Automated Sliced Gibbs sampling for Arbitrary Kernels

NC STATE

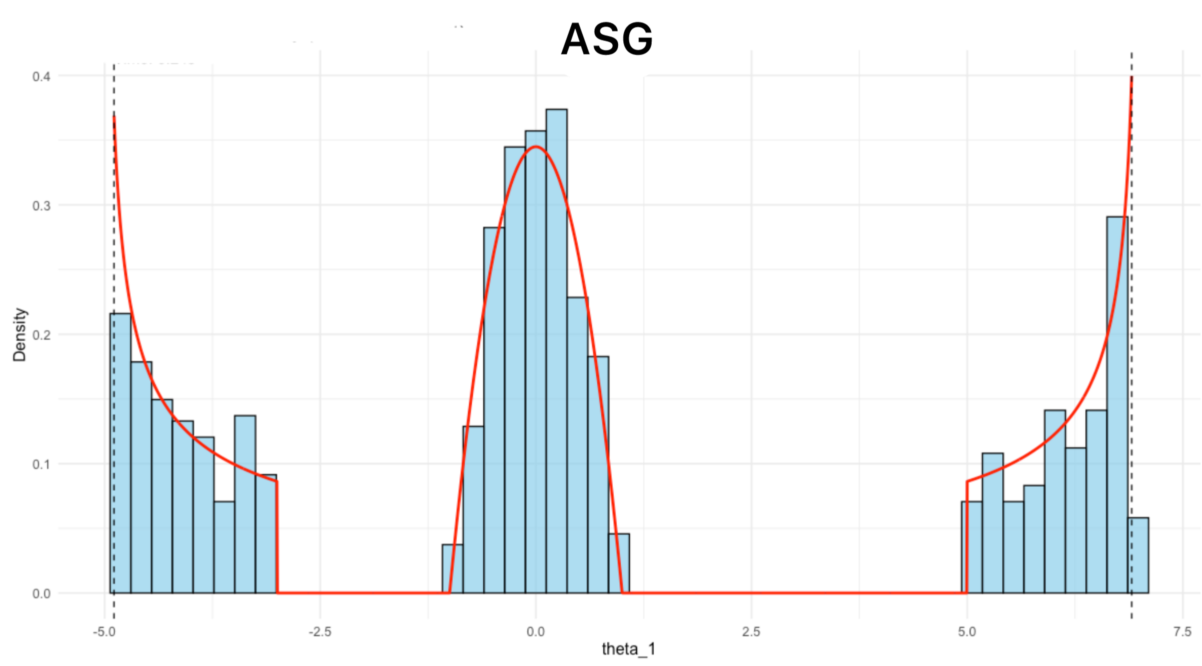
Prithwish Ghosh [1] Sujit Ghosh [1]

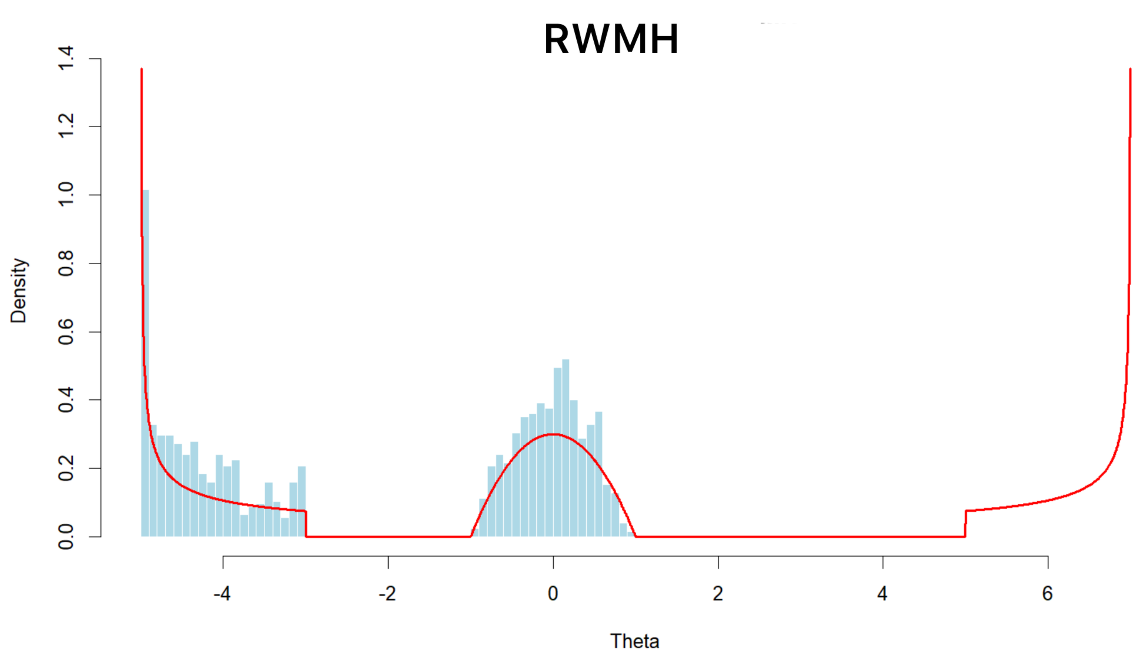
[1]Department of Statistics, North Carolina State University

Motivation

- RW–MH:
  - Needs proposal tuning, Gets stuck in local modes,Very low ESS in multimodal targets.
- Gibbs Sampling:
  - Requires analytic full conditionals,Needs known supports for all coordinates, Fails for curved or irregular densities.
- Slice Sampling:
  - Still needs user-chosen slice brackets, Manual tuning per dimension, Not scalable to high-dimensional problems.
- Overall:
  - All require manual tuning,Not automatic for multimodal or non-concave densities.

1. Multimodal 1D mixture (Beta mixture)

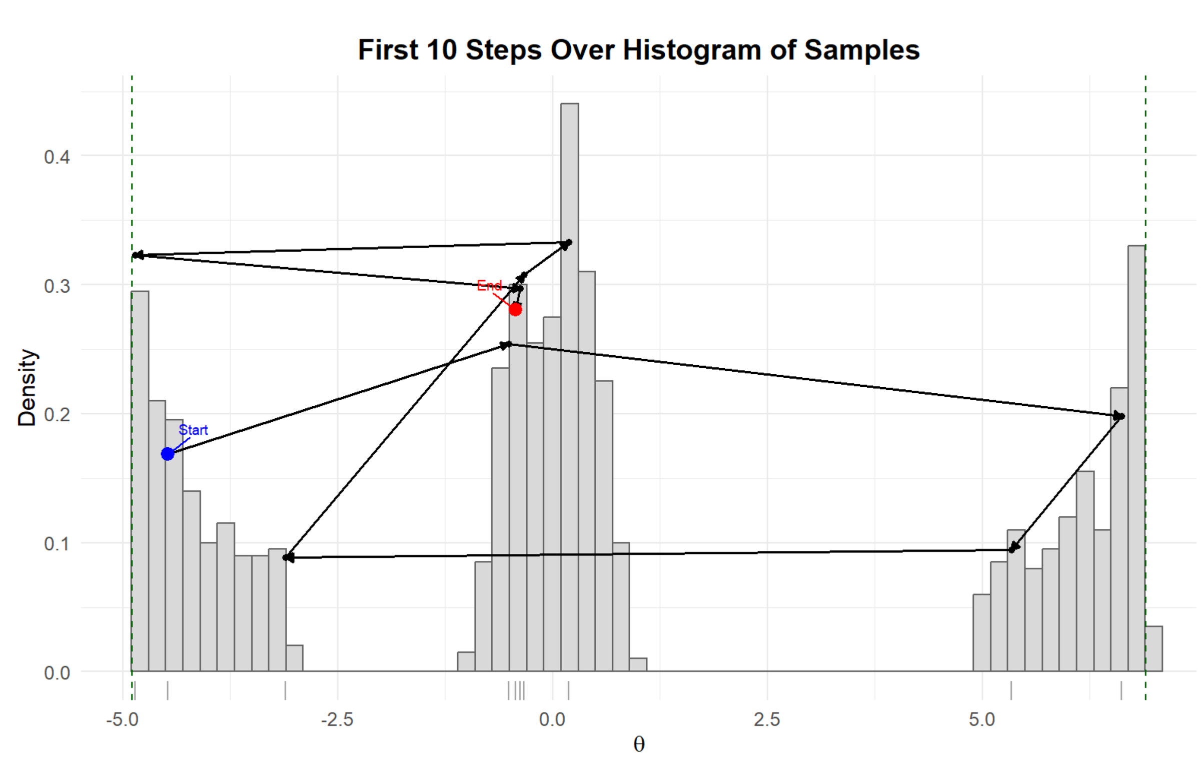




RW-MH Completely misses samples from one component of the mixture.

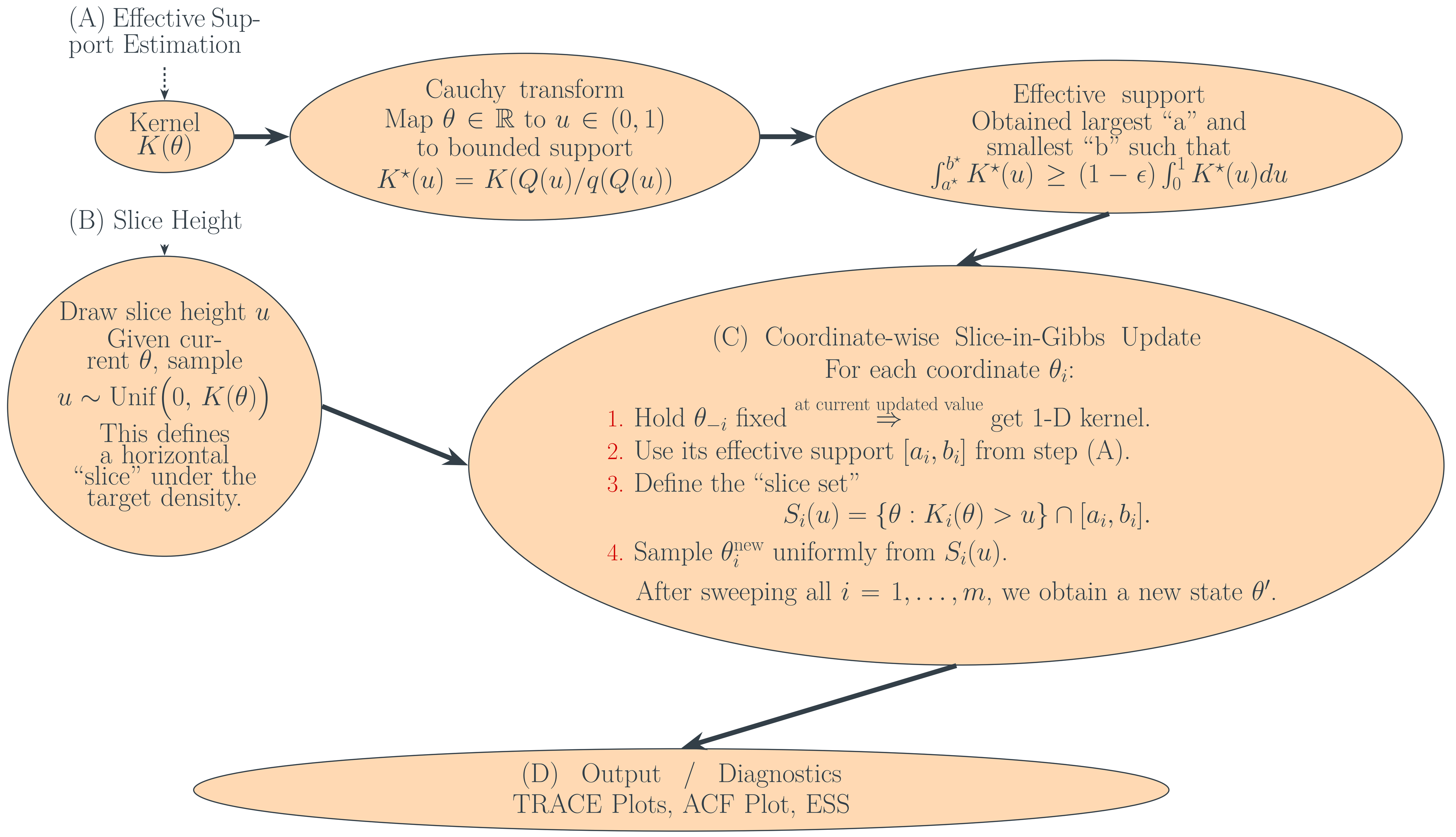
$$K(\theta) = \frac{0.3}{2}\text{Beta}\left(\frac{\theta+5}{2};0.5,1\right) + \frac{0.4}{2}\text{Beta}\left(\frac{\theta+1}{2};2,2\right) + \frac{0.3}{2}\text{Beta}\left(\frac{\theta-5}{2};1,0.5\right).$$

2. Why the ASG Works?



Amazingly with first few iterates the ASG sampler rapidly explores all major modes of the Beta-mixture target, then stabilizes within high-density regions—illustrating fast global mixing and accurate effective-support estimation.

Proposed method: ASG

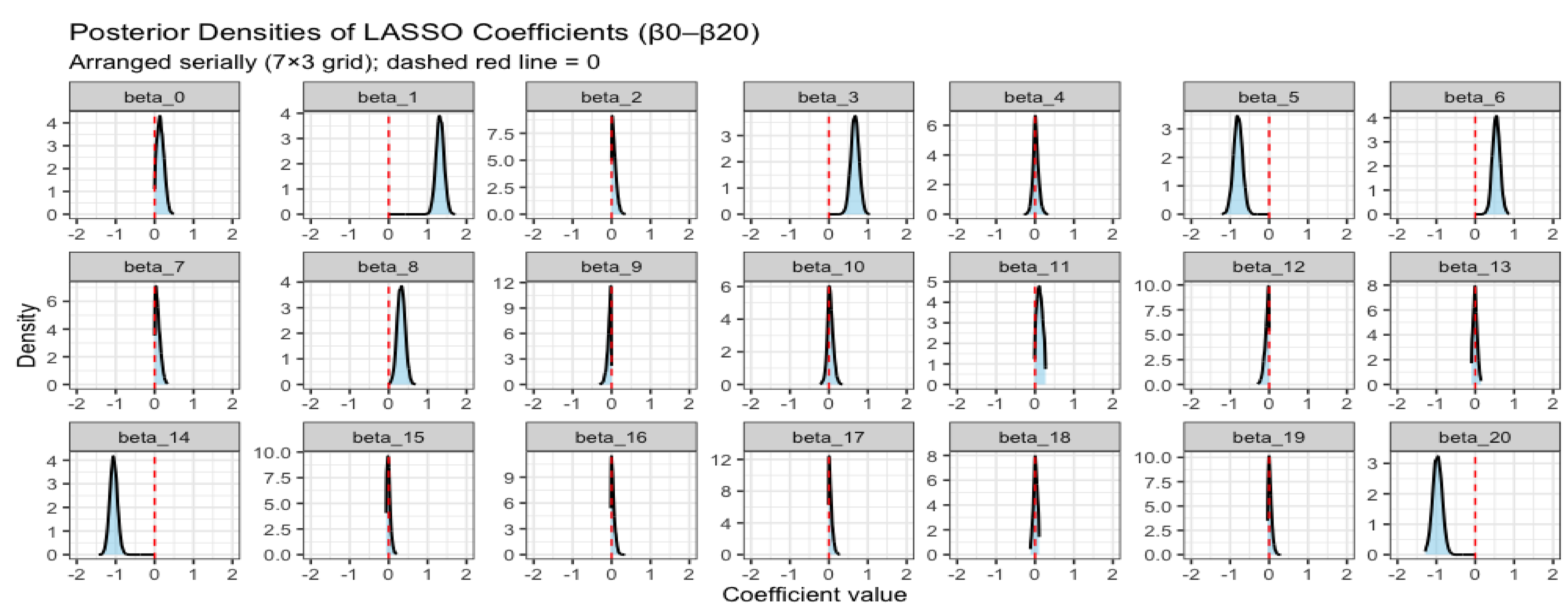


LASSO Demo (MC Samples = 100000, burn in = 2500, thin = 1)

We sample in $\theta = (\beta_0, \beta_1, \cdots, \beta_{20})$ from the unnormalized density $K(\theta) \propto \exp\left\{-\left[\frac{1}{2}\sum_{i=1}^N\left(y_i - \beta_0 - \mathbf{x}_i^\top\beta\right)^2 + N\lambda\sum_{j=1}^m|\beta_j|\right]\right\}$, with $\beta = (\beta_1, \beta_2, \cdots, \beta_{20})^\top$, standardized columns from **QuickStartExample** dataset, $N = 100$, and $\lambda = 0.1$ by glmnet cv

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
Posterior Mode	0.1332	1.3150	0.0011	0.6614	0.0023	-0.8035	0.5434	0.0168	0.3230	-0.0001
Lasso Estimate	0.1500	1.3282	0.0000	0.6735	0	-0.8050	0.5269	0.0038	0.3254	0

β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}	β_{18}	β_{19}	β_{20}
0.0049	0.0972	-0.0108	-0.0025	-1.0594	-0.0075	-0.0005	0.0029	0.0015	0.0000	-0.9595
0.0000	0.1393	0.0000	0.0000	-1.0684	0.0000	0.0000	0.0000	0.0000	0.0000	-1.0030



The close agreement between the posterior modes and the LASSO estimates indicates that proposed sampler accurately identifies the dominant coefficients while adaptively shrinking the negligible ones toward zero. Unlike deterministic LASSO optimization, the proposed sampling method captures full posterior uncertainty and density structure. The average

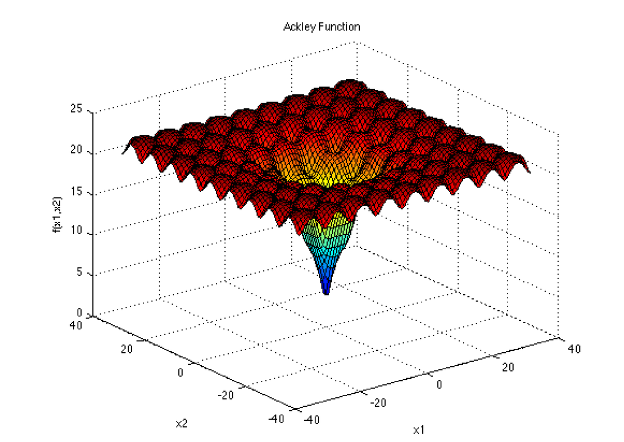
Avg. ESS = 66883 (+/- 17160). Total time = 165s

Non Concave Densities (MC Samples = 1000, Burn In = 250, Thin = 1)

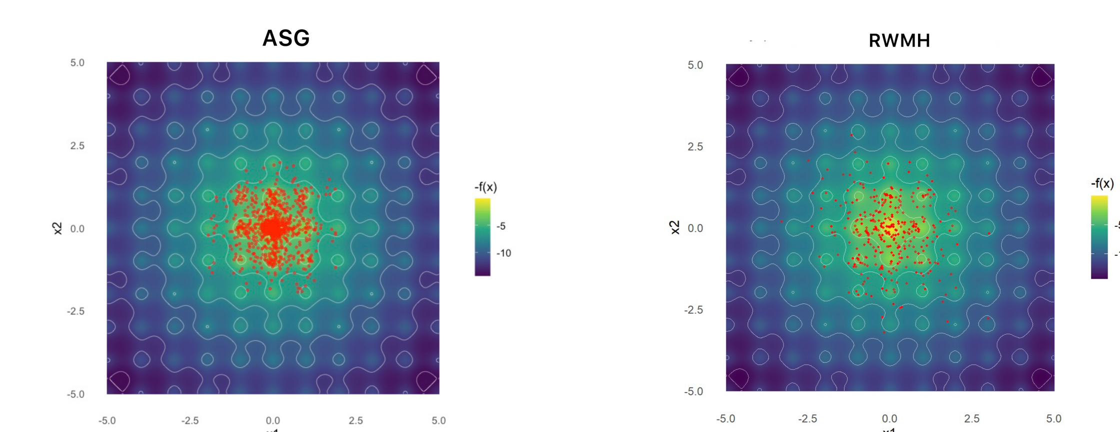
Ackley kernel The Ackley function is widely used for testing optimization algorithms. The function has multiple local optima.

$$-\log K(\boldsymbol{\theta}) = \left[-t\exp\left(-v\sqrt{\frac{1}{2}\sum_{i=1}^{2}\theta_i^2}\right) - \exp\left(\frac{1}{2}\sum_{i=1}^{2}\cos(c\,\theta_i)\right) + t + e\right], \quad \boldsymbol{\theta} \in [a,b]^2,$$

In the below figure we took t = 20, v = 0.2, c = 2π

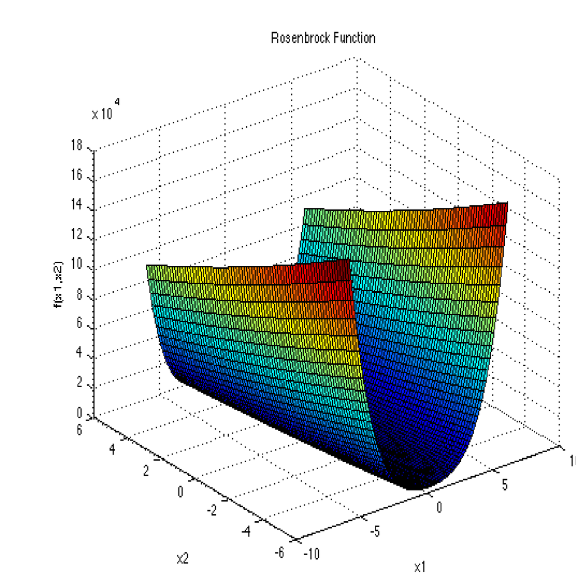


Name	Time Taken(in Sec)	ESS
ASG sampler (Dim 1):	1.99	1118
ASG sampler (Dim 2):	1.99	1000
MH Sampler (Dim 1):	0.17	143
MH Sampler (Dim 2):	0.17	163



Rosenbrock Kernel The Rosenbrock function, also referred as Banana function, is a popular test problem for gradient-based optimization algorithms.

$$-\log K(\boldsymbol{\theta}) = \frac{\theta_1^2}{10} + \frac{\theta_2^2}{10} + 2\left(\theta_2 - \theta_1^2\right)^2$$



Name	Time Taken(in Sec)	ESS
Custom ASG sampler (Dim 1):	0.104	1000
Custom ASG sampler (Dim 2):	0.104	730
MH Sampler (Dim 1):	0.16	17
MH Sampler (Dim 2):	0.16	25

Table 1. Comparison Table for Banana shaped Distribution Kernel