Introduction to inverse problems and Markov chain Monte Carlo methods

School of Mathematics, University of Edinburgh

Lorenzo Gabriel Stigliano

24th April 2021

Outline of todays presentation

1. What are inverse problems?
   - Optimization Approach
   - Bayesian Inference Approach
2. Markov chain Monte Carlo Methods (MCMC)
   - Markov chain theory
   - Motivation to use MCMC methods
   - Reversible Samplers
   - Nonreversible Samplers
3. Takeaways

1. What are inverse problems?

- Concerned with determining casual factors from observed data.
- Mathematically: determine function **inputs** based on function **outputs**.

$$y = G(u) + \eta$$

- These problems appear in, for example:
  - Computational imaging

- Why are these problems hard to solve?
  1. We do not know the value of η.
  2. We cannot simply invert G.

1. Example - Image De-noising


Observation $y = u + \eta$


Original image u

- Each pixel $u_{n,m}$ has some additive noise from some $N(0, \sigma^2 I)$.

1.1 Optimization Approach

- Minimize an objective function.

$$u^* = arg\ min_{u^* \in U}\ ||y - G(u)||_2^2$$

- Leads to:
  - **Ill-posed** problem:
    - there could not be a unique solution u^*
    - u^* does not depend continuously on y
  - **Ill-conditioned** problem: small changes in y can lead to large changes in u^*

- We can introduce a regularization term.

$$u^* = arg\ min_{u^* \in U}\ ||y - G(u)||_2^2 + \alpha R(u)$$

- We produce a point estimate.

1.2 Bayesian inference approach

- Probabilistic approach.
- Solution: Probability density function $p(u|y)$.

- This allows for:
  - **uncertainty quantification**.
  - incorporation of **prior** knowledge.

- The inverse problem is **well-posed**:
  - A **unique** posterior distribution.
  - The posterior distribution depends **continuously** on y.

1.2 Bayesian inference approach

- The framework:
  - **Prior probability density** $\pi_{pr}(u)$.

  - **Likelihood** function $L(y|u)$.
    - $L(y|u) \sim N(G(u), \lambda^2 I)$, if we know the noise $\eta \sim N(0, \lambda^2 I)$.

  - **Bayes Theorem** $p(u|y) \propto \mathrm{L(y|u)}\pi_{pr}(u)$

- Point estimates:
  - **Conditional Mean**: $u^* = E[u|y] = \int u\, p(\mathrm{u|y})du$
  - **Maximum A-posterior estimator**: $u^* = arg\ max_{u^* \in U}\ p(u|y)$

- We can estimate CM expression by $E[u|y] \approx \frac{1}{N}\sum_{i=1}^{N} u$ where $u \sim p(\mathrm{u|y})$

2. Markov chain Monte Carlo Methods (MCMC)

$$E[\varphi] = \int \varphi(u)\pi(\mathrm{u})du \textbf{\textit{ is estimated by }} \frac{1}{N}\sum_{i=1}^{N} \varphi(u) \text{ where } u \sim \pi$$

Some definitions…

- **MCMC estimator**: $E_N^{MCMC} = \frac{1}{N}\sum_{i=1}^{N} \varphi(u_i)$ where $\{u_i\}_{i=1}^{N}$ is a Markov chain
- **Markov property :** A sequence of random variables $\{u_i\}_{i=1}^{\infty}$ is a Markov chain if:

$$P[u_i = x_i | u_{i-1} = x_{i-1}, \ldots, u_1 = x_1] = P[u_i = x_i | u_{i-1} = x_{i-1}]$$

- **Stationary distribution** π:

$$\pi(x) = \int_{\Omega} \pi(x)k(x,y)\,dy$$

2.1. Why use MCMC Methods

- We do **not** need to know the explicit form of the target distribution.
- Linking back… $p(u|y) \propto \mathrm{L(y|u)}\pi_{pr}(u)$.

From Markov chain theory:

1. **Stationary distribution**: we can create a chain $\{u_i\}_{i=1}^{\infty}$ such that $u_i \sim \pi$ as $i \to \infty$
2. The **Ergodic average** $\frac{1}{N}\sum_{i=1}^{N}\varphi(u_i) \to E[\varphi]$ as N → ∞

Central limit theorem:

- The estimator $\frac{1}{N}\sum_{i=1}^{N}\varphi(u_i) \to N(E[\varphi], \frac{\sigma^2}{N})$
- The **asymptotic variance**: $\sigma^2 = Var[\varphi(u_1)] + 2\sum_{i=1}^{\infty} Cov[\varphi(u_1), \varphi(u_i)]$

2.3. Implementing MCMC – Reversible Samplers

- **Reversible chain**: $\pi(y)k(y,x) = \pi(x)k(x,y), \forall x, y \in \Omega$

Metropolis Hastings Algorithm:

1. Choose u_0 with $\pi(u_0) > 0$

Loop:

2. At state u_i , sample a proposal u' from density $q(u'|u_i)$
3. Accept sample u' with probability:

$$\alpha(u'|u_i) = \min\left(1, \frac{\pi(u')q(u_i|u')}{\pi(u_i)\, q(u'|u_i)}\right)$$

- We accept a proposal sample u', when $\frac{\pi(u')}{\pi(u_i)}$ and $\frac{q(u_i|u')}{q(u'|u_i)}$ are large.
- We need to properly tune the parameters in $q(u'|u_i)$

2.4. Example – Reversible Sampler

- **Target distribution**: $\pi = N(10,2)$
- **Algorithm**: Metropolis Hastings algorithm with a random walk proposal density.
- **Random walk proposal**: $q(u'|u_i) \sim \; N(u_i, \beta^2 I)$; $u' = u_i + \beta\varepsilon \; where \; \varepsilon \sim N(0, I)$.

- We are interested at the **rate** at which the autocorrelation function decays.
- **Autocorrelation function :** $\delta(\tau) = \frac{cov(u_i, u_{i+\tau})}{cov(u_1, u_1)}$

1. Importance of tuning β.

- Chains ran for 4000 iterations using small and large β.

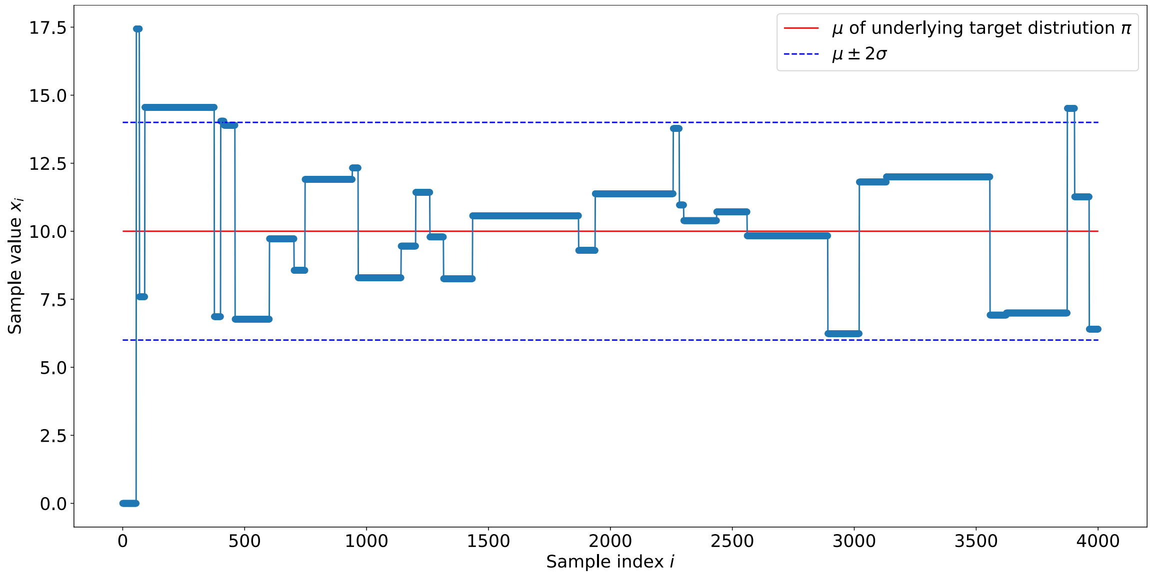
- *Top figure*:

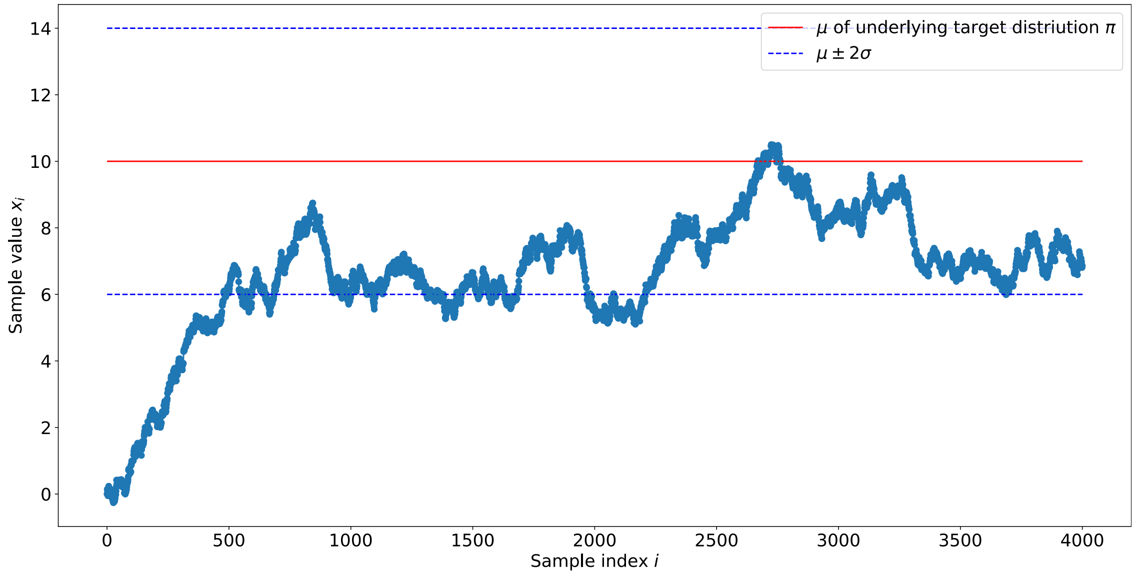
  $$\beta = 20.$$

- *Bottom figure*:

  $$\beta = 0.3.$$

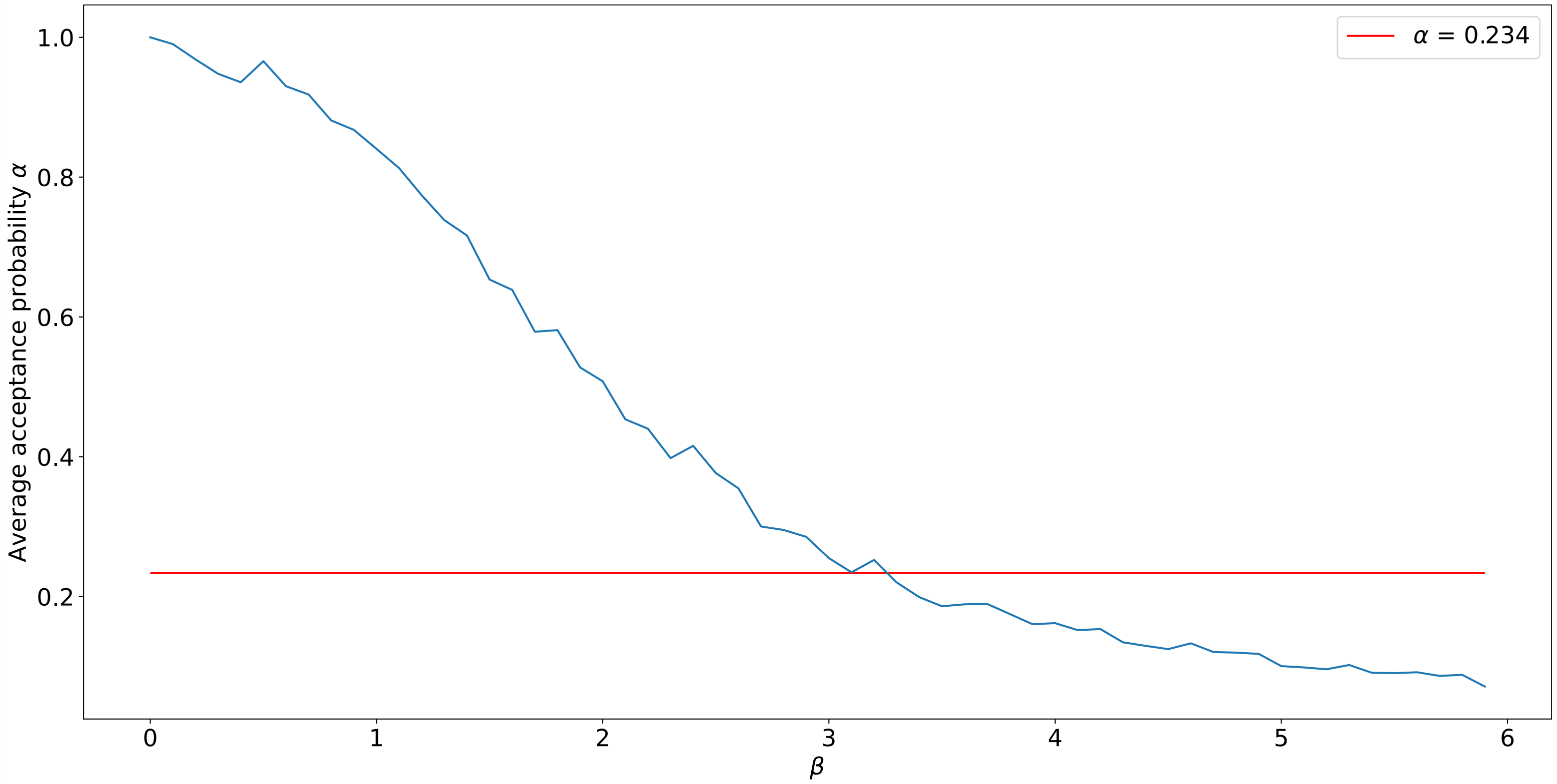
- $\alpha(u'|u_i) = \min\left(1, \frac{\pi(u')q(u_i|u')}{\pi(u_i)q(u'|u_i)}\right)$





2. How to tune β.

- Tuning β with such that average acceptance probability ≈ 0.234.
- This is rule of thumb is for random walk proposals.



2. How to tune β.

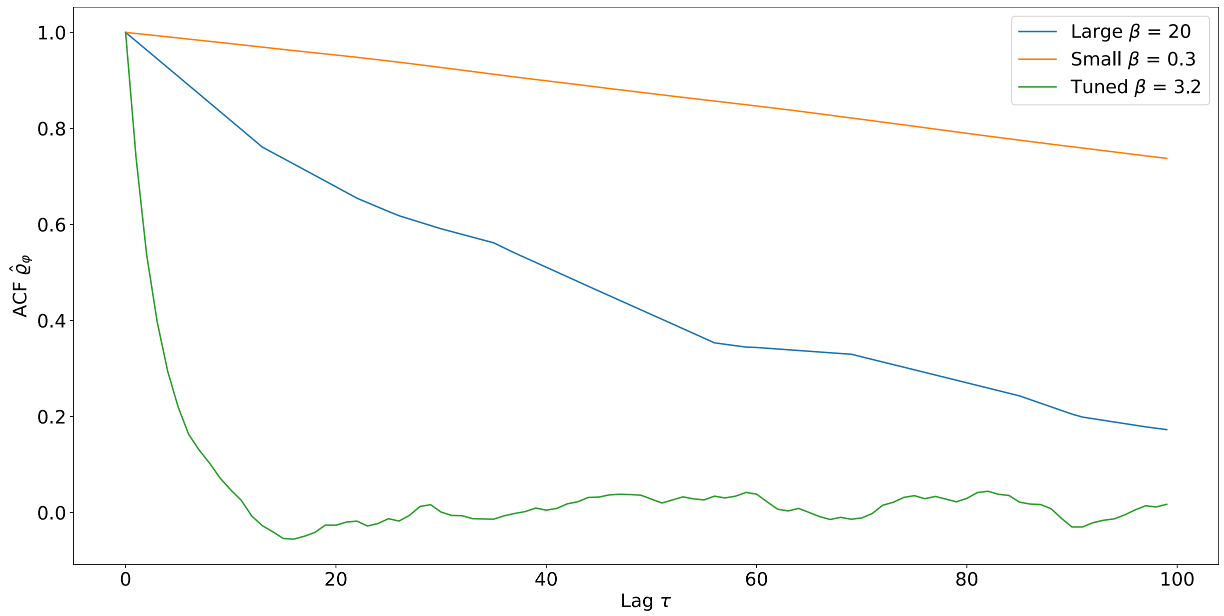
- *Top graph* :

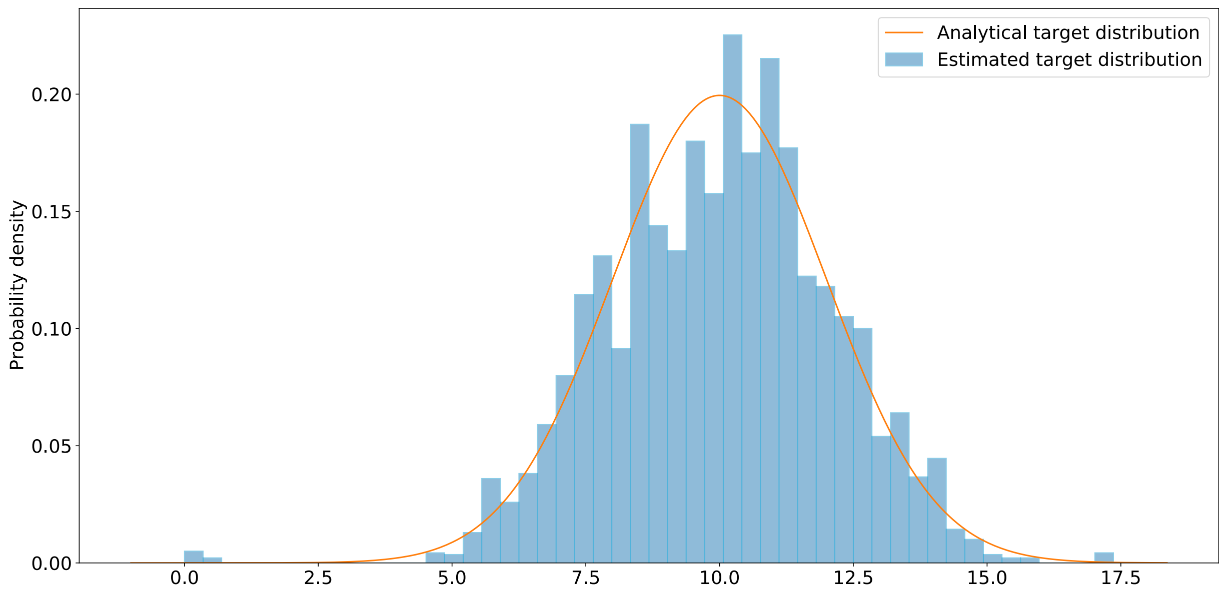
  Autocorrelation function against **lag.**

- *Bottom graph*:

  Estimated target distribution.

  4000 iterations, $\beta = 3.2$.





3. Problems with small β.

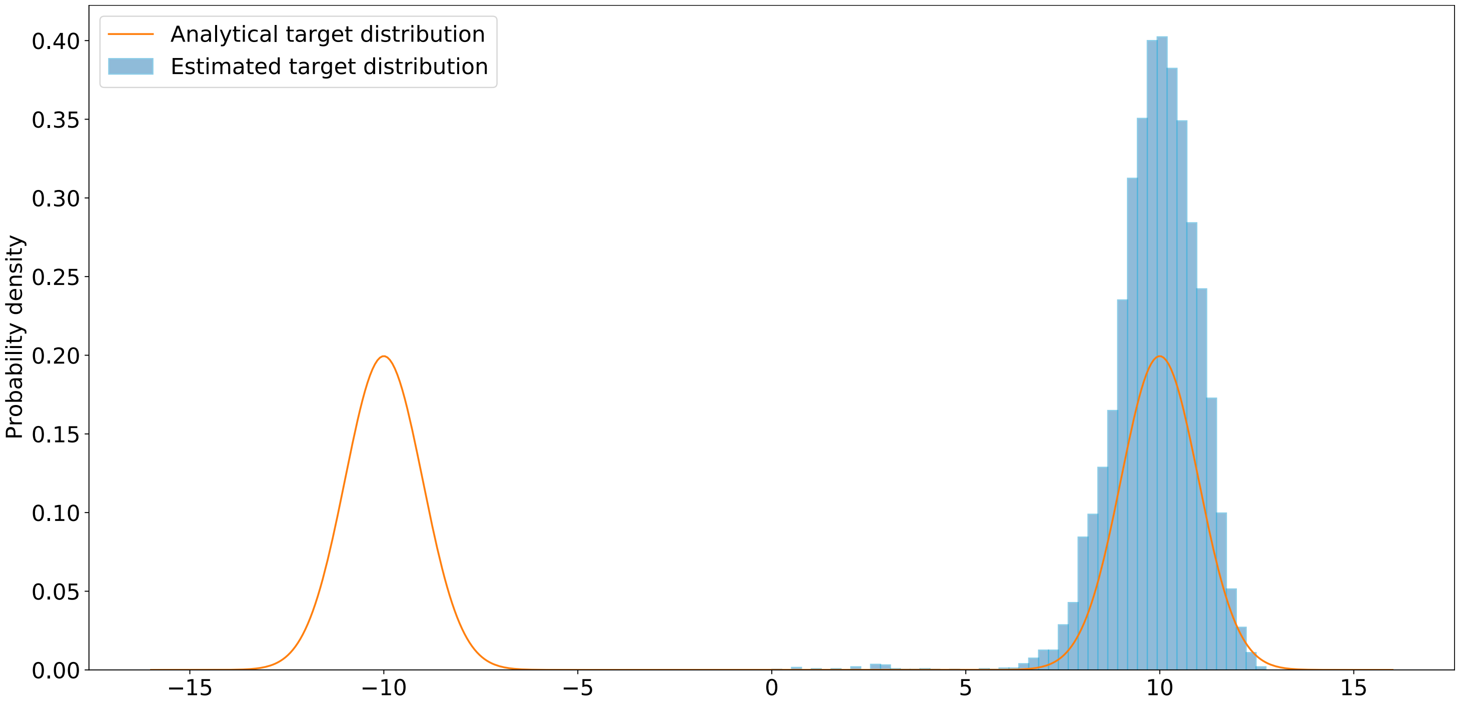
- These distributions are given by:

$$p(x) = \sum_{i=1}^{K} w_i N(x|\mu_i, \sigma_i)$$

- **Target distribution**:

$$\pi(x) = \frac{1}{2}N(x|-10,1)+\frac{1}{2}N(x|10,1)$$

- $\alpha(u'|u_i) = \min\left(1, \frac{\pi(u')q(u_i|u')}{\pi(u_i)q(u'|u_i)}\right)$



2.5. Challenges - Reversible Samplers

1. How do we determine that we have reached our desired stationary distribution?

2. The samples are usually correlated.

3. Which proposal density do we choose?

2.6. Non-Reversible Samplers

- Non-reversible samplers create **non-reversible** chains.

- They are useful since they have the same properties as reversible chains:
  1. **Stationary distribution**
  2. The **Ergodic average**

Advantages over reversible samplers:

- **Faster convergence to equilibrium**.
- **Reduction of asymptotic variance**.

2.7. Implementing MCMC – I-Jump Sampler

- The **I-Jump sampler** is a **nonreversible** sampler.
- Like the MH algorithm but with **two** key differences.

1. Use of two one-sided proposal densities.
    - $f(u'|u_i): u' = u_i + \xi$
    - $g(u'|u_i): u' = u_i - \xi$
    - $\xi \sim \Gamma(a, b)$.
2. Introduction of an indicator variable $\boldsymbol{z}$.

I-Jump Sampler:

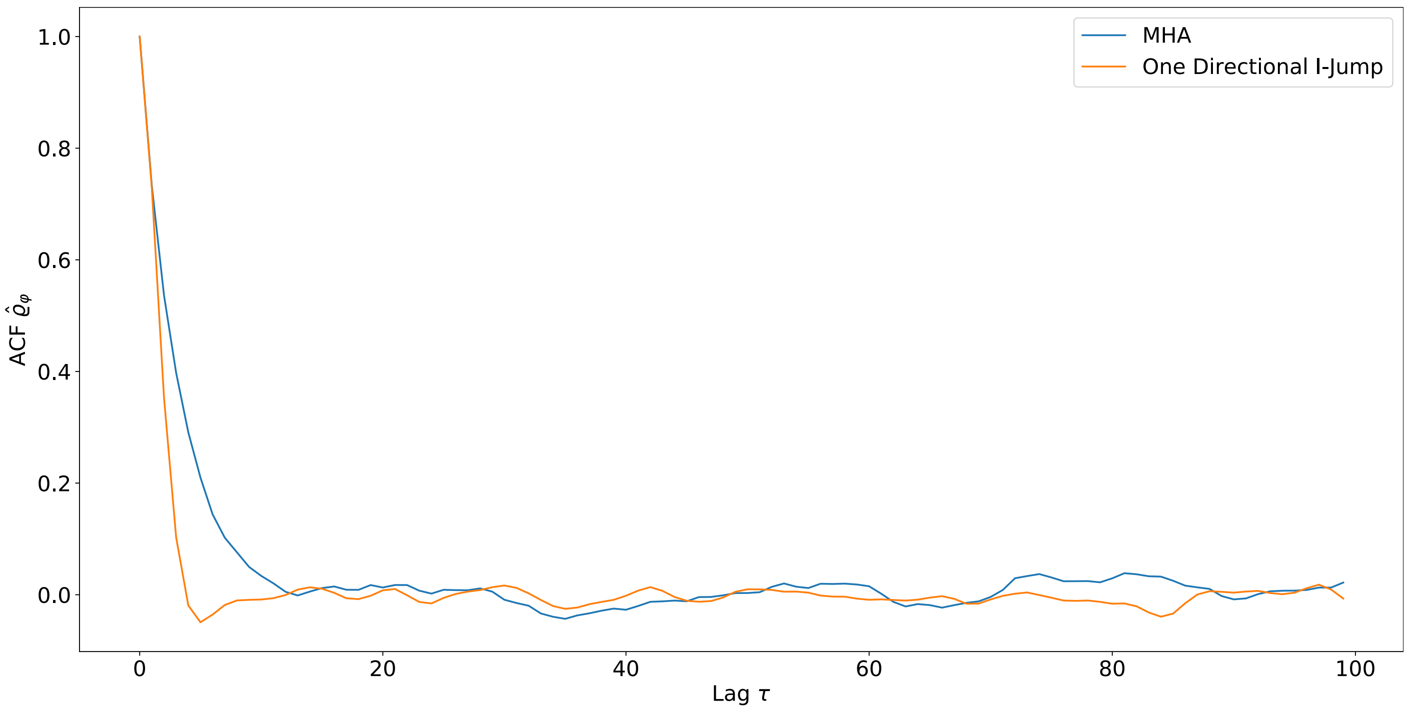
1. Choose u_0 with $\pi(u_0) > 0$
2. Randomly pick $\boldsymbol{z} \in \{-1,1\}$

Loop:

3. If $\boldsymbol{z} > 0$:
4. At state u_i, sample u' from $f(u'|u_i)$
5. $\alpha(u'|u_i) = \min\left(1, \frac{\pi(u')g(u_i|u')}{\pi(u_i)f(u'|u_i)}\right)$
6. else:
7. At state u_i, sample u' from $g(u'|u_i)$
8. $\alpha(u'|u_i) = \min\left(1, \frac{\pi(u')f(u_i|u')}{\pi(u_i)g(u'|u_i)}\right)$
9. Accept sample u' with probability $\alpha(u'|u_i)$
10. if we reject : $-\boldsymbol{z}$

1. Comparing against MH algorithm.

- **Target distribution**: $\pi = N(10,2)$
- **Autocorrelation function** against **lag**.
- MH algorithm the **same** one we tuned before.
- Here $a = 1,\ b = 0.4$.



3. What we can takeaway from this presentation

1. Inverse problems are **hard** to solve.
2. Why we would like to use Bayesian inference approach.
3. The **need** and **advantages** of using of MCMC methods.
4. The **importance** of appropriately **tuning** MCMC parameters.
5. The **benefits** of using nonreversible samplers over reversible samplers.

Thank you for listening.
Any questions?