Supplementary Material for Slice Sampling on Hamiltonian Trajectories

1. Hamiltonian Dynamics for Non-Gaussian Distributions

Assume we have some prior \( p \) for \( f \), where

\[
- \log p(f|\alpha) = \log Z - g(f|\alpha),
\]

for some hyperparameters \( \alpha \) and where \( Z \) is the normalizing constant. Then we can again set up a Hamiltonian,

\[
H(f, p) = g(f|\alpha) + \frac{1}{2} p^T M^{-1} p.
\]

Hamilton’s equations yield the system of differential equations

\[
\nabla_t^2 f = -M^{-1} \nabla_t g(f|\alpha).
\]

This can be solved exactly for some cases of \( g(f|\alpha) \); the Gaussian example above is one such case.

When \( M = \text{diag}(m_1, \ldots, m_d) \) and \( g(f|\alpha) = \sum_{i=1}^d h_i(f_i) \), the system is an uncoupled system and often has an analytic solution. In particular, \( f_i(t) \) is the solution to

\[
\frac{1}{2} \left( \int \left[ c_1 - \frac{1}{m_i} h_i(f_i) \right]^{-1/2} df_i \right)^2 = (t + c_2)^2,
\]

where \( c_2 \) is determined by \( f_i(0) \) and

\[
c_1 = \frac{1}{2} \left( \dot{f}_i(0)^2 + \frac{1}{m_i} h_i(f_i(0)) \right),
\]

where \( \dot{f}_i(t) \) denotes the time-derivative of \( f_i \) at time \( t \). For certain distributions, (4) has an analytic solution.

1.1. Exp(\( \lambda \)) and Laplace(\( \lambda \))

Working with the general form of the solution can be difficult, especially when the sample space is constrained. The exponential distribution is one such example, but it induces a potential for which solutions are easy to obtain directly. Let the prior be such that the components of \( f \) are mutually independent and \( f_i \sim \text{Exp}(\lambda_i) \) so that

\[
- \log p(f_i|\lambda_i) = -\log(\lambda_i) + \lambda_i \cdot f_i, \quad f_i > 0.
\]

Hamilton’s equations are particularly simple in this case, which describes one-dimensional projectile motion, e.g. a bouncing ball in a constant gravitational potential. The solution to Hamilton’s equations is

\[
f_i(t) = \frac{\lambda_i}{2m_i} t^2 + \dot{f}_i(0) t + f_i(0), \quad 0 \leq t \leq T_0.
\]

An example of such a trajectory is shown in Figure 1a. \( T_0 > 0 \) is the time at which the particle has position coordinate equal to 0, at which point its momentum changes signs, i.e. it “bounces.” For \( t > T_0 \), the particle repeatedly traces out the same trajectory. We find \( T_0 \) as

\[
T_0 = \frac{m_i}{\lambda_i} \dot{f}_i(0) + \sqrt{\frac{m_i^2}{\lambda_i^2} \dot{f}_i(0)^2 + \frac{2m_i}{\lambda_i} f_i(0)}.
\]

This yields the period of the trajectory,

\[
T = 2 \sqrt{\frac{m_i^2}{\lambda_i^2} \dot{f}_i(0)^2 + \frac{2m_i}{\lambda_i} f_i(0)}.
\]

Every time at which the particle reaches zero is then \( z_j = T_0 + (j-1)T \). Hamilton’s equations also yield the momentum,

\[
p_i(t) = -\lambda_i t + m_i \dot{f}_i(0),
\]

which we can use to find the momentum at the reflection point \( T_0 \), but before reflection,

\[
p_i(T_0^-) = -\sqrt{\frac{m_i^2}{\lambda_i^2} \dot{f}_i(0)^2 + 2m_i \lambda_i f_i(0)}.
\]

After the first reflection, we have \( p_i(T_0^+) = -p_i(T_0^-) \), and the dynamics proceed according to the equation

\[
f_i(t) = -\frac{\lambda_i}{2m_i} (t - z_j)^2 + \frac{p_i(T_0^+)}{m_i} (t - z_j), \quad z_j \leq t \leq z_{j+1}.
\]

We can use slightly different equations to describe the dynamics under a Laplace prior All that is required is some bookkeeping on the sign of the motion because the particle is not reflected at \( f_i = 0 \), but the sign on the potential switches. An example is shown in Figure 1b.

1.2. Pareto(\( x_m, \alpha \)) and GPD(\( \mu, \sigma, \xi \)) via transformation

The Pareto and Generalized Pareto (denoted GPD) distributions are typically used to model processes with heavy
tails. The density of the Pareto distribution is

\[
\pi(f_i|x_m, \alpha) = \alpha x_m^\alpha f_i^{-(\alpha + 1)},
\]

\[f_i \geq x_m, \quad x_m > 0, \quad \alpha > 0. \quad (13)\]

We can show that the random variable \(y_i := \log f_i - \log x_m\) is distributed as \(\text{Exp}(\alpha)\), for \(y_i > 0\). Using this fact, we can generate analytic trajectories as in subsection 1.1 for \(y_i\) and slice sample from the resulting curves \(f_i(t) = x_m e^{y_i(t)}\). An example is shown in Figure 1c.

The GPD has the density

\[
\pi(f_i|\mu, \sigma, \xi) = \frac{1}{\sigma} \left(1 + \xi \frac{f_i - \mu}{\sigma}\right)^{-(1+\xi^{-1})},
\]

\[f_i \geq \mu, \quad \xi \geq 0. \quad (14)\]

(The GPD is also defined for \(\xi < 0\), in which case \(\mu \leq f_i \leq \mu - \sigma/\xi\). We focus on the \(\xi \geq 0\) case for now.) The random variable \(y_i := \log \left(1 + \xi \frac{f_i - \mu}{\sigma}\right)\) is distributed as \(\text{Exp}(\xi^{-1})\) for \(y_i > 0\), and we can again use the calculations from subsection 1.1 to slice sample from the curve \(f_i(t) = \mu + \frac{\sigma}{\xi} (e^{y_i(t)} - 1)\). An example is shown in Figure 1d.
Figure 1. Example trajectories under different priors.