Supplementary material: Visualization in Bayesian workflow

Jonah Gabry†

Department of Statistics and ISERP, Columbia University, New York, USA.

E-mail: jgabry@gmail.com

Daniel Simpson†

Department of Statistical Sciences, University of Toronto, Canada.

Aki Vehtari

Department of Computer Science, Aalto University, Espoo, Finland.

Michael Betancourt

ISERP, Columbia University, and Symplectomorphic, LLC, New York, USA.

Andrew Gelman

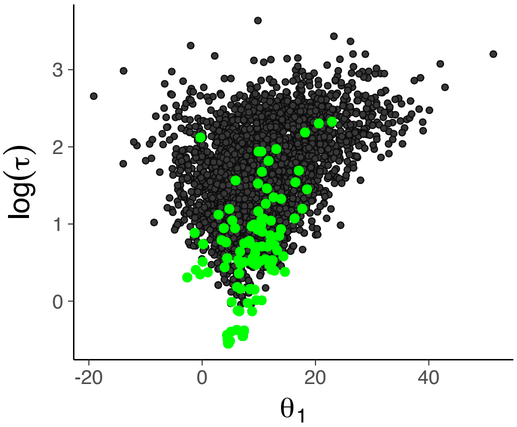
Departments of Statistics and Political Science, Columbia University, New York, USA.

1. The 8-schools problem and the visualization of divergent trajectories

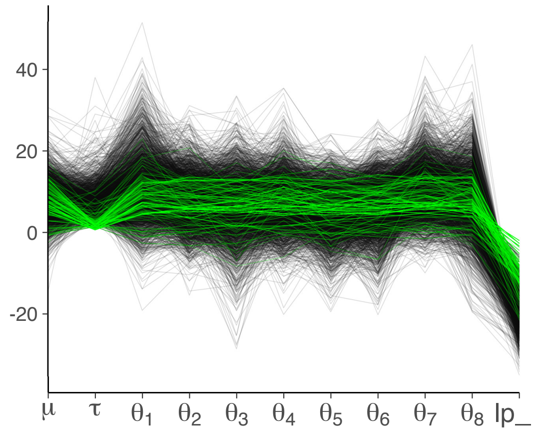
Consider the hierarchical 8-schools problem outlined in (Rubin, 1981; Gelman et al., 2013). Figure 1a shows a scatterplot of the log standard deviation of the school-specific parameters (τ, y-axis) against the parameter representing the mean for the first school (θ_1, x-axis). The starting points of divergent transitions, shown in green, concentrate in a particular region which is evidence of a geometric pathology in parameter space. Figure 1b gives a different perspective on the divergences. It is a parallel coordinates plot including *all* parameters from the 8-schools example with divergent iterations also highlighted in green. We can see in both the bivariate plot and the parallel coordinates plot that the divergences tend to occur when the hierarchical standard deviation τ goes to 0 and the values of the θ_j's are nearly constant. These problems can be fixed by re-parameterization (Betancourt and Girolami, 2015).

Now that we know precisely what part of the parameter space is causing problems, we can fix it. Funnels in the parameter space can be resolved through a reparameterization that fattens out the problem area. The standard tool for fixing funnels caused by hierarchical models is moving to a non-centered parameterization, where the narrowest coordinate is made a priori independent of the other coordinates in the funnel (Betancourt and Girolami, 2015). This will typically fatten out the funnel and remove the cluster of divergences.

†Joint first author



(a) A bivariate plot of the log standard deviation of school-level effects ($\log(\tau)$, y-axis) against the mean for the first school (θ_1, x-axis) for the 8-schools problem. The green dots indicate starting points of divergent transitions. The pile up of divergences in a corner of the samples (in this case the neck of the funnel shape) strongly indicates that there is a problem with this part of the parameter space. This plot can be made using `mcmc_scatter` in `bayesplot`.



(b) Parallel coordinates plot for the 8-schools problem showing the school-specific parameters ($\theta_1, \ldots, \theta_8$) and their prior mean and standard deviation (μ, τ). The green lines indicate the starting points of divergent transitions. In this case it is clear that all of the divergent paths have a small value of τ, which results in little variability in the θ_j's (the green lines are flat). This plot can be made using `mcmc_parcoord` in `bayesplot`.

Fig. 1: *Several different diagnostic plots for Hamiltonian Monte Carlo. Models were fit using the RStan interface to Stan 2.17 (Stan Development Team, 2017).*

References

Betancourt, M. and M. Girolami (2015). Hamiltonian Monte Carlo for hierarchical models. In S. K. Upadhyay, U. Singh, D. K. Dey, and A. Loganathan (Eds.), *Current Trends in Bayesian Methodology with Applications*, pp. 79–101. Chapman & Hall. arXiv:1312.0906.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis* (Third ed.). Chapman & Hall/CRC. Chapter 6, Section "Marginal predictive checks".

Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics 6*, 377–401.

Stan Development Team (2017). RStan: the R interface to Stan, version 2.16.1. `http://mc-stan.org`.