Astrostatistics: Fri 26 Feb 2020

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2020

- Today: continue Bayesian computation / MCMC
    - MacKay: Ch 29-30; Bishop: Ch 11; Gelman BDA
    - Givens & Hoeting. "Computational Statistics"
      (Free online through Cambridge Library iDiscover)
- Lecture MCMC code examples online: lecture_codes/
- Example Class 3: currently Thu, 5 March 3:30pm —> Move to:
    - Thursday, 12 March, 3:30pm (room?)
    - Friday, 13 Mar, 12pm, MR5
    - Wed, 11 Mar, pm (MR11)?

Overview of
Markov Chain Monte Carlo (MCMC)

- Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings algorithm
  - Gibbs sampling
  - Metropolis-within-Gibbs
- Comparing performance of MCMC algorithms
  - Autocorrelation time
  - Effective Sample Size
- Gibbs Sampling as Metropolis-Hastings
- Detailed Balance & theoretical considerations

Metropolis-within-Gibbs

$$\boldsymbol{\theta} = (\theta_1 \ldots \theta_d) \qquad\qquad \boldsymbol{\theta}_{-j} \equiv (\theta_1, \ldots \theta_{j-1}, \theta_{j+1} \ldots \theta_d)$$

- When you can't solve for tractable conditional distributions for all θ_j: $P(\theta_j | \boldsymbol{\theta}_{-j}, \boldsymbol{D})$
- Replace each substep for updating each jth parameter θ_j with a separate Metropolis rule, compute Metropolis ratio, and accept/reject
- Cycle through all parameters, and repeat all for N MCMC steps

d-dim Metropolis-within-Gibbs Sampler

$$\boldsymbol{\theta} = (\theta_1 \ldots \theta_d) \qquad \boldsymbol{\theta}^t_{-j} \equiv (\theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^t, \ldots, \theta_d^t)$$

1. Choose a random starting point θ_0

2. At cycle t = 1…N, cycle through the d-parameters:

   A. For each j = 1… d, propose a new j-th parameter value from a 1-Dimensional Gaussian: $\theta_j^* \sim N(\theta_j^t, \tau_j^2)$

   B. Evaluate ratio of posteriors at proposed vs current values:

   $$\begin{aligned} r &= P(\theta_j^*, \boldsymbol{\theta}^t_{-j} \,|\, \boldsymbol{D}) / P(\theta_j^t, \boldsymbol{\theta}^t_{-j} |, \boldsymbol{D}) \\ &= P(\theta_j^* \,|\, \boldsymbol{\theta}^t_{-j}, \boldsymbol{D}) / P(\theta_j^t \,|\, \boldsymbol{\theta}^t_{-j}, \boldsymbol{D}) \end{aligned}$$

   C. Accept $\theta_j^{t+1} = \theta_j^*$ with prob min(r,1), otherwise $\theta_j^{t+1} = \theta_j^t$

3. After full cycle, record current values $\boldsymbol{\theta}^{t+1}$

4. Repeat steps 2 for all parameters until convergence and enough samples

Metropolis-within-Gibbs Sampling: Example: Gelman BDA Section 11.1)

Likelihood: $$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \qquad \rho \text{ known}$$

Priors: $$P(\theta_1) = P(\theta_2) \propto 1$$

Posterior: $$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Big| \boldsymbol{y} \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Suppose we can't solve for the Conditional Posteriors.

Metropolis-within-Gibbs Sampling: Code Demo metropolisgibbs_example.m

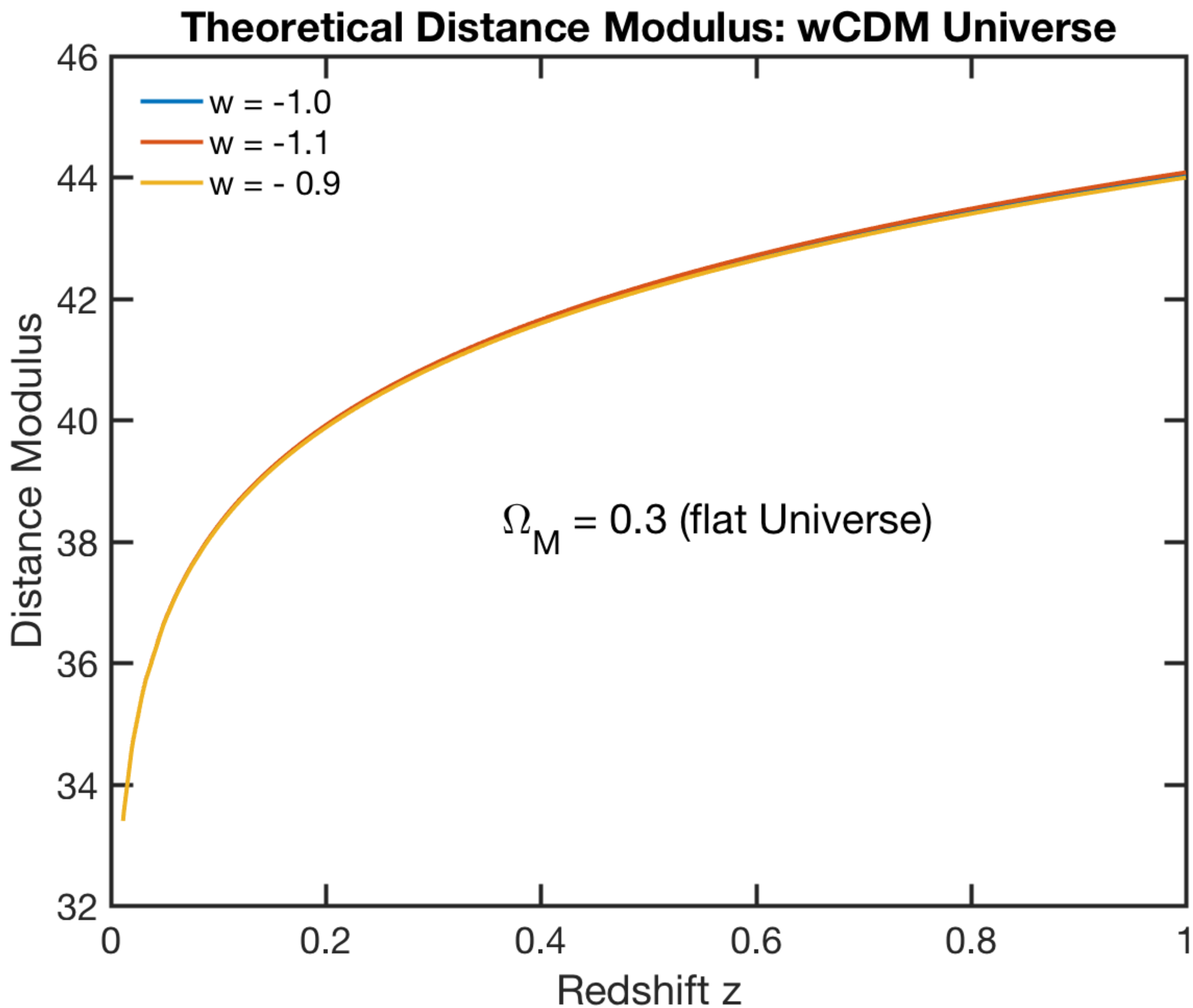
demo different ρ's

Supernova Cosmology Case Study:

Now assume flat Universe

$$\Omega_L = 1 - \Omega_M$$

but unknown $\mathcal{M}_0, \sigma^2_{\text{int}}, \Omega_M, w$

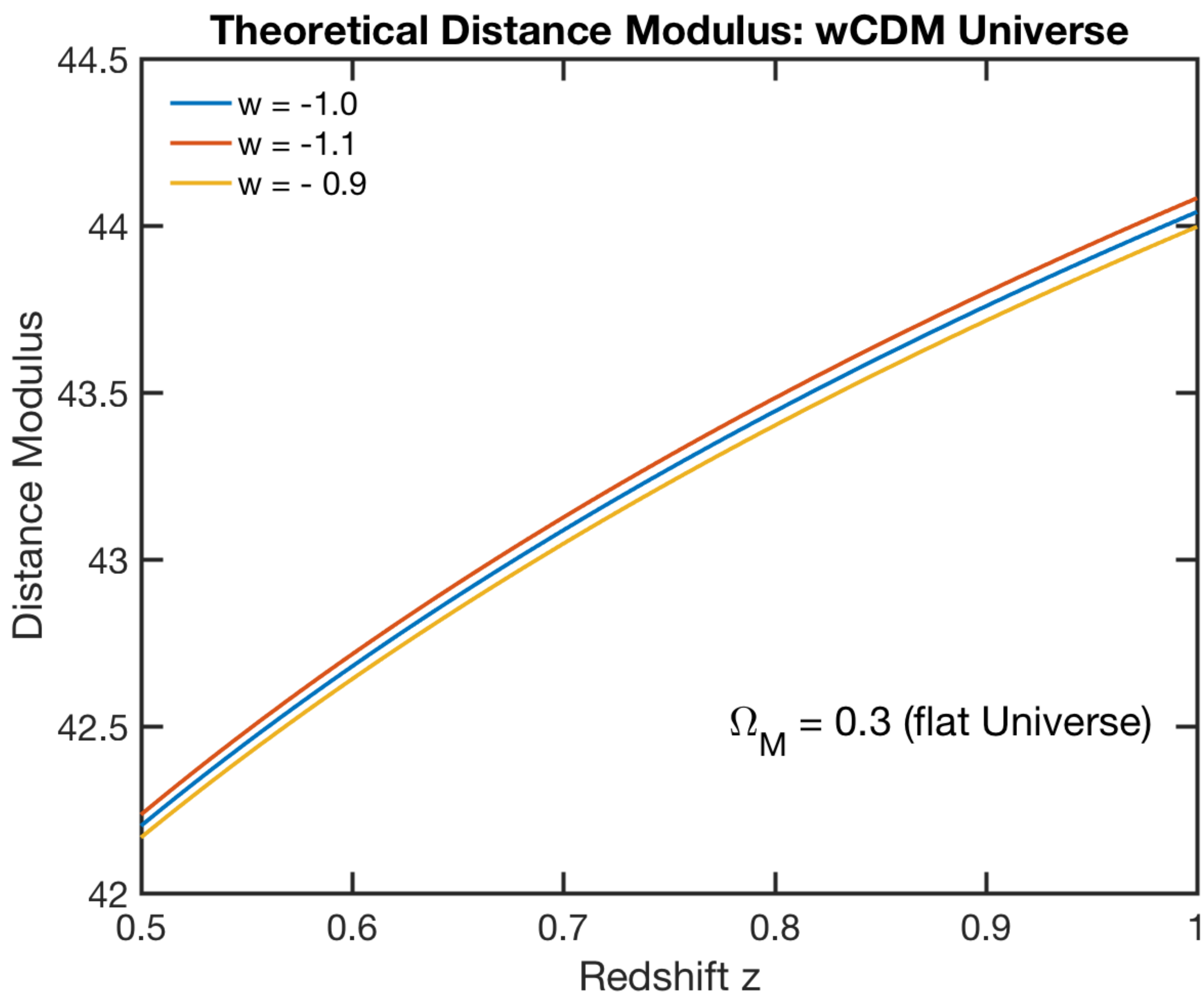


Supernova Cosmology Case Study:

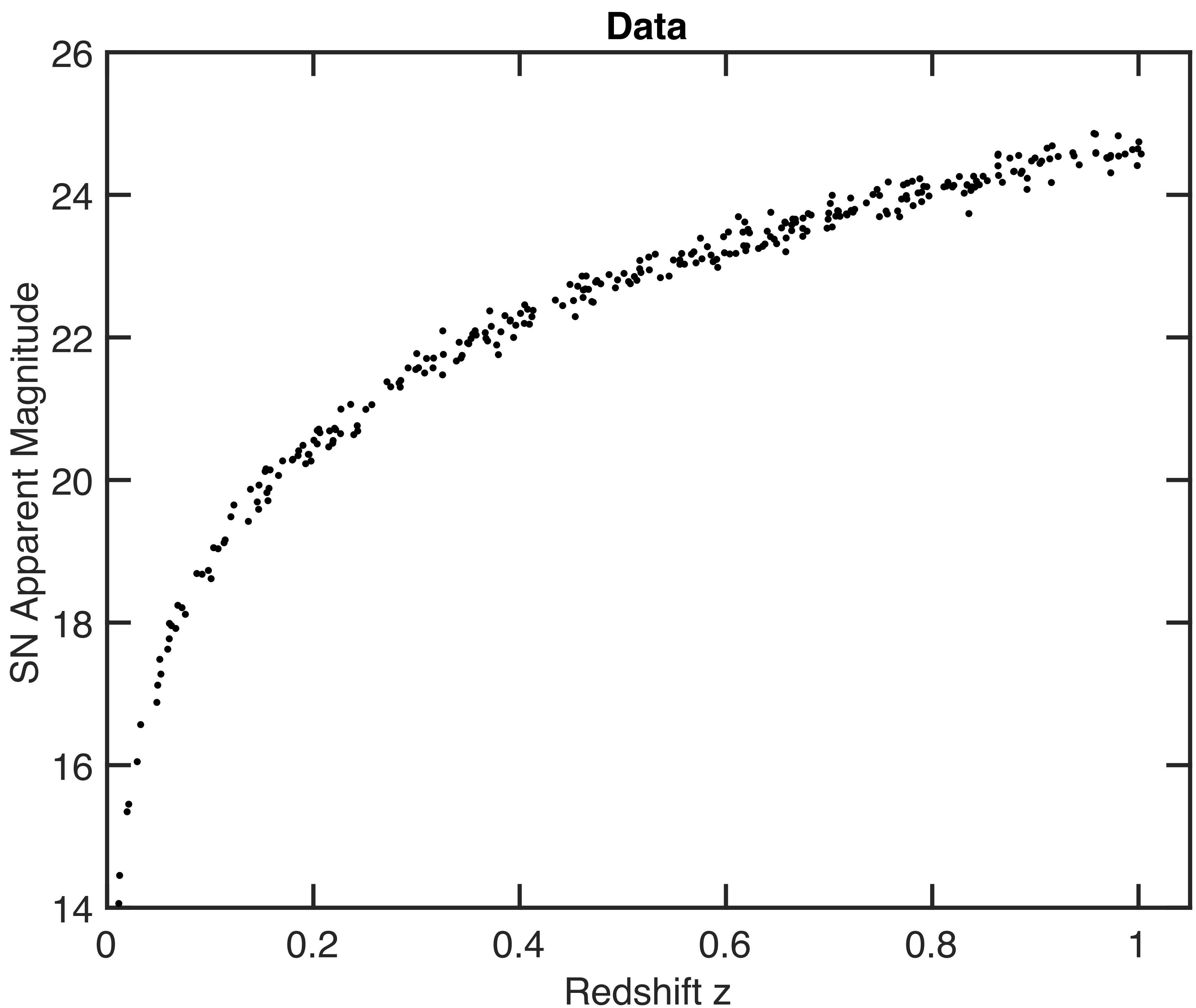
Now assume flat Universe

$$\Omega_L = 1 - \Omega_M$$

but unknown $\mathcal{M}_0, \sigma^2_{\rm int}, \Omega_M, w$

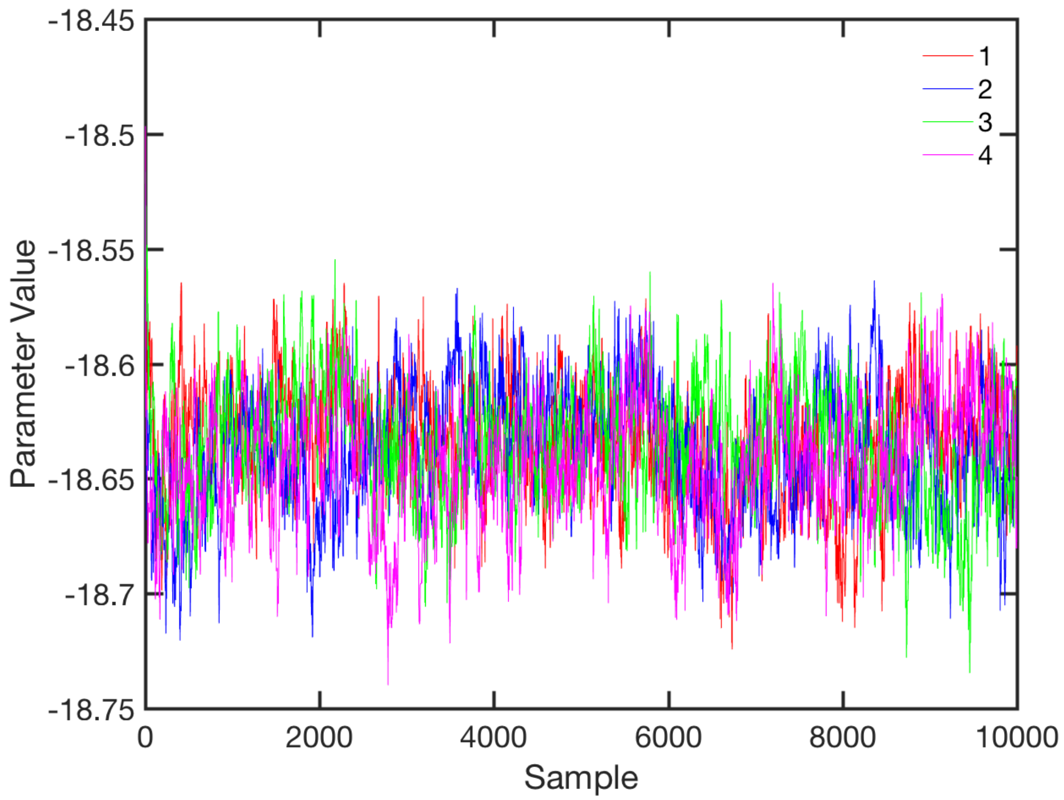


Idealised supernova dataset

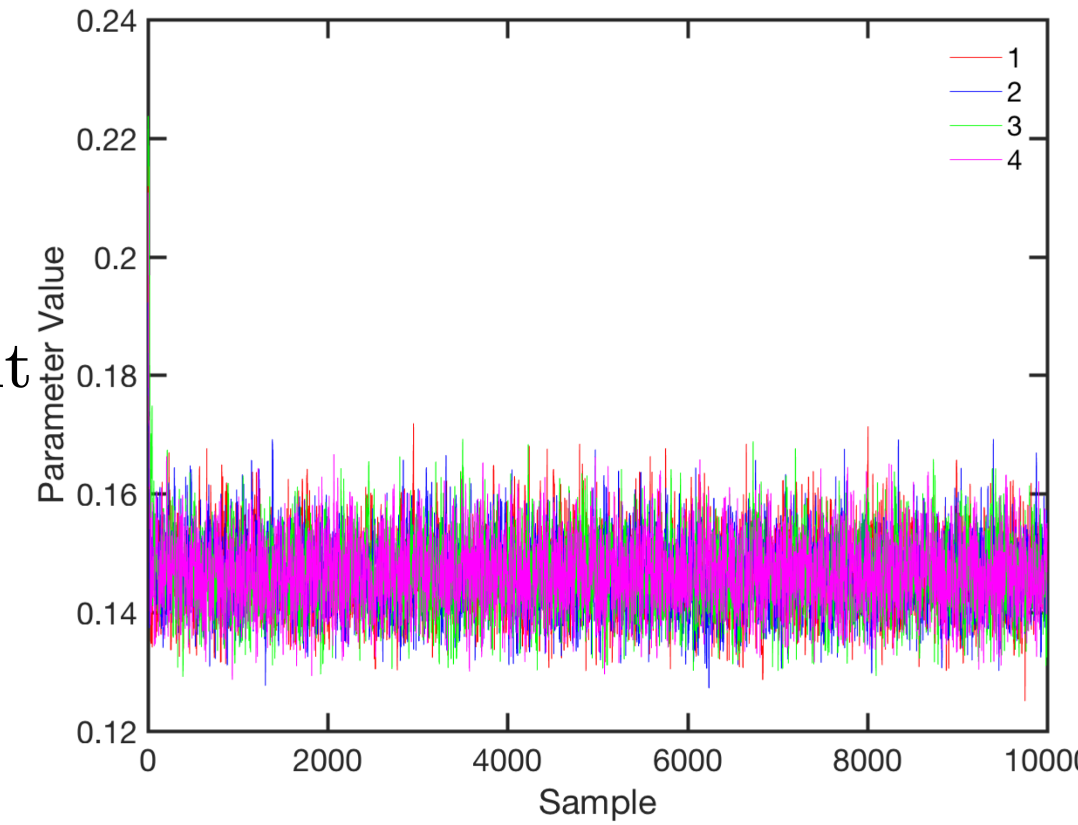


4x1D Metropolis-within-Gibbs: Trace paths 10k cycles, 4 chains

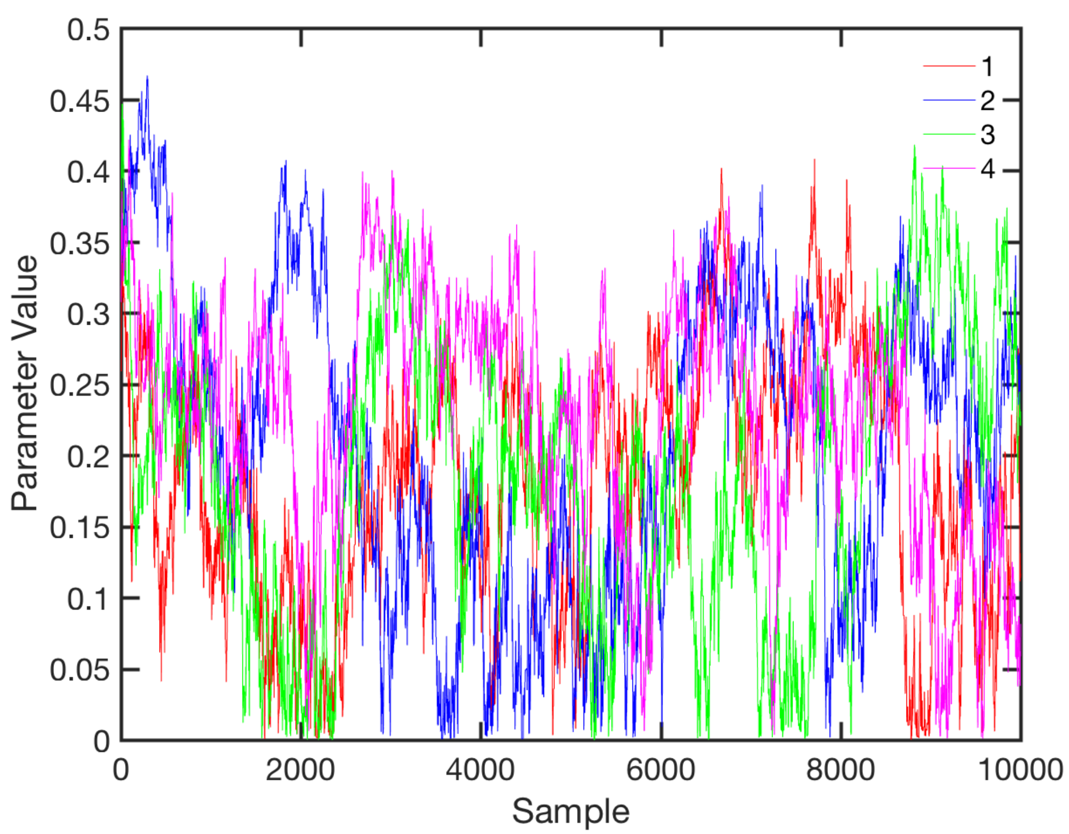
$\mathcal{M}_0$



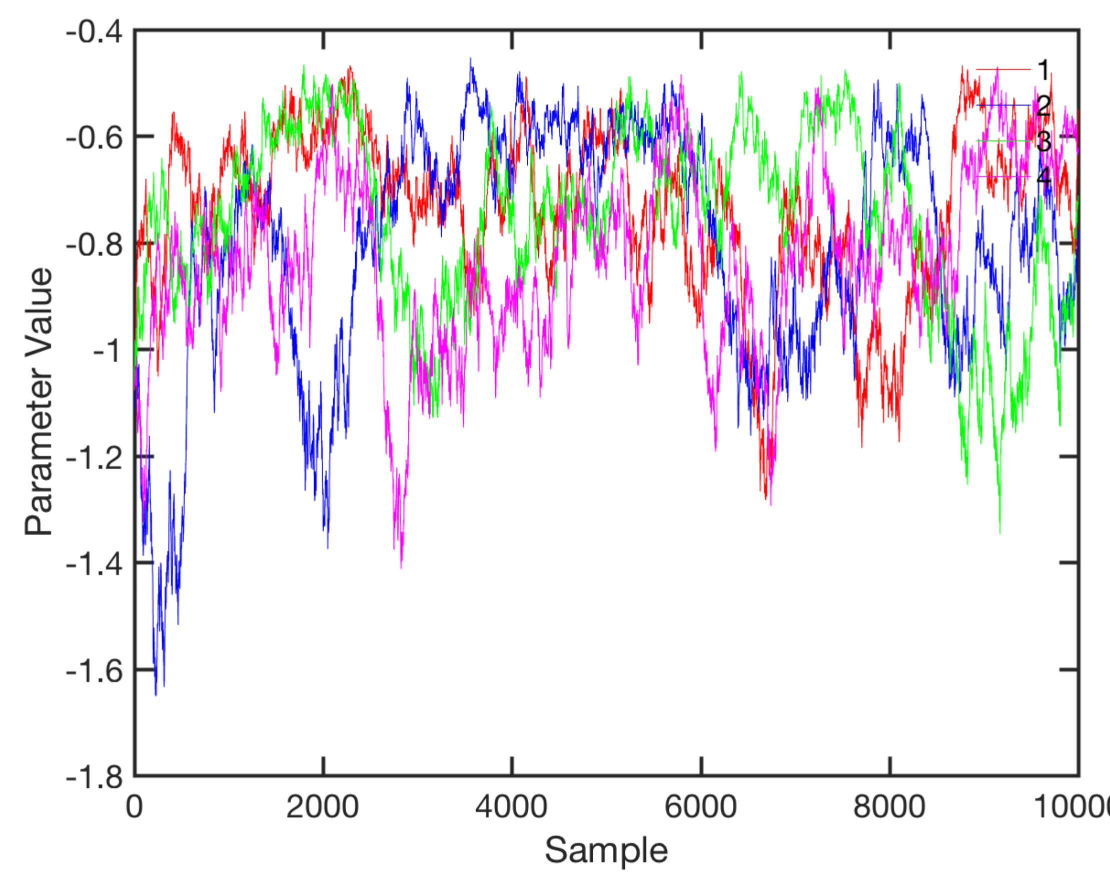
σ_{int}



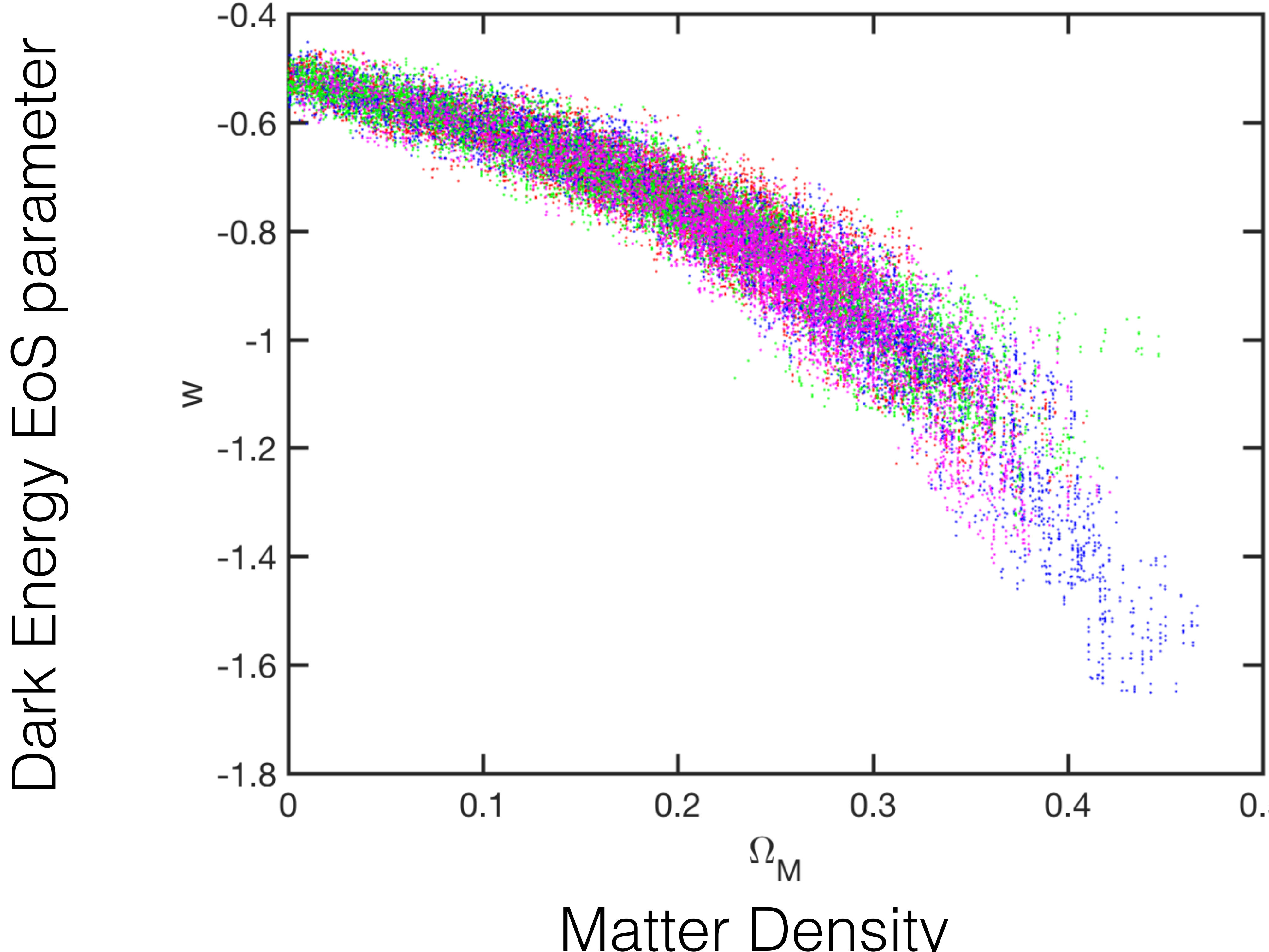
Ω_M



w



Metropolis-within-Gibbs: 2D trace plot 10,000 steps, 4 chain



Highly correlated parameters

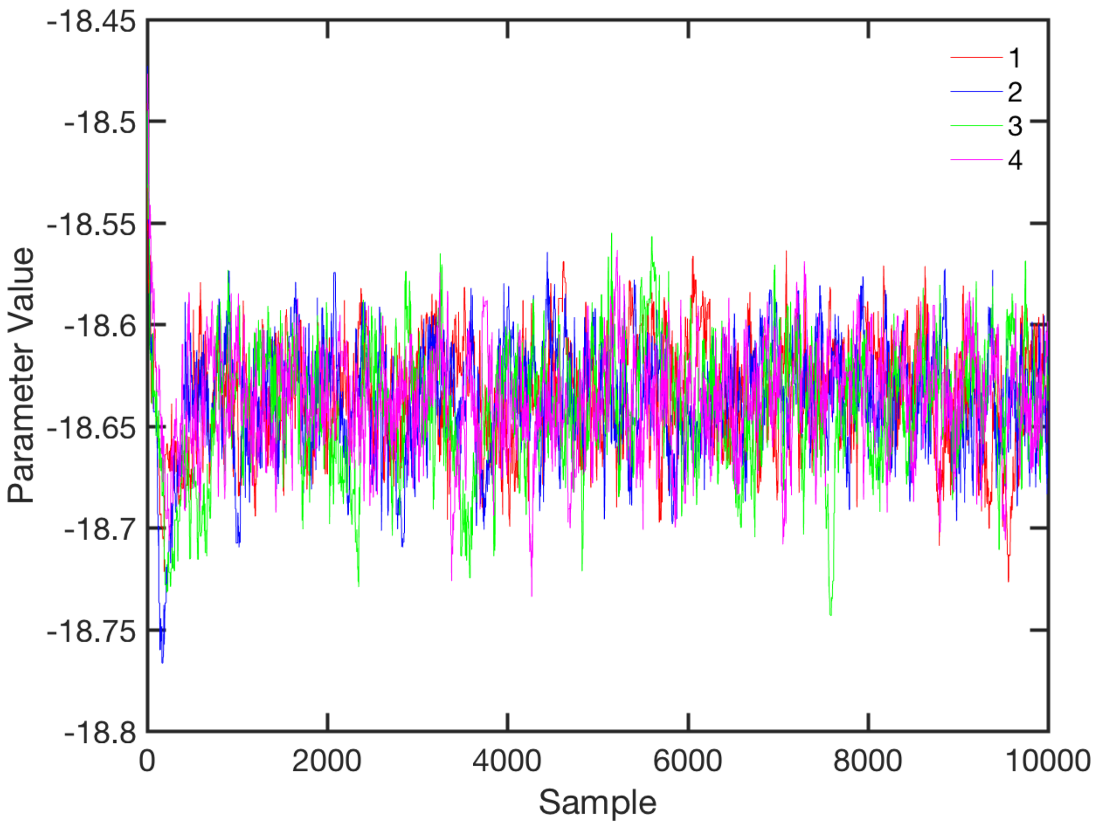
```
>> std(mc)

ans =

    0.0249    0.0061    0.0953    0.1742

>> corr(mc)

ans =

    1.0000   -0.0132   -0.4534    0.6653
   -0.0132    1.0000    0.0290   -0.0282
   -0.4534    0.0290    1.0000   -0.9337
    0.6653   -0.0282   -0.9337    1.0000
```

Strong anti-correlation btw Ω_M, w

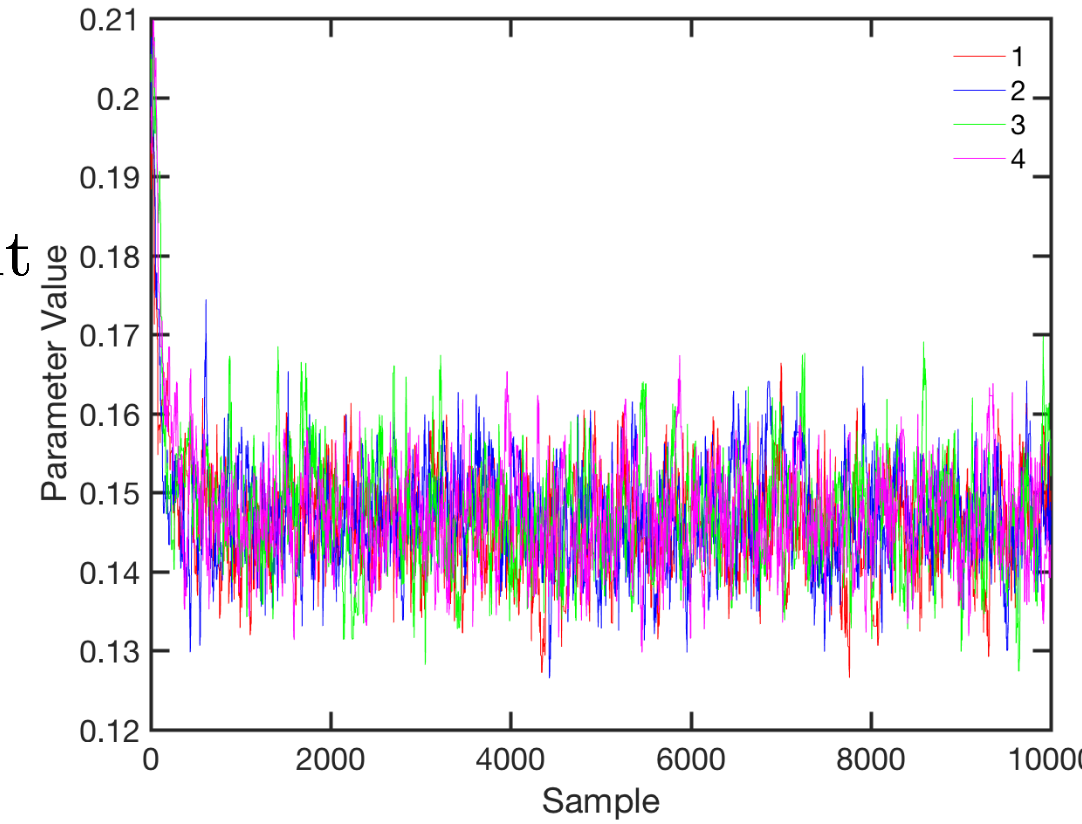
But can use as a correlated proposal distribution in 4D Metropolis! (show code, why better?)

4D Metropolis: Trace paths 10k cycles, 4 chains Better!

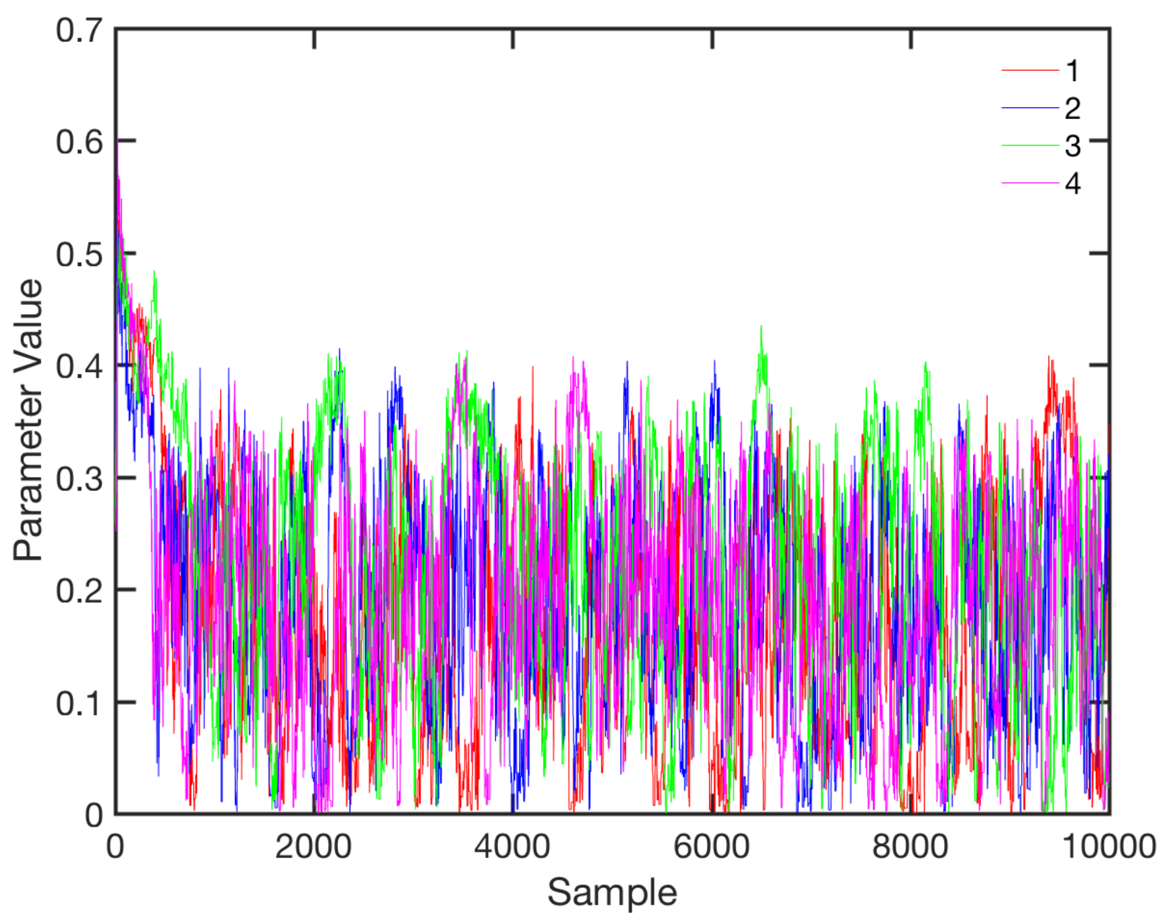
$\mathcal{M}_0$



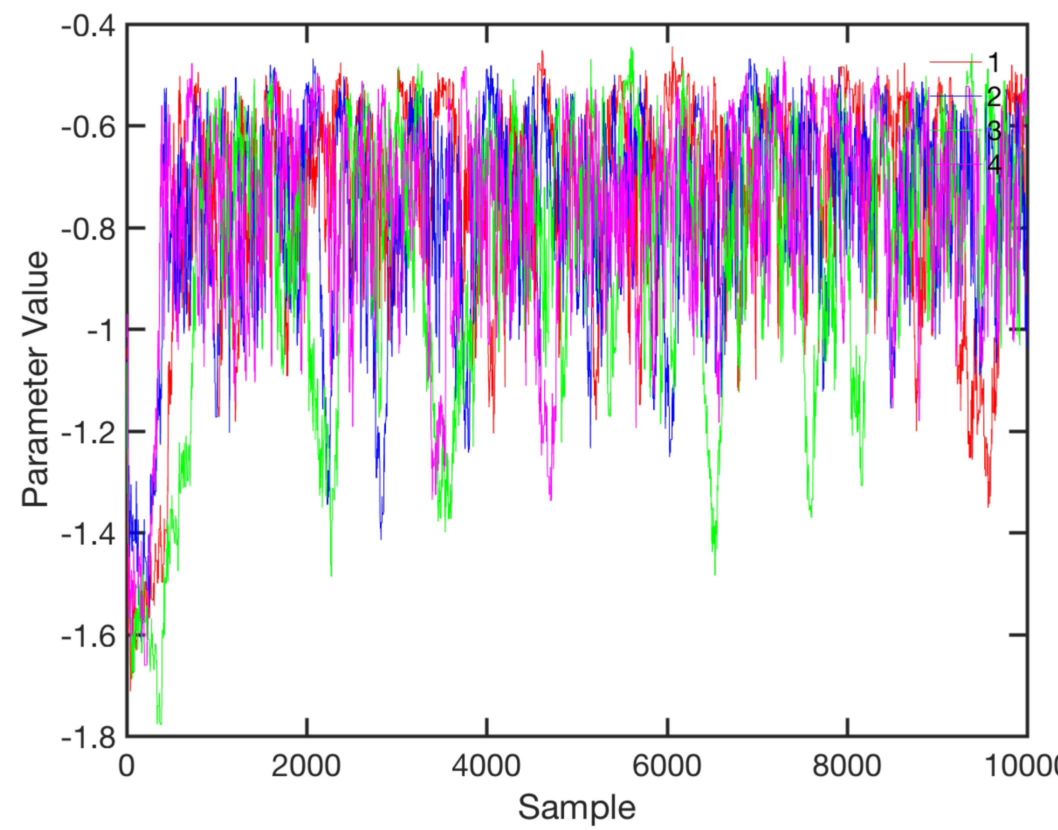
σ_{int}



Ω_M



w



Autocorrelation function

For each scalar parameter θ

$\hat{\rho}_t = C_t / C_0$ Sample Autocorrelation of lag t

$$C_t = \frac{1}{N-t-1} \sum_{i=1}^{N-t} (\theta_i - \bar{\theta})(\theta_{i+t} - \bar{\theta})$$

$$C_0 = \text{Sample Variance of } \theta$$

$$\tau = 1 + 2 \sum_{t=1}^{\infty} \rho_t$$

Autocorrelation time

Effective Sample Size

$$\hat{\tau} = 1 + 2\sum_{t=1}^{T}\hat{\rho}_t$$ Estimated Autocorr time

$$\text{Truncate at } T \text{ lags, so that } \hat{\rho}_T \approx 0.1$$

Effective number of independent samples

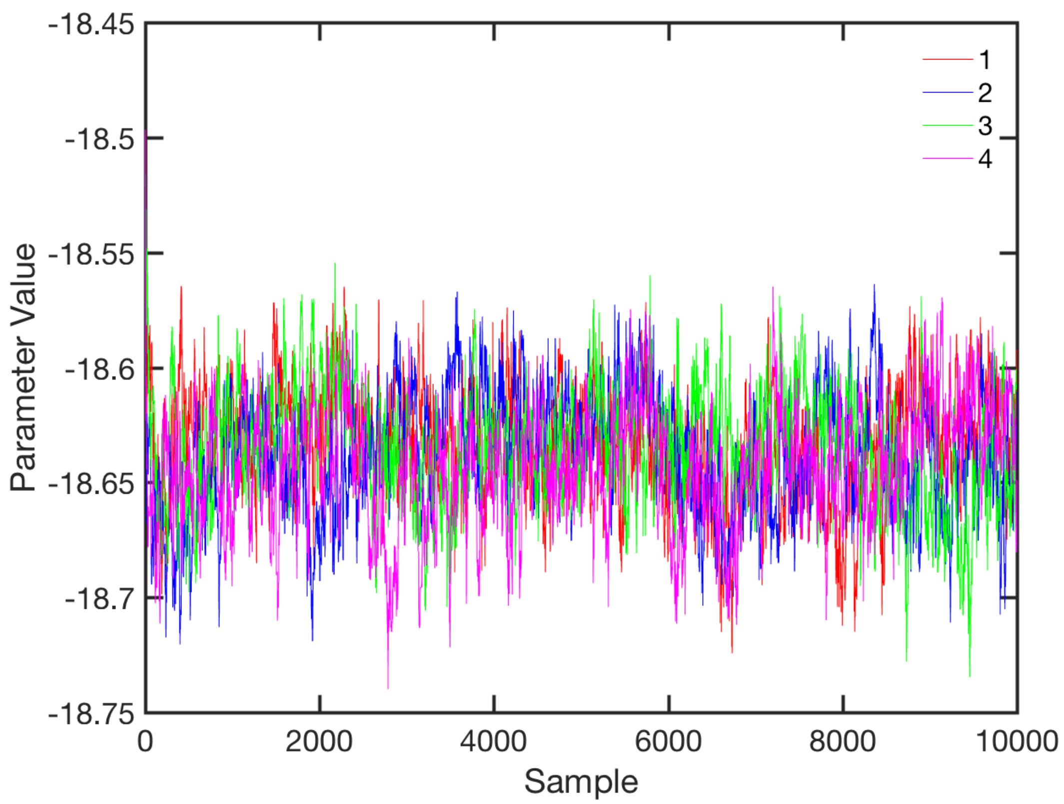
$$N_{\text{eff}} = N/\hat{\tau}$$

Can be computed for a single chain
(see Gelman BDA 3rd for multi-chain generalisation)

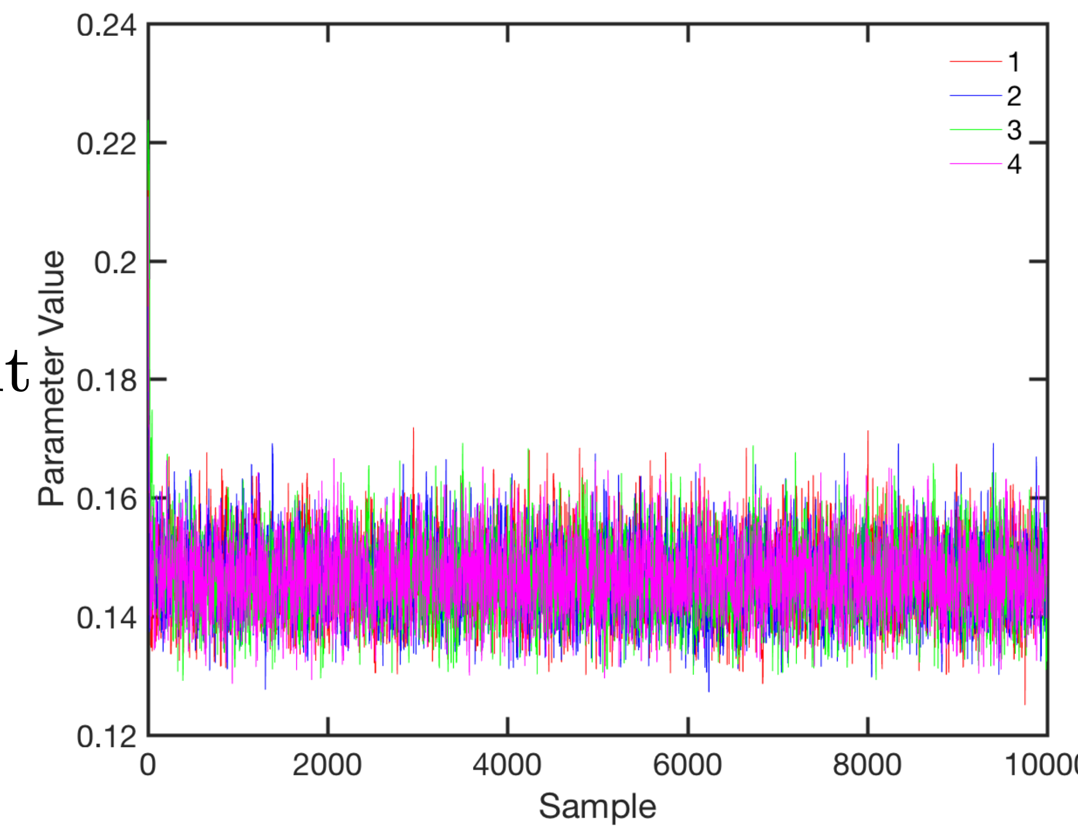
Slowest parameter is the limiting one!

MwG: Trace paths 10k cycles, 4 chains slow convergence (especially w)!

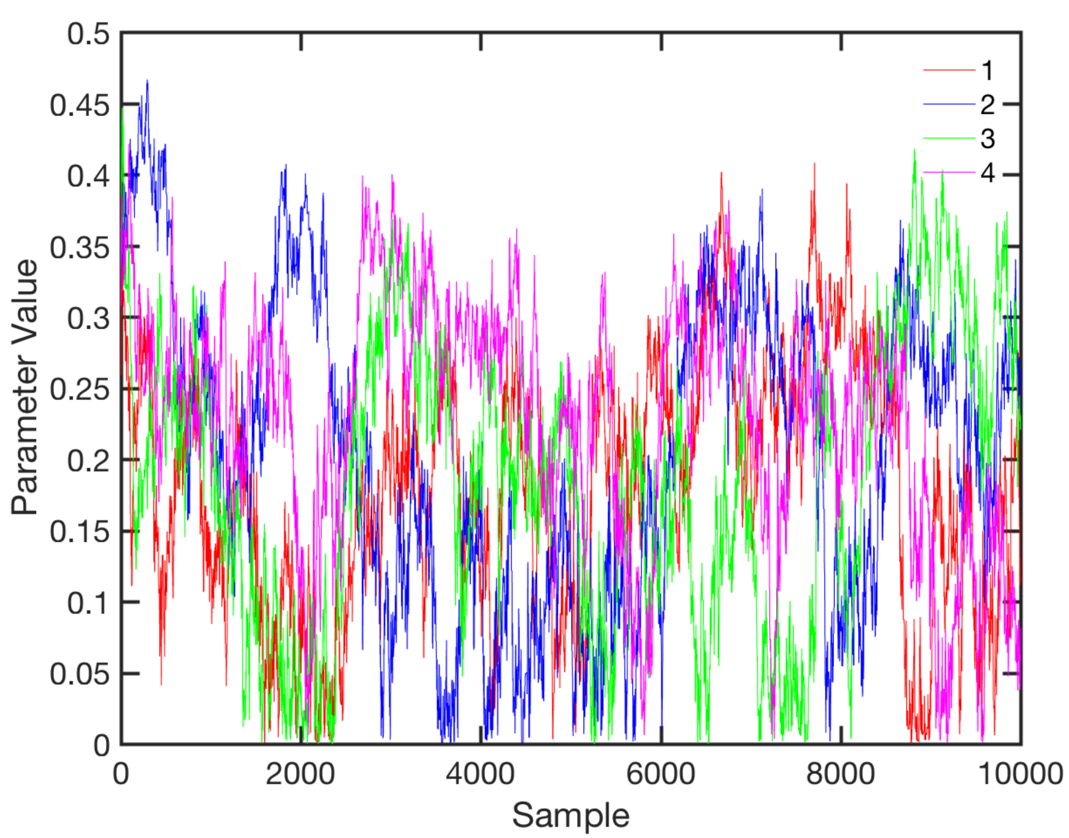
$\mathcal{M}_0$



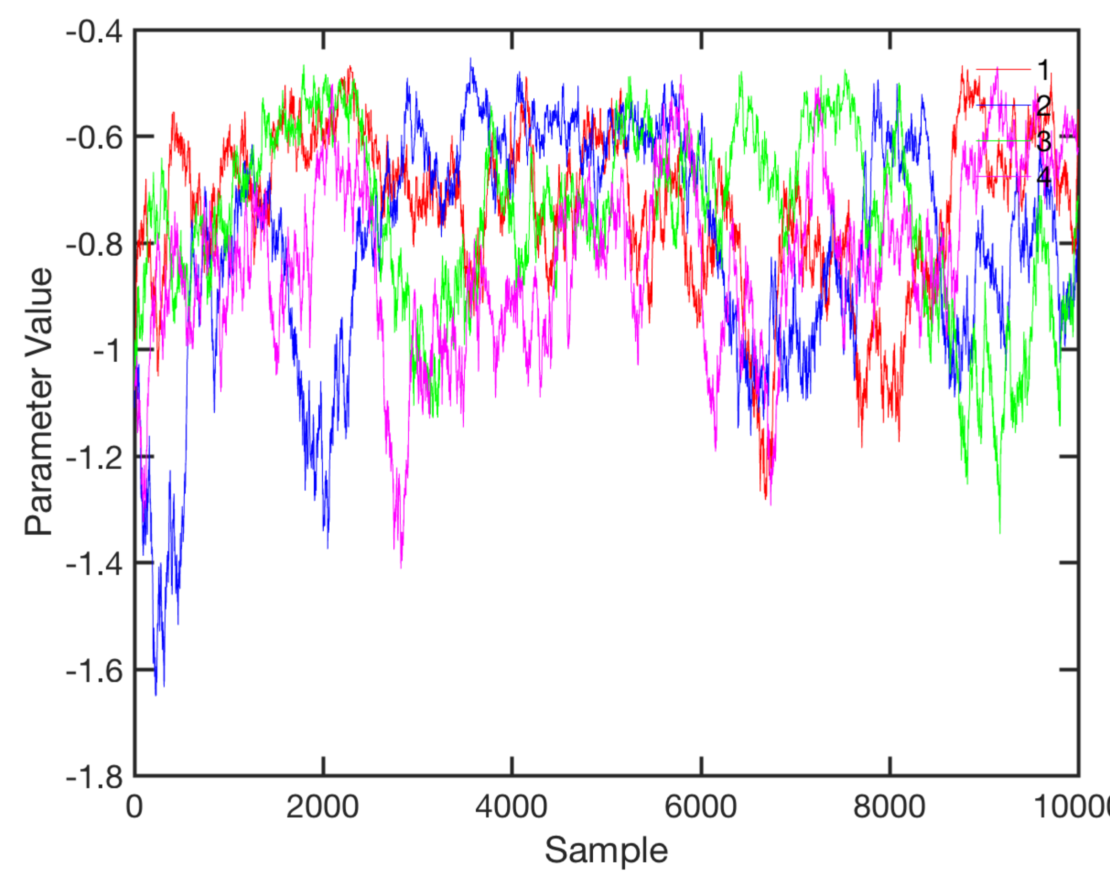
σ_{int}



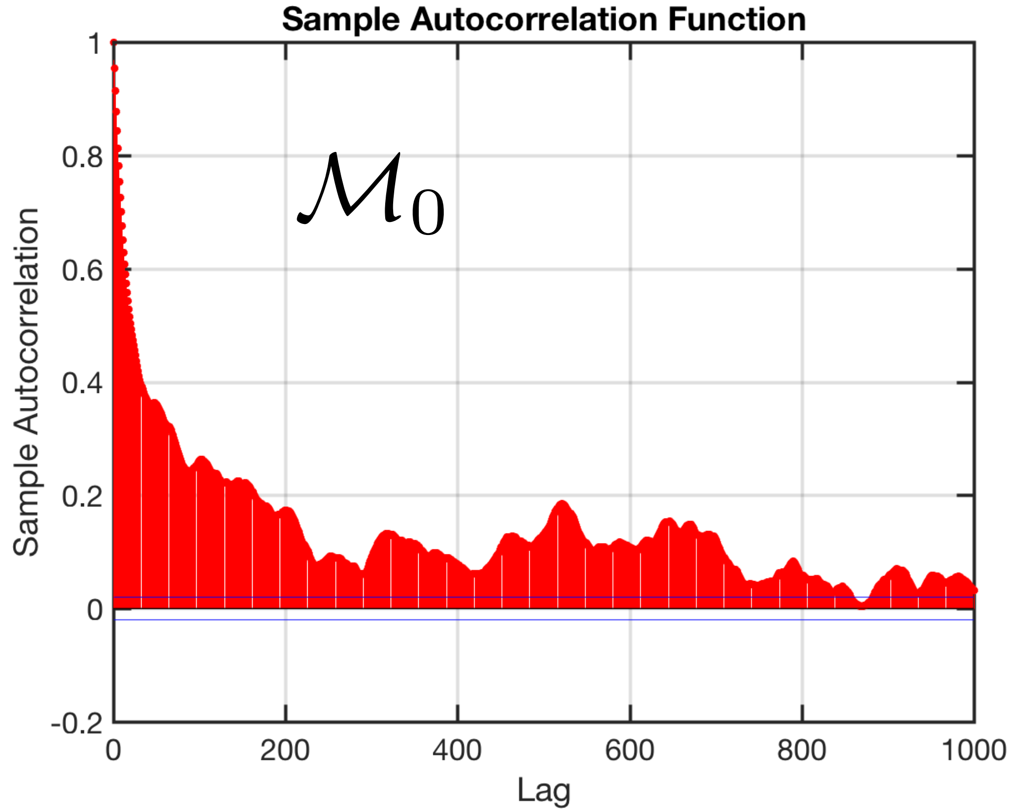
Ω_M

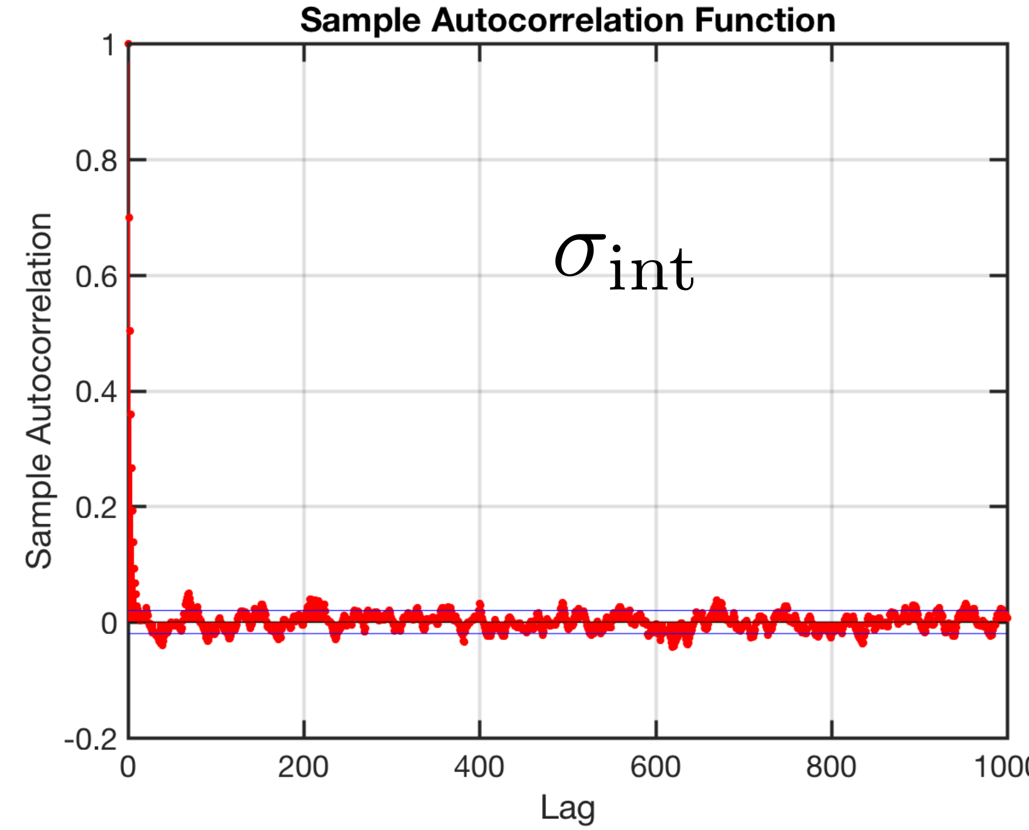


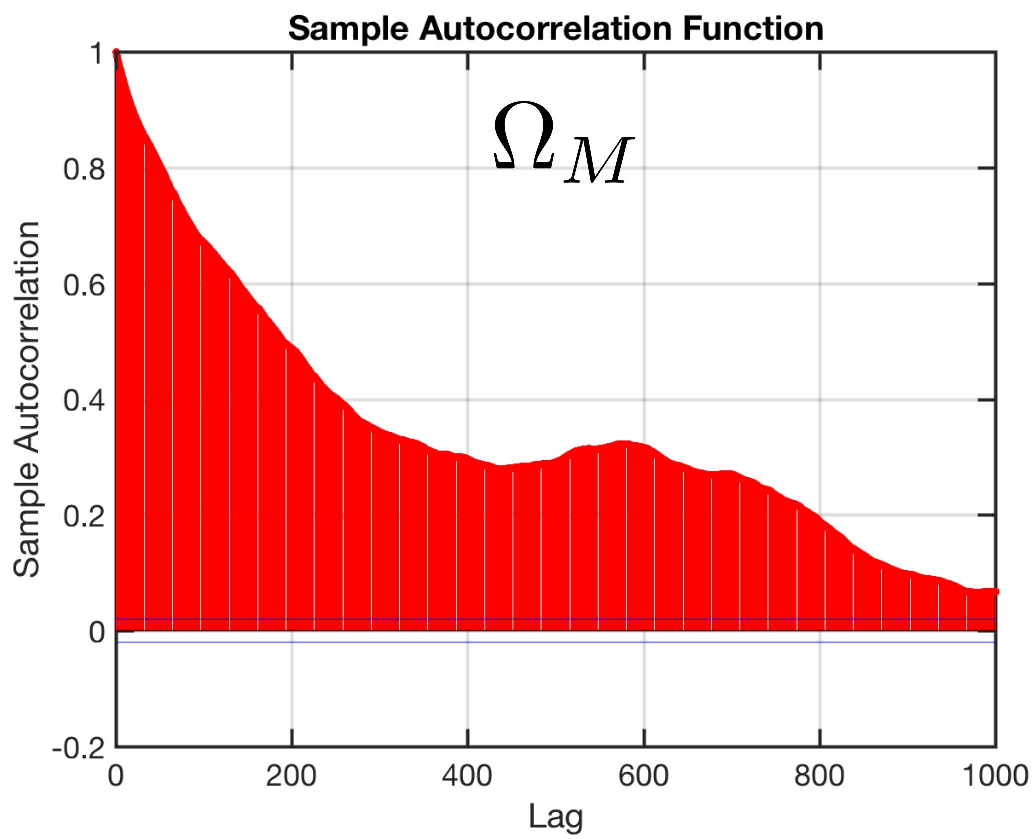
w

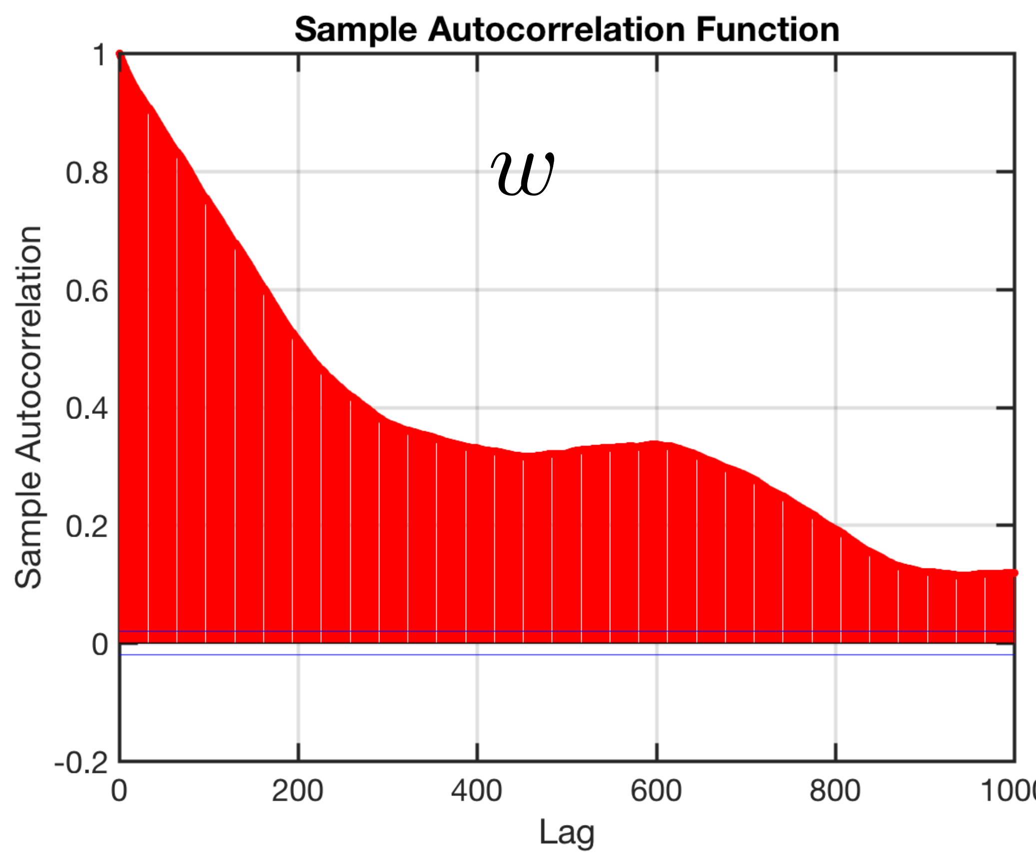


Corrected MwG: Autocorr over 1000 lags

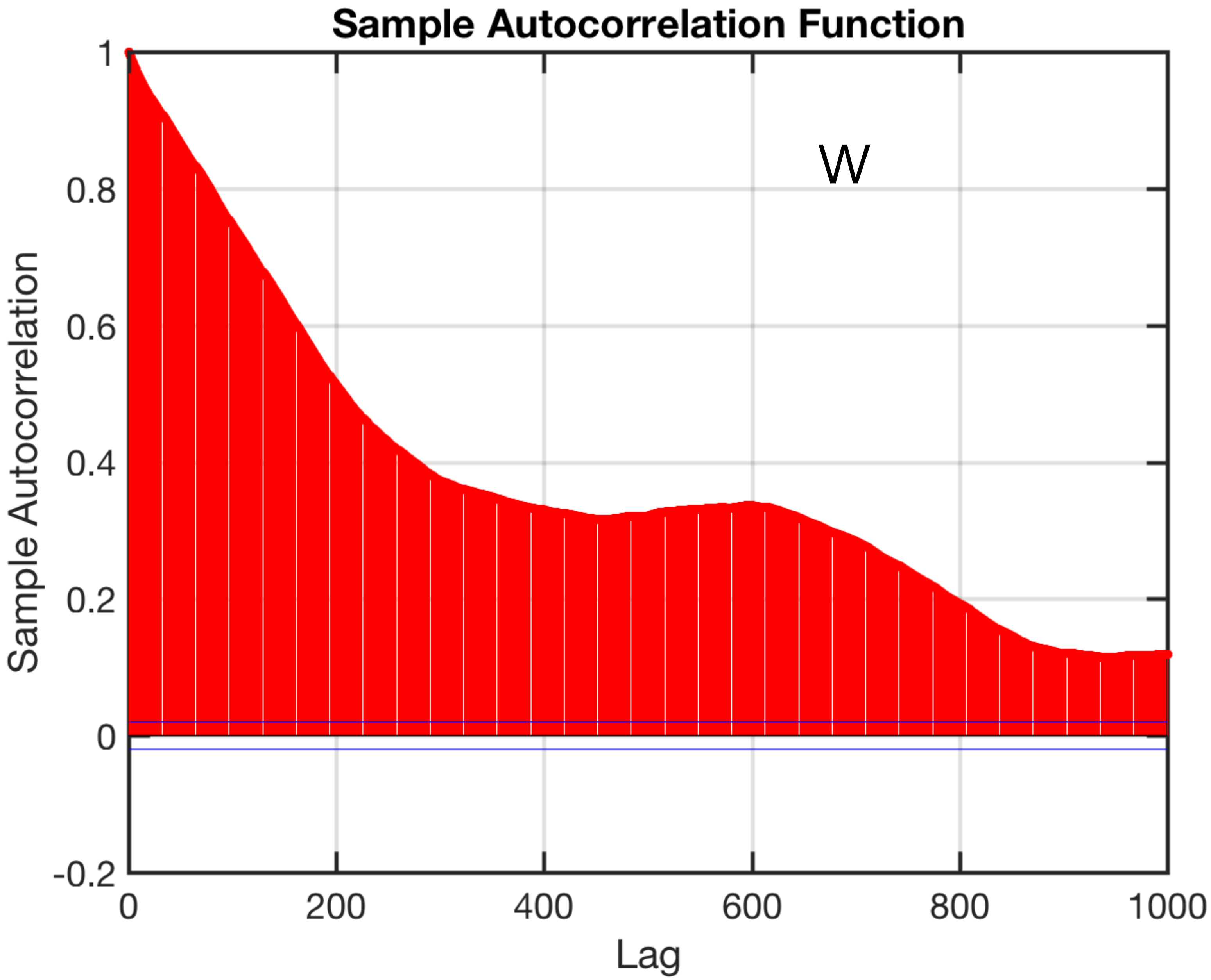








Metrop-w/in-Gibbs: autocorrelation fcn of w (slowest mixing parameter)



$$\hat{\tau} = 1 + 2\sum_{t=1}^{T} \hat{\rho}_t$$

$$\hat{\tau} = 751 \text{ cycles}$$

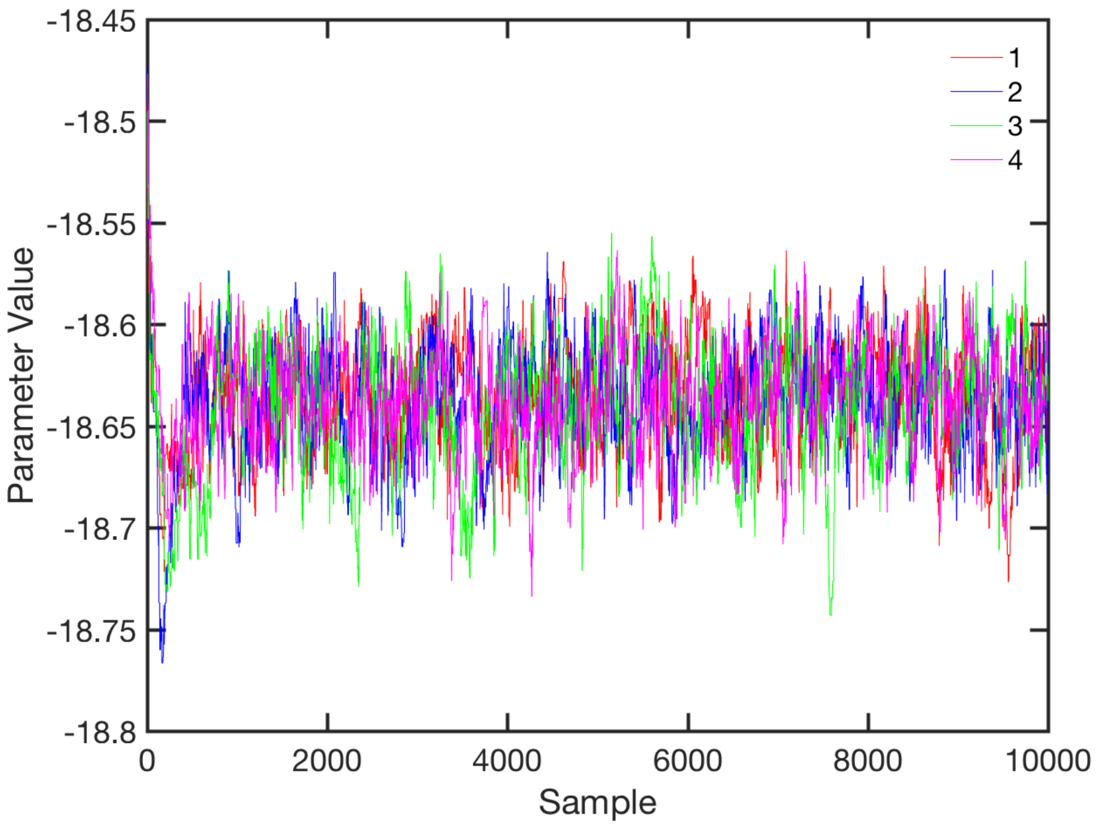
(single chain)

$$N_{\text{eff}} = 10,000/751 = 13 \text{ indep. samples}$$

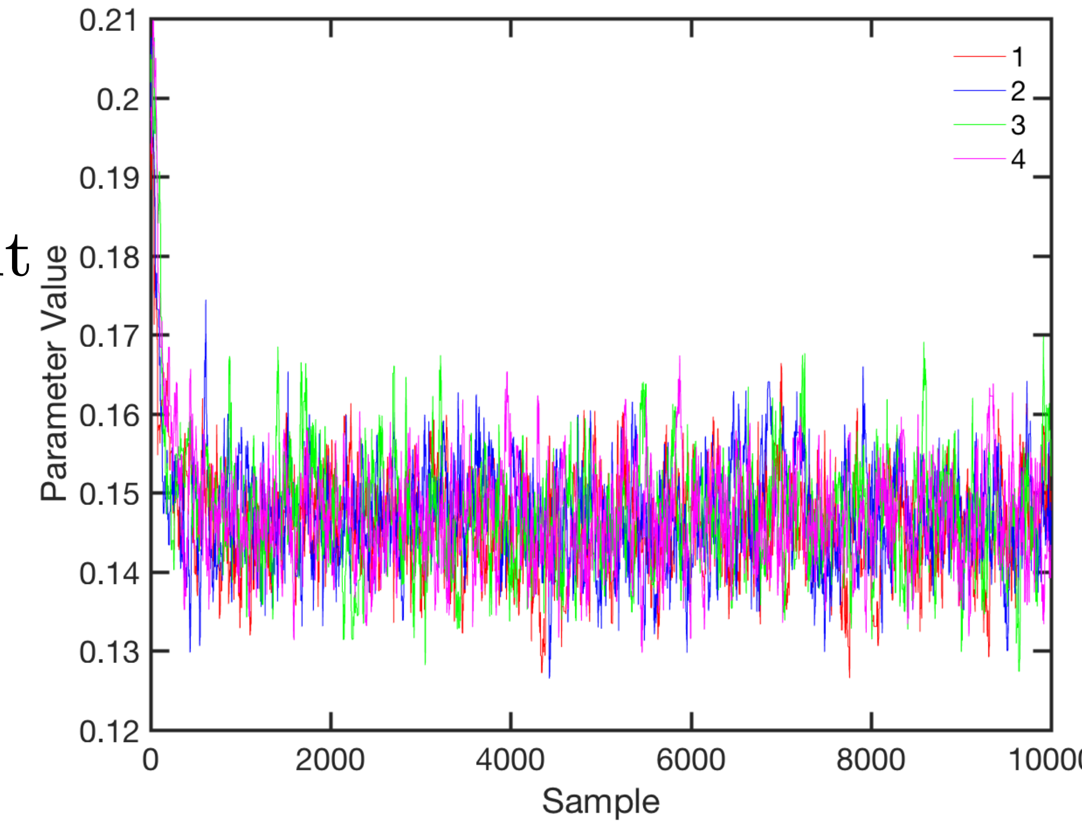
Rate = 552 sec / 13 = 42.5 sec per independent sample

4D Metropolis: Trace paths 10k cycles, 4 chains
Better!

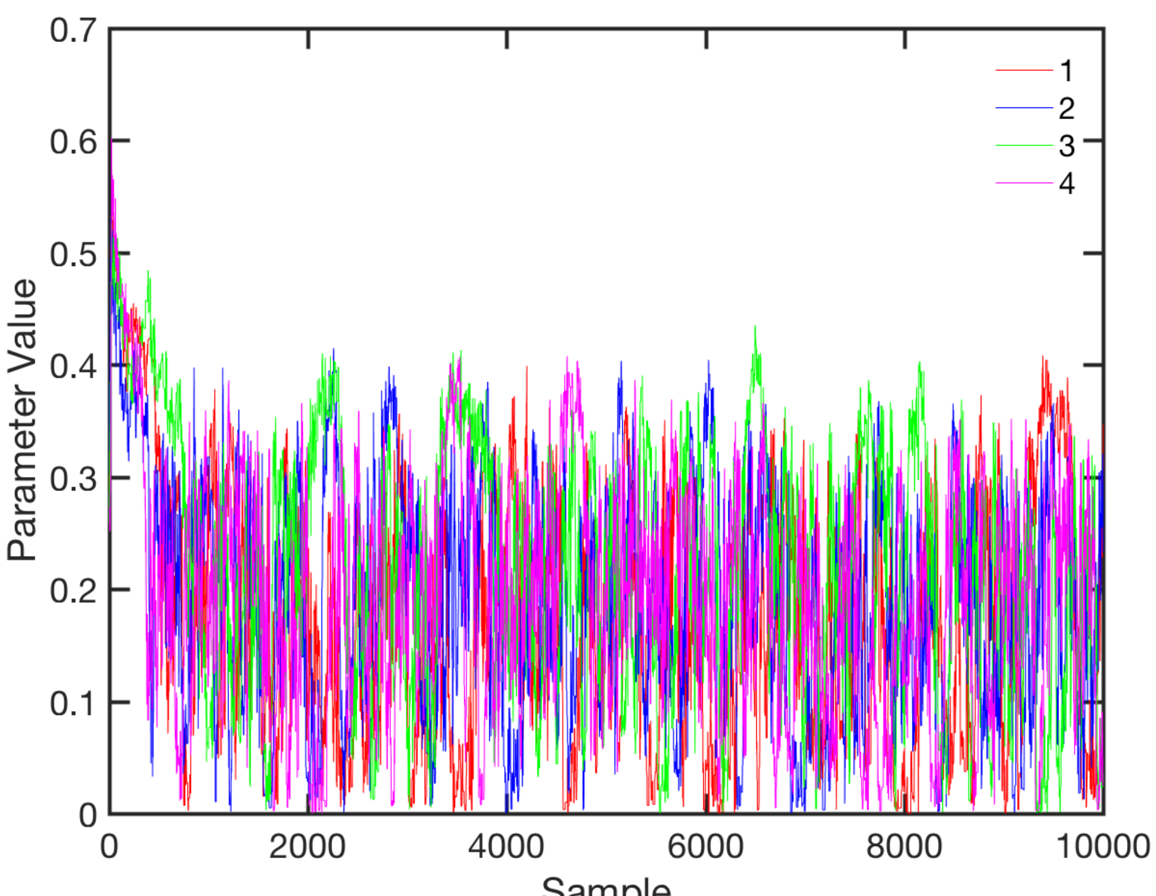
$\mathcal{M}_0$



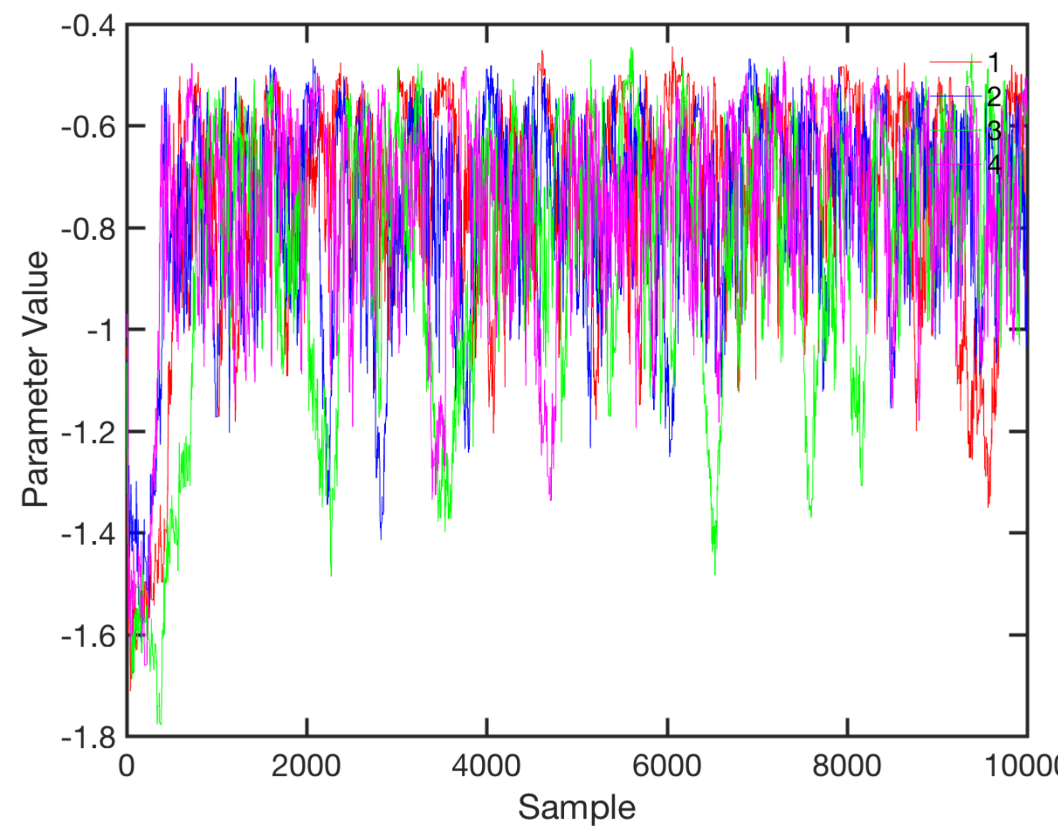
$\sigma_{\rm int}$



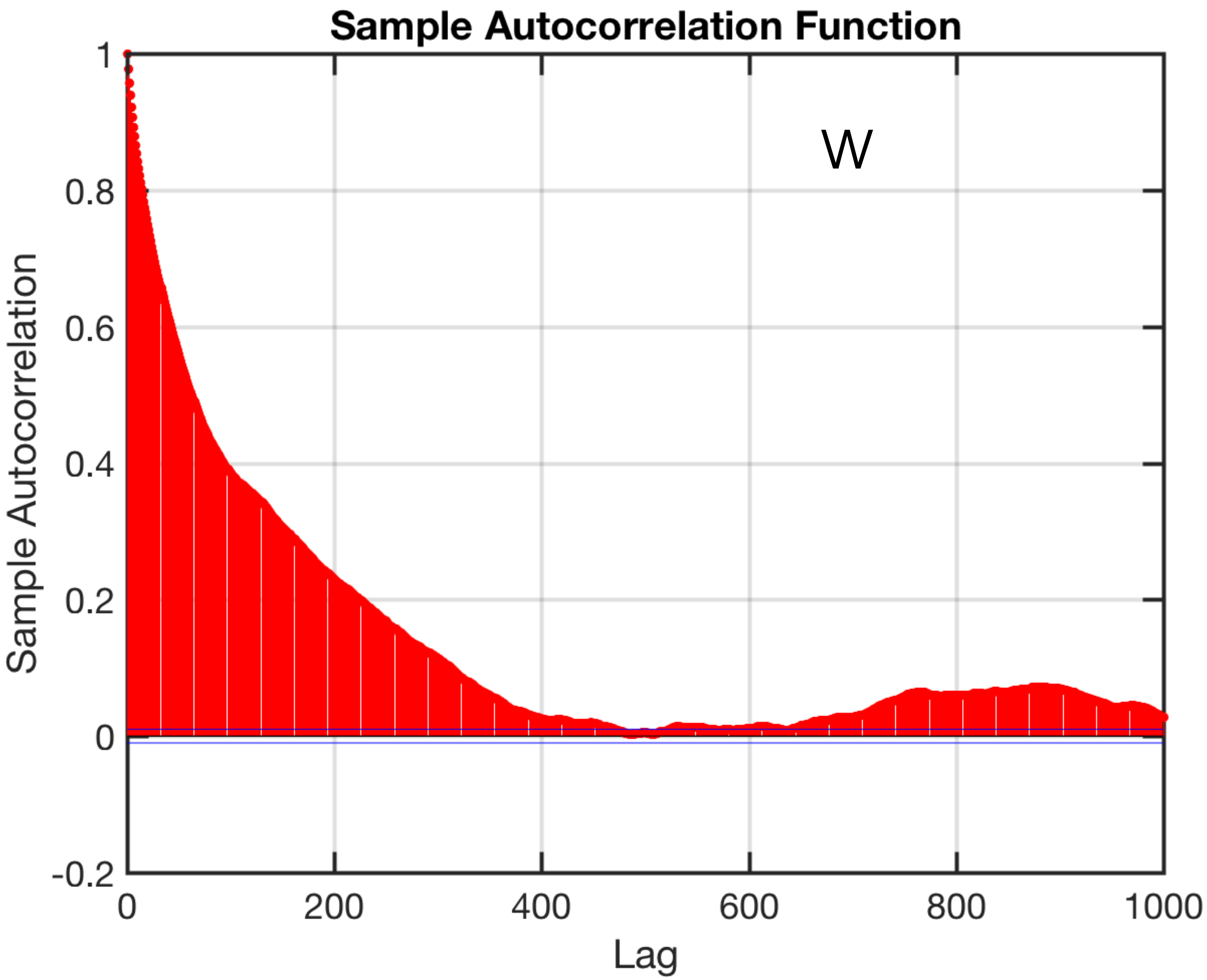
Ω_M



w



4D Metropolis: autocorrelation fcn of w (slowest mixing parameter)



$$\hat{\tau} = 1 + 2\sum_{t=1}^{T} \hat{\rho}_t$$

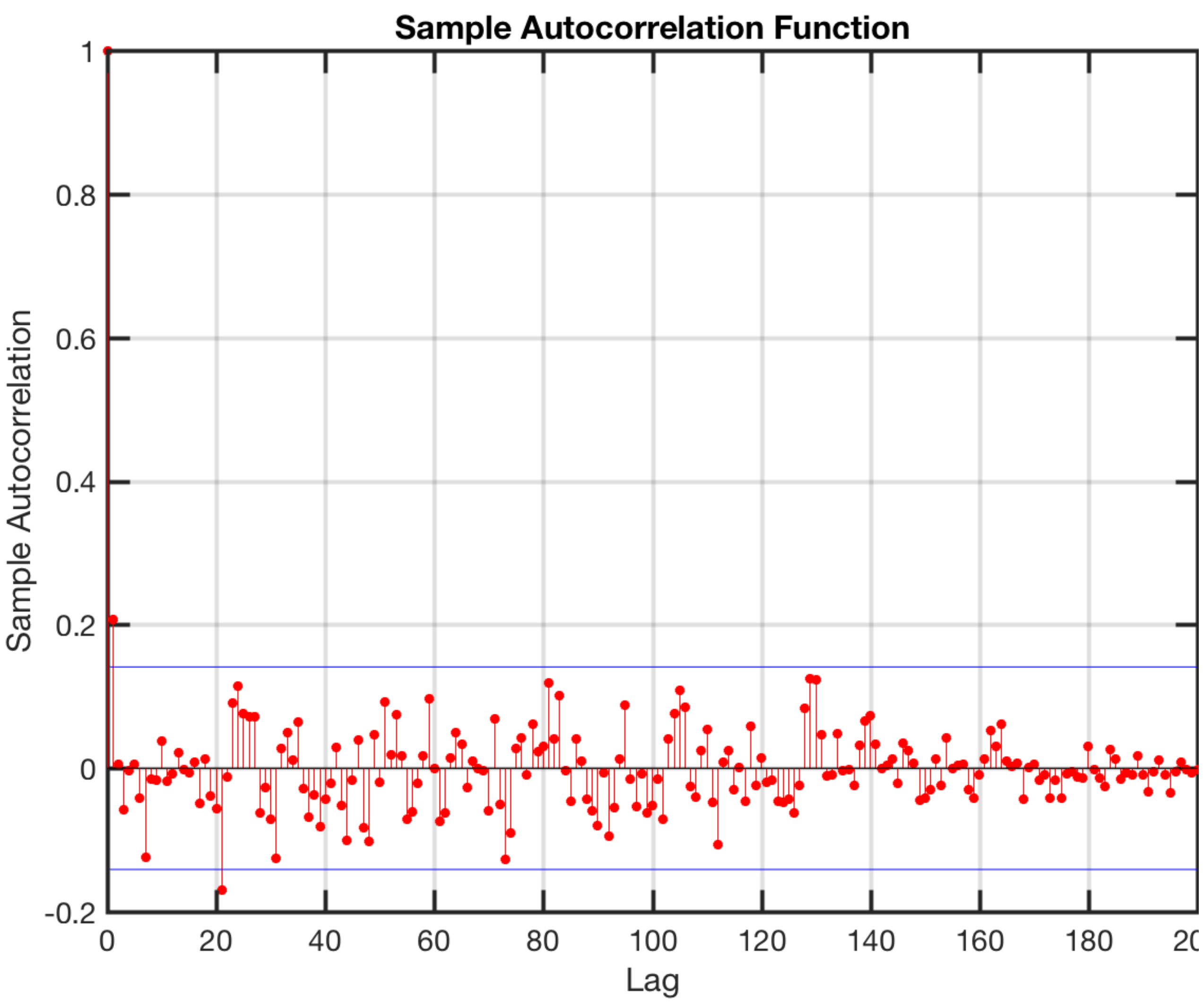
$$\hat{\tau} = 215 \text{ iterations}$$

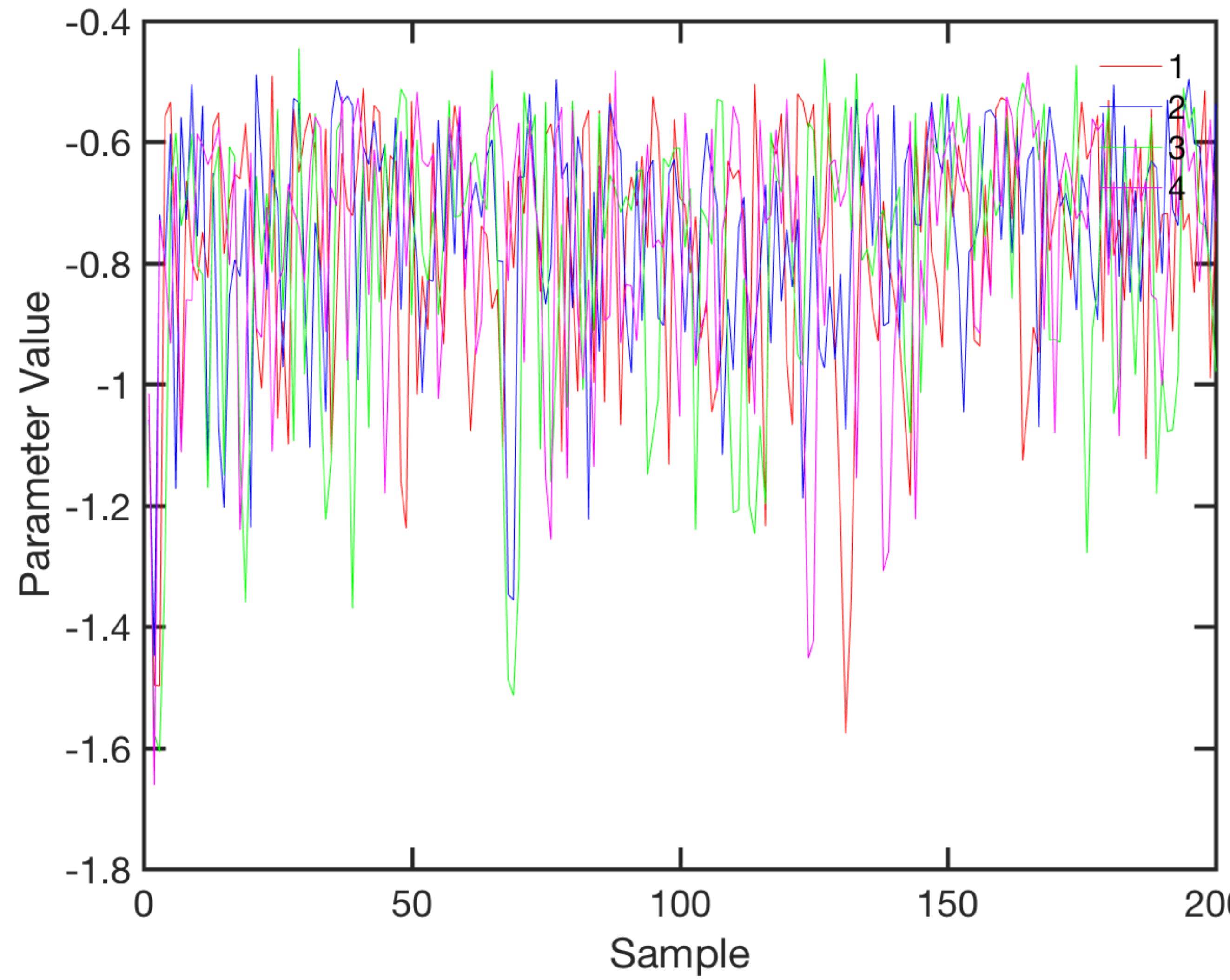
(single chain)

$$N_{\text{eff}} = 10,000/215 = 46 \text{ indep. samples}$$

Rate = 140 sec / 46 = 3.0 sec per independent sample
14x faster than MwG!

4D Metropolis: autocorrelation fcn of w after thinning by 200

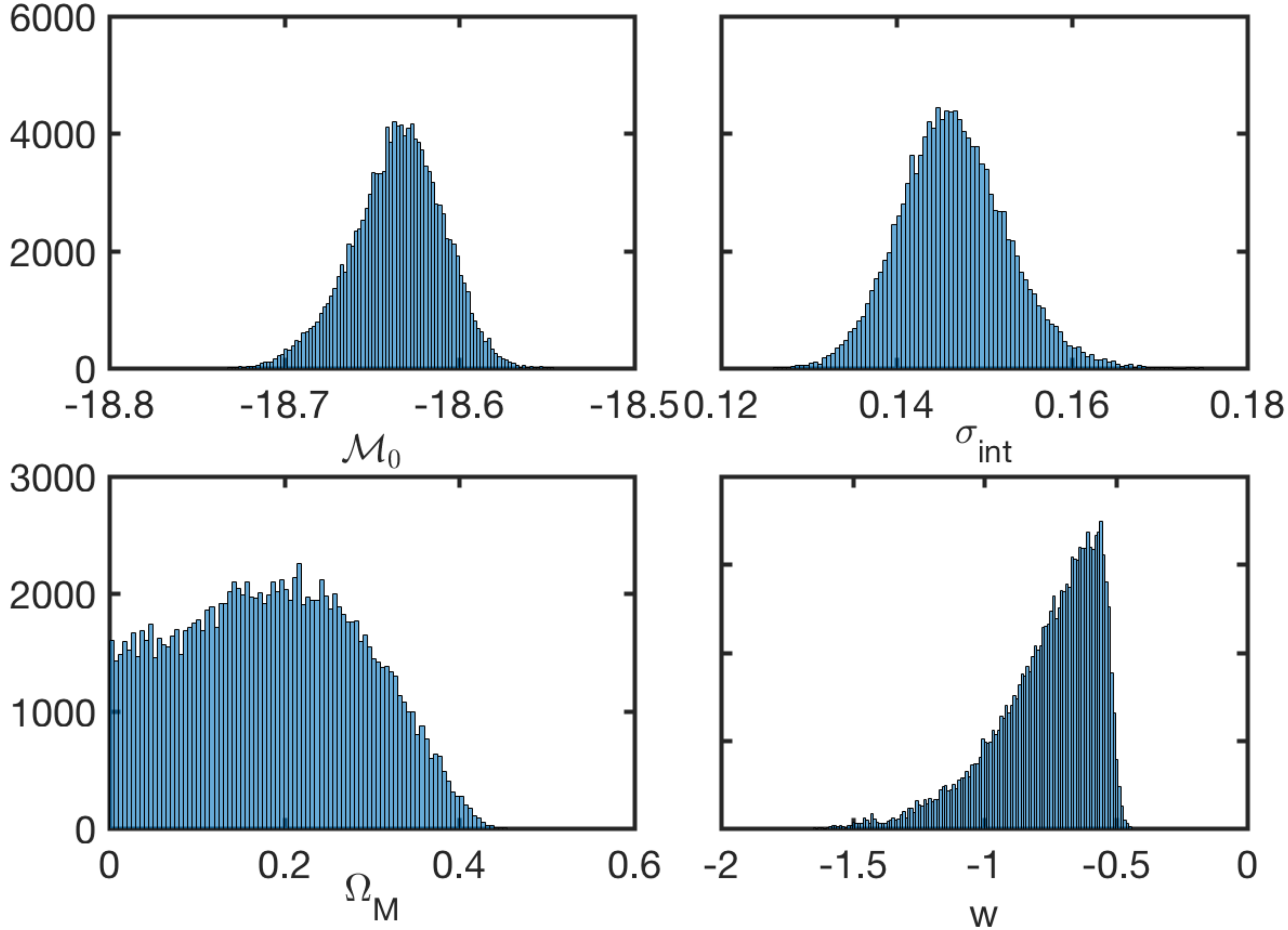




W

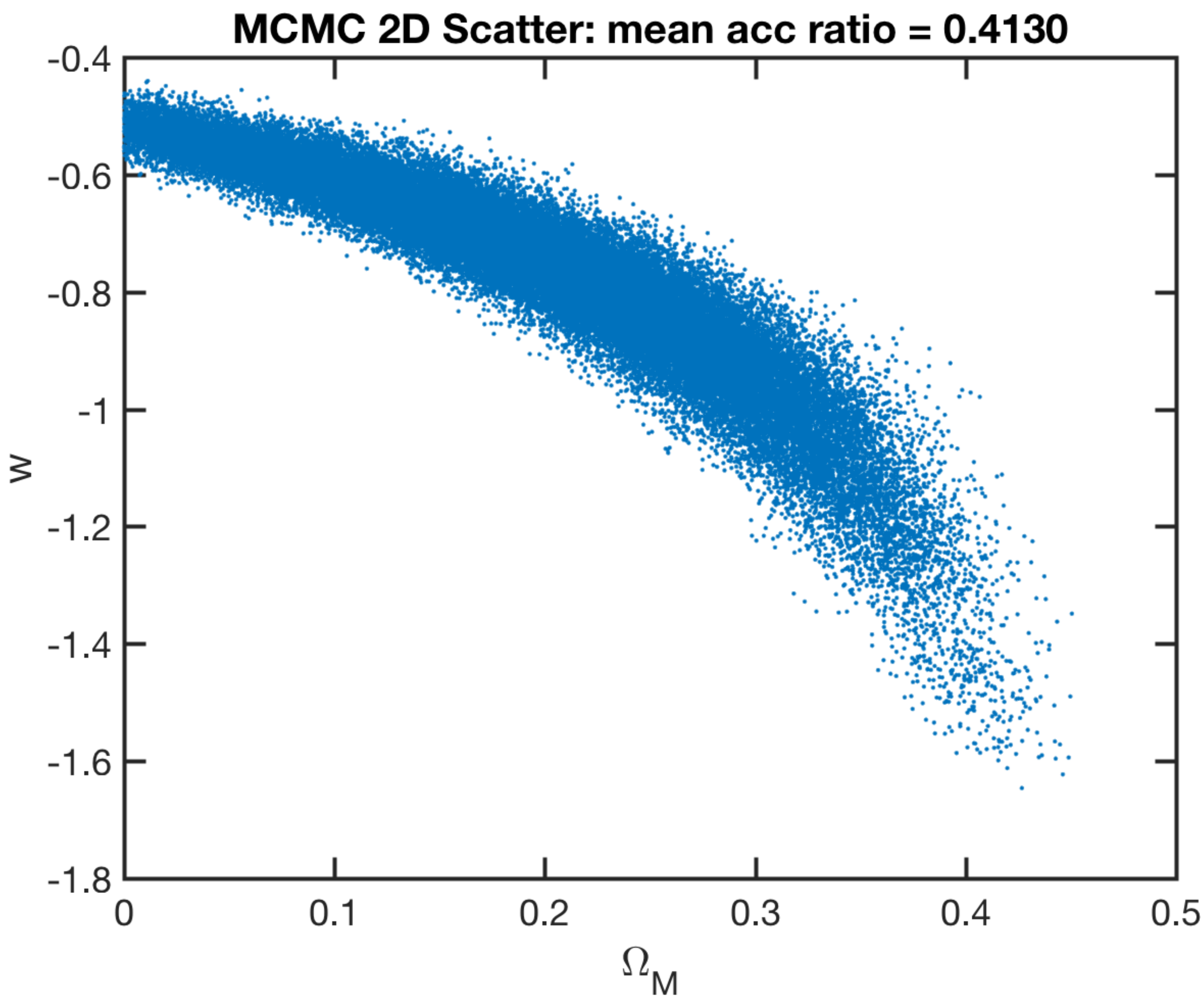
4D Metropolis run with 4x10k chains

Marginal Posterior Histograms



4D Metropolis run with 4x10k chains

Dark Energy EoS parameter



Matter Density

4D Metropolis run with 4x10k chains