Proposal Distribution

Proposal Distribution: Two basic approaches

- Choose a fixed proposal function
  - o Independent Metropolis
- Random walk (local Metropolis)

$$\theta^* = \theta^{k-1} + Rz$$

o Two (of several) choices: $Z \sim N(0,1)$

$$\text{(i)}\ R = cI \Rightarrow \theta^* \sim N(\theta^{k-1}, cI)$$

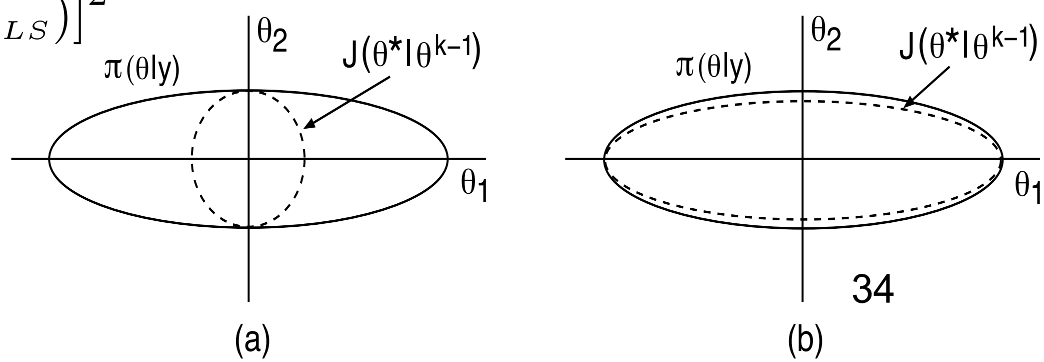
$$\text{(ii)}\ R = \text{chol}(V) \Rightarrow \theta^* \sim N(\theta^{k-1}, V)$$

where

Sensitivity Matrix

$$V = \sigma_{OLS}^2 \left[\mathcal{X}^T(\theta_{OLS})\mathcal{X}(\theta_{OLS})\right]^{-1}$$

$$\sigma_{OLS}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left[y_i - f_i(\theta_{OLS})\right]^2$$



Metropolis Algorithm

Metropolis Algorithm: [Metropolis and Ulam, 1949]

1. Initialization: Choose an initial parameter value θ^0 that satisfies $\pi(\theta^0|y) > 0$.
2. For $k = 1, \cdots, M$

   (a) For $z \sim N(0,1)$, construct the candidate

   $$\theta^* = \theta^{k-1} + Rz$$

   where R is the Cholesky decomposition of V or D. This ensures that

   $$\theta^* \sim N(\theta^{k-1}, V) \text{ or } \theta^* \sim N(\theta^{k-1}, D).$$

   (b) Compute the ratio

   $$r(\theta^*|\theta^{k-1}) = \frac{\pi(\theta^*|y)}{\pi(\theta^{k-1}|y)} = \frac{\pi(y|\theta^*)\pi_0(\theta^*)}{\pi(y|theta^{k-1})\pi_0(\theta^{k-1})}. \quad (1)$$

   (c) Set

   $$\theta^k = \begin{cases} \theta^* & , \text{ with probability } \alpha = \min(1, r) \\ \theta^{k-1} & , \text{ else.} \end{cases}$$

   That is, we accept θ^* with probability 1 if $r \geqslant 1$ and we accept it with probability r if $r < 1$.

Metropolis-Hastings Algorithm

Metropolis-Hastings Algorithm: $J(\theta^*|\theta^{k-1})$ does not have to be symmetric

- Acceptance Ratio: $$r(\theta^*|\theta^{k-1}) = \frac{\pi(\theta^*|y)/J(\theta^*|\theta^{k-1})}{\pi(\theta^{k-1}|y)/J(\theta^{k-1}|\theta^*)} = \frac{\pi(y|\theta^*)\pi_0(\theta^*)J(\theta^{k-1}|\theta^*)}{\pi(y|\theta^{k-1})\pi_0(\theta^{k-1})J(\theta^*|\theta^{k-1})}.$$

Examples:

- Cauchy distribution: $J(\theta^*|\theta^{k-1}) = \frac{1}{\pi[1+(\theta^*)^2]}$
- $\chi^2(k)$ distribution: $J(\theta^*|\theta^{k-1}) = \kappa(\theta^*)^{k/2-1}e^{\theta^*/2}$

Note: Considered one of top 10 algorithms of 20th century

Random Walk Metropolis Algorithm for Parameter Estimation

1. Set number of chain elements M and design parameters n_s, σ_s
2. Determine $\theta^0 = \arg\min_\theta \sum_{i=1}^N [y_i - f_i(\theta)]^2$
3. Set $SS_{\theta^0} = \sum_{i=1}^N [y_i - f_i(\theta^0)]^2$
4. Compute initial variance estimate: $s_0^2 = \frac{SS_{\theta^0}}{n-p}$
5. Construct covariance estimate $V = s_0^2[\mathcal{X}^T(\theta^0)\mathcal{X}(\theta^0)]^{-1}$ and $R = \mathrm{chol}(V)$
6. For $k = 1, \cdots, M$
   (a) Sample $z_k \sim N(0,1)$
   (b) Construct candidate $\theta^* = \theta^{k-1} + Rz_k$
   (c) Sample $u_\alpha \sim \mathcal{U}(0,1)$
   (d) Compute $SS_{\theta^*} = \sum_{i=1}^N [y_i - f_i(\theta^*)]^2$
   (e) Compute

$$\alpha(\theta^*|\theta^{k-1}) = \min\left(1, e^{-[SS_{\theta^*} - SS_{\theta^{k-1}}]/2s_{k-1}^2}\right)$$

   (f) If $u_\alpha < \alpha$,

   Set $\theta^k = \theta^*$, $SS_{\theta^k} = SS_{\theta^*}$

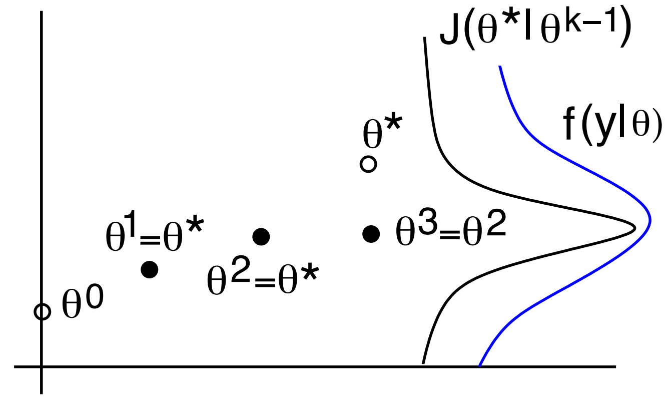
   else

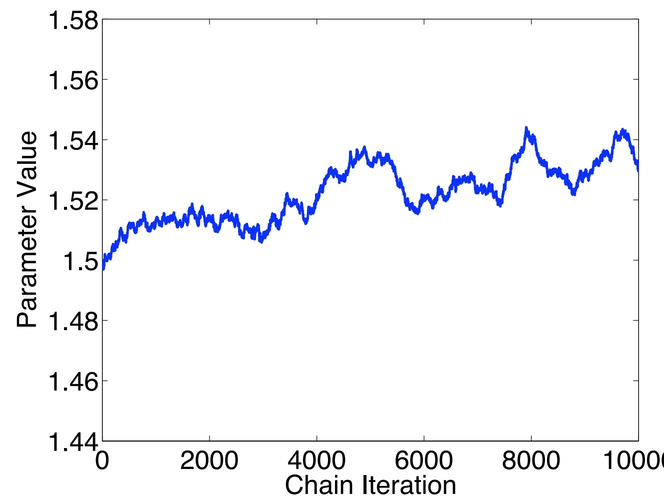
   Set $\theta^k = \theta^{k-1}$, $SS_{\theta^k} = SS_{\theta^{k-1}}$

   endif

   (g) Update $s_k \sim \text{Inv-gamma}(a_{val}, b_{val})$ where

   $a_{val} = 0.5(n_s + n)$, $b_{val} = 0.5(n_s\sigma_s^2 + SS_{\theta^k})$





Sampling Error Variance

Strategy: Treat error variance σ^2 as parameter to be sampled.

Definition: The property that the prior and posterior distributions have the same parametric form is termed *conjugacy.*

Note: The likelihood

$$f(y, \theta|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_\theta/2\sigma^2}$$

has the conjugate prior

$$\pi_0(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{\beta/\sigma^2}$$

The posterior is

$$\pi(\sigma^2|\theta, y) \propto (\sigma^2)^{-(\alpha+1+n/2)} e^{-(\beta+SS_\theta/2)/\sigma^2}$$

so that

$$\sigma^2|(y,\theta) \sim \text{Inv-gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{SS_q}{2}\right)$$

or

$$\sigma^2|(y,\theta) \sim \text{Inv-gamma}\left(\frac{n_s + n}{2}, \frac{n_s\sigma_s^2 + SS_q}{2}\right)$$

Note:

- n_0 taken small;
  e.g., $n_0 = 1$ or $n_0 = .01$
- Take $\sigma_s^2 = s_{k-1}^2 = \frac{R_{k-1}^T R_{k-1}}{n-p}$

Random Walk Metropolis

Example: We revisit the spring model

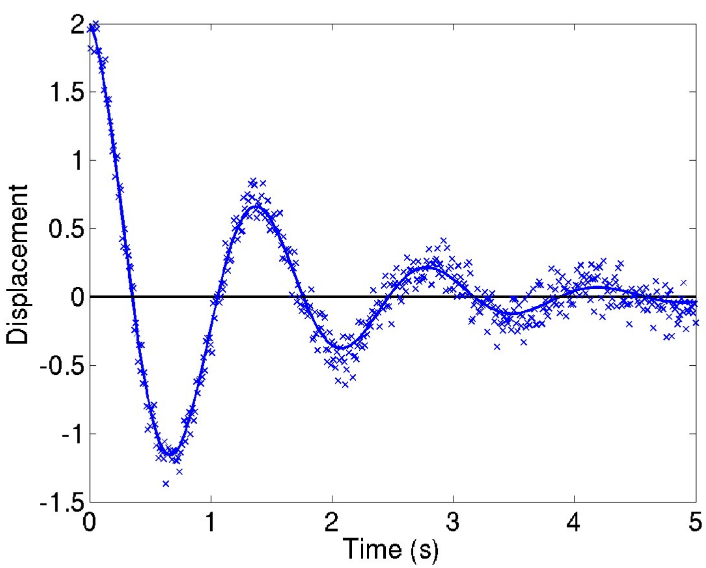
$$\ddot{z} + C\dot{z} + Kz = 0$$

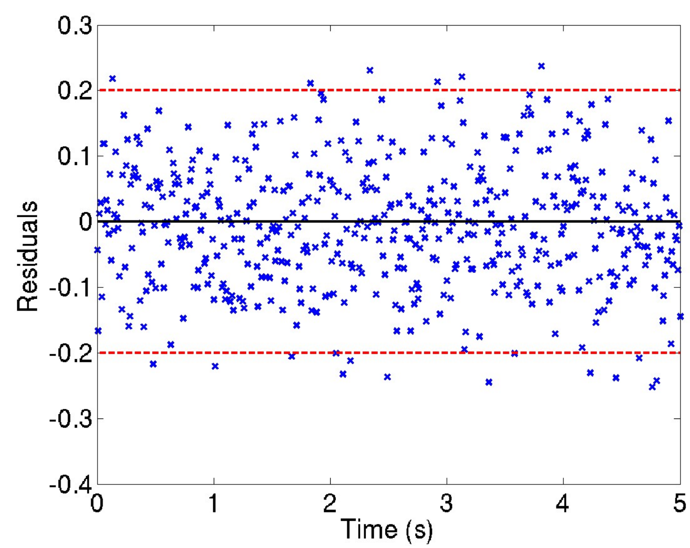
$$z(0) = 2\,,\ \dot{z}(0) = -C$$

which has the solution

$$z(t) = 2e^{-Ct/2}\cos(\sqrt{K - C^2/4} \cdot t)$$

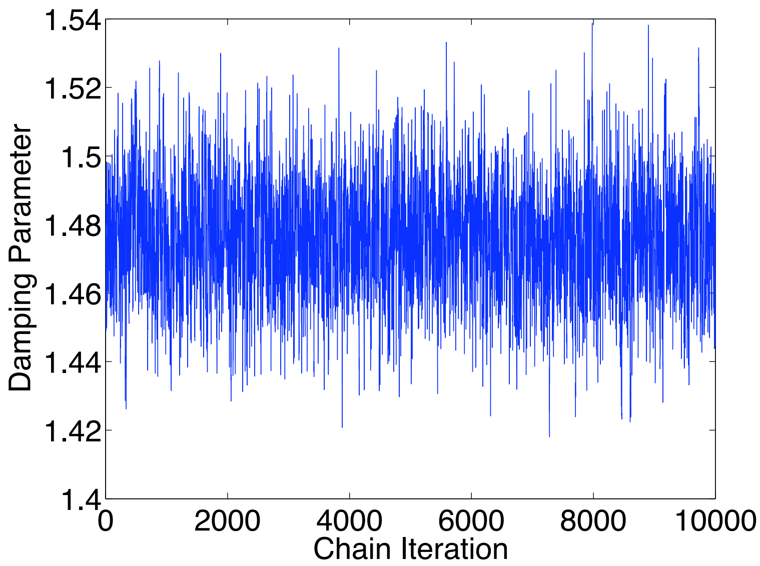
We assume that $\varepsilon_i \sim N(0, \sigma_0^2)$ where $\sigma_0 = 0.1$.

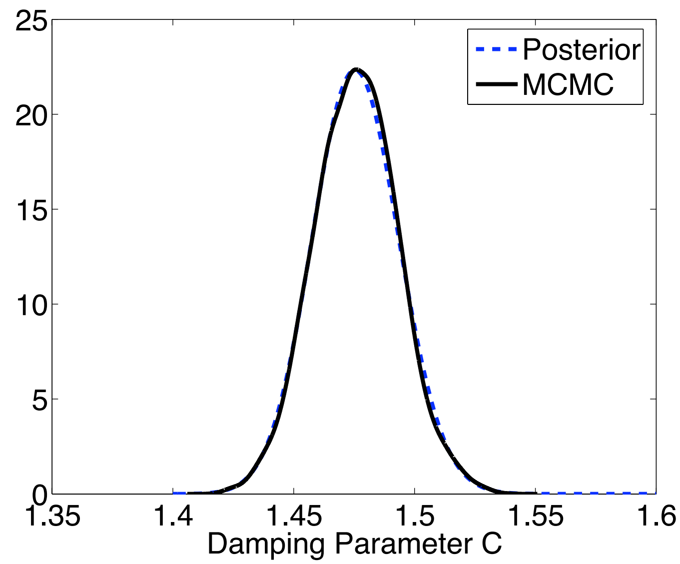




Random Walk Metropolis

Case i: Take K =20.5 and $\theta = [C, \sigma^2]$



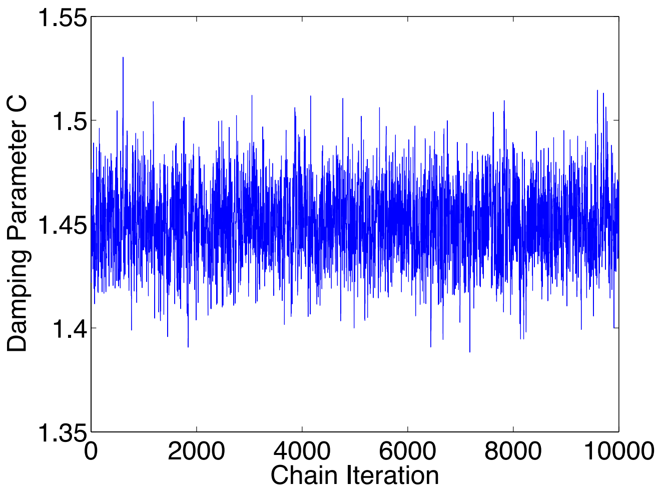


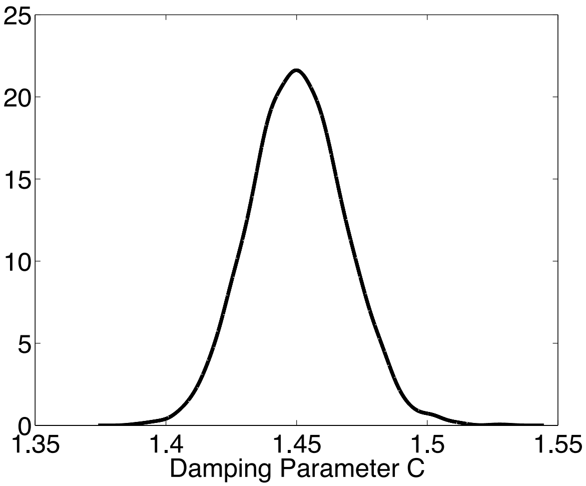
Note: Kernel density estimator (KDE) used to construct density.

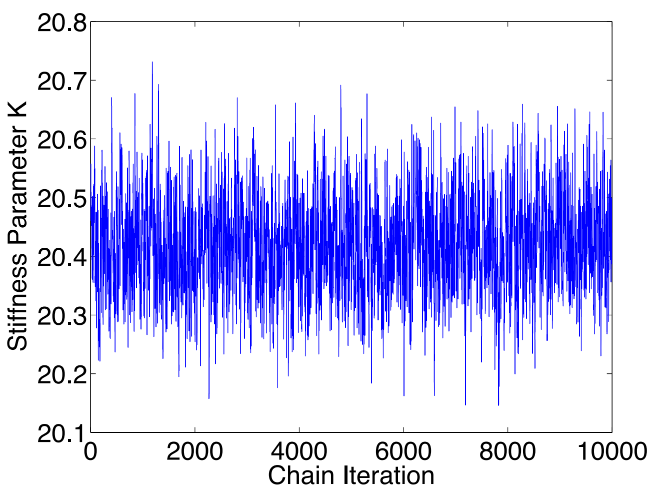
Random Walk Metropolis

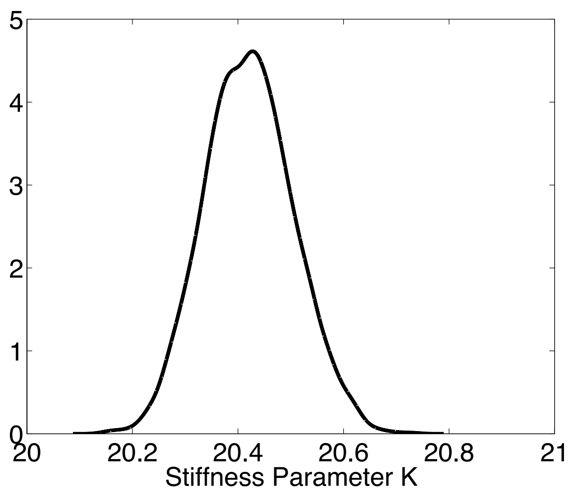
Case ii: Take $\theta = [C, K, \sigma^2]$ with $J(\theta^*|\theta^{k-1}) = N(\theta^{k-1}, V)$ and

$$V = \begin{bmatrix} 0.000345 & 0.000268 \\ 0.000268 & 0.007071 \end{bmatrix}$$









Note:

$$2\sigma_C \approx 0.04$$

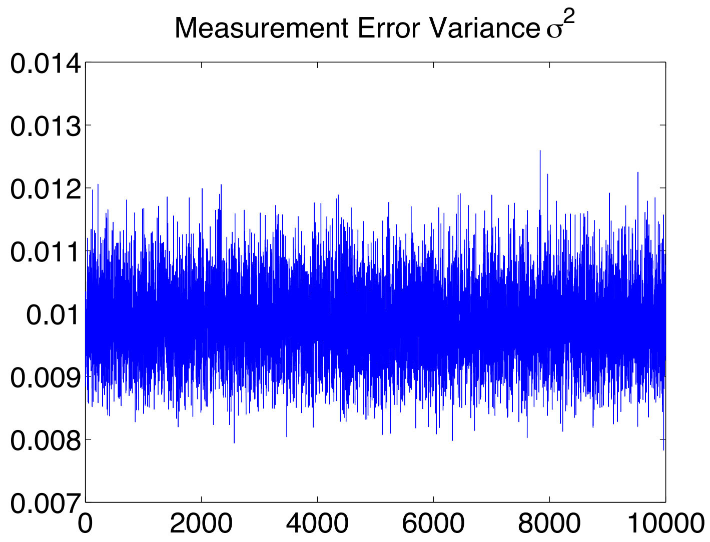
$$\Rightarrow \sigma_C^2 \approx 0.4 \times 10^{-3}$$

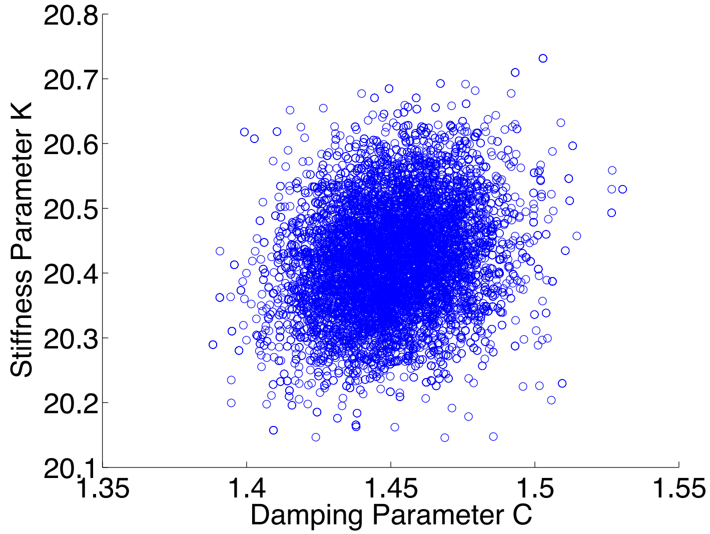
$$2\sigma_K \approx 0.18$$

$$\Rightarrow \sigma_K^2 \approx 0.0081$$

Random Walk Metropolis

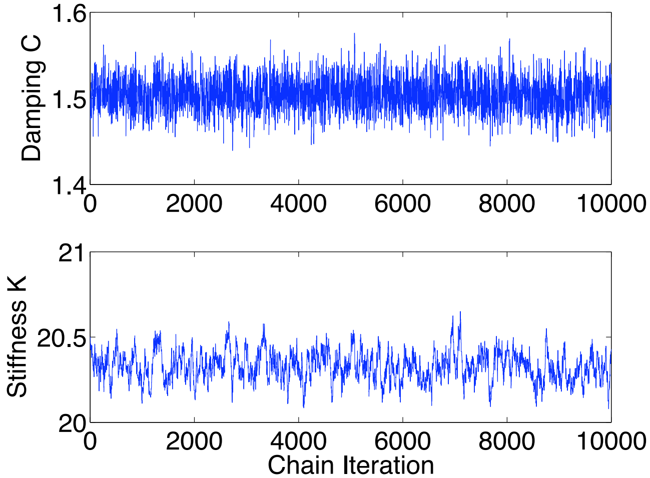
Case ii: Measurement error variance and joint samples



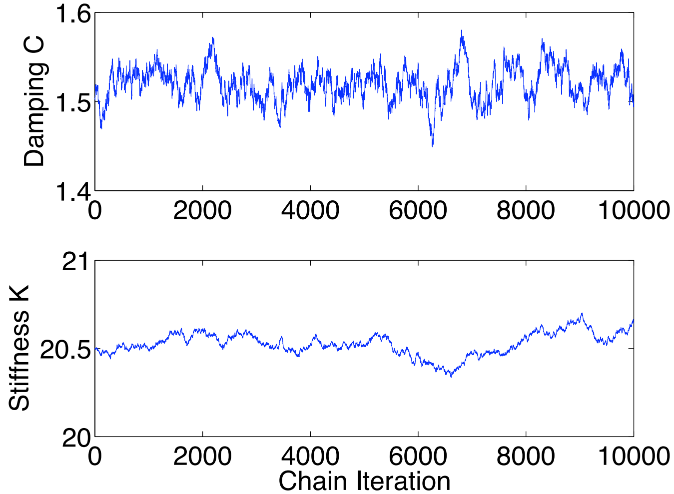


Random Walk Metropolis

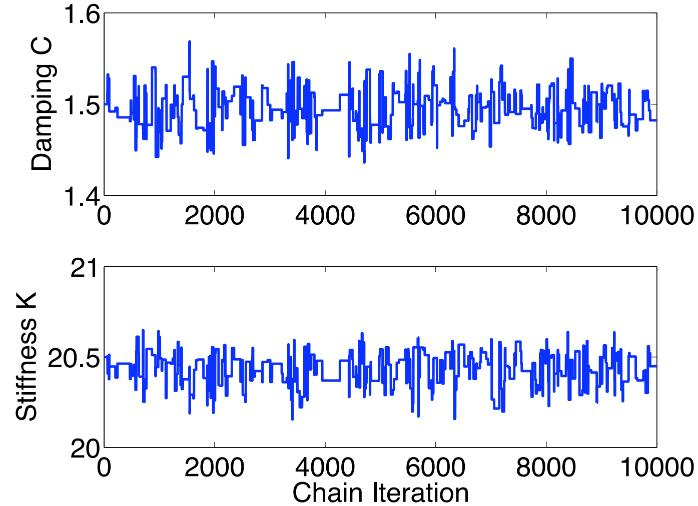
Case iii: Isotropic proposal function $J(\theta^* | \theta^{k-1}) = N(\theta^{k-1}, sI)$



$s = 9 \times 10^{-4}$



$s = 9 \times 10^{-6}$



$s = 9 \times 10^{-2}$

Stationary Distribution and Convergence Criteria

Here

$$\begin{aligned} p_{k-1,k} &= P(X_k = \theta^k | X_{k-1} = \theta^{k-1}) \\ &= P(\text{proposing } \theta^k) P(\text{accepting } \theta^k) \\ &= J(\theta^k|\theta^{k-1}) \alpha(\theta^k|\theta^{k-1}) \\ &= J(\theta^k|\theta^{k-1}) \min\left(1, \frac{\pi(\theta^k|y) J(\theta^{k-1}|\theta^k)}{\pi(\theta^{k-1}|y) J(\theta^k|\theta^{k-1})}\right) \end{aligned}$$

Detailed Balance Condition:

$$\pi_{k-1} p_{k-1,k} = \pi_k p_{k,k-1}$$

$$\Rightarrow \pi(\theta^{k-1}|y) p_{k-1,k} = \pi(\theta^k|y) p_{k,k-1}$$

From relation

$$y \min(1, x/y) = \min(x, y) = x \min(1, y/x)$$

it follows that

$$\begin{aligned} \pi(\theta^{k-1}|y) p_{k-1,k} &= \pi(\theta^{k-1}|y) J(\theta^k|\theta^{k-1}) \min\left(1, \frac{\pi(\theta^k|y) J(\theta^{k-1}|\theta^k)}{\pi(\theta^{k-1}|y) J(\theta^k|\theta^{k-1})}\right) \\ &= \pi(\theta^k|y) J(\theta^{k-1}|\theta^k) \min\left(1, \frac{\pi(\theta^{k-1}|y) J(\theta^k|\theta^{k-1})}{\pi(\theta^k|y) J(\theta^{k-1}|\theta^k)}\right) \\ &= \pi(\theta^k|y) p_{k,k-1} \end{aligned}$$

Delayed Rejection Adaptive Metropolis (DRAM)

Adaptive Metropolis:

• Update chain covariance matrix as chain values are accepted.

$$V_k = s_p \mathrm{cov}(q^0, q^1, \cdots, q^{k-1}) + \varepsilon I_p$$

• *Diminishing adaptation* and *bounded convergenc*e required since no longer Markov chain.

• Employ recursive relations

$$\begin{aligned} \bar{\theta}^{k+1} &= \frac{1}{k+1} \sum_{i=0}^{k} \theta^i \\ &= \frac{k}{k+1} \cdot \frac{1}{k} \sum_{i=0}^{k-1} \theta^i + \frac{1}{k+1} \theta^k \\ &= \frac{k}{k+1} \bar{\theta}^k + \frac{1}{k+1} \theta^k \end{aligned}$$

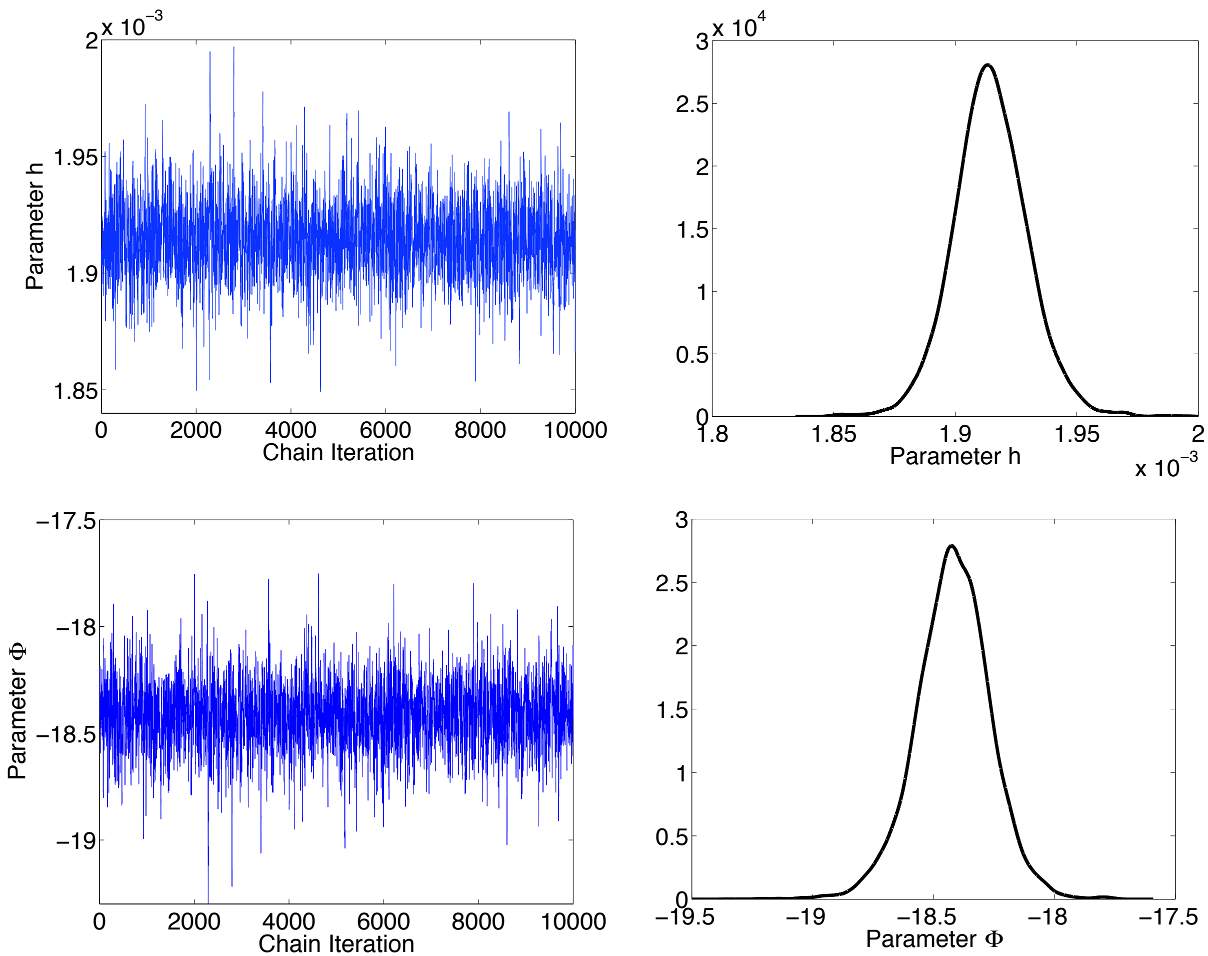
$$V_{k+1} = \frac{k-1}{k} V_k + \frac{s_p}{k} \left[k\bar{\theta}^{k-1}(\bar{\theta}^{k-1})^T - (k+1)\bar{\theta}^k(\bar{\theta}^k)^T + \theta^k(\theta^k)^T + \varepsilon I_p \right]$$

Delayed Rejection Adaptive Metropolis (DRAM)

Example: Heat model

$$\frac{d^2 T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}\left[T_{amb} - T_s(L)\right]$$



Bayesian Analysis

$\sigma = 0.2604$

$\sigma_\Phi = 0.1552$

$\sigma_h = 1.5450 \times 10^{-5}$

Frequentist Analysis

$\sigma = 0.2504$

$\sigma_\Phi = 0.1450$

$\sigma_h = 1.4482 \times 10^{-5}$

Delayed Rejection Adaptive Metropolis (DRAM)

Example: HIV model

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

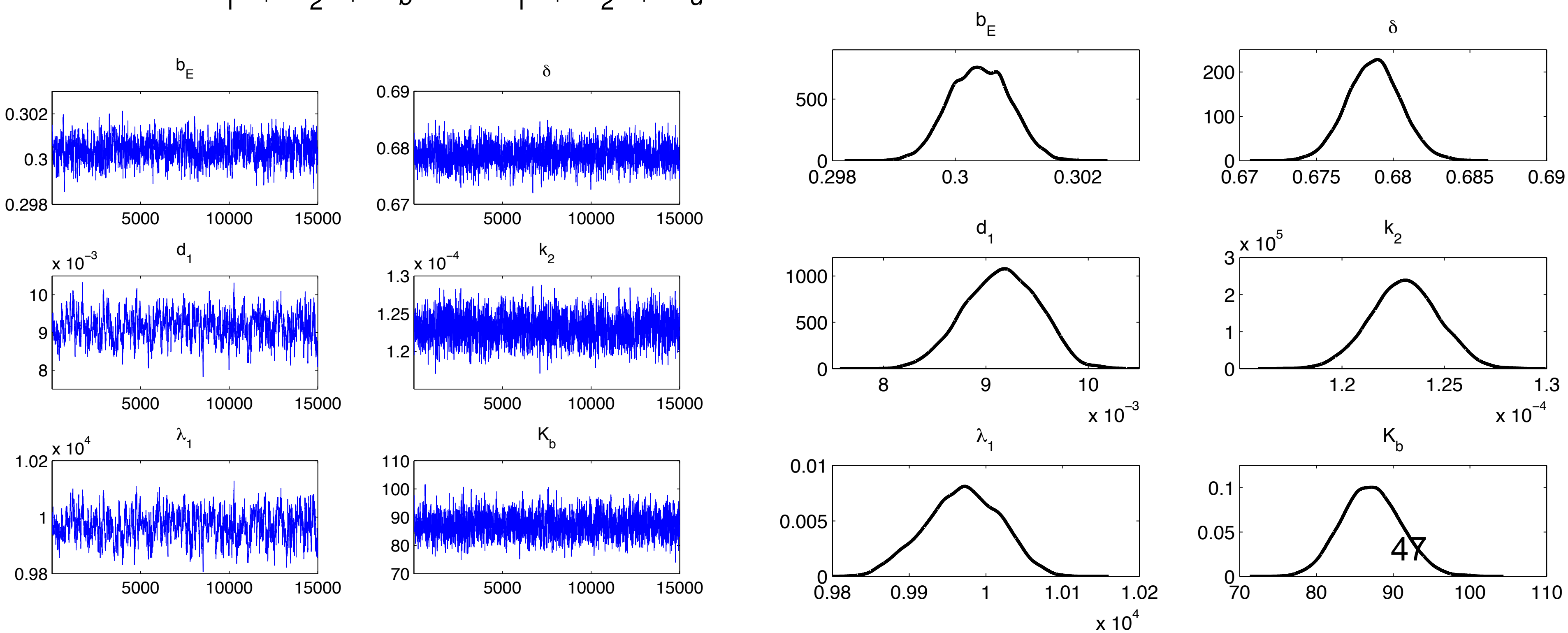
$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

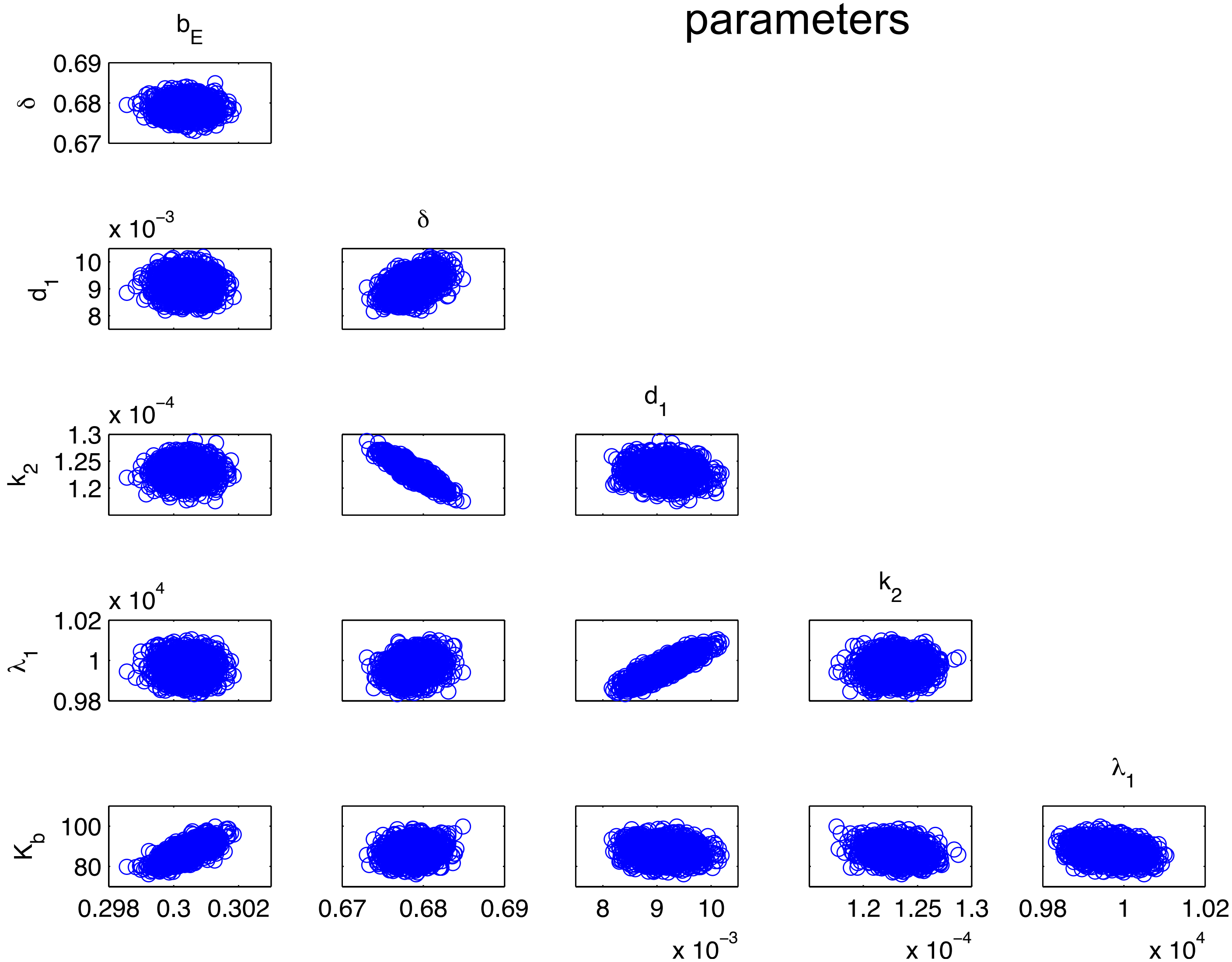
$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E.$$



Delayed Rejection Adaptive Metropolis (DRAM)

Example: HIV model

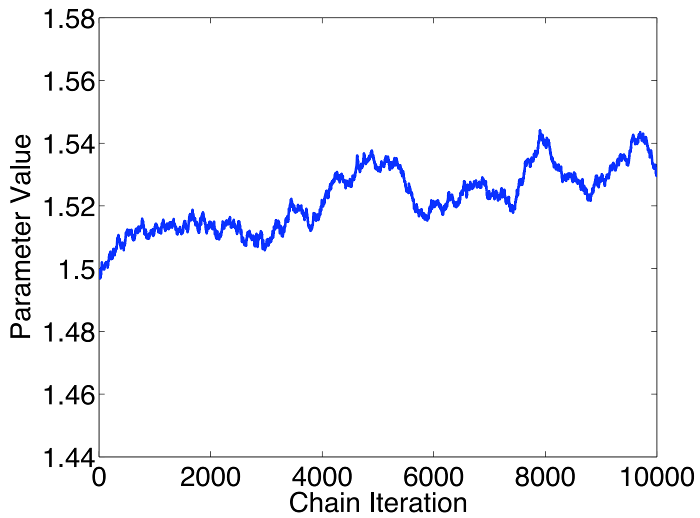
Note: Correlated versus nonidentifable parameters



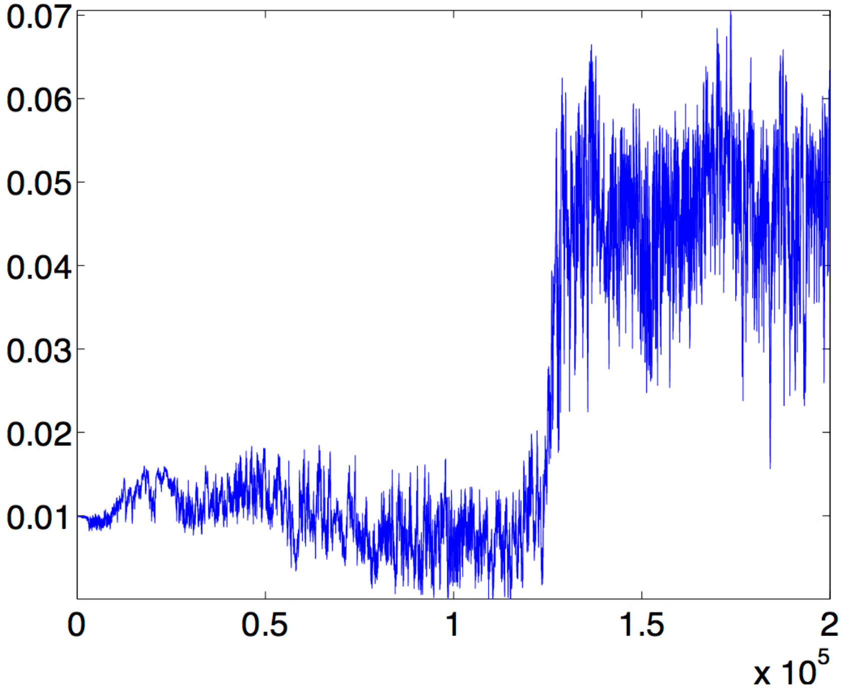
Chain Convergence (Burn-In)

Techniques:

- Visually check chains
- Statistical tests
- Often abused in the literature



Chain not converged



Chain for nonidentifiable parameter

Effects of Parameter Non-identifiability: Section 12.4.2

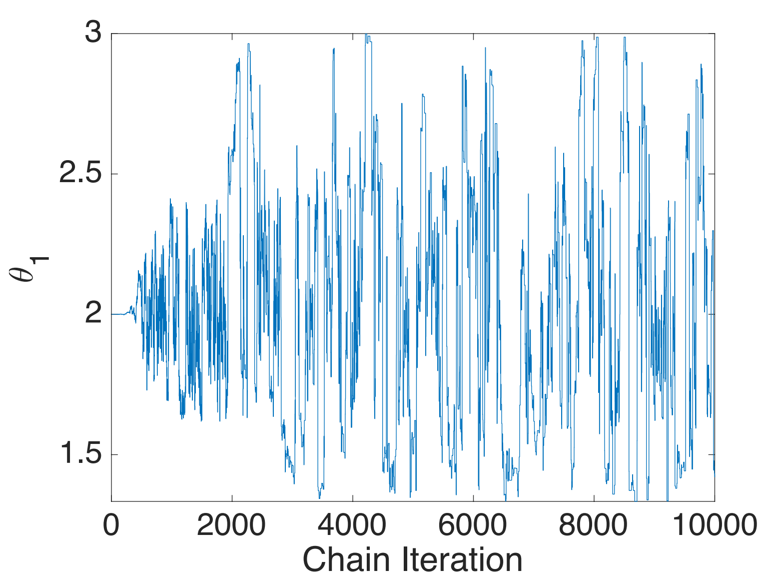
Example 12.13:

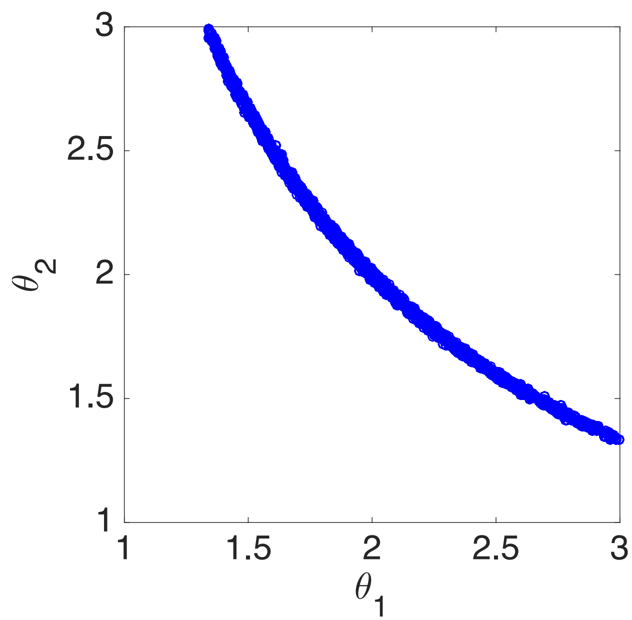
$$y_i = \theta_1 \theta_2 t_i + \varepsilon_i \ , \ \ i = 1, \ldots, n$$

Parameter values: $\theta_1 = \theta_2 = 2$

Times: $t_i \in [0, 1]$

Prior: $\mathcal{U}^2(4/3, 3)$





Note: Non-identifiable on manifold $h(\theta_{sub}) = 4 - \theta_1 \theta_2$

Effects of Parameter Non-identifiability: Section 12.4.2

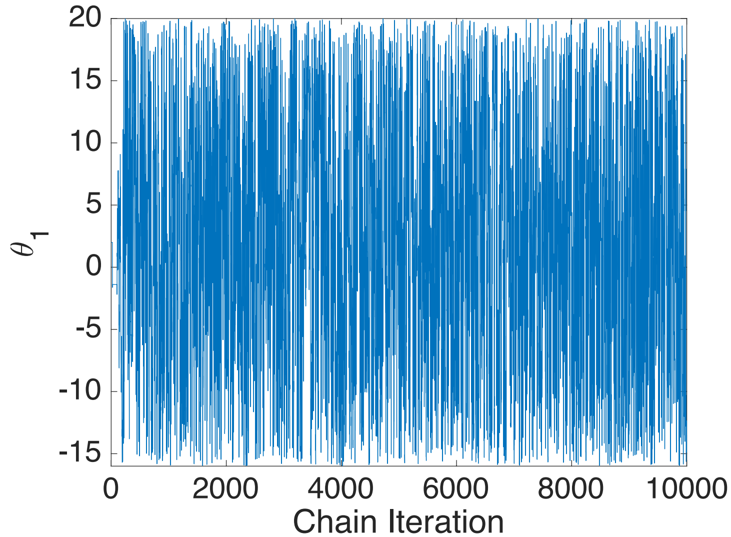
Example 12.13:

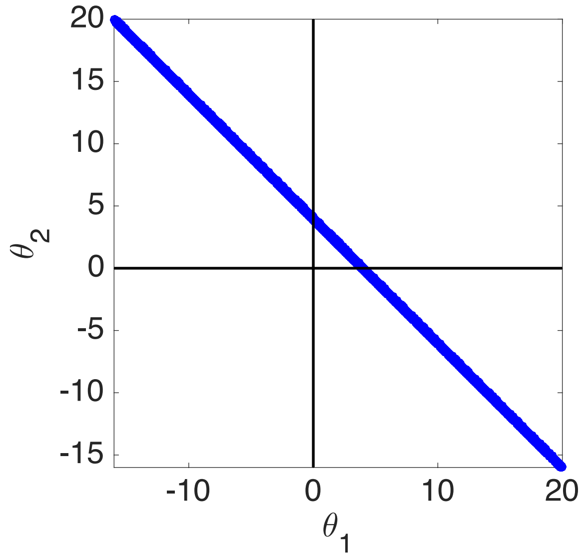
$$y_i = (\theta_1 + \theta_2)t_i + \varepsilon_i \;,\; i = 1, \ldots, n$$

Parameter values: $\theta_1 = \theta_2 = 2$

Times: $t_i \in [0, 1]$

Prior: $\mathcal{U}^2(-17, 20)$ so that $\theta_1 + \theta_2 = 4$ at endpoints.





Note: Non-identifiable on linear manifold $h(\theta_{sub}) = 4 - (\theta_1 + \theta_2)$

Large-Scale Example: Wetland Methane Emission Model

Example 12.22: [Susiluoto et al, ``Calibrating the sqHIMMELI v.1.0 wetland methane emission model with hierarchical modeling and adaptive MCMC," Geoscientific Model Development, 11, pp.1199--1228, 2018].

Compartment Model:

$$\frac{\partial [CH_4]}{\partial t}(t,z) = -T_{CH_4} + R_{CH_4}^{\text{exu}} + R_{CH_4}^{\text{peat}} - R_{CH_4}^{\text{oxid}}$$

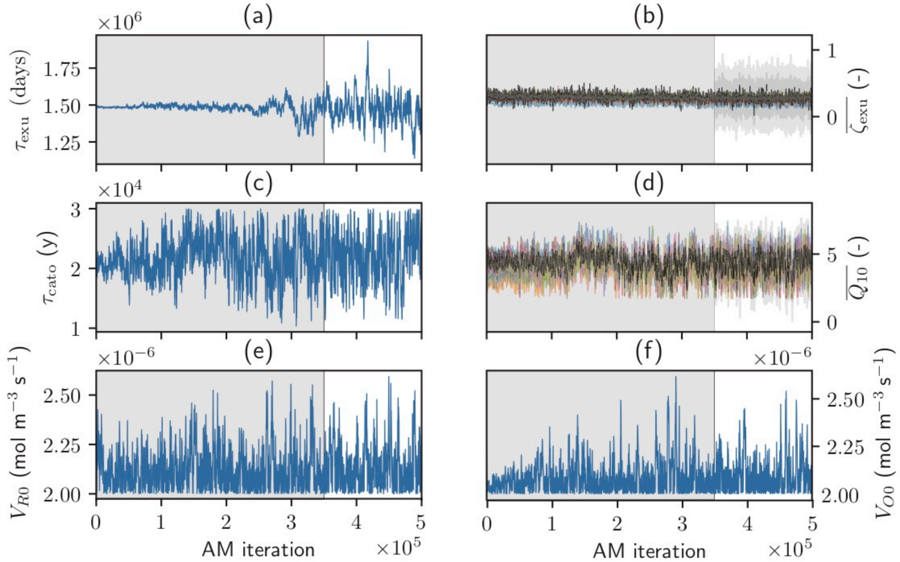
$$\frac{\partial [O_2]}{\partial t}(t,z) = -T_{O_2} - R_{\text{aerob}}^{\text{peat}} - R_{CO_2}^{\text{exu}} - 2R_{CH_4}^{\text{oxid}}$$

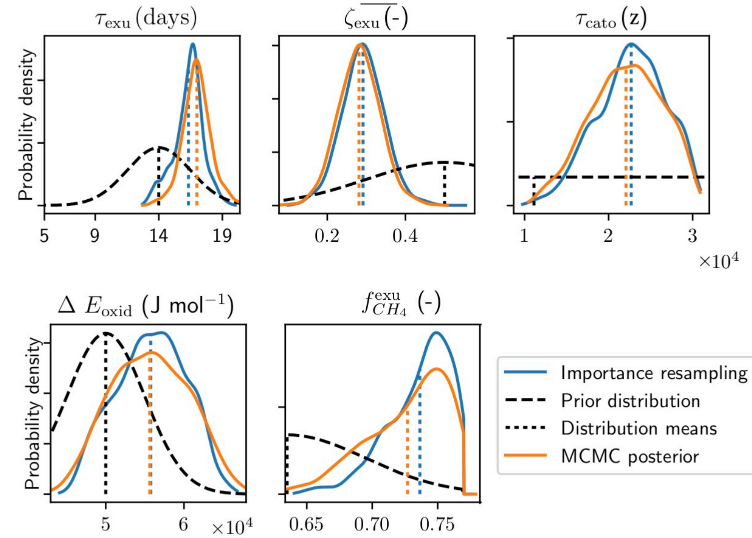
$$\frac{\partial [CO_2]}{\partial t}(t,z) = -T_{CO_2} + R_{CO_2}^{\text{exu}} + R_{CO_2}^{\text{peat}} + R_{CH_4}^{\text{oxid}} + R_{\text{aerob}}^{\text{peat}}$$

Representative Constitutive Relation:

$$R_{CH_4}^{\text{peat}}(z) = k_{\text{cato}}(z) g_{CH_4}^{Q_{10}} \frac{\rho_{\text{cato}} f_{C_{\text{cato}}}}{M_C}$$

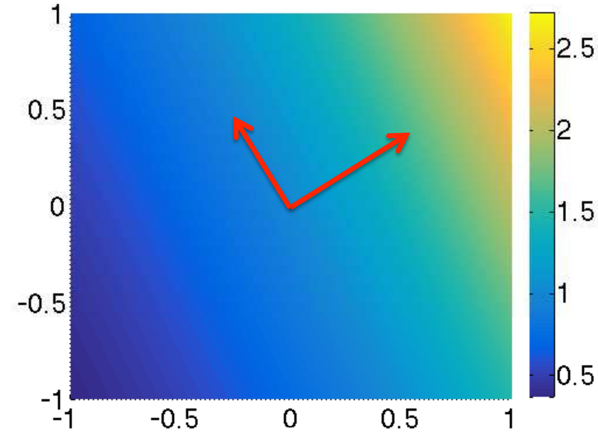
Large-Scale Example: Wetland Methane Emission Model





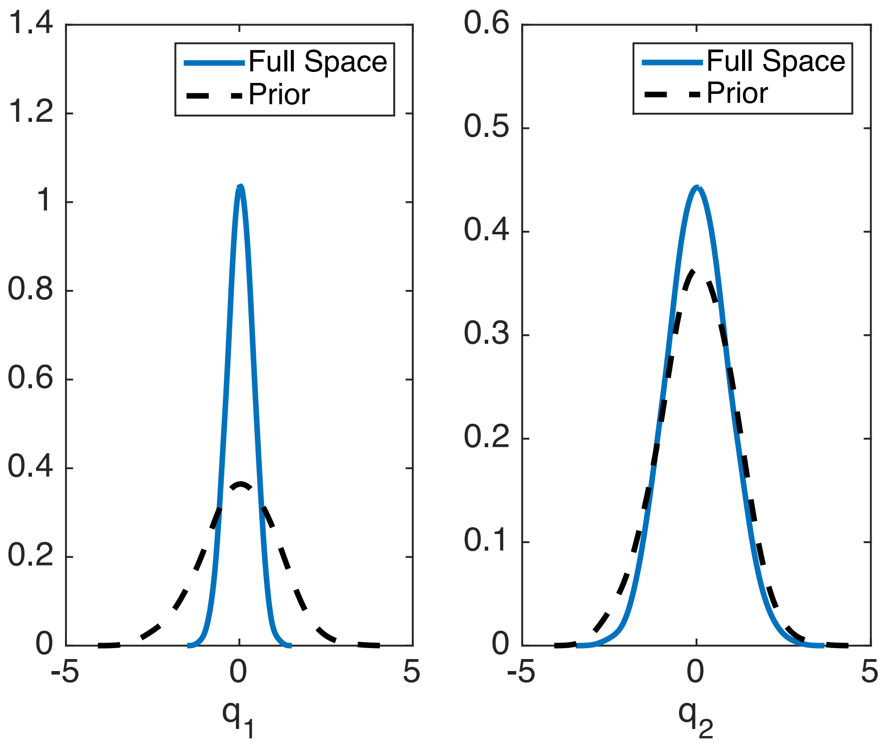
Bayesian Inference on Active Subspace

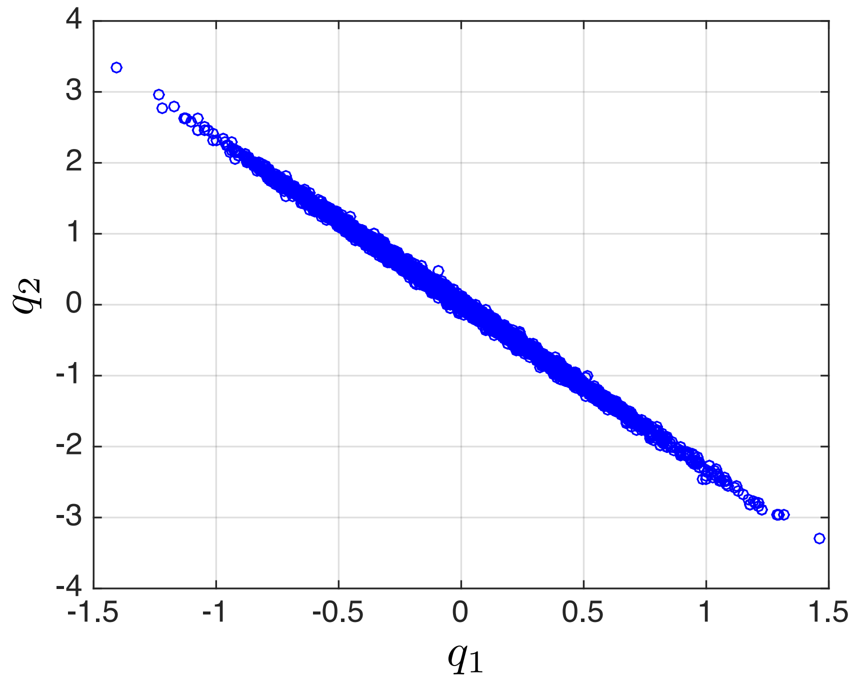
Example: $y = \exp(0.7\theta_1 + 0.3\theta_2)$



Full Space Inference:

- Parameters not jointly identifiable
- Result: Prior for 2nd parameter is minimally informed.
- Goal: Use active subspace to quantify parameter sensitivity and guide inference.





Bayesian Inference on Active Subspaces

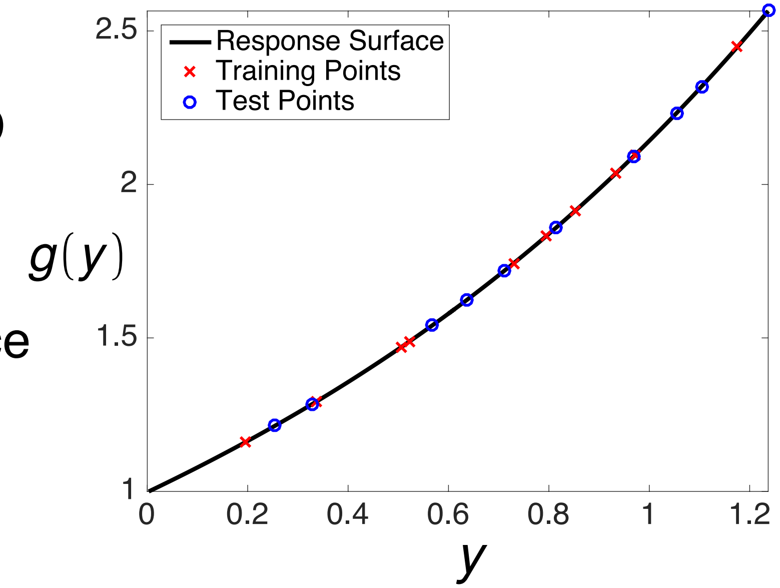
Example: $y = \exp(0.7\theta_1 + 0.3\theta_2)$

Active Subspace: For gradient matrix G, form SVD

$$G = U \Lambda V^T$$

Eigenvalue spectrum indicates 1-D active subspace with basis

$$U(:,1) = [0.91\ ,\ 0.39]$$

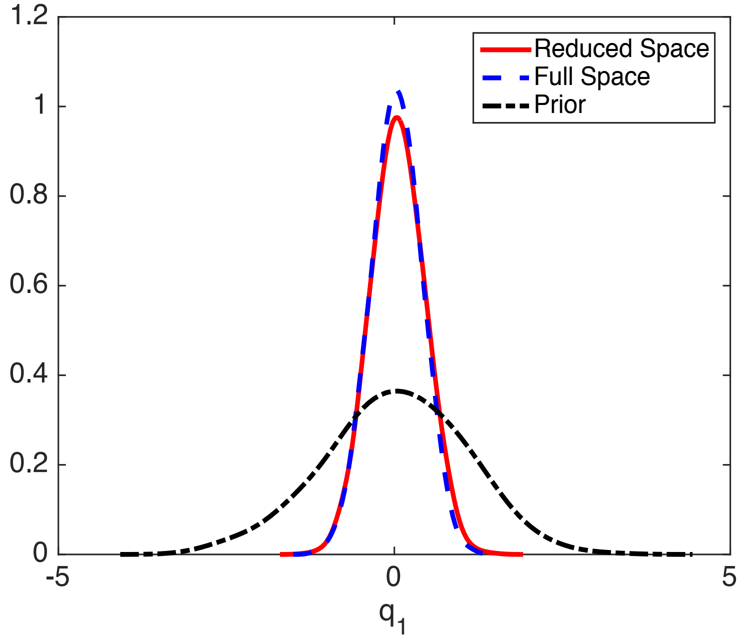


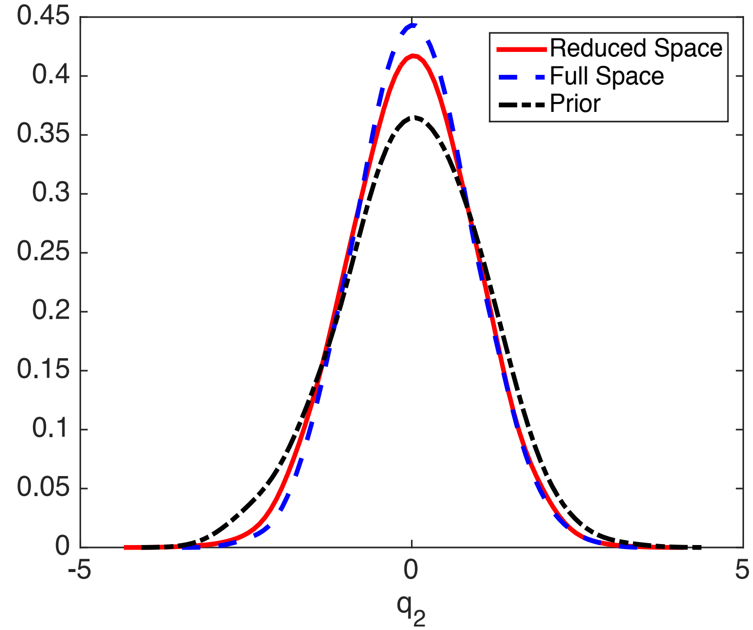
Strategy: Inference based on active subspace

- For values $\{\theta^j\}_{j=1}^N$, compute $y^j = U(:,1)^T\theta^j$ and fit response surface
- Perform Bayesian inference for y
- Because model is "invariant" to $z = U(:,2)^T\theta$, draw $\{z^j\} \sim N(0,1)$
- Transform to $\theta^j = U(:,1)y^j + U(:,2)z^j$ to obtain posterior densities for physical parameters

Bayesian Inference on Active Subspaces

Results: Inference based on active subspace





Global Sensitivity: For active subspace of dimension n, consider vector of activity scores

$$\alpha_i(n) = \sum_{j=1}^{n} \lambda_j w_{i,j}^2 \;,\; i = 1, \ldots, p$$

Present Example: Here n = 1 and $w_1^2 = U(:,1) .* U(:,1) = [0.91^2 \;,\; 0.39^2]$

Conclusion: First parameter is more influential and better informed during Bayesian inference.

Bayesian Inference on Active Subspaces

Example: Family of elliptic PDE's

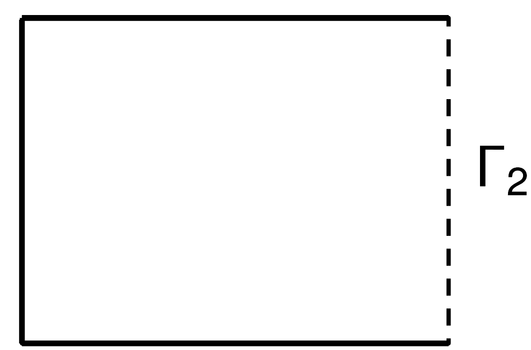
$$-\nabla_s \cdot (a(q,s,\ell)\nabla_s u(s,a(q,s,\ell))) = 1 \ , \ s = [0,1]^2 \ , \ \ell = 1,\cdots,n$$

with the random field representations

$$a(q,s,\ell) = a_{\min} + e^{\overline{a}(s,\ell)+\sum_{i=1}^{p} q_k^\ell \gamma_i \phi_i(s)}$$

Quantity of interest: e.g., strain along edge at n levels

$$f\left(\mathbf{q}^1,\ldots,\mathbf{q}^n\right) \approx \sum_{\ell=1}^{n} \frac{1}{|\Gamma_2|}\int_{\Gamma_2} u(q,s,\ell)ds$$

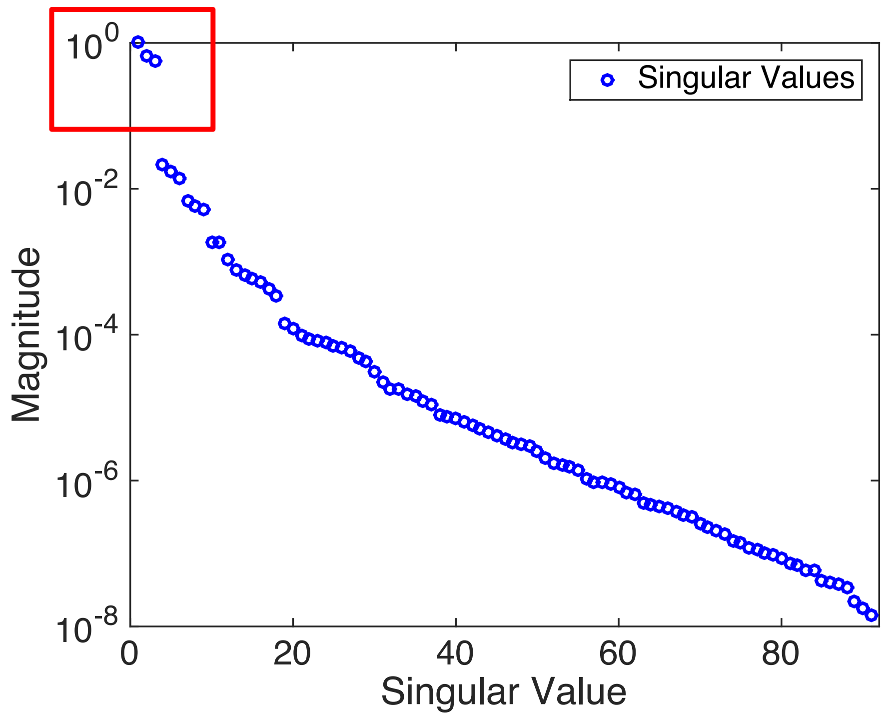


Problem Dimensions:

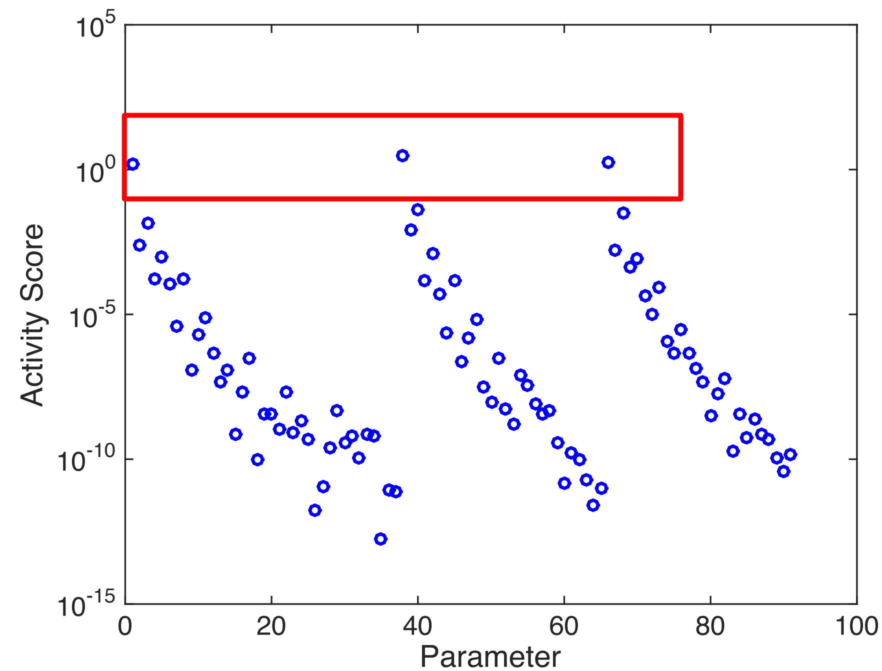
- Parameter dimension: p = 91
- Active subspace dimension: n = 3
- Finite element approximation

Bayesian Inference on Active Subspaces

Singular Values: Recall n = 3



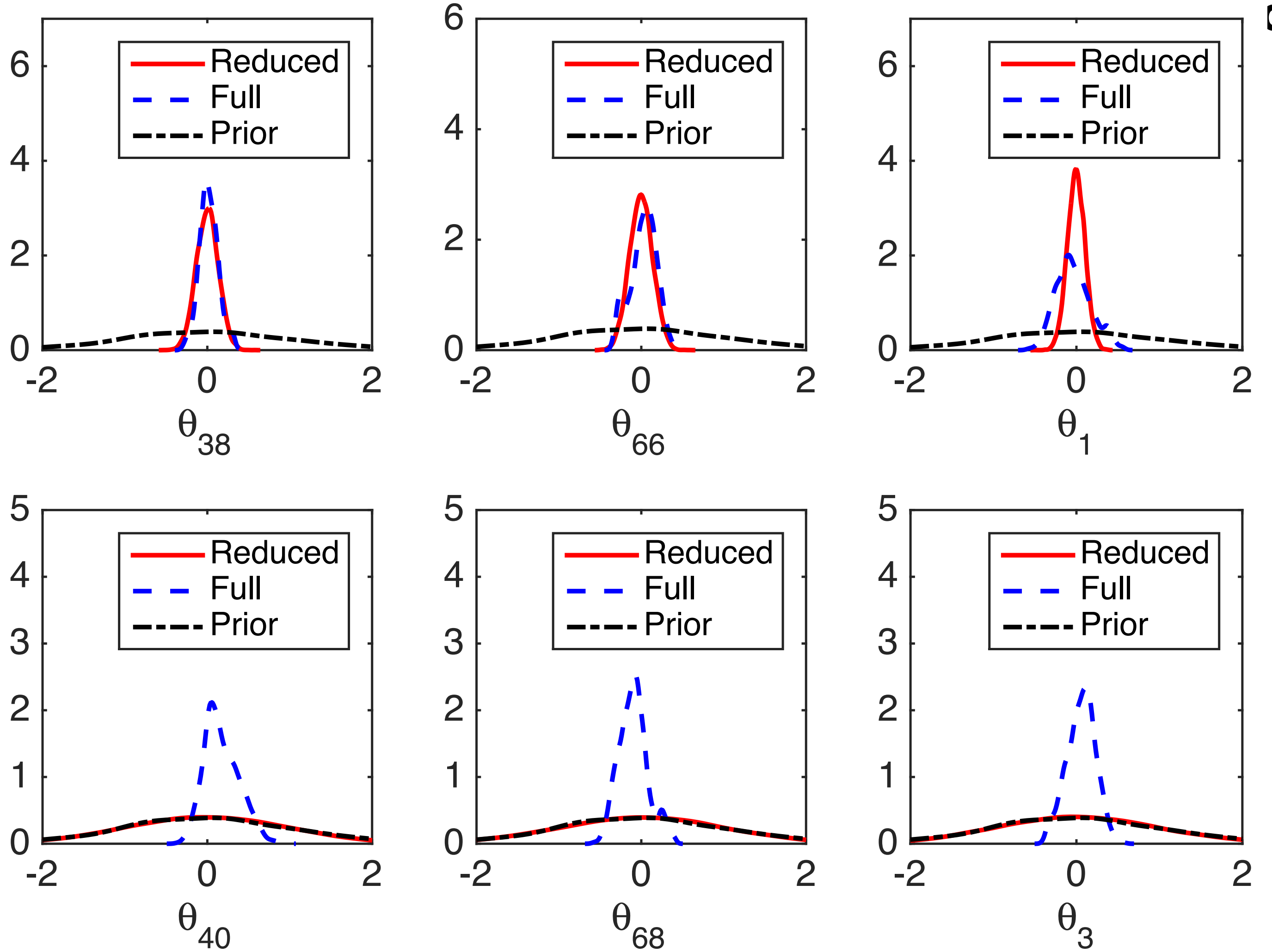
Activity Scores: Quantify global sensitivity



Conclusion: Parameters 1, 38, 66 are most influential and will be primarily informed during Bayesian inference

Bayesian Inference on Active Subspaces

Recall: Parameters 1, 38, 66 are most influential and will be primarily informed during Bayesian inference



ote:

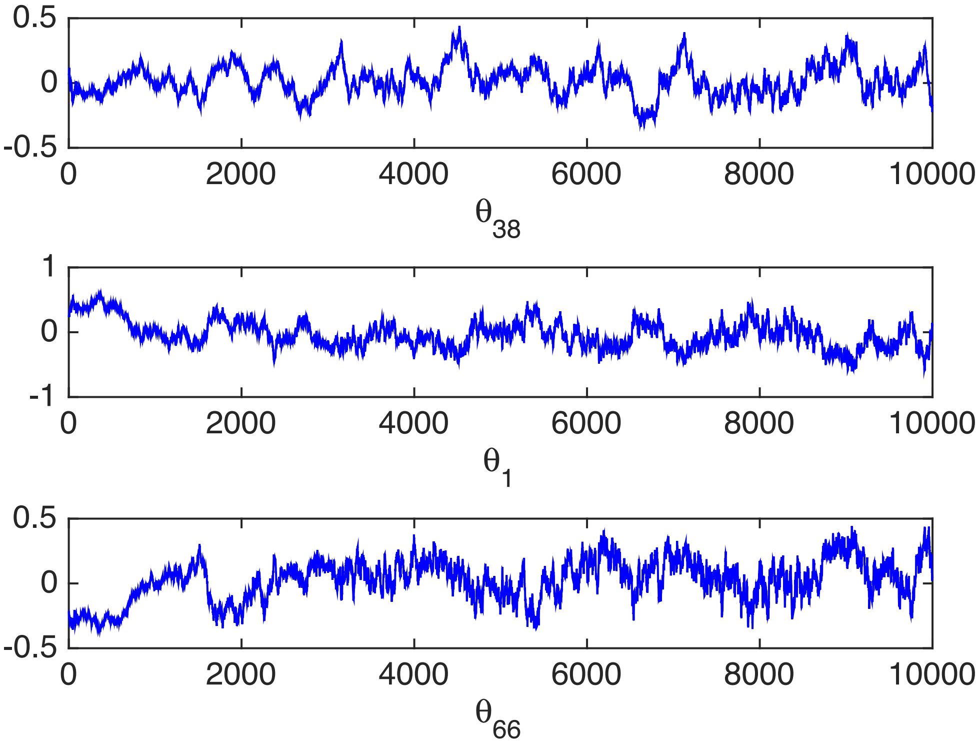
Full space: 18 hours

Reduced: 20 seconds

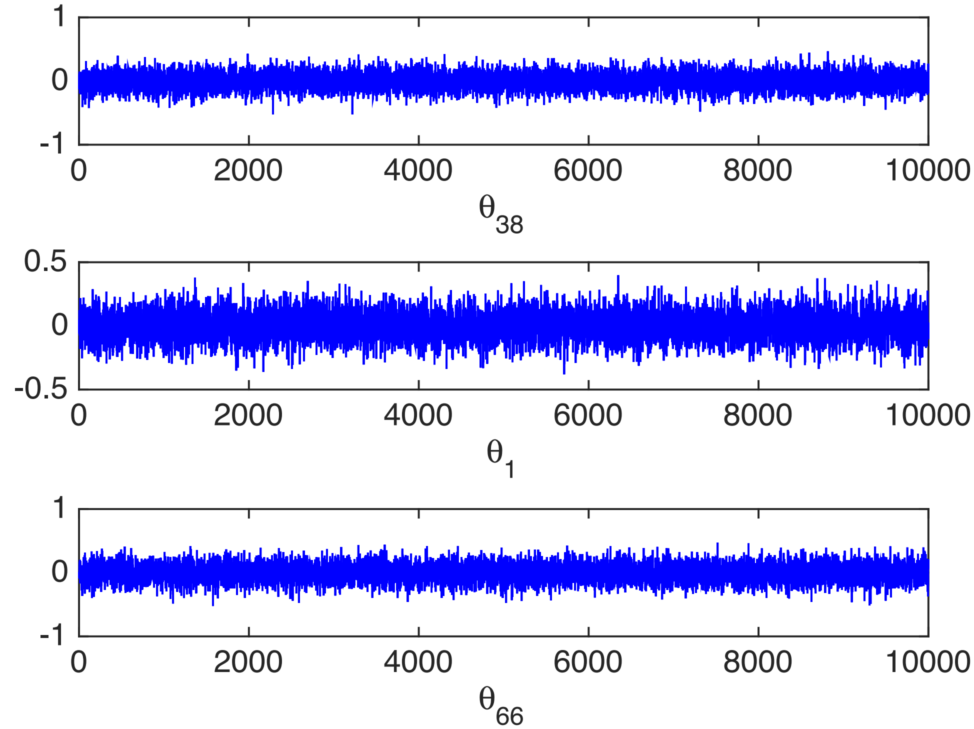
Bayesian Inference on Active Subspaces

Note:

- Chains for full space not converging well due to parameter nonidentifiability
- Hence full space inference is less reliable



Full Space



Active Subspace

Delayed Rejection Adaptive Metropolis (DRAM)

Websites

- http://www4.ncsu.edu/~rsmith/UQ_TIA/CHAPTER8/index_chapter8.html
- http://helios.fmi.fi/~lainema/mcmc/

Examples

- Examples on using the toolbox for some statistical problems.

Delayed Rejection Adaptive Metropolis (DRAM)

We fit the Monod model

$$y = \theta_1 \frac{1}{\theta_2 + 1} + \epsilon \quad , \quad \epsilon \sim N(0, I\sigma^2)$$

to observations

x (mg / L COD): 28 55 83 110 138 225 375

y (1 / h): 0.053 0.060 0.112 0.105 0.099 0.122 0.125

First clear some variables from possible previous runs.

```
clear data model options
```

Next, create a data structure for the observations and control variables. Typically one could make a structure data that contains fields xdata and ydata.

```
data.xdata = [28 55 83 110 138 225 375]'; % x (mg / L COD)
data.ydata = [0.053 0.060 0.112 0.105 0.099 0.122 0.125]'; % y (1 / h)
```

Construct model

```
modelfun = @(x,theta) theta(1)*x./(theta(2)+x);
ssfun = @(theta,data) sum((data.ydata-modelfun(data.xdata,theta)).^2);
model.ssfun = ssfun;
model.sigma2 = 0.01^2;
```

Delayed Rejection Adaptive Metropolis (DRAM)

Input parameters

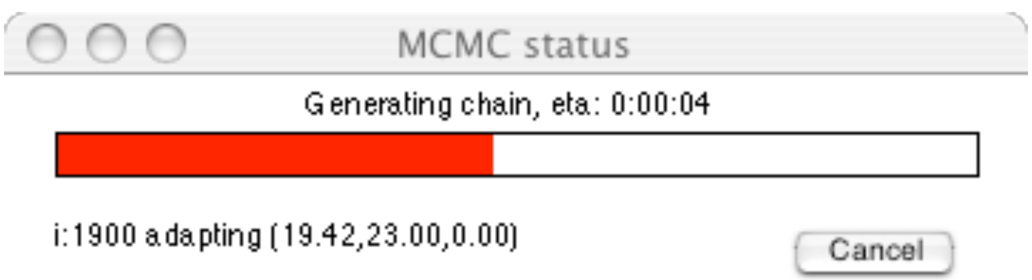
```
params = {
    {'theta1', tmin(1), 0}
    {'theta2', tmin(2), 0} };
```

and set options

```
options.nsimu = 4000;
options.updatesigma = 1;
options.qcov = tcov;
```

Run code

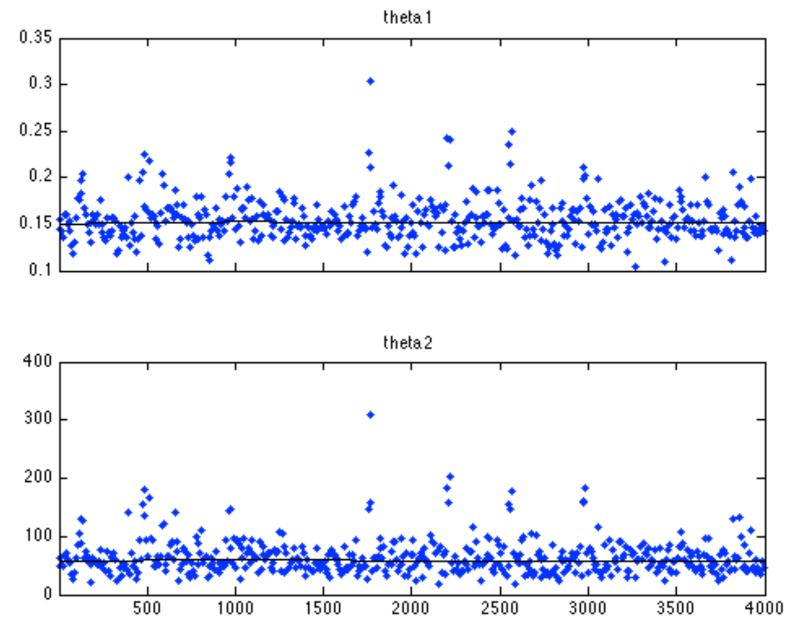
```
[res,chain,s2chain] = mcmcrun(model,data,params,options);
```

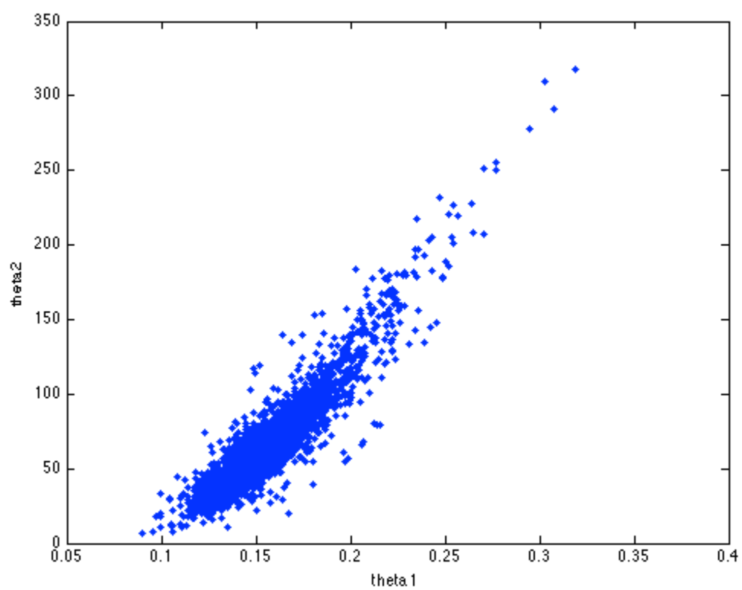


Delayed Rejection Adaptive Metropolis (DRAM)

Plot results

```
figure(2); clf
mcmcplot(chain,[],res,'chainpanel');
figure(3); clf
mcmcplot(chain,[],res,'pairs');
```





Examples:

- Several available in MCMC_EXAMPLES
- ODE solver illustrated in algae example

Delayed Rejection Adaptive Metropolis (DRAM)

Construct credible and prediction intervals

```
figure(5); clf
out = mcmcpred(res,chain,[],x,modelfun);
mcmcpredplot(out);
hold on
plot(data.xdata,data.ydata,'s'); % add data points to the plot
xlabel('x [mg/L COD]');
ylabel('y [1/h]');
hold off
title('Predictive envelopes of the model')
```