

Introduction

Hamiltonian Monte Carlo, binned among the most effective MCMC methods^[1] in statistical computing, struggles with sampling parameters with a constrained support. It is standard to create mappings from a constrained to unconstrained space for these parameters^[3], instead of modifying the sampler. For composed transforms such as correlation matrices (positive definite and unit diagonal) or simplexes etc., adjusting the sampler is relatively challenging. Although, the latter is feasible for rudimentary constraints, such as unit vector or bound ranges based on scalars. We look for a bijective, smooth function (with simpler Jacobian computations, in a relative sense) for these transformations. In this paper, we attempt to evaluate the computational and statistical efficiency of such commonplace transforms in statistical modeling.

Constraining Transforms

We look at surjective constraining transforms $f: \mathbf{Y} \rightarrow \mathbf{X}$ from an unconstrained space $\mathbf{Y} = \mathbb{R}^M$ onto a constrained space $\mathbf{X} \subset \mathbb{R}^N$. Along with each transform, we have an additional (perhaps improper) density function $h(y)$ defined over $y \in \mathbf{Y}$ such that if $y \sim h(\cdot)$, then $f(x) \sim \text{uniform}(\mathbf{X})$.

Unit Simplex

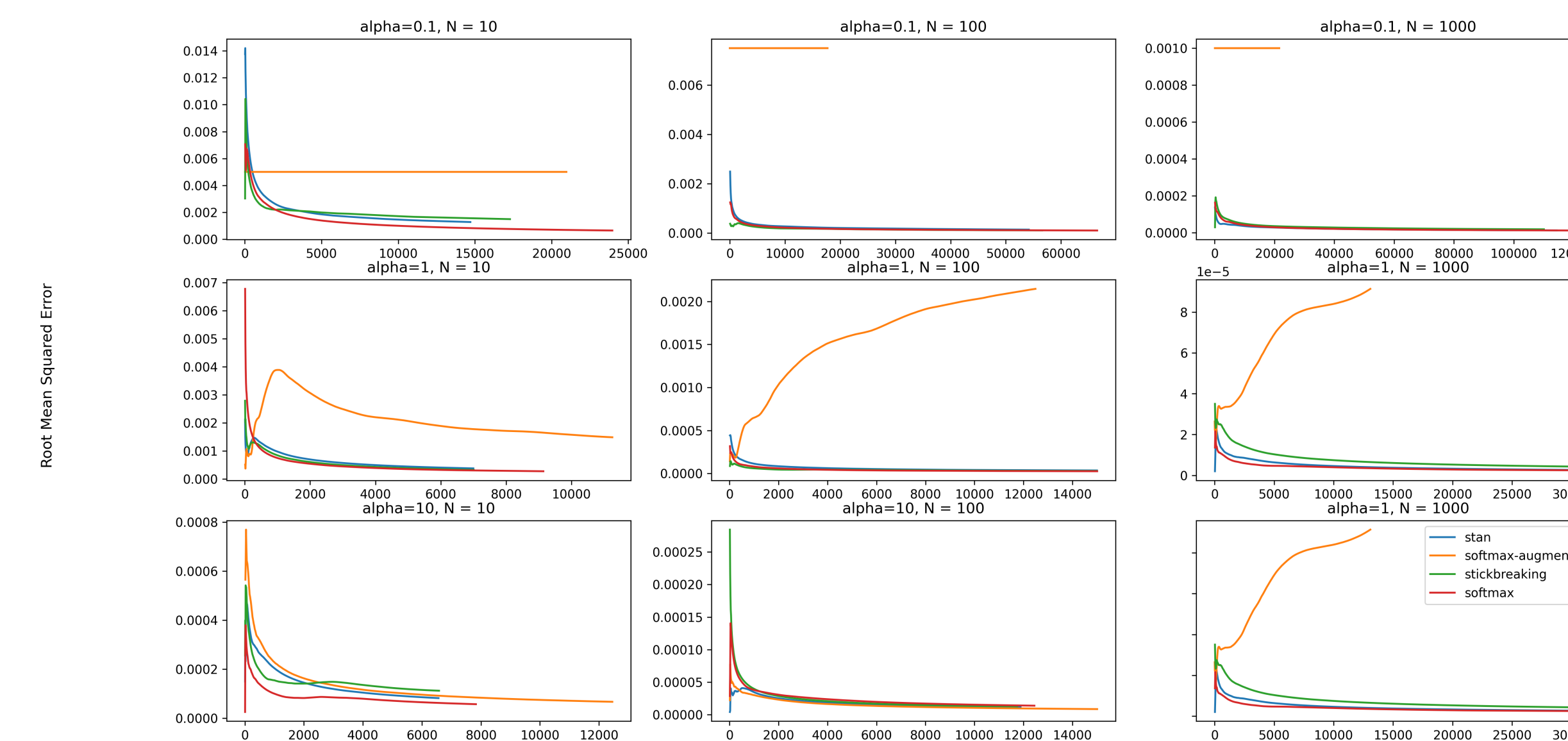
A unit N -simplex is an $N+1$ -dimensional vector of non-negative values that sums to one. Simplexes are useful for representations of multinomial probabilities (e.g., probabilities of categories in a classification problem). We evaluate four transforms on the simplex-Additive log ratio transform(softmax with first value pinned to 0), augmented-softmax parametrization and the Stick-Breaking process as a Stan^[2] program along with the actual Stan implementation.

Evaluation

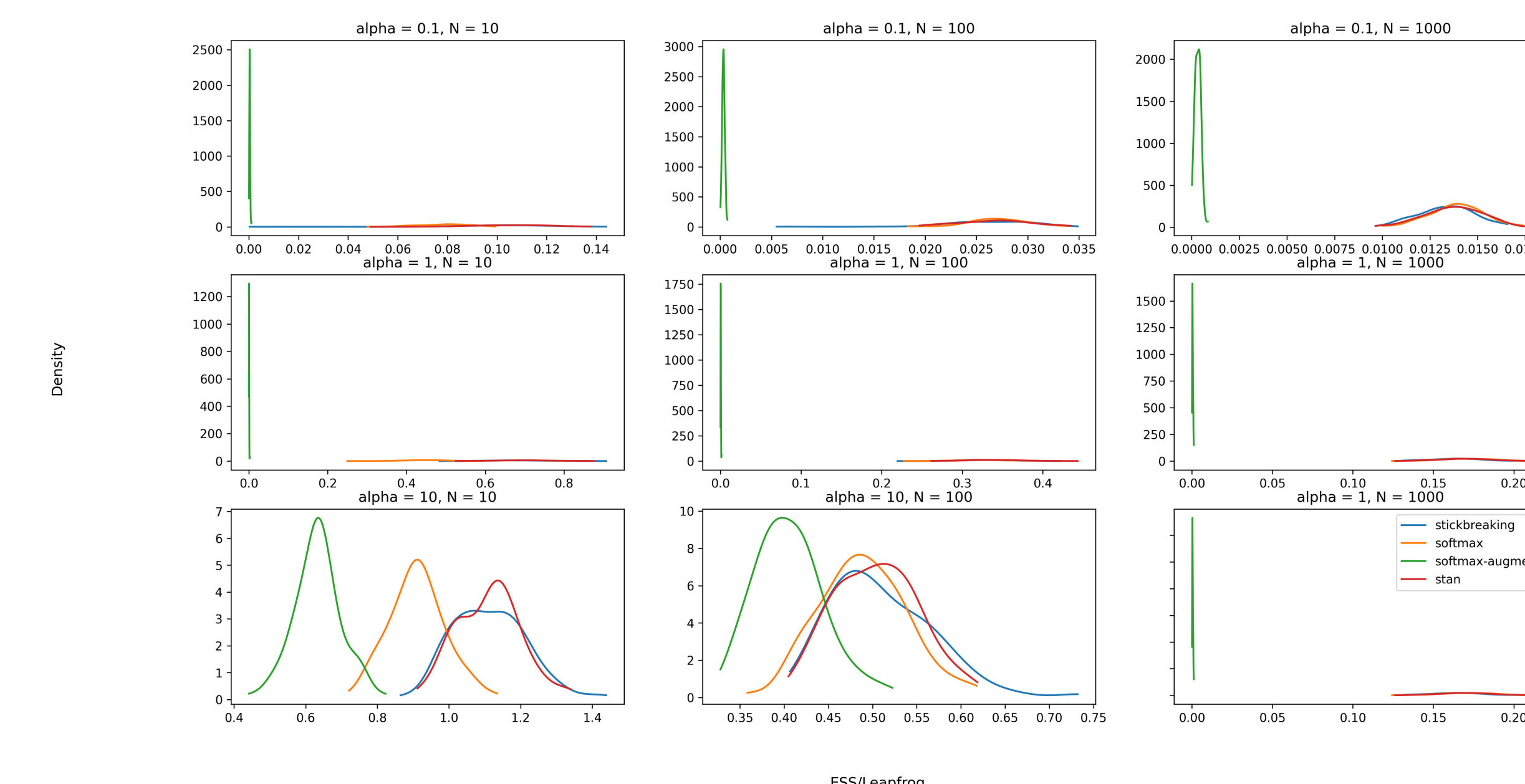
We are trying to evaluate transforms in the head, body, and tail distributions in terms of computational efficiency of both the transform, its Jacobian determinant and associated gradients, the geometry induced in the unconstrained space (e.g., log convexity and conditioning), as well as Hamiltonian Monte Carlo sampling efficiency.

Discussion

Considering a symmetric Dirichlet in the model likelihood, we evaluate RMSE for the four simplex transforms. In this case, the Stick-breaking transform in Stan seems to be working better. Also, note the odd behavior of the augmented-softmax transform, this is caused by a scaling parameter p for the log-likelihood term.



Similar trends can be seen in the Probability Density Function plot



Ongoing Work

Now, we are evaluating Jacobians of the inverse transforms for other constrained parameter types like vectors with sum-to-zero or unit length constraints, and matrices (or their Cholesky factors) constrained to be positive definite (e.g., covariance matrices), positive definite with unit diagonal (e.g., correlation matrices), or orthonormal. To be followed by analyzing efficiency and stability of these transforms. A strong motivation behind this project is to help Probabilistic Programming language users choose suitable transforms depending on their model and use case.

References

- [1] Radford Neal. MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, , and Xiao-Li Meng, editors, Handbook of Markov Chain Monte Carlo, chapter 5. Chapman and Hall/CRC, 2011.
- [2] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. Journal of Statistical Software, 76(1):1–32, 2017.
- [3] Tor Erlend Fjelde, Kai Xu, Mohamed Tarek, Sharan Yalburgi, and Hong Ge. Bijectors.jl: Flexible transformations for probability distributions. In Symposium on Advances in Approximate Bayesian Inference, pages 1–17. PMLR, 2020

Acknowledgements

This work was graciously supported by the Simons Foundation. Gratitude to Bob Carpenter for his astute insights and memorable mentorship.