

Different MCMC Algorithms

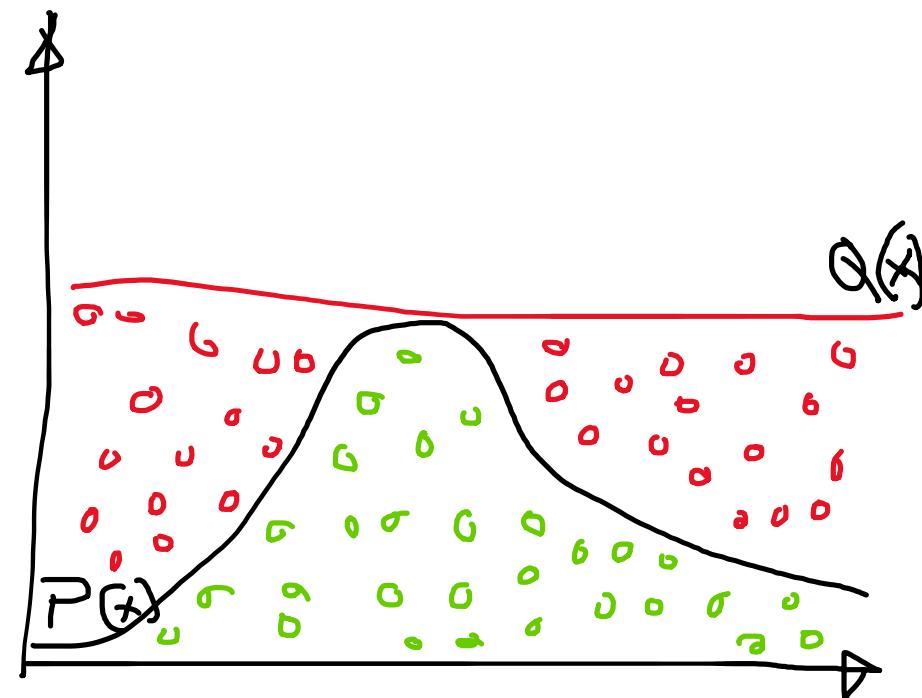
Supranta S. Boruah

MCMC Algorithms beyond Metropolis-Hastings

- In the last two seminars, we saw the MH algorithm which we used for a different inference tasks
- Today we will study some other Monte-Carlo algorithms. Some of these are:
 - Rejection Sampling (just Monte-Carlo, not MCMC)
 - Gibbs Sampling
 - Slice Sampling
 - Hamiltonian Monte Carlo
 - Ensemble Sampler (Is it really Markov Chain?)

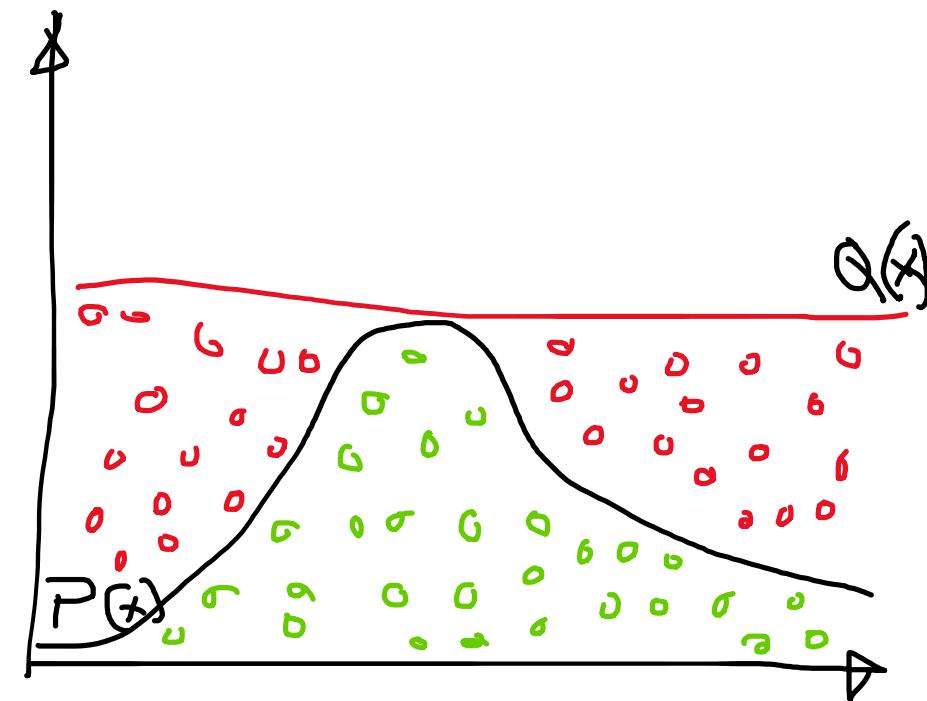
Rejection Sampling

- Rejection sampling is used to sample from an arbitrary 1D distribution, $P(x)$.
- Algorithm:
 - i. Choose an enclosing distribution, $Q(x)$ (Let's choose uniform for simplicity)
 - ii. Draw a point, x_p from $Q(x)$.
 - iii. Draw another point, $y \sim \text{Unif}(0, Q(x_p))$. If $y > P(x_p)$, reject the point. Otherwise accept it.
 - iv. Repeat steps (i)-(iii) until you have N samples



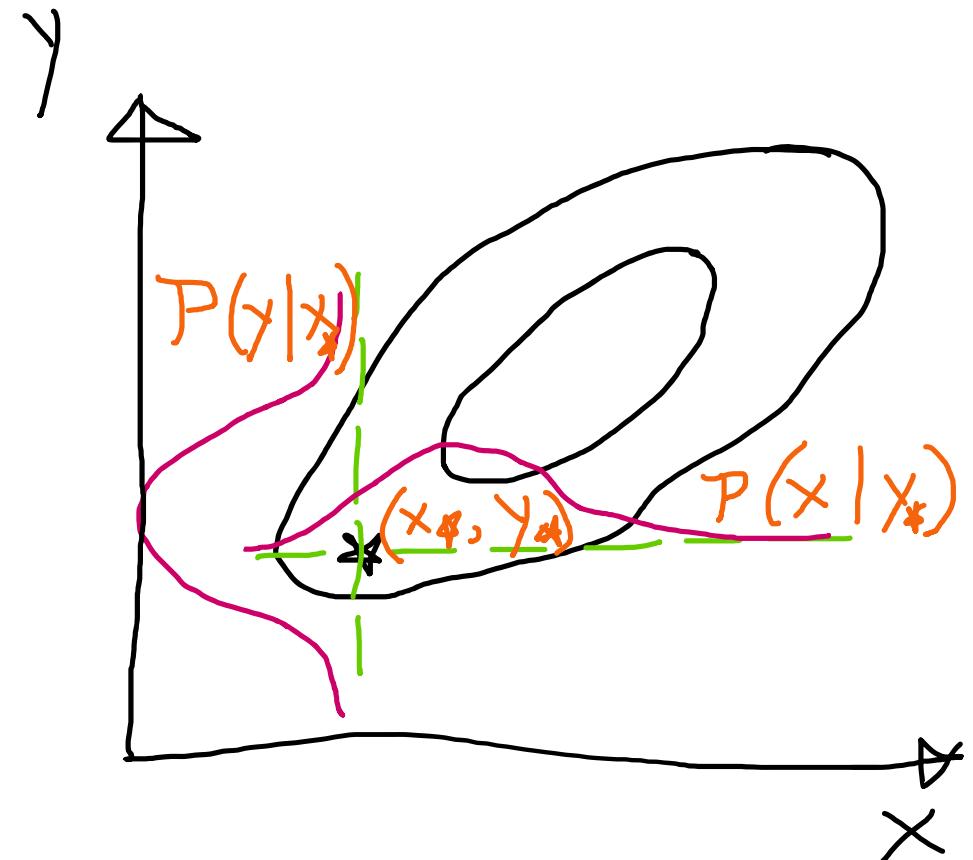
Rejection Sampling

- Properties and practical considerations:
 - $P(x)$ and $Q(x)$ need not be normalized. All we need is that $P(x)$ is enclosed by $Q(x)$
 - If $Q(x)$ envelope over P is too big, most of the samples would be rejected, thus, is computationally inefficient. Nevertheless, it give the correct distribution for $P(x)$.



Gibbs and Slice Sampling

- Now, let's move to scenarios where we need to sample from a multidimensional probability, e.g., from a 2D Gaussian with some covariance matrix.
- Both Gibbs and Slice sampling work with the same principle: To sample from the joint probability, sample from a conditional probability for each dimension

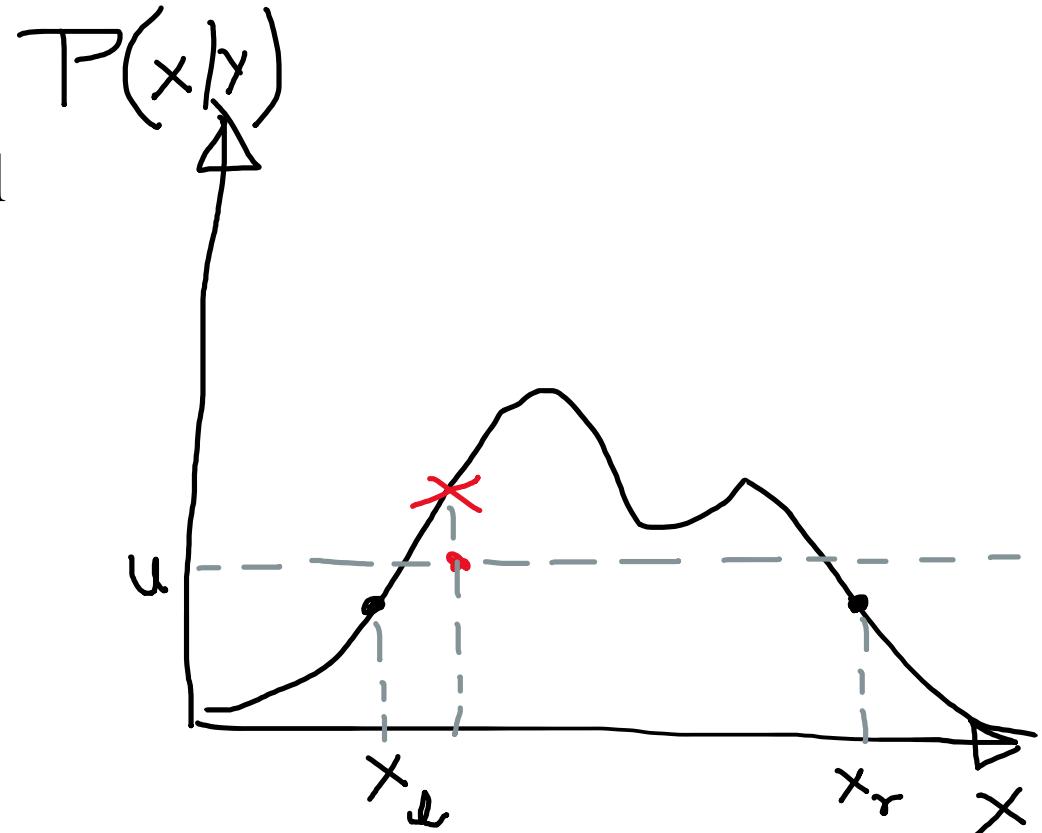


Gibbs Sampling

- If the conditional probabilities are analytical and simple, e.g., Gaussian, you can sample directly from the conditional distribution.
- Algorithm:
 - The point at the t -th time step is $(x^{(t)}, y^{(t)})$.
 - Sample $x^{(t+1)} \sim P(x|y^{(t)})$.
 - Then sample $y^{(t+1)} \sim P(y|x^{(t+1)})$.
 - $(x^{(t+1)}, y^{(t+1)})$ is then the $(t + 1)$ -th point.
- Note that Gibbs sampling is also Markov Chain Monte Carlo method

Slice Sampling

- Slice sampling follows the same algorithm as above, but for slice sampling, the conditional probabilities can be arbitrary. Sample 1D conditional distributions (similar to rejection sampling)
- Algorithm:
 - Evaluate $P(x)$. Draw $u \sim \text{Unif}(0, P(x))$.
 - Create an interval (x_l, x_r) enclosing x . Step out from x with a step of w until $P(x_l/x_r) < u$
 - Loop:
 - Draw $x' \sim \text{Unif}(x_l, x_r)$. If $P(x') > u$, accept x' and break out of the loop
 - Else, modify the interval (x_l, x_r) such that one of them is x' .



Hamiltonian Monte Carlo

- Most of the methods we discussed above does well when there are just a handful parameters to sample from. Sampling from very high dimensional parameter space is difficult for most MCMC algorithms.
- Hamiltonian Monte Carlo (HMC) method used concepts of Hamiltonian dynamics to suppress the random walk behavior of the Markov Chain.
- Main Idea:
 - For each parameter, x_i , introduce an auxiliary momentum variable, p_i .
 - Define a Hamiltonian system with potential with, $\psi(x) = \ln(P(x))$. The Hamiltonian is given by, $H(x, p) = \frac{p^2}{2m} + \psi(x)$. Then the motion of (x_i, p_i) is a Hamiltonian motion.
 - Sample from $\exp(-H(x, p)) = \exp\left(-\frac{p^2}{2m}\right) P(x)$. But since we are only interested in the distribution of x , we look at the marginal distribution of x , (which is $P(x)$).

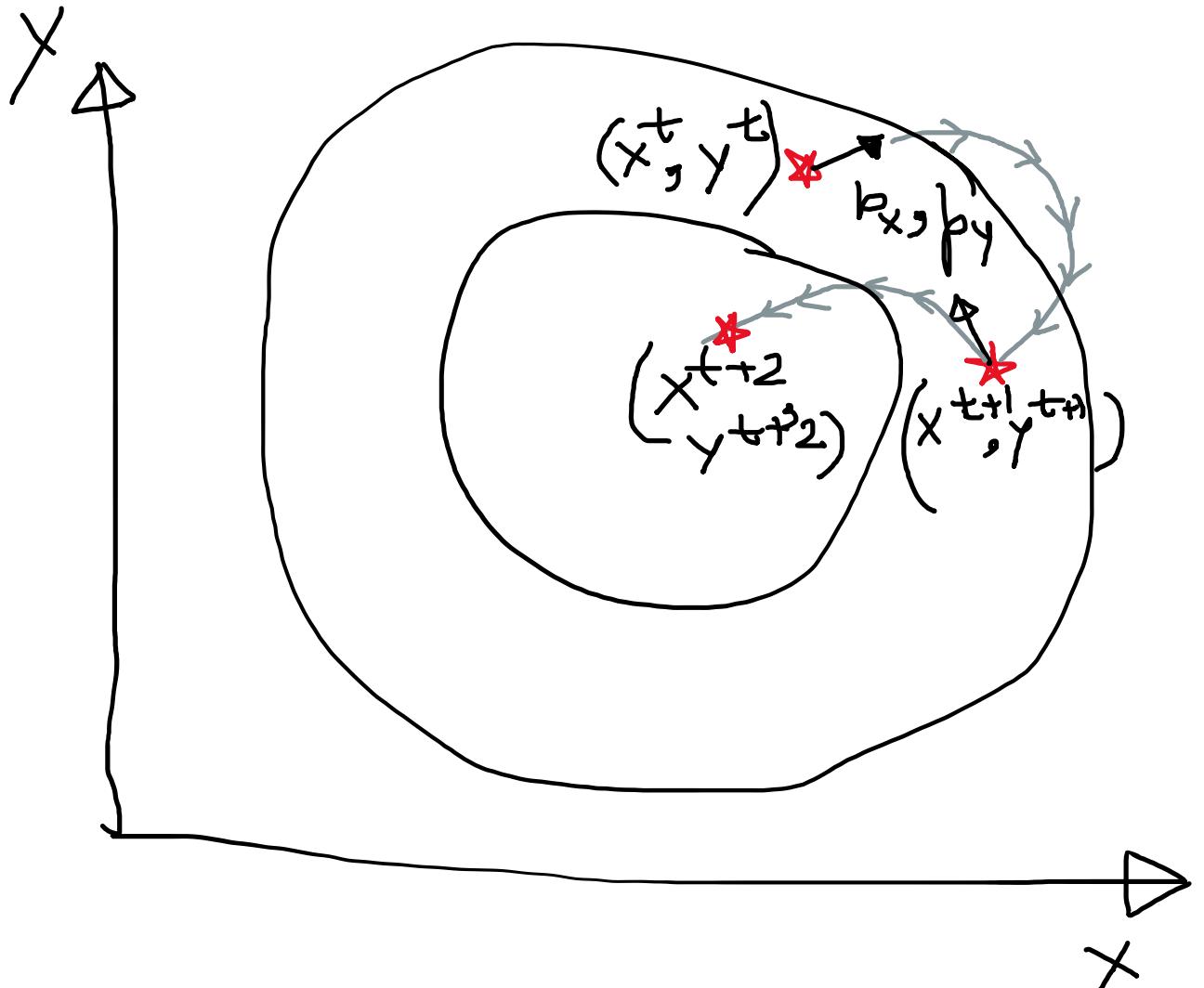
Hamiltonian Monte Carlo

- Algorithm:
 - Sample (x, p) from the $\exp(H(x, p))$ in two steps. First, sample p from a Gaussian $\exp(-p^2/2m)$. The mass matrix, m , plays a role akin to that of step size in MH algorithm and is a hyperparameter.
 - For a given p , follow the trajectory of x according to Hamiltonian equations of motion, $\frac{dx}{dt} = \frac{p}{m}$, $\frac{dp}{dt} = -\frac{d\psi}{dx}$. Therefore, **we need the gradient of the log-probability as well**.
 - This gives a new sample, (x_N, p_N) . Evaluate the Hamiltonian $H(x_N, p_N)$. Accept or reject the new point with probability $\exp(-\Delta H)$. Note that in a perfect Hamiltonian motion, ΔH is zero. But it is non-zero because of numerical error.
 - At each time step, we sample a new momentum variable.

Hamiltonian Monte Carlo

- Points to note:

- We sample a new p at each time step.
- Need to evaluate the gradient of the log-posterior to determine how the momentum changes.
- Typically, we use a leapfrog integration scheme to integrate the Hamiltonian equations of motion.



Hamiltonian Monte Carlo Example: BORG

- Forward modelled initial density reconstruction algorithm **BORG**.
- Reconstruct the density field in the Universe from galaxies in a 256^3 cartesian coordinate grid.
- # of parameters: 256^3 density on a grid.

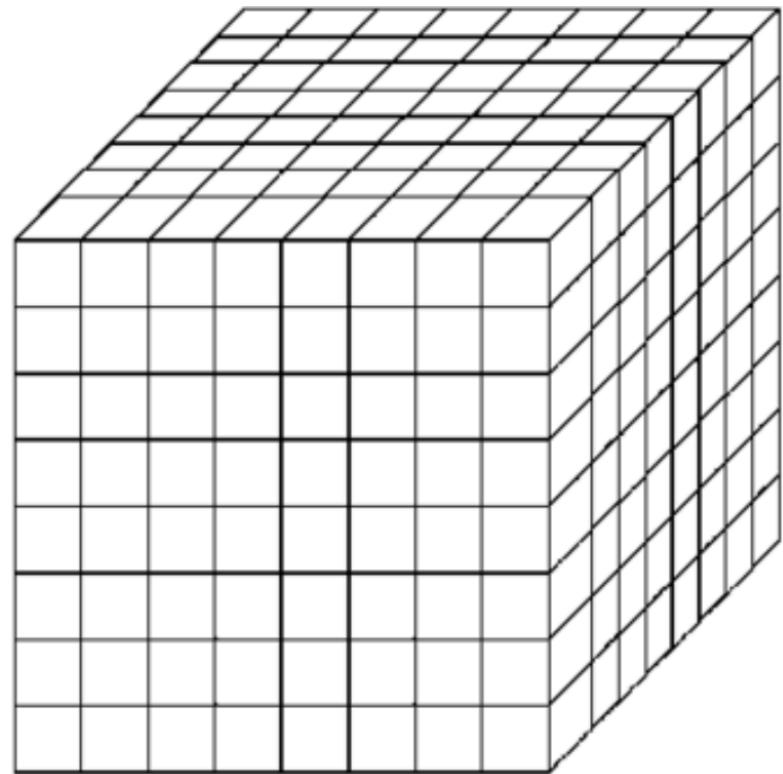
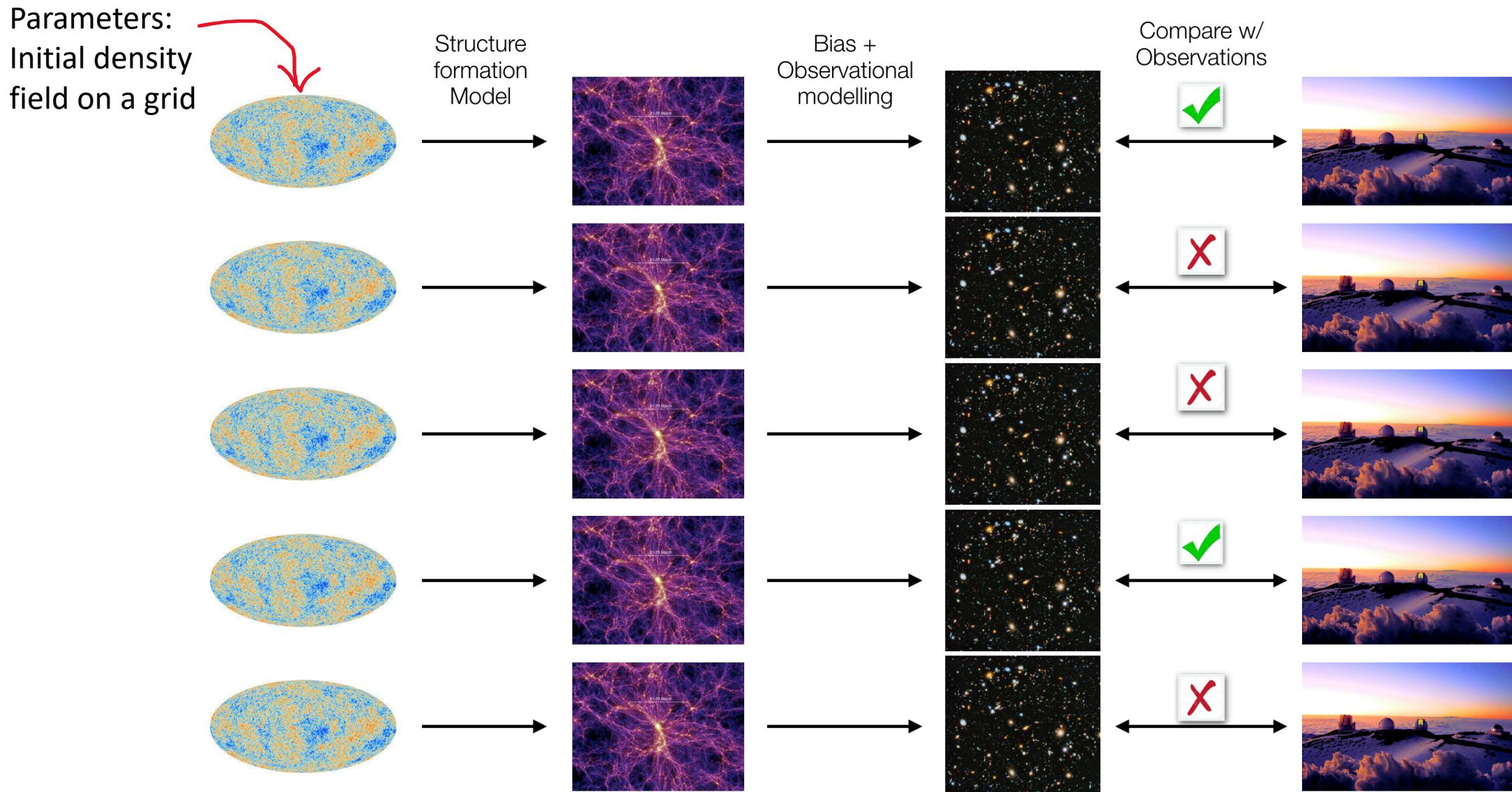


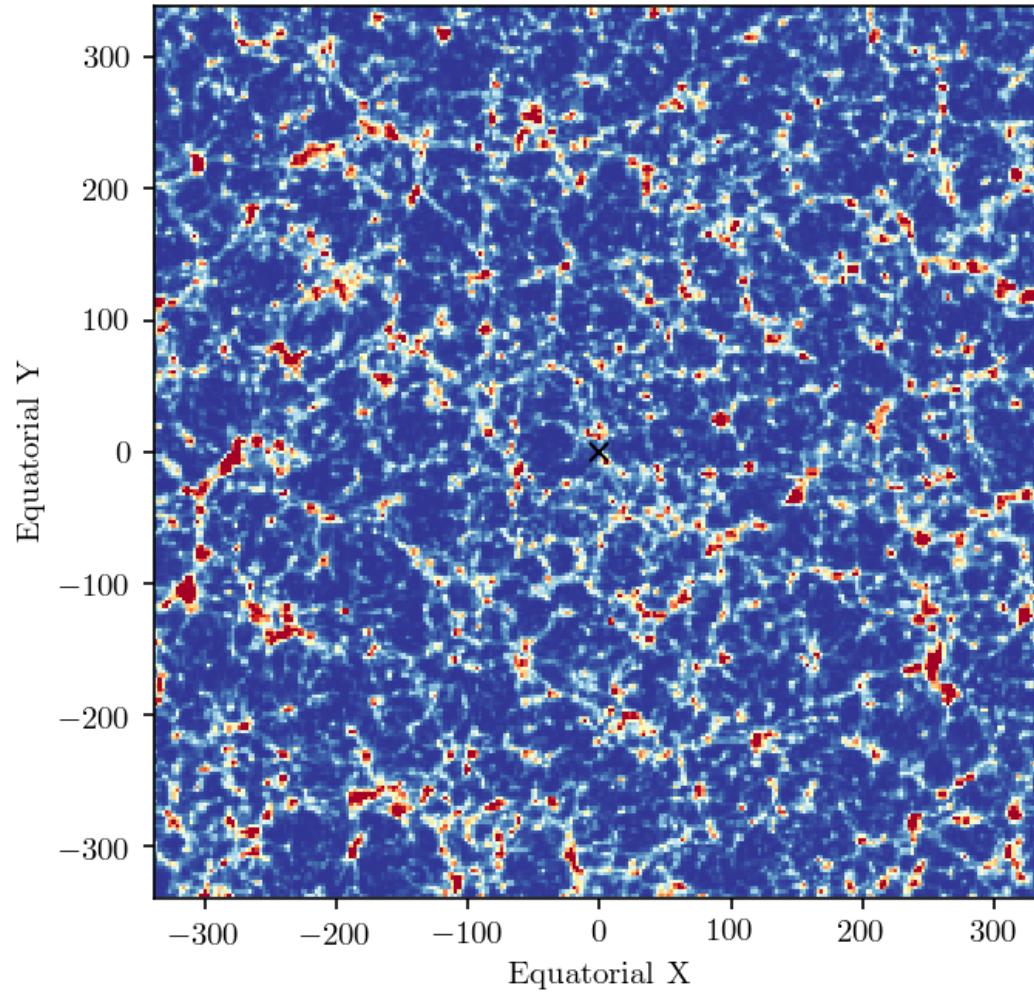
Image Credit: Giatili et. al.

Hamiltonian Monte Carlo



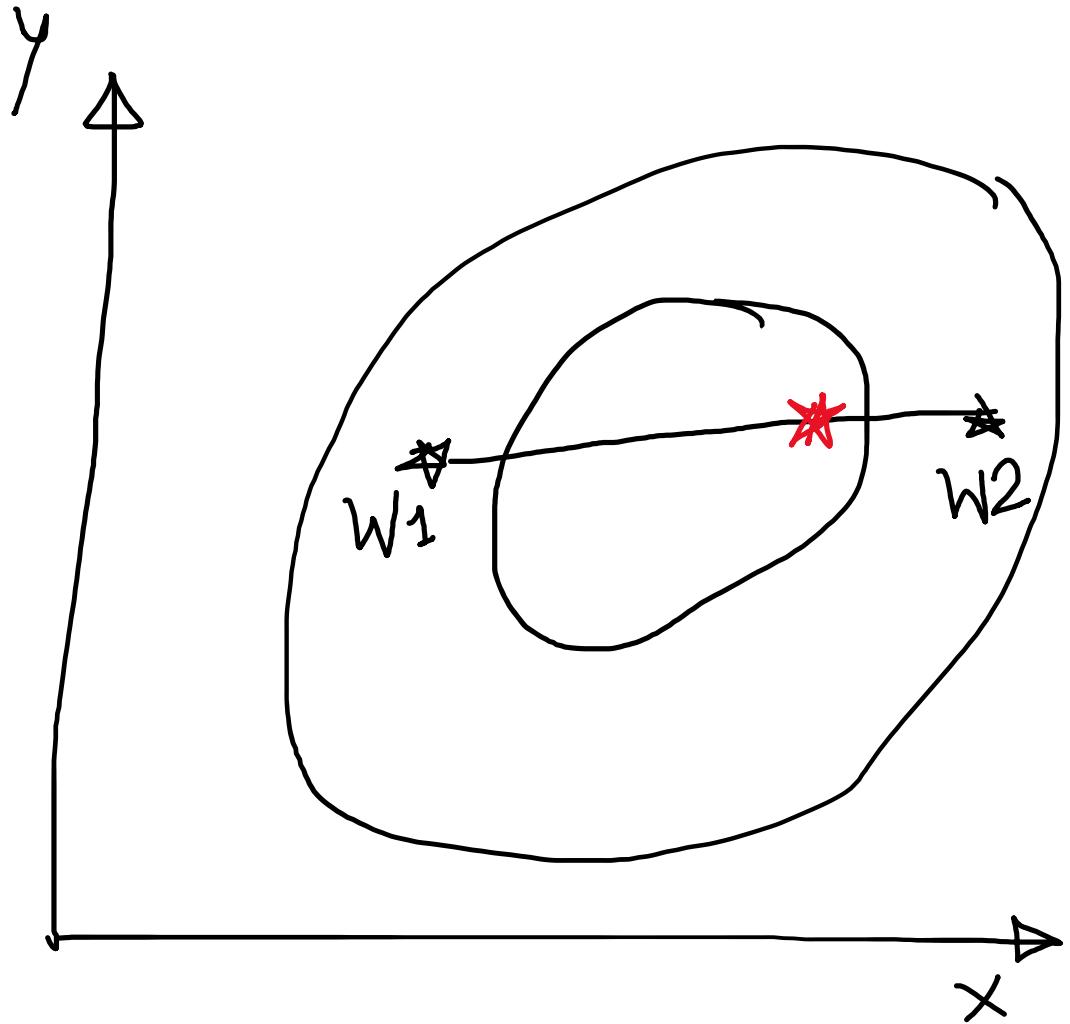
Hamiltonian Monte Carlo

Animation of the final density field



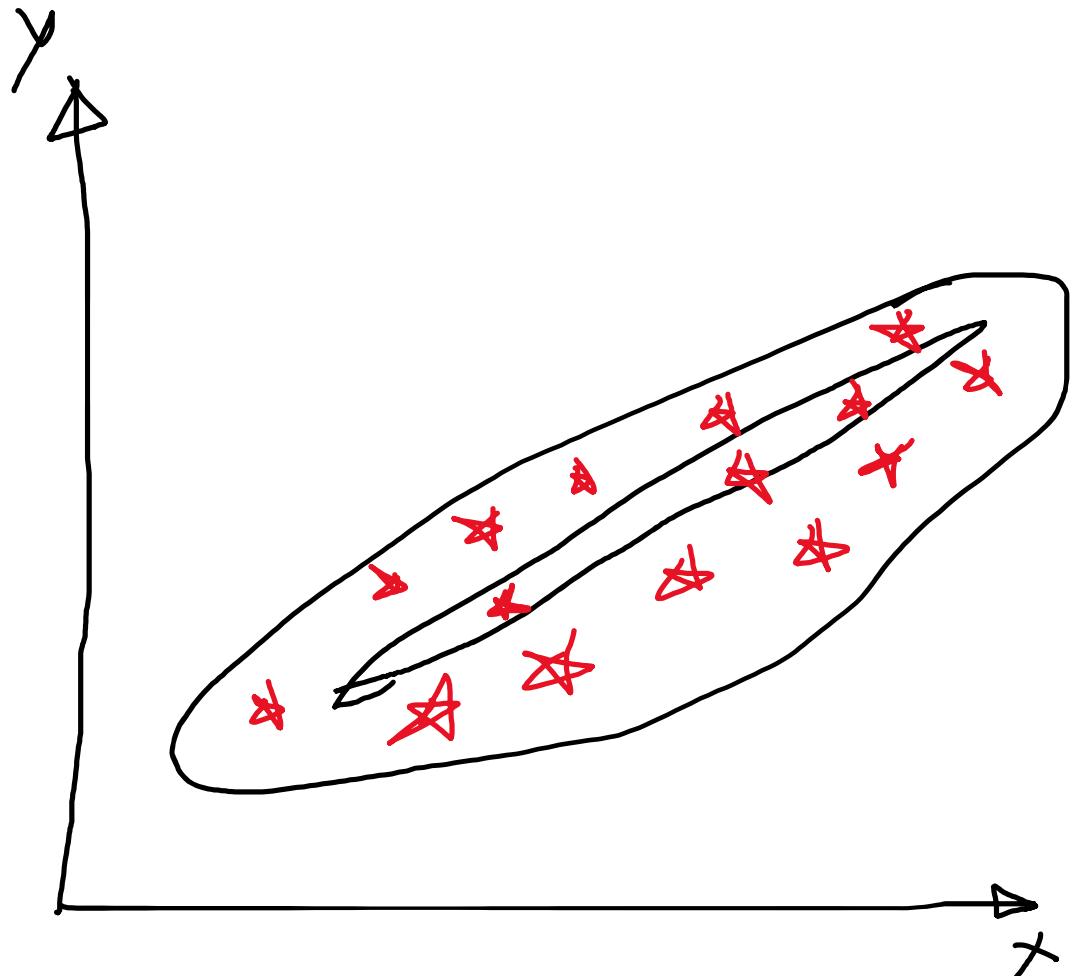
Ensemble Samplers

- Ensemble Samplers have become immensely popular in astrophysical data analysis (mainly because of the package emcee).
- In this algorithm, one initializes multiple walkers. And then proposes new moves for each chain based on the other chains. For the k -th chain, the $(t + 1)$ -th step is given by, $X_k^{(prop)} = X_k^{(t)} + z (X_s^{(t)} - X_k^{(t)})$, where the s -th chain is drawn at random from the complementary set of data, and $z \sim g(z) \propto \frac{1}{\sqrt{z}}$.
- The proposed point is accepted/rejected to maintain detailed balance



Ensemble Samplers

- Ensemble Samplers are particularly useful if one has a high degeneracy in a distribution. One doesn't have to tune / adapt the covariance matrix.
- After burn-in, the walkers fill the posterior volume quickly.
- On average, the proposals are of order, $\mathcal{O}(\sigma)$ and in the direction of the covariance.



Ensemble Samplers

- Example:
- 64 walkers

