Multiproposal Elliptical Slice Sampling

Guillermina Senn[1,2]

with N. Glatt-Holtz[1], G. Carigi[1], A. Holbrook[3], H. Tjelmeland[2]

[1]Indiana University Bloomington

[2]NTNU

[3]UCLA

SIAM Uncertainty Quantification

Minneapolis, 2026 March 24

NTNU
Norwegian University of Science and Technology

Summary

Goal: Scalable MCMC for high-dimensional posteriors with form

$$\pi(x|y) \propto \mathcal{L}(y|x)\,\mathcal{N}(x;0,C).$$

Summary

Goal: Scalable MCMC for high-dimensional posteriors with form

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x; 0, C).$$

Preconditioned Crank–Nicolson → Elliptical Slice Sampling → Multiproposal Elliptical Slice Sampling → High-dimensional Bayesian inversion

no gradients

dimension-robust

self-tuning

parallelizable

Summary

Goal: Scalable MCMC for high-dimensional posteriors with form

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x;0,C).$$

Preconditioned Crank–Nicolson → Elliptical Slice Sampling → Multiproposal Elliptical Slice Sampling → High-dimensional Bayesian inversion

Preconditioned Crank–Nicolson: no gradients, dimension-robust

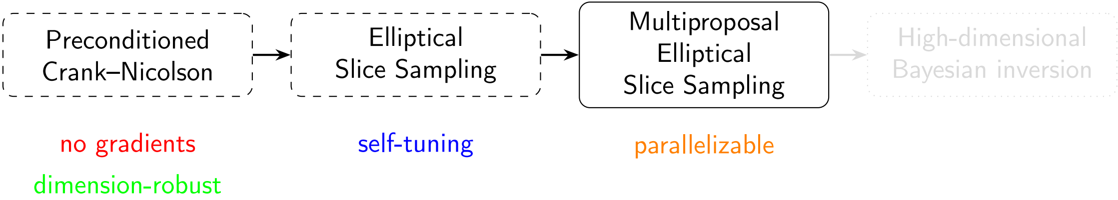
Elliptical Slice Sampling: self-tuning

Multiproposal Elliptical Slice Sampling: parallelizable

Summary

Goal: Scalable MCMC for high-dimensional posteriors with form

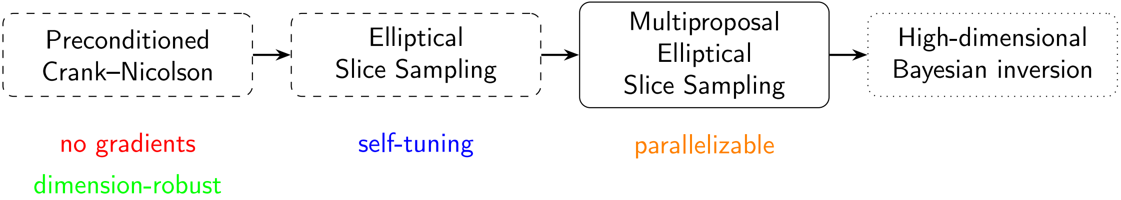
$$\pi(x|y) \propto \mathcal{L}(y|x)\,\mathcal{N}(x; 0, C).$$



Summary

Goal: Scalable MCMC for high-dimensional posteriors with form

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x; 0, C).$$



Motivation

- Bayesian inverse problem

$$\pi(x|\mathrm{y}) \propto \pi(\mathrm{y}|x)\,\pi(x), \quad \mathrm{y} \in \mathbb{R}^k \text{ (observations)}, \quad x \in \mathbb{R}^d \text{ (parameters)}$$

- Often non-parametric:

$$\{x(s) : s \in D\}, \quad D \subset \mathbb{R}^d.$$

- Sampling with Markov models:
  - Construct ergodic Markov chain $X = \{X_1, X_2, \ldots, X_N\}$ with π as invariant measure.
  - Approximate expectations:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(X_i) = \hat{\mathbb{E}}_\pi[f(X)]$$

Motivation

- Bayesian inverse problem

$$\pi(x|\mathrm{y}) \propto \pi(\mathrm{y}|x)\,\pi(x), \quad \mathrm{y} \in \mathbb{R}^k \text{ (observations)}, \quad x \in \mathbb{R}^d \text{ (parameters)}$$

- Often non-parametric:

$$\{x(s) : s \in D\}, \quad D \subset \mathbb{R}^d.$$

- Sampling with Markov models:
  - Construct ergodic Markov chain $X = \{X_1, X_2, \ldots, X_N\}$ with π as invariant measure.
  - Approximate expectations:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \hat{\mathbb{E}}_\pi[f(X)]$$

Motivation

- Bayesian inverse problem

$$\pi(x|\mathrm{y}) \propto \pi(\mathrm{y}|x)\,\pi(x), \quad \mathrm{y} \in \mathbb{R}^k \text{ (observations)}, \quad x \in \mathbb{R}^d \text{ (parameters)}$$

- Often non-parametric:

$$\{x(s) : s \in D\}, \quad D \subset \mathbb{R}^d.$$

- Sampling with Markov models:
  - Construct ergodic Markov chain $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$ with π as invariant measure.
  - Approximate expectations:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \hat{\mathbb{E}}_\pi[f(X)]$$

Motivation

- Bayesian inverse problem

$$\pi(x|\mathrm{y}) \propto \pi(\mathrm{y}|x)\,\pi(x), \quad \mathrm{y} \in \mathbb{R}^k \text{ (observations)}, \quad x \in \mathbb{R}^d \text{ (parameters)}$$

- Often non-parametric:

$$\{x(s) : s \in D\}, \quad D \subset \mathbb{R}^d.$$

- Sampling with Markov models:
  - Construct ergodic Markov chain $\boldsymbol{X} = \{X_1, X_2, \dots, X_N\}$ with π as invariant measure.
  - Approximate expectations:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \hat{\mathbb{E}}_\pi[f(X)]$$

Motivation

- Bayesian inverse problem

$$\pi(x|\mathrm{y}) \propto \pi(\mathrm{y}|x)\,\pi(x), \quad \mathrm{y} \in \mathbb{R}^k \text{ (observations)}, \quad x \in \mathbb{R}^d \text{ (parameters)}$$

- Often non-parametric:

$$\{x(s) : s \in D\}, \quad D \subset \mathbb{R}^d.$$

- Sampling with Markov models:
  - Construct ergodic Markov chain $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$ with π as invariant measure.
  - Approximate expectations:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \hat{\mathbb{E}}_\pi[f(X)]$$

Markov chain Monte Carlo (MCMC)

Algorithm 1: Metropolis-Hastings (Hastings, 1970)

Input: Current state x_i, target density $\pi(x)$, proposal density $q(\cdot|x)$.

1. Sample $x^\star \sim q(\cdot\,|x_i)$
2. Compute acceptance probability

$$\alpha(x_i, x^\star) = \min\{1, \frac{\mathcal{L}(\mathrm{y}|x^\star)}{\mathcal{L}(\mathrm{y}|x_i)} \times \frac{\pi(x^\star)}{\pi(x_i)} \times \frac{q(x_i|x^\star)}{q(x^\star|x_i)}\}$$

3. Accept $x^\star$ with probability $\alpha(x_i, x^\star)$, otherwise reject.

$$\text{total cost}_\varepsilon = \text{cost per iteration} \times \text{number of iterations}_\varepsilon.$$

- Both increase with input size.

Pre-conditioned Crank-Nicolson (pCN)

Target[1]:

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x;0,C).$$

Algorithm 2: pCN (Cotter et al., 2013; Neal, 1999)

Input: Current state x, hyperpar $\rho \in (0,1)$.

1. Propose $x^\star = \rho\, x + \sqrt{1-\rho^2}\,\nu, \quad \nu \sim \mathcal{N}(0, C)$ no gradients
2. Compute $\alpha(x, x^\star) = \min\{1, \mathcal{L}(\mathrm{y}|x^\star)/\mathcal{L}(\mathrm{y}|x)\}$
3. Accept $x^\star$ with probability $\alpha(x, x^\star)$, otherwise reject.

Proposal is reversible with respect to the prior:

$$\mathcal{N}(x;0,C)\mathcal{N}(x^\star;\rho\,x,(1-\rho^2)C) = \mathcal{N}(x^\star;0,C)\mathcal{N}(x;\rho\,x^\star,(1-\rho^2)C) \qquad \Rightarrow \text{dimension-free}$$

[1] Stuart (2010); Dashti and Stuart (2013)

Pre-conditioned Crank-Nicolson (pCN)

Target[1]:

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x; 0, C).$$

Algorithm 2: pCN (Cotter et al., 2013; Neal, 1999)

Input: Current state x, hyperpar $\rho \in (0,1)$.

1. Propose $x^\star = \rho\, x + \sqrt{1-\rho^2}\,\nu, \quad \nu \sim \mathcal{N}(0, C)$ no gradients
2. Compute $\alpha(x, x^\star) = \min\{1, \mathcal{L}(\mathrm{y}|x^\star)/\mathcal{L}(\mathrm{y}|x)\}$
3. Accept $x^\star$ with probability $\alpha(x, x^\star)$, otherwise reject.

Proposal is reversible with respect to the prior:

$$\mathcal{N}(x; 0, C)\mathcal{N}(x^\star; \rho\, x, (1-\rho^2)C) = \mathcal{N}(x^\star; 0, C)\mathcal{N}(x; \rho\, x^\star, (1-\rho^2)C) \qquad \Rightarrow \text{dimension-free}$$

[1] Stuart (2010); Dashti and Stuart (2013)

Pre-conditioned Crank-Nicolson (pCN)

Target[1]:

$$\pi(x|\mathrm{y}) \propto \mathcal{L}(\mathrm{y}|x)\,\mathcal{N}(x; 0, C).$$

Algorithm 2: pCN (Cotter et al., 2013; Neal, 1999)

Input: Current state x, hyperpar $\rho \in (0, 1)$.

1. Propose $x^\star = \rho\, x + \sqrt{1-\rho^2}\,\nu, \quad \nu \sim \mathcal{N}(0, C)$ no gradients
2. Compute $\alpha(x, x^\star) = \min\{1, \mathcal{L}(\mathrm{y}|x^\star)/\mathcal{L}(\mathrm{y}|x)\}$
3. Accept $x^\star$ with probability $\alpha(x, x^\star)$, otherwise reject.

Proposal is reversible with respect to the prior:

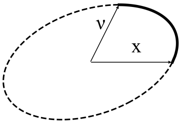
$$\mathcal{N}(x; 0, C)\mathcal{N}(x^\star; \rho\, x, (1-\rho^2)C) = \mathcal{N}(x^\star; 0, C)\mathcal{N}(x; \rho\, x^\star, (1-\rho^2)C) \qquad \Rightarrow \text{dimension-free}$$

[1] Stuart (2010); Dashti and Stuart (2013)

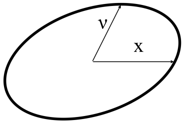
Elliptical Slice Sampling (ESS)[2]

- Elliptical proposal:

$$x^\star = x\cos\theta + \nu\sin\theta, \quad \nu \sim \mathcal{N}(0, C), \quad \theta \in (0, 2\pi].$$



(a) pCN: $x^\star = \rho x + \sqrt{1-\rho^2}\,\nu, \quad \rho \in (0,1)$



(b) ESS proposal

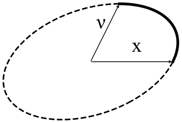
- Proposal holds the prior invariant.
- We can do better than elliptical MH $\rightarrow$ self-tuning.

[2] Murray et al. (2010)

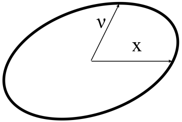
Elliptical Slice Sampling (ESS)[2]

- Elliptical proposal:

$$x^\star = x\cos\theta + \nu\sin\theta, \quad \nu \sim \mathcal{N}(0, C), \quad \theta \in (0, 2\pi].$$



(a) pCN: $x^\star = \rho x + \sqrt{1-\rho^2}\,\nu, \quad \rho \in (0,1)$



(b) ESS proposal

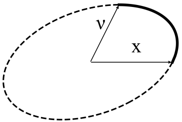
- Proposal holds the prior invariant.
- We can do better than elliptical MH → self-tuning.

[2] Murray et al. (2010)

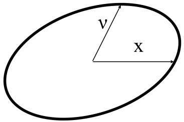
Elliptical Slice Sampling (ESS)[2]

- Elliptical proposal:

$$x^\star = x\cos\theta + \nu\sin\theta, \quad \nu \sim \mathcal{N}(0, C), \quad \theta \in (0, 2\pi].$$



(a) pCN: $x^\star = \rho x + \sqrt{1-\rho^2}\,\nu, \quad \rho \in (0,1)$



(b) ESS proposal

- Proposal holds the prior invariant.
- We can do better than elliptical MH → self-tuning.

[2]Murray et al. (2010)

Slice Sampling

- Augmented variable method.
- Idea:

$$\pi(x,y) = \begin{cases} \frac{1}{\int f(x)\,dx} & 0 < y < f(x) \\ 0 & \text{otherwise} \end{cases}$$

Algorithm 2: Slice sampling (Neal, 2003)

1. Sample $y \sim \mathcal{U}(0, f(x))$, defining *slice* $\mathcal{S} = \{x : f(x) \geq y\}$
2. Sample $x \sim \mathcal{U}(\mathcal{S})$

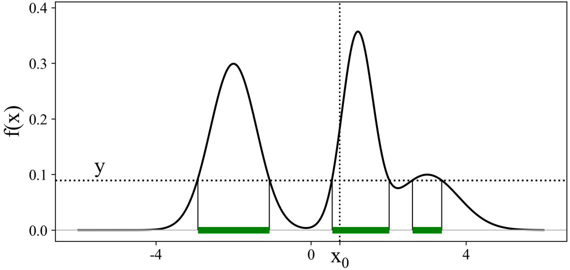


Figure: Slice sampling in 1D.

Slice Sampling

- Augmented variable method.
- Idea:

$$\pi(x,y) = \begin{cases} \frac{1}{\int f(x)\,dx} & 0 < y < f(x) \\ 0 & \text{otherwise} \end{cases}$$

Algorithm 3: Slice sampling (Neal, 2003)

1. Sample $y \sim \mathcal{U}(0, f(x))$, defining *slice* $\mathcal{S} = \{x : f(x) \geq y\}$
2. Sample $x \sim \mathcal{U}(\mathcal{S})$

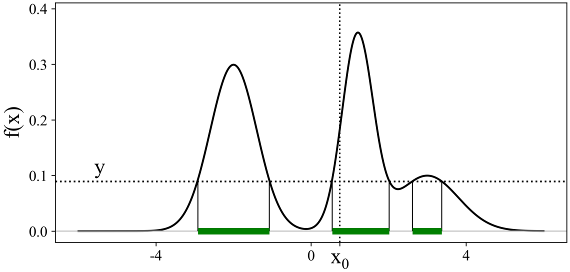


Figure: Slice sampling in 1D.

Slice Sampling

- Augmented variable method.
- Idea:

$$\pi(x, y) = \begin{cases} \frac{1}{\int f(x)\, dx} & 0 < y < f(x) \\ 0 & \text{otherwise} \end{cases}$$

Algorithm 3: Slice sampling (Neal, 2003)

1. Sample $y \sim \mathcal{U}(0, f(x))$, defining *slice* $\mathcal{S} = \{x : f(x) \geq y\}$
2. Sample $x \sim \mathcal{U}(\mathcal{S})$

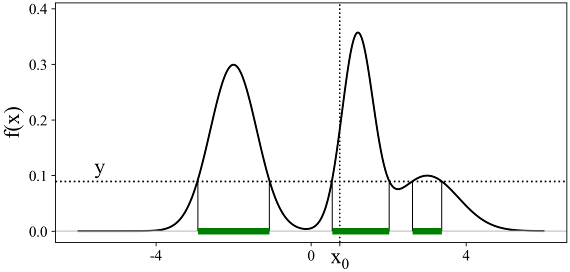
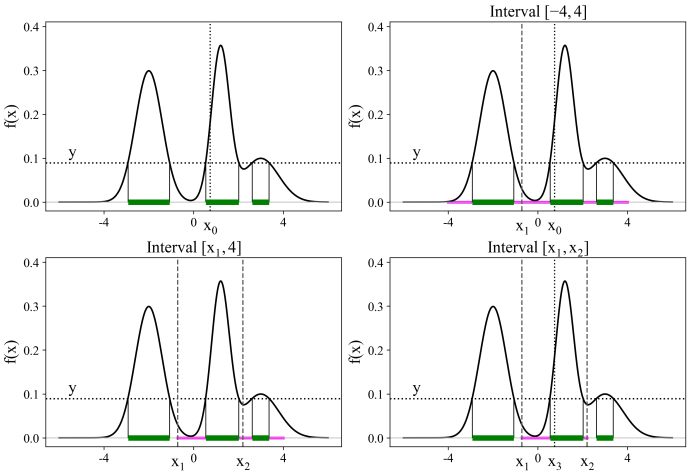


Figure: Slice sampling in 1D.

Shrinking

- How to sample $x \sim \mathcal{U}(\mathcal{S})$ when $\mathcal{S}$ is unknown?



Elliptical Slice Sampling (continued)

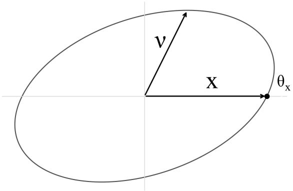
Algorithm 4: ESS (single iteration)

Input: Current state x.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\min} \leftarrow 0$, $\theta_{\max} \leftarrow 2\pi$
while *true* **do**
    Sample $\theta \sim \mathcal{U}(\theta_{\min}, \theta_{\max})$
    Propose $x^{\star} = x \cos\theta + \nu \sin\theta$
    if $\mathcal{L}(\mathrm{y}|x^{\star}) \geq y$ **then**
        accept $x \leftarrow x^{\star}$ and stop
    else
        shrink $(\theta_{\min}, \theta_{\max})$ around θ_x and continue self-tuning

Elliptical Slice Sampling (continued)

Algorithm 4: ESS (single iteration)

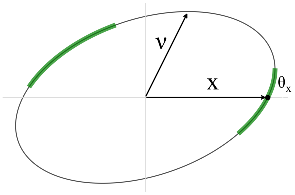
Input: Current state x.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$
while *true* **do**
    Sample $\theta \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}})$
    Propose $x^\star = x \cos\theta + \nu \sin\theta$
    if $\mathcal{L}(\mathrm{y}|x^\star) \geq y$ **then**
        accept $x \leftarrow x^\star$ and stop
    else
        shrink $(\theta_{\text{min}}, \theta_{\text{max}})$ around θ_x and continue self-tuning



Elliptical Slice Sampling (continued)

Algorithm 4: ESS (single iteration)

Input: Current state x.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$
while *true* **do**
  Sample $\theta \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}})$
  Propose $x^\star = x \cos\theta + \nu \sin\theta$
  if $\mathcal{L}(\mathrm{y}|x^\star) \geq y$ **then**
    accept $x \leftarrow x^\star$ and stop
  else
    shrink $(\theta_{\text{min}}, \theta_{\text{max}})$ around θ_x and continue self-tuning



Elliptical Slice Sampling (continued)

Algorithm 4: ESS (single iteration)

Input: Current state x.

Draw $\nu \sim \mathcal{N}(0, C)$

Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$

Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$

Set $\theta_{\mathrm{min}} \leftarrow 0$, $\theta_{\mathrm{max}} \leftarrow 2\pi$

while *true* **do**

  Sample $\theta \sim \mathcal{U}(\theta_{\mathrm{min}}, \theta_{\mathrm{max}})$

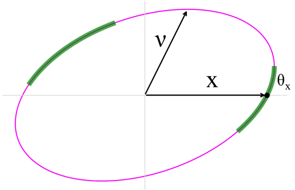
  Propose $x^\star = x \cos\theta + \nu \sin\theta$

  if $\mathcal{L}(\mathrm{y}|x^\star) \geq y$ **then**

    accept $x \leftarrow x^\star$ and stop

  else

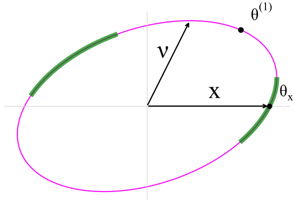
    shrink $(\theta_{\mathrm{min}}, \theta_{\mathrm{max}})$ around θ_x and continue self-tuning



Elliptical Slice Sampling (continued)

Algorithm 4: ESS (single iteration)

Input: Current state x.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$
while *true* **do**
    Sample $\theta \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}})$
    Propose $x^\star = x \cos\theta + \nu \sin\theta$
    if $\mathcal{L}(\mathrm{y}|x^\star) \geq y$ **then**
        accept $x \leftarrow x^\star$ and stop
    else
        shrink $(\theta_{\text{min}}, \theta_{\text{max}})$ around θ_x and continue self-tuning



One iteration of Elliptical Slice Sampling

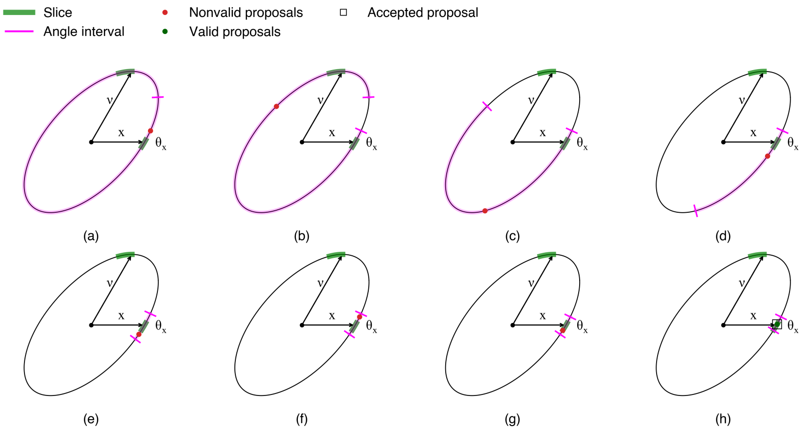


Figure: One iteration of ESS.

Multiproposal Elliptical Slice Sampling

Algorithm 5: MESS (single iteration)

Input: Current state x, number of proposals M.

Draw $\nu \sim \mathcal{N}(0, C)$

Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$

Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$

Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$

while *true* **do**

- Draw $\theta^{(j)} \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}}), \quad j = 1, \ldots, M$ parallelizable
- Propose $x^{(j)} = x \cos\theta^{(j)} + \nu \sin\theta^{(j)}$
- Compute $\mathcal{A} = \{j : \mathcal{L}(\mathrm{y}|x^{(j)}) \geq y\}$
- **if** $|\mathcal{A}| = 0$ **then**
  - shrink using all M angles and continue
- **else**
  - select $m \in \mathcal{A}$, set $x \leftarrow x^{(m)}$ and stop.

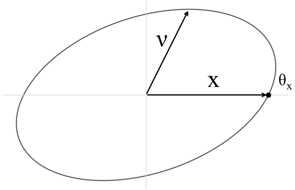


Figure: MESS ($M = 5$).

Multiproposal Elliptical Slice Sampling

Algorithm 5: MESS (single iteration)

Input: Current state x, number of proposals M.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$
while *true* **do**
  Draw $\theta^{(j)} \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}}), \quad j = 1, \ldots, M$ parallelizable
  Propose $x^{(j)} = x \cos\theta^{(j)} + \nu \sin\theta^{(j)}$
  Compute $\mathcal{A} = \{j : \mathcal{L}(\mathrm{y}|x^{(j)}) \geq y\}$
  if $|\mathcal{A}| = 0$ **then**
    shrink using all M angles and continue
  else
    select $m \in \mathcal{A}$, set $x \leftarrow x^{(m)}$ and stop.

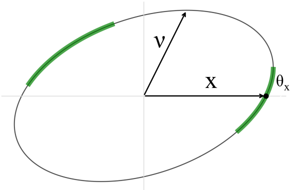


Figure: MESS ($M = 5$).

Multiproposal Elliptical Slice Sampling

Algorithm 5: MESS (single iteration)

Input: Current state x, number of proposals M.
Draw $\nu \sim \mathcal{N}(0, C)$
Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$
Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$
Set $\theta_{\min} \leftarrow 0$, $\theta_{\max} \leftarrow 2\pi$
while *true* **do**
  Draw $\theta^{(j)} \sim \mathcal{U}(\theta_{\min}, \theta_{\max}), \quad j = 1, \ldots, M$ parallelizable
  Propose $x^{(j)} = x \cos\theta^{(j)} + \nu \sin\theta^{(j)}$
  Compute $\mathcal{A} = \{j : \mathcal{L}(\mathrm{y}|x^{(j)}) \geq y\}$
  if $|\mathcal{A}| = 0$ **then**
    shrink using all M angles and continue
  else
    select $m \in \mathcal{A}$, set $x \leftarrow x^{(m)}$ and stop.

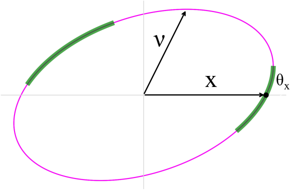


Figure: MESS ($M = 5$).

Multiproposal Elliptical Slice Sampling

Algorithm 5: MESS (single iteration)

Input: Current state x, number of proposals M.

Draw $\nu \sim \mathcal{N}(0, C)$

Draw $\theta_x \sim \mathcal{U}(0, 2\pi)$

Draw slice level $y \sim \mathcal{U}(0, \mathcal{L}(\mathrm{y}|x))$

Set $\theta_{\text{min}} \leftarrow 0$, $\theta_{\text{max}} \leftarrow 2\pi$

while *true* **do**

- Draw $\theta^{(j)} \sim \mathcal{U}(\theta_{\text{min}}, \theta_{\text{max}})$, $\quad j = 1, \ldots, M$ parallelizable
- Propose $x^{(j)} = x \cos \theta^{(j)} + \nu \sin \theta^{(j)}$
- Compute $\mathcal{A} = \{j : \mathcal{L}(\mathrm{y}|x^{(j)}) \geq y\}$
- **if** $|\mathcal{A}| = 0$ **then**
  - shrink using all M angles and continue
- **else**
  - select $m \in \mathcal{A}$, set $x \leftarrow x^{(m)}$ and stop.

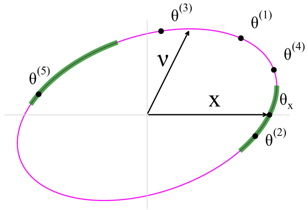


Figure: MESS ($M = 5$).

Mess in action

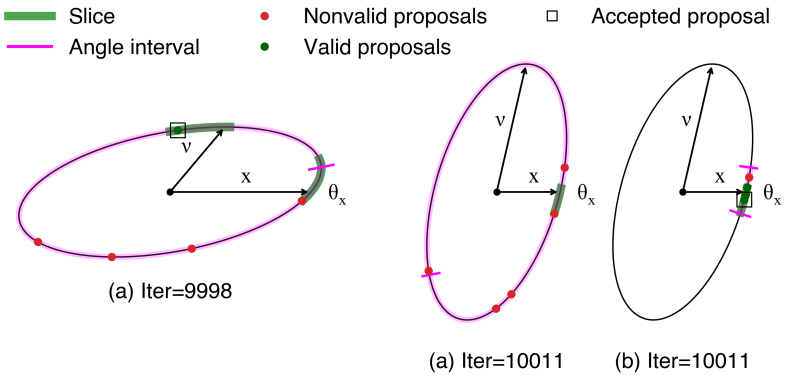


Figure: Two iterations of MESS with $M = 5$.

Acceptance step with transition matrix

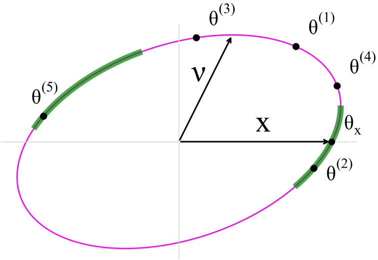
At subiteration k:

- Sample $\theta^{(1)}, \ldots, \theta^{(5)}$.
- Valid indices: $\mathcal{A} = \{j : \mathcal{L}(\theta^{(j)}) \geq y\} = \{2, 5\}$.
- Ordered valid angles:

$$(\theta^{(2)}, \theta_x, \theta^{(5)}) \text{ with labels } (0, 1, 2).$$

- Transition matrix with entries

$$\boldsymbol{P} = \begin{bmatrix} 0 & P_{01} & P_{02} \\ P_{10} & 0 & P_{12} \\ P_{20} & P_{21} & 0 \end{bmatrix}$$



Acceptance step with transition matrix

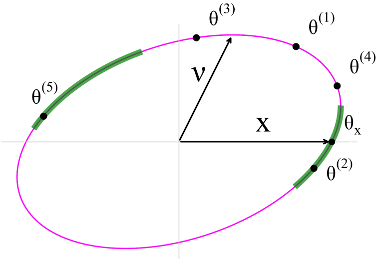
At subiteration k:

- Sample $\theta^{(1)}, \ldots, \theta^{(5)}$.
- Valid indices: $\mathcal{A} = \{j : \mathcal{L}(\theta^{(j)}) \geq y\} = \{2, 5\}$.
- Ordered valid angles:

$$(\theta^{(2)}, \theta_x, \theta^{(5)}) \text{ with labels } (0, 1, 2).$$

- Transition matrix with entries

$$\boldsymbol{P} = \begin{bmatrix} 0 & P_{01} & P_{02} \\ P_{10} & 0 & P_{12} \\ P_{20} & P_{21} & 0 \end{bmatrix}$$



Acceptance step with transition matrix

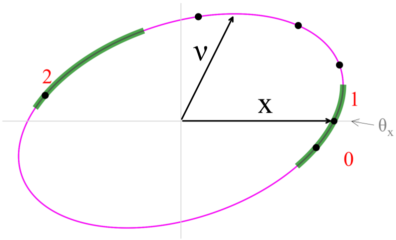
At subiteration k:

- Sample $\theta^{(1)}, \ldots, \theta^{(5)}$.
- Valid indices: $\mathcal{A} = \{j : \mathcal{L}(\theta^{(j)}) \geq y\} = \{2, 5\}$.
- Ordered valid angles:

$$(\theta^{(2)}, \theta_x, \theta^{(5)}) \text{ with labels } (0, 1, 2).$$

- Transition matrix with entries

$$\boldsymbol{P} = \begin{bmatrix} 0 & P_{01} & P_{02} \\ P_{10} & 0 & P_{12} \\ P_{20} & P_{21} & 0 \end{bmatrix}$$



Acceptance step with transition matrix

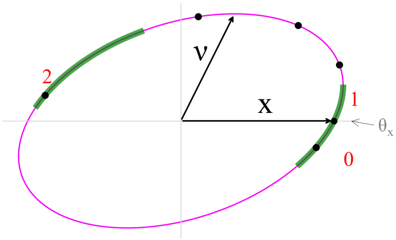
At subiteration k:

- Sample $\theta^{(1)}, \ldots, \theta^{(5)}$.
- Valid indices: $\mathcal{A} = \{j : \mathcal{L}(\theta^{(j)}) \geq y\} = \{2, 5\}$.
- Ordered valid angles:

$$(\theta^{(2)}, \theta_x, \theta^{(5)}) \text{ with labels } (0, 1, 2).$$

- Transition matrix with entries

$$\boldsymbol{P} = \begin{bmatrix} 0 & P_{01} & P_{02} \\ P_{10} & 0 & P_{12} \\ P_{20} & P_{21} & 0 \end{bmatrix}$$



Constructing the transition matrix

1. Naive choice (uniform):

$$\boldsymbol{P}^{\text{unif}} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

2. Use angular distances to bias transitions:

$$\max_{\boldsymbol{P}} \sum_{r=1}^{3} \sum_{s=1}^{3} D_{ang}(r,s) P_{rs}$$

subject to double-stochastic $\boldsymbol{P}$ with zero diagonal.

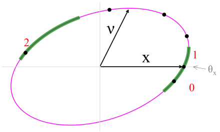


Figure: Labels of ordered valid angles.

- Angular distance matrix:

$$D_{ang} = \begin{bmatrix} 0 & 0.3 & 3 \\ 0.3 & 0 & 2.7 \\ 3 & 2.7 & 0 \end{bmatrix} \quad \text{(rad)}$$

- Example optimized matrix:

$$\boldsymbol{P}^{\star} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Constructing the transition matrix

1. Naive choice (uniform):

$$P^{\text{unif}} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

2. Use angular distances to bias transitions:

$$\max_{P} \sum_{r=1}^{3} \sum_{s=1}^{3} D_{ang}(r,s) P_{rs}$$

subject to double-stochastic P with zero diagonal.

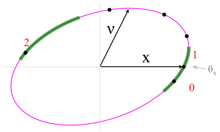


Figure: Labels of ordered valid angles.

- Angular distance matrix:

$$D_{ang} = \begin{bmatrix} 0 & 0.3 & 3 \\ 0.3 & 0 & 2.7 \\ 3 & 2.7 & 0 \end{bmatrix} \quad \text{(rad)}$$

- Example optimized matrix:

$$P^{\star} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Constructing the transition matrix

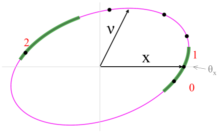


Figure: Labels of ordered valid angles.

1. Naive choice (uniform):

$$\boldsymbol{P}^{\text{unif}} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

2. Use angular distances to bias transitions:

$$\max_{\boldsymbol{P}} \sum_{r=1}^{3} \sum_{s=1}^{3} D_{ang}(r,s) P_{rs}$$

subject to double-stochastic $\boldsymbol{P}$ with zero diagonal.

- Angular distance matrix:

$$D_{ang} = \begin{bmatrix} 0 & 0.3 & 3 \\ 0.3 & 0 & 2.7 \\ 3 & 2.7 & 0 \end{bmatrix} \quad \text{(rad)}$$

- Example optimized matrix:

$$\boldsymbol{P}^{\star} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Constructing the transition matrix

1. Naive choice (uniform):

$$\boldsymbol{P}^{\text{unif}} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

2. Use angular distances to bias transitions:

$$\max_{\boldsymbol{P}} \sum_{r=1}^{3} \sum_{s=1}^{3} D_{ang}(r,s) P_{rs}$$

subject to double-stochastic $\boldsymbol{P}$ with zero diagonal.

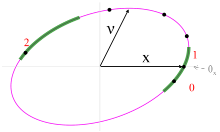


Figure: Labels of ordered valid angles.

- Angular distance matrix:

$$D_{ang} = \begin{bmatrix} 0 & 0.3 & 3 \\ 0.3 & 0 & 2.7 \\ 3 & 2.7 & 0 \end{bmatrix} \quad \text{(rad)}$$

- Example optimized matrix:

$$\boldsymbol{P}^{\star} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Mixing improvement with uniform MESS

- Average number of shrinking steps decays with M:

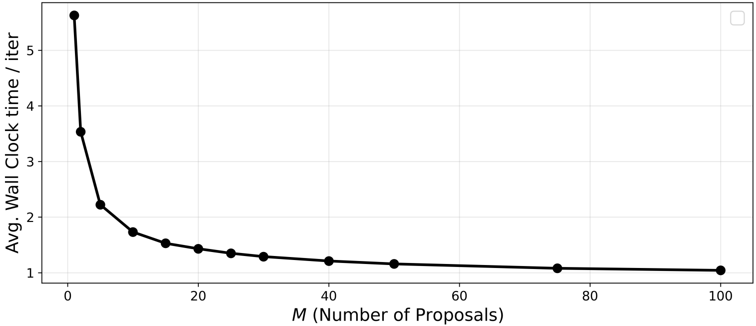


Figure: Logistic regression model from Murray et al. (2010).

- With M cores, wall-clock time decreases ≈ shrinking steps.

Mixing improvement with uniform MESS

- Average number of shrinking steps decays with M:

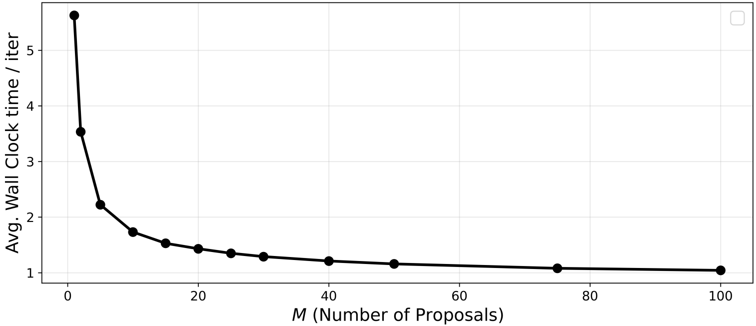


Figure: Logistic regression model from Murray et al. (2010).

- With M cores, wall-clock time decreases ≈ shrinking steps.

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3] Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3] Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3]Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3] Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3]Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3]Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d\times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3] Model from Glatt-Holtz et al. (2024).

Toy solute transport model[3]

- Mimics transportation of a solute by a fluid under damping and external forcing.

$$\frac{d}{dt}\Theta + (A + \kappa I)\Theta = g.$$

  - $\Theta \in \mathbb{R}^d$: Solute concentration (in Fourier basis).
  - $\kappa \in \mathbb{R}$: Damping coefficient.
  - $g \in \mathbb{R}^d$: Forcing term $g = (1, 0, \ldots, 0)$.
  - $A \in \mathbb{R}^{d \times d}$: Spectral transport matrix.
    - Zero-diagonal.
    - Antisymmetric.
- Steady state:

$$\Theta(A) = (A + \kappa I)^{-1} g.$$

Figure: Example.

[3] Model from Glatt-Holtz et al. (2024).

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

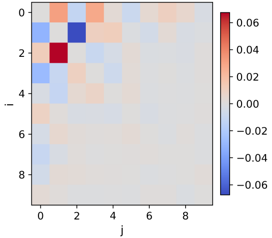
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

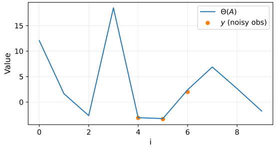
$$y = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$y_k = \mathcal{P}(y) \in \mathbb{R}^k, k \leq d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

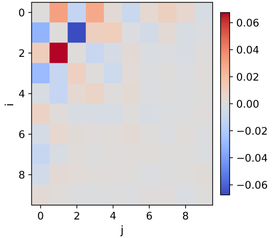
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

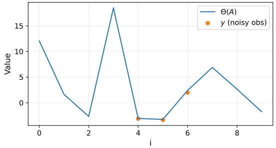
$$y = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$y_k = \mathcal{P}(y) \in \mathbb{R}^k, k \le d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

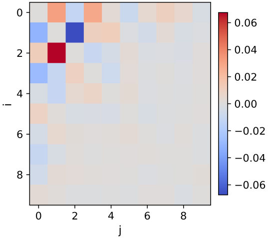
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

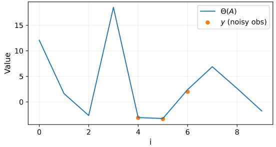
$$y = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$y_k = \mathcal{P}(y) \in \mathbb{R}^k, k \leq d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

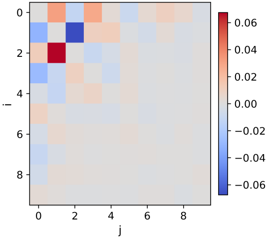
$$A \sim N(0, C), \quad C = \operatorname{diag}(\{\operatorname{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

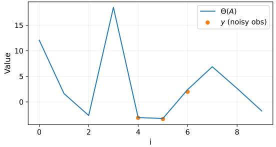
$$y = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$y_k = \mathcal{P}(y) \in \mathbb{R}^k, k \leq d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

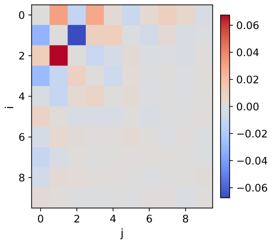
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

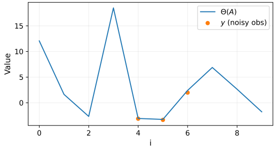
$$\mathrm{y} = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$\mathrm{y}_k = \mathcal{P}(\mathrm{y}) \in \mathbb{R}^k, k \leq d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

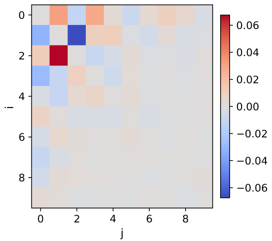
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

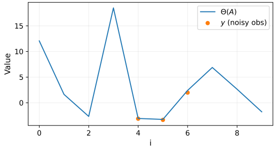
$$\mathrm{y} = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$\mathrm{y}_k = \mathcal{P}(\mathrm{y}) \in \mathbb{R}^k, k \leq d$$



(a) A_{10}.



(b) $\Theta(A_{10})$ and y_3.

Bayesian inverse problem

- Gaussian prior on the non-zero elements of A:

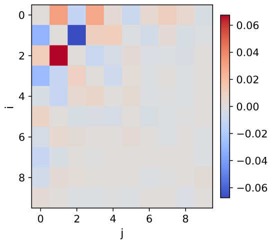
$$A \sim N(0, C), \quad C = \mathrm{diag}(\{\mathrm{Var}(a_{ij})\}).$$

  - $m = d(d-1)/2$ elements in upper triangle.
  - Contains Fourier coefficients of the concentration field.
  - C penalizes long range interactions and high frequencies.
- Gaussian likelihood:

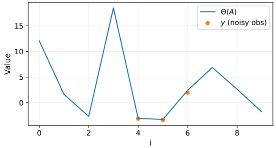
$$\mathrm{y} = \Theta(A) + \eta, \quad \eta \sim N(0, \sigma^2 I_d).$$

  - Optional projection $\mathcal{P}$:

$$\mathrm{y}_k = \mathcal{P}(\mathrm{y}) \in \mathbb{R}^k, k \leq d$$

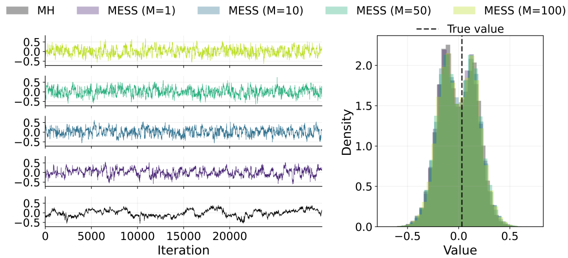


(a) A_{10}.

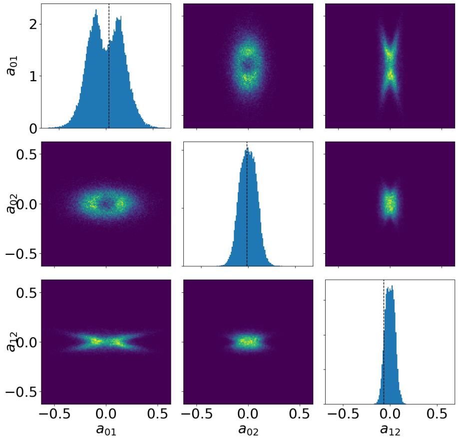


(b) $\Theta(A_{10})$ and y_3.

Posterior distribution for $d = 10$

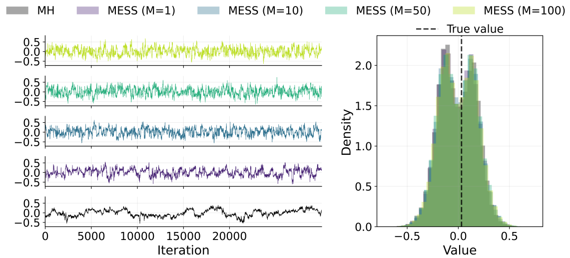


(a) Traceplots for a_{01}.

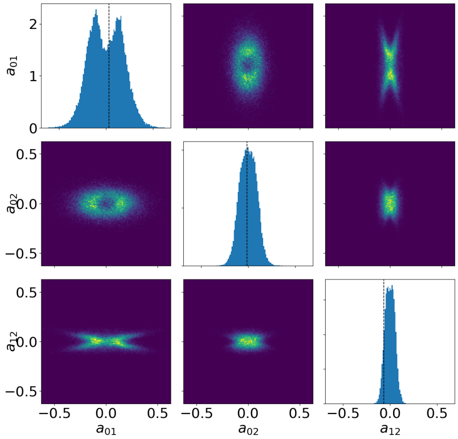


(b) Posterior marginals and pairwise density plots at $d = 10$.

Posterior distribution for $d = 10$



(a) Traceplots for a_{01}.



(b) Posterior marginals and pairwise density plots at $d = 10$.

Scaling with dimension

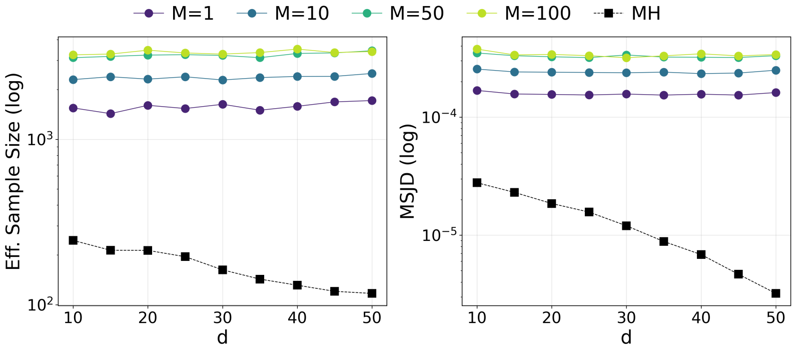


Figure: MESS mixing rate does not degrade with dimension.

Conclusions

Multiproposal Elliptical Slice Sampling (MESS) is

Self-tuning

no gradients

dimension-robust

parallelizable.

Useful for large-scale and/or non-parametric Bayesian estimation with Gaussian priors.

References I

Cotter, S. L., G. O. Roberts, A. M. Stuart, and D. White (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science 28*(3), 424–446.

Dashti, M. and A. M. Stuart (2013). The bayesian approach to inverse problems. *arXiv: Probability*.

Glatt-Holtz, N. E., A. J. Holbrook, J. A. Krometis, and C. F. Mondaini (2024). Parallel MCMC Algorithms: Theoretical Foundations, Algorithm Design, Case Studies. *Transactions of Mathematics and Its Applications 8*(2).

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*(1), 97–109.

Murray, I., R. P. Adams, and D. J. C. MacKay (2010). Elliptical Slice Sampling. In *Proc. of the 13th AISTATS*, Volume 9 of *PMLR*, pp. 541–548.

Neal, R. M. (1999). Regression and Classification Using Gaussian Process Priors. In *Bayesian Statistics 6: Proc. of the Sixth Valencia International Meeting June 6-10, 1998*. Oxford University Press.

Neal, R. M. (2003). Slice Sampling. *The Annals of Statistics 31*(3), 705–741.

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica 19*, 451–559.

Comparison with Gibbs and HMC: blind deconvolution inverse problem

- Observation model:

$$d = w \star c + e, \quad w \sim N(0, C_w), \quad c \sim N(0, C_c), \quad e \sim N(0, C_e).$$

- Posterior is multimodal (sign-shift symmetry).