Advanced statistical methods and Bayesian inference in scientific research

Lecture 11

W. Dębski

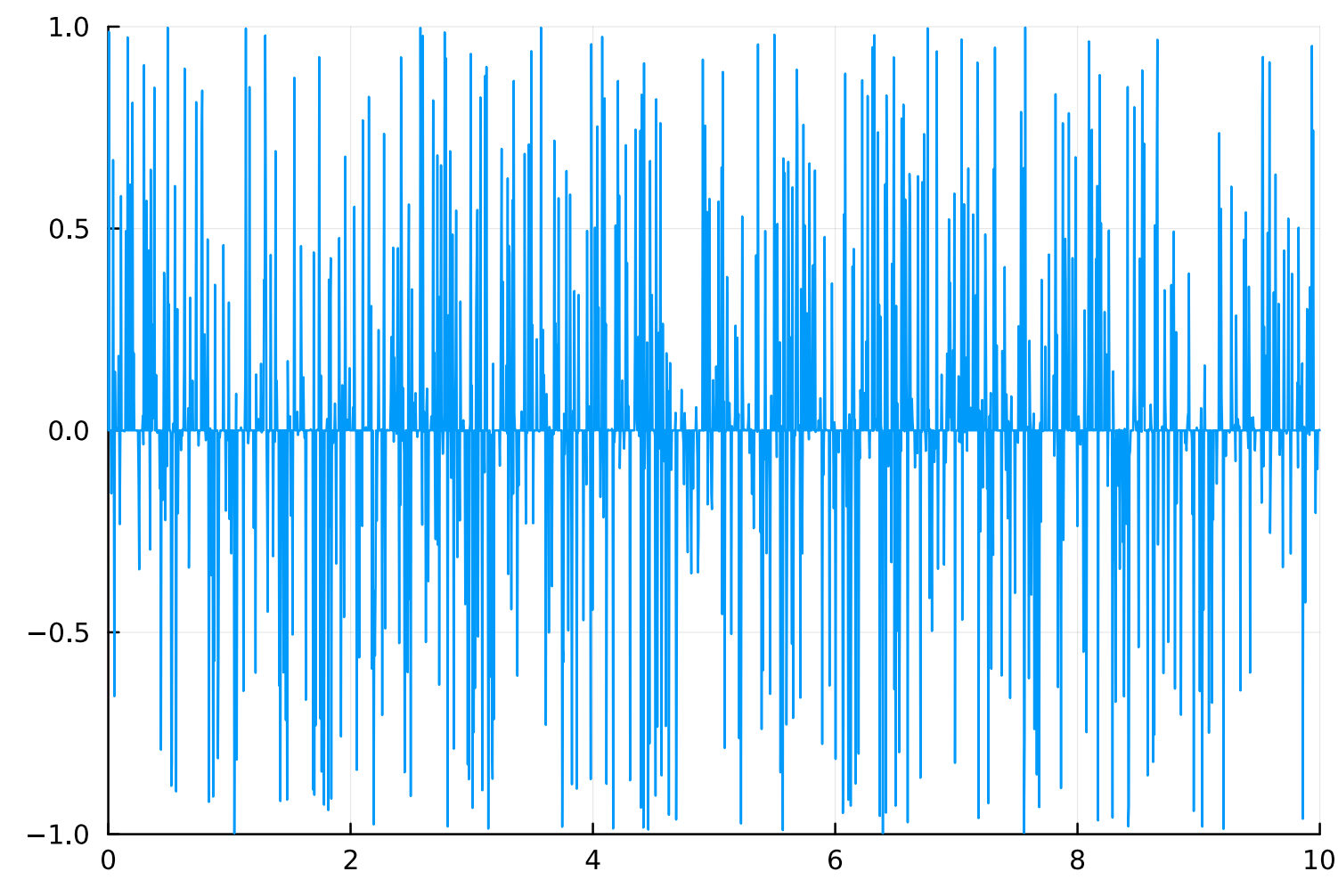
10.06.2024

Metropolisa-Hastings (MH) algorithm

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

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

Samples correlation



Hamiltonian MCMC

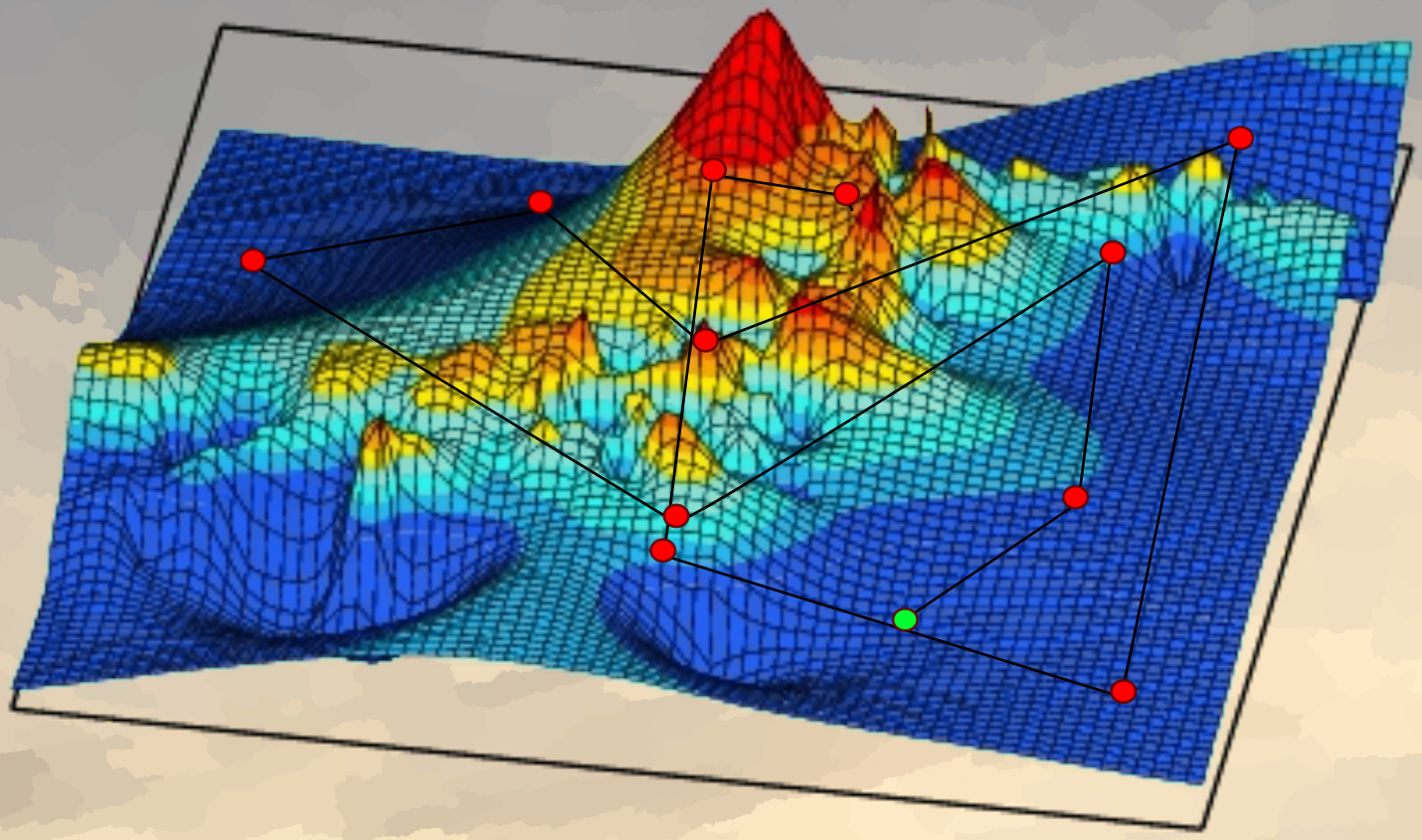
Original Metropolis algorithm is actually performing purely random walk over $\mathcal{M}$

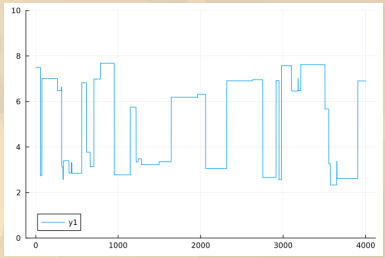
$$\mathbf{x}^{\beta} = \mathbf{x}^{\alpha} + \delta\mathbf{x} \qquad \delta\mathbf{x} = \mathbf{x}_o \; + \; \delta \times rand(-1, 1)$$

(filtered next by Metropolis selection Kernel)

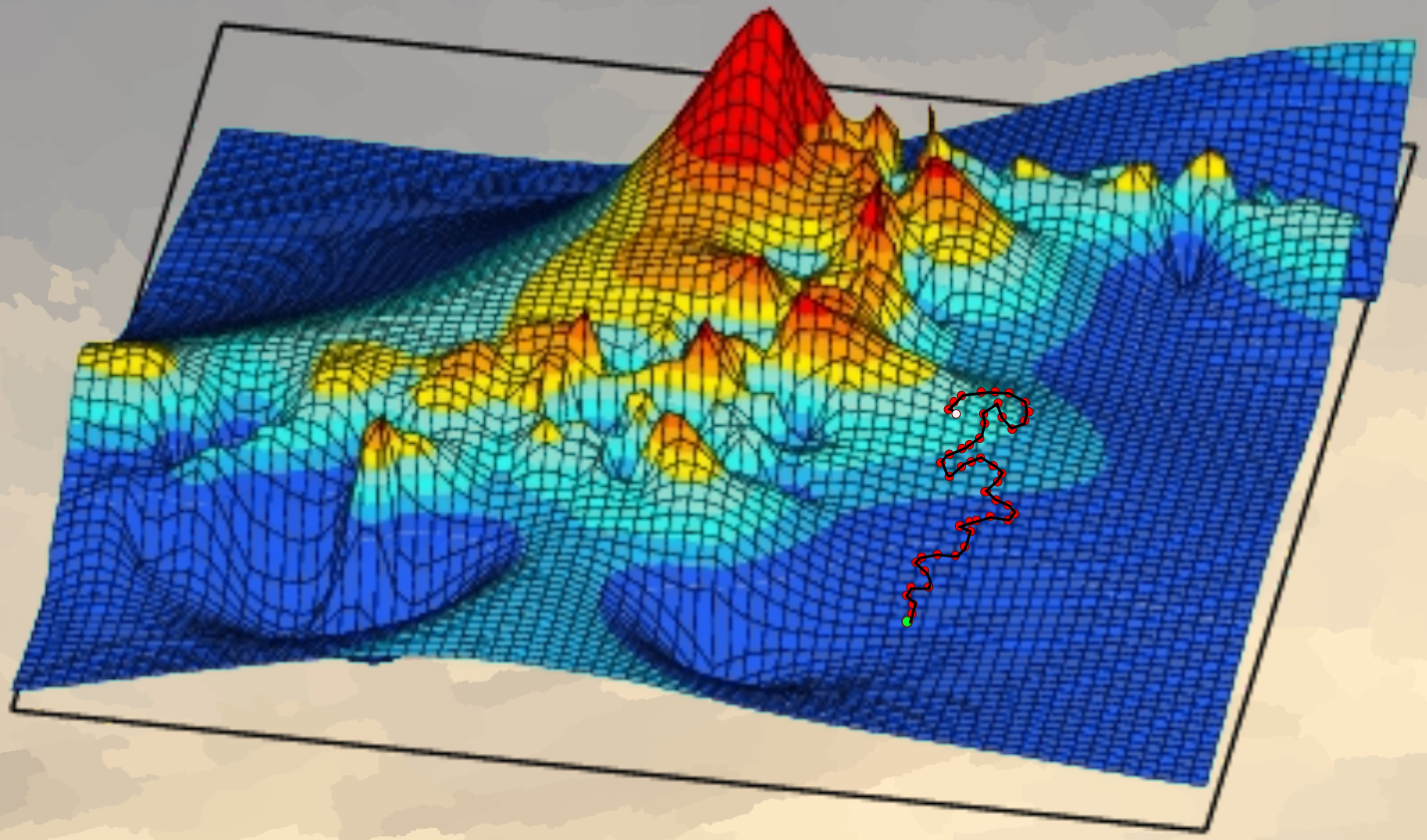
For this reason the algorithm requires an extremely long runs and subsequent samples are strongly correlated

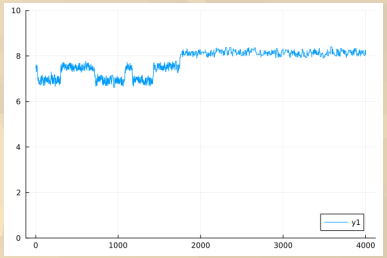
MH - too large steps



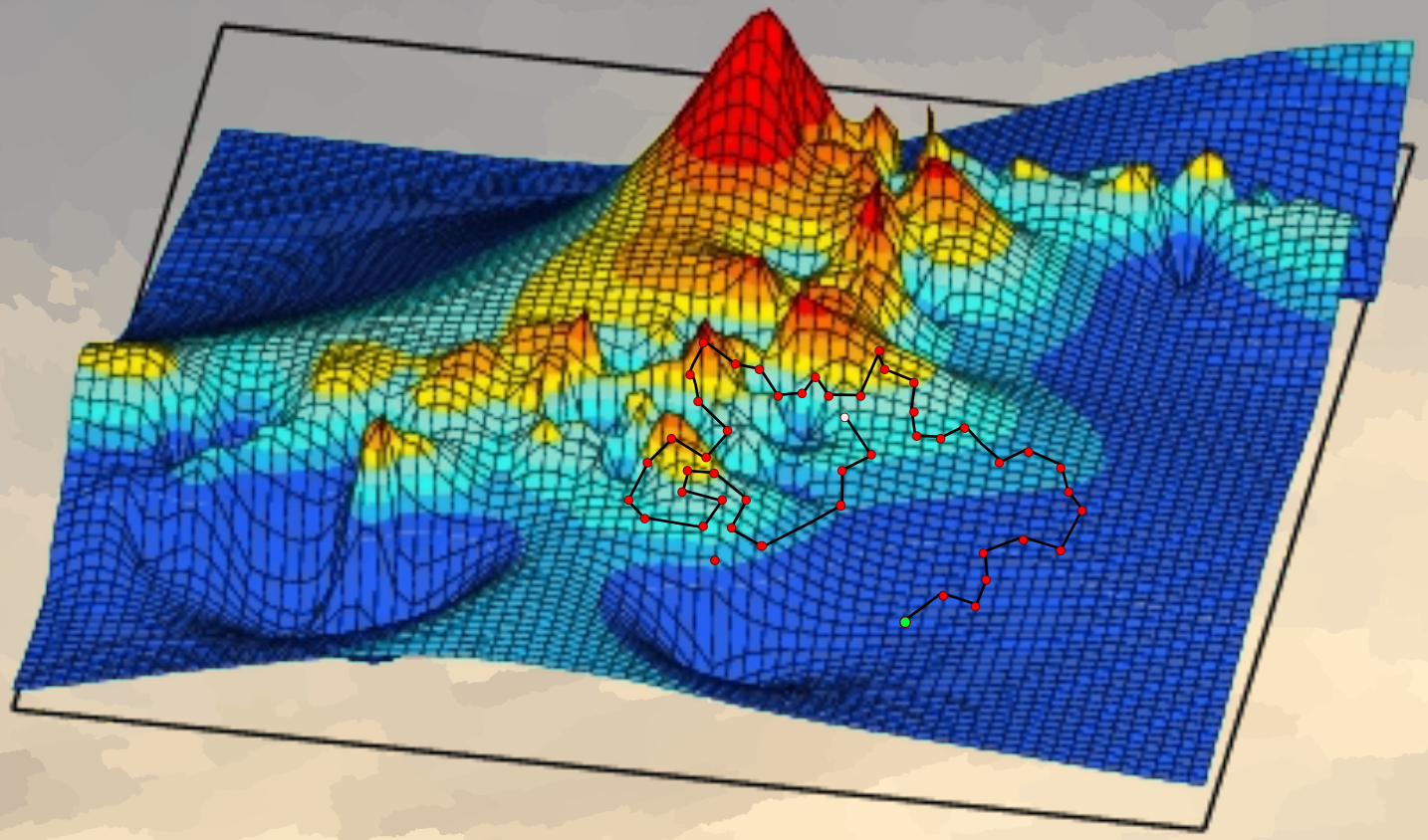


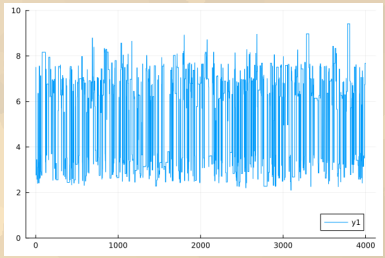
MH - too small steps





MH - oiptimium steps





Hamiltonian MCMC

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

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

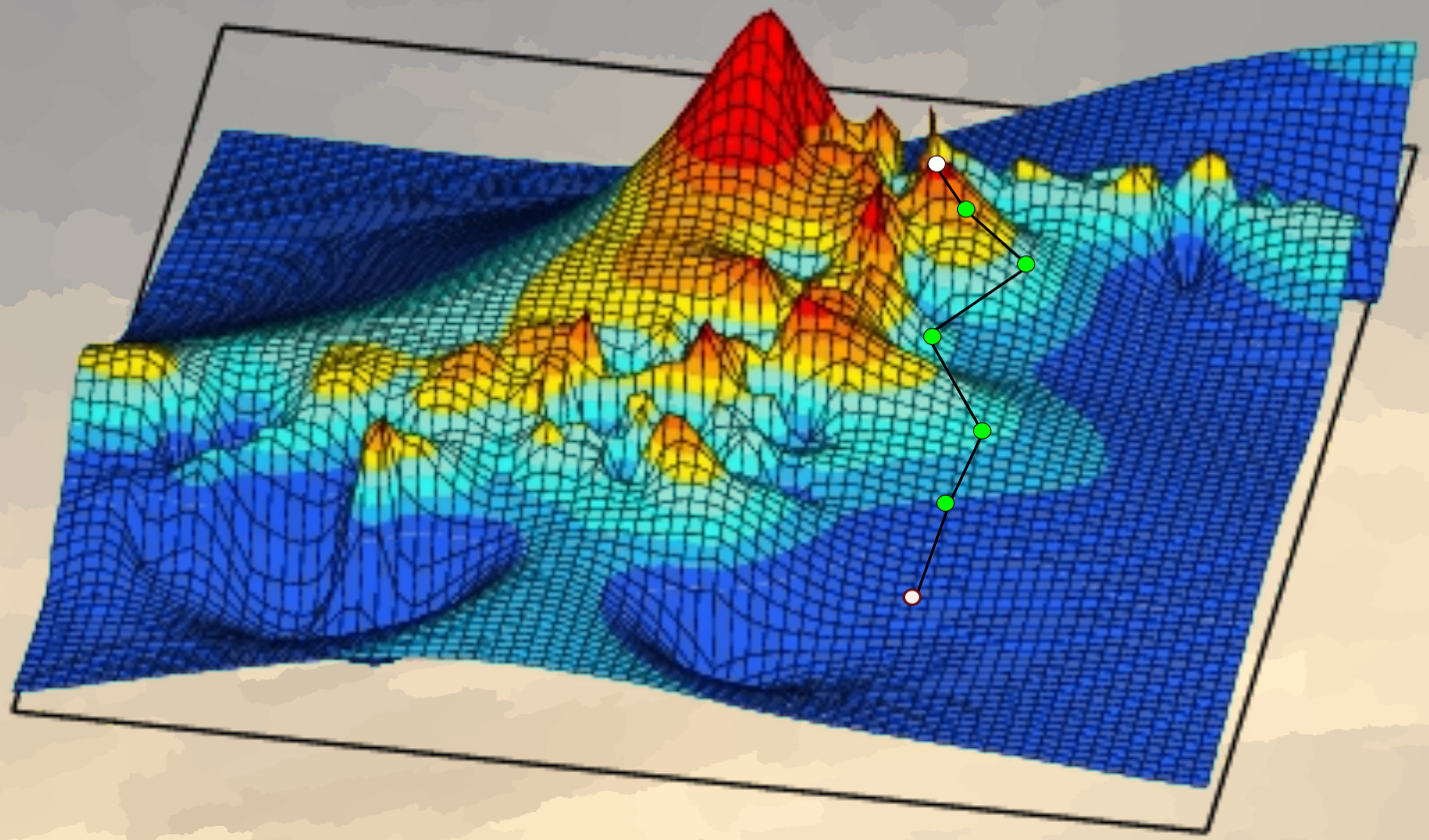
and one can calculate gradient of $S(\mathbf{x})$ with respect to $\mathbf{x}$

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

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

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

MCMC $\Longrightarrow$ Gradient optimization



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 - example

Variables (x, p) with hamiltonian $H(x, p) = \frac{1}{2}kx^2 + \frac{p^2}{2m}$

Hamilton equations

$$\frac{dx}{dt} = p/m \qquad \frac{dp}{dt} = -kx$$

from which

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

$$x(t) = \sin(\omega t) \qquad p(t) = m\omega\cos(\omega t)$$

Trajectory:

$$x^2 + (p/(m\omega))^2 - 1 = 0$$

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: $\sigma(\mathbf{x}) = k \exp(-S(\mathbf{x}))$

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

- Choose starting point $\mathbf{x}_0$ and auxiliary $\mathbf{p}$ e.g. from $\mathcal{N}(0, I_M)$
- Repeat sampling from $\sigma(\mathbf{x}, p) = e^{-H(x,p)} = \sigma(x)e^{-1/2\mathbf{p}^T M^{-1}\mathbf{p}}$

★ **evolve current state $(\mathbf{x}^\alpha, \mathbf{p}^\alpha)$ over time τ according to Hamiltonian equations to get a proposal MCMC state $\mathbf{x}^* = \mathbf{x}^\alpha(\tau)$**

★ accept (reject) $\mathbf{x}^*$ with Metropolis Hasting acceptance kernel

Hamiltonian Monte Carlo - "deterministic" evolution part

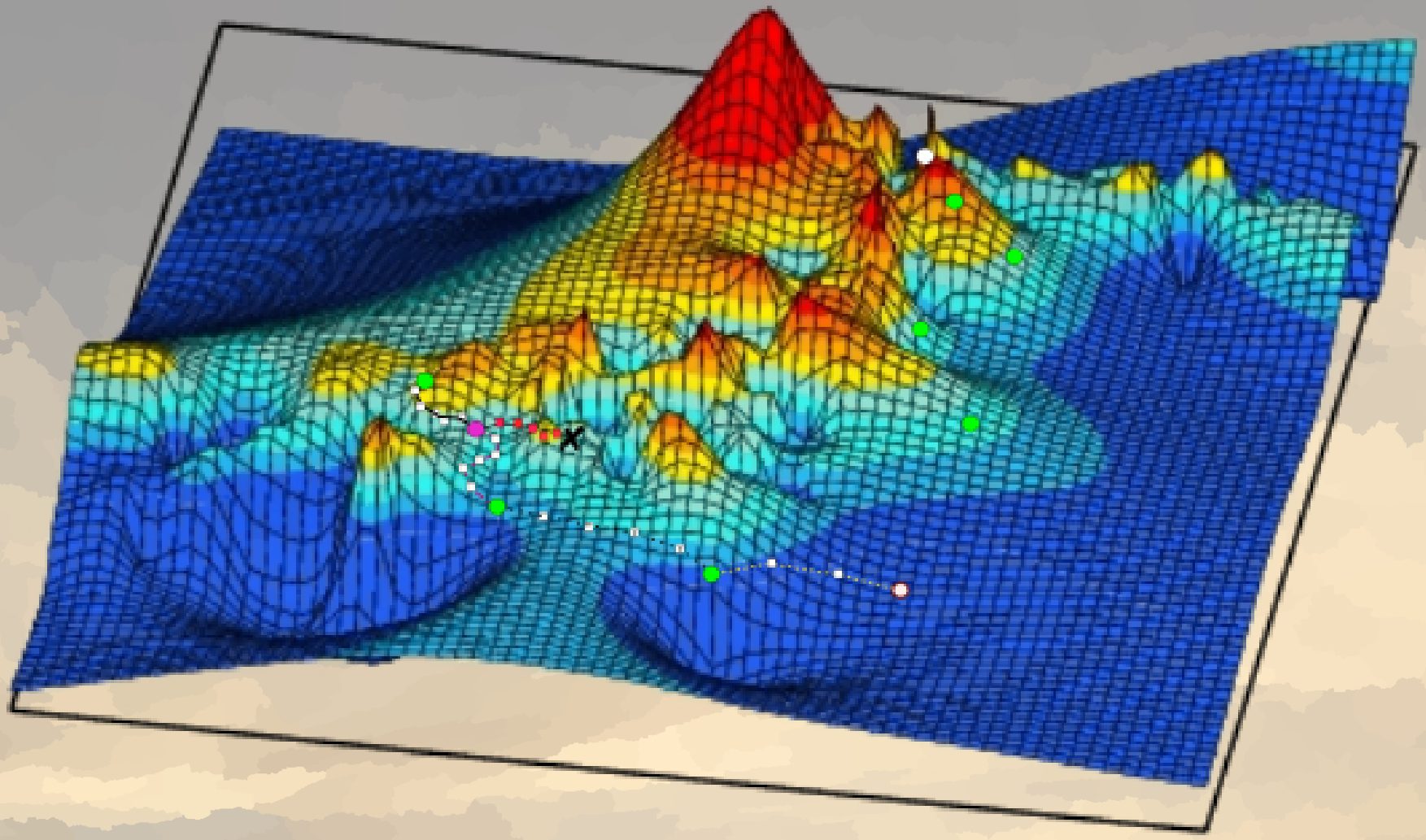
$$\frac{dx}{dt} = \frac{\partial H}{\partial p}, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial x}$$

Discretized version (with leapfrog integration schemata) with time step ϵ

$$\begin{aligned} p_i(t+\epsilon/2) &= p_i(t) - (\epsilon/2)\,\frac{\partial S}{\partial x_i}(x(t)) \\ x_i(t+\epsilon/2) &= x_i(t) + \epsilon\,\frac{p_i(t+\epsilon/2)}{m_i} \\ p_i(t+\epsilon) &= p_i(t+\epsilon/2) - (\epsilon/2)\,\frac{\partial S}{\partial x_i}(x(t+\epsilon)) \end{aligned}$$

Metropolis vs. Hamiltonian MC

Hamiltonian MCMC almost summary



HMCMC - single step - Langevin method

$$H(\mathbf{x},\mathbf{p}) \;=\; S(\mathbf{x}) + \frac{1}{2}\sum_i p_i^2$$

- sample p_i from $\mathcal{N}(0, I)$
- generate proposal x^*

$$x_i^* \;=\; x_i - \frac{\epsilon^2}{2}\,\frac{\partial S}{\partial x_i}(x_i) \;+\; \epsilon\, p_i$$

$$p_i^* \;=\; p_i - \frac{\epsilon}{2}\,\frac{\partial S}{\partial x_i}(x_i) - \frac{\epsilon}{2}\,\frac{\partial S}{\partial x_i}(x_i^*)$$

- accept or reject x^*

$$P(x \to x^*) \;=\; \min\left\{1, \exp\left(-(H(x^*,p^*) - H(x,p)\right)\right\}$$

Langevin MCMC

First equation:

$$x_i^* \;=\; x_i - \frac{\epsilon^2}{2}\,\frac{\partial S}{\partial x_i}(x_i) \;+\; \epsilon p_i$$

is a discretized version of the overdamped Langevin equation

$$\frac{dx}{dt} \;=\; -\,\lambda\,\frac{\partial S}{\partial x} \;+\; \eta(t)$$

particularly if $x \equiv v$ $\qquad S(v) = 1/2\,v^2; \quad \sigma(v) = \exp(-1/2\,v^2)$

$$\frac{dv}{dt} \;=\; -\,\lambda\,v \;+\; \eta(t)$$

describes Brovian motion.

See you June, 17th