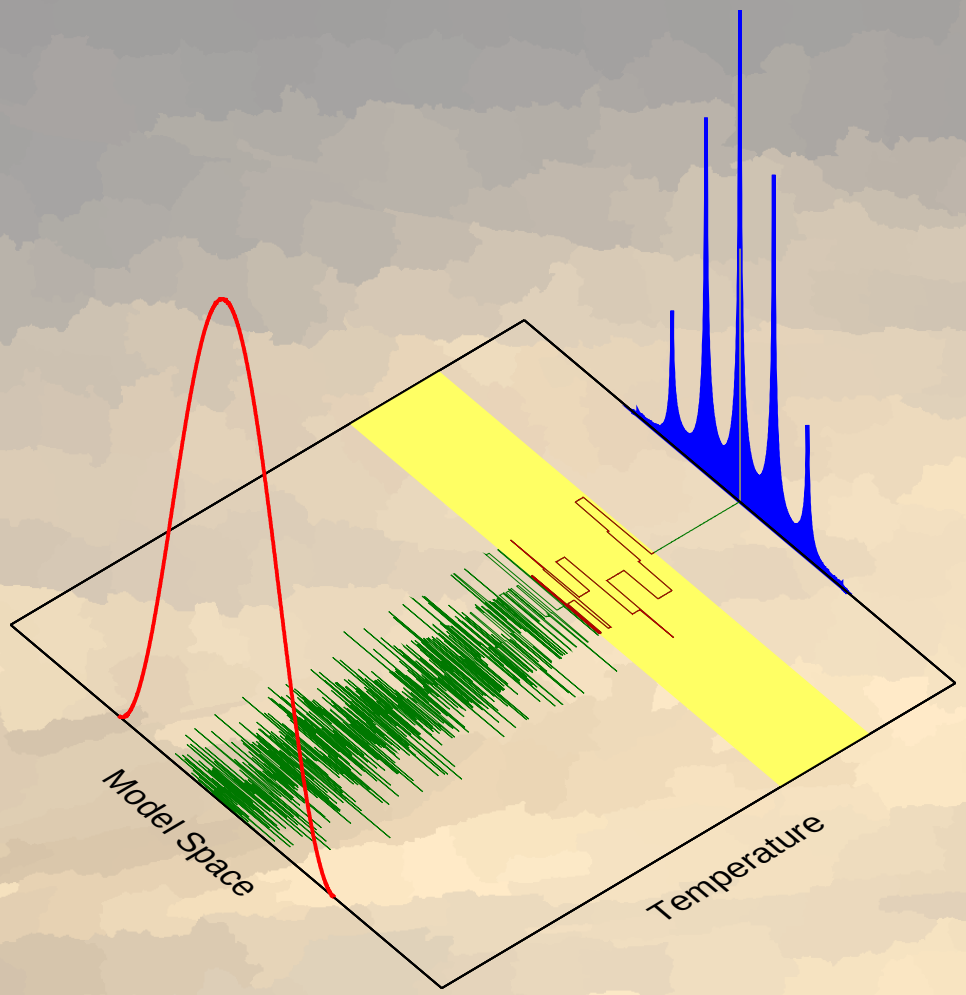
However Ingberg in 70'th had proposed to change sample generation with T

$$x_{i+1} = x_i + d(T) \times rand(0,1)$$

achiving very fast cooling spead

$$T(k) \sim \frac{1}{k}$$

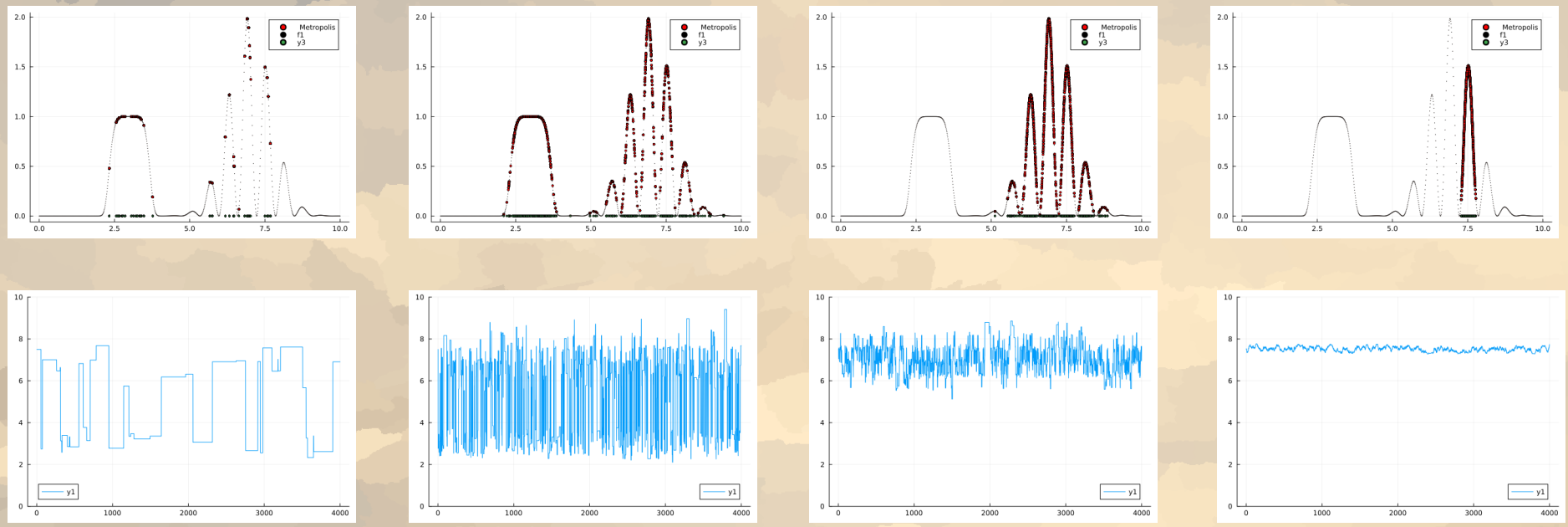
Simulated annealing - non-stationary sampler



Improving test sample generation

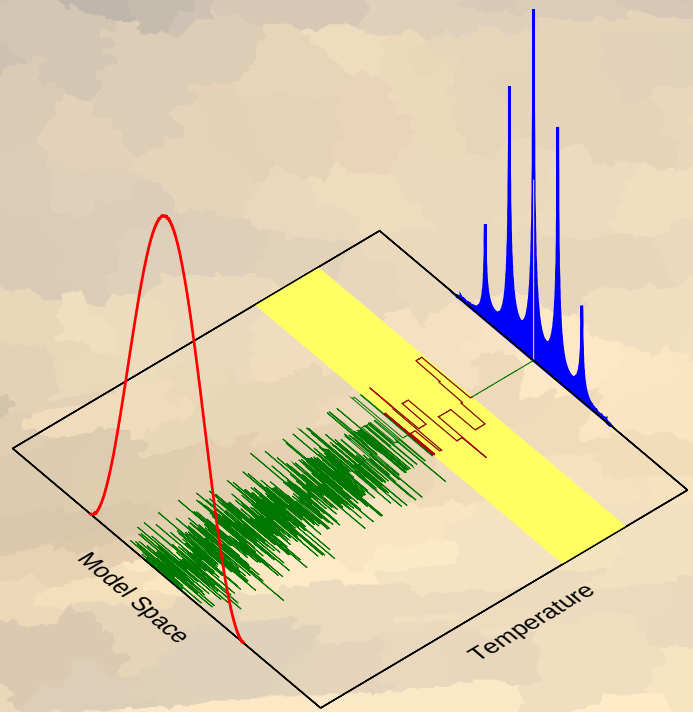
Non-stationarity of MH chain can also be caused/achived by changing the method of sample generation (like Ingberg had done). What result can be?

100 5 1 0.05



Can we do it?

If it can be accepted that our sampling will not be “perfect” but slightly contaminated/biased by slow nonstation arity of the chain we may profit it by significantly speeding up sampling but a price of loosing accuracy



Hamiltonian MCMC - extension Metropolis idea

$$x_{i+1} = x_i + \delta x_i$$

$$K(x_i; x_{i+1}) = \min\left\{1, \frac{p(x_{i+1})}{p(x_i)}\right\}$$

Hamiltonian MCMC

Original Metropolis algorithm is actually performing purely random walk over $\mathcal{M}$

$$\mathbf{x}^{\beta} = \mathbf{x}^{\alpha} + \delta\mathbf{x} \qquad \delta\mathbf{x} = \mathbf{x}_o \; + \; \delta \times rand(-1, 1)$$

(filtered next by Metropolis selection Kernel)

For this reason the algorithm requires an extremely long runs and subsequent samples are strongly correlated

Hamiltonian MCMC

Hastings has extended the Metropolis algorithm by allowing generating test samples from auxiliary (*a priori*) distribution what has improved (sometimes not) convergence of the original algorithm.

Many other attemps have also been undertaken to improve the situation (convergence of the chain) with varying results. Usually proposed extensions work better for some classes of problems but fails in others

NO FREE LUNCH THEOREM

Hamiltonian MCMC

Significant improvements of this situation requires additional information on sampled $\sigma(\mathbf{x})$. Particular (often met) situation occurs when

$$\sigma(\mathbf{x}) \;=\; k \exp\left(-L(\mathbf{x}, \mathbf{d})\right)$$

and one can calculate gradient of $L(\mathbf{x}, \mathbf{d})$ with respect to $\mathbf{x}$

$$\nabla_i L \;=\; \frac{\partial L(\mathbf{x}, \mathbf{d})}{\partial x_i}$$

One can use this information for a more efficient generation of test samples

$$\mathbf{x}^{\beta} = \mathbf{x}^{\alpha} + \delta\mathbf{x}$$

Hamiltonian MCMC - idea

The most popular method of including $\nabla_i L$ information into sampling algorithm is so called

Hamiltonian Monte Carlo

It's basic idea is avoiding generation of test sample $\mathbf{x}^\beta = \mathbf{x}^\alpha + \delta\mathbf{x}$ more or less randomly but rather to take it as a result of a deterministic evolution from a given chain state $\mathbf{x}^\alpha$.

This evolution is based on the Hamiltonian dynamics in an extended phase space (X, V)

Hamiltonian dynamics

Let **P** be a particle of a unit mass moving in R^3 space due to a potential energy E_p. Let

- x - position of particle
- v - its velocity

One can define the Hamiltonian (energy) functional

$$H(x,v) \;=\; E_p \;+\; \frac{1}{2}||v||^2$$

over the phase space $\mathcal{P} = \{p(t) := (x(t), v(t))\}$

Hamiltonian dynamics

Evolution equations:

$$\frac{dx}{dt} = \frac{\partial H}{\partial v}, \qquad \frac{dv}{dt} = -\frac{\partial H}{\partial x}$$

- Evolution conserves "energy":

$$\frac{dH}{dt} = 0, \qquad H(x(t), v(t)) \; = \; H(x(0), v(0))$$

- Volume in phase space is conserved:

$$\vec{F} = (\frac{dx}{dt}, \frac{dv}{dt}); \quad \nabla \cdot \vec{F} = 0$$

Hamiltonian dynamics

Hamiltonian's equation describe evolution of the system in the phase space $\mathcal{P}$ parameterized by time t

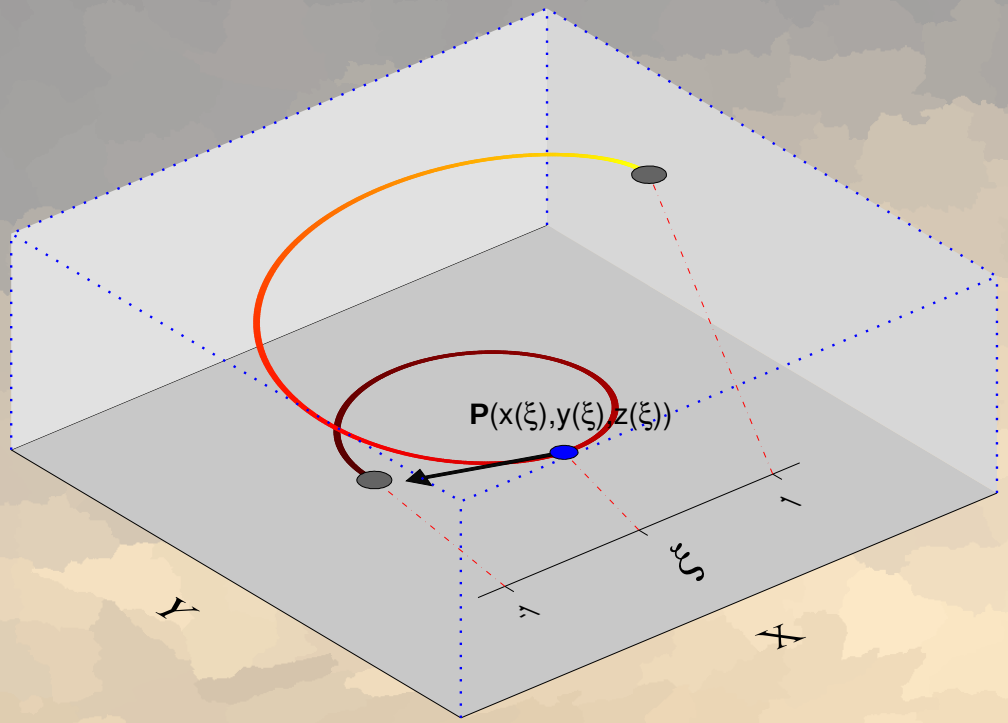
$$p(t) \;=\; (x(t), v(t)))$$

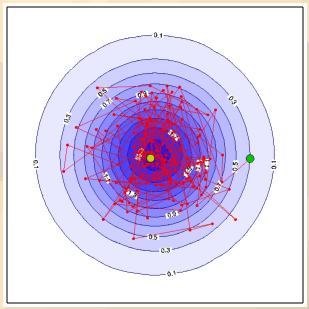
However, from geometrical point of view, after removing "parameter " t, this evolution is represented by given trajectory in $\mathcal{P}$ space

$$\Gamma(x, v) \;=\; 0$$

Energy, and "volume" are conserved along this trajectory!

Hamiltonian dynamics phase space trajectory





Hamiltonian MCMC

If we properly define "hamiltonian" for our sampled pdf $\sigma(x)$ then, for any point $(x,v_i) \in \mathcal{P}$ one can run "dynamic evolution" to get trajectory $\Gamma(x,v) = 0$ and pick up any point of this trajectory as a test sample for metropolis acceptance step.

$$x^* = x_\Gamma : \quad \Gamma(x_\Gamma, v_\Gamma) = 0$$

Energy conservation assures (if H is properly defined) that

$$\sigma(x^*) \approx \sigma(x_i)$$

So, it is very probably that the test sample will be accepted by MH selection rule

$$x_{i+1} = x^*$$

Hamiltonian Monte Carlo

- Extend the model parameter space: $\mathcal{X} \rightarrow (\mathcal{X}, \mathcal{P})$
- Build the Hamiltonian functional

$$H(\mathbf{x}, \mathbf{p}) \;=\; -\ln(\sigma(\mathbf{x})) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}$$

- Choose starting point $\mathbf{x}_0$ and auxiliary $\mathbf{p}$ e.g. from $\mathcal{N}(0, I_M)$
- Repeat

> ★ evolve current state $(\mathbf{x}^\alpha, \mathbf{p}^\alpha)$ over time τ according to Hamiltonian equations to get a proposal MCMC state $\mathbf{x}^\beta = \mathbf{x}^\alpha(\tau)$
>
> ★ accept (reject) $\mathbf{x}^\beta$ with Metropolis Hasting acceptance kernel

See you June, 10th