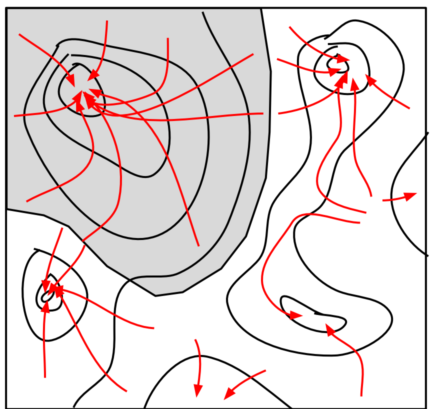
Adaptive initialization regeneration for generic search and optimization-based MH proposals

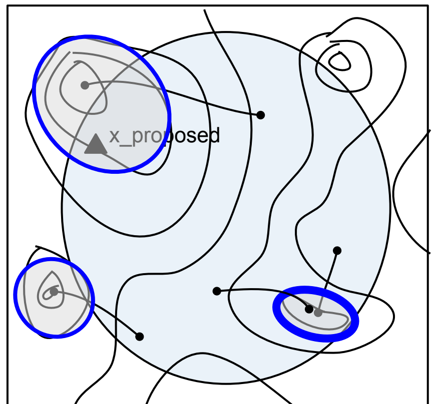
Suppose we have access to deterministic or stochastic local search/optimization algorithms that find modes of a posterior distribution, by attempting to maximize the unnormalized posterior (1), conditioned on an initial point x_init. We use such algorithms to construct mode-jumping proposal distributions using the meta-inference MH framework (2). The key challenge is obtaining low-variance estimates of the proposal density, which reduces to finding the conditional distribution on the search/opt. algorithm initialization given the search/opt. algorithm output. We learn this distribution adaptively over the course of the Markov chain by fitting the initial and proposed values using a Dirichlet process mixture model that is trained during the execution of the meta-inference MH chain to adaptively improve iteration. (3,4,5). We can also use multiple regenerations (multiple importance samples) to improve the accuracy further (6).

1



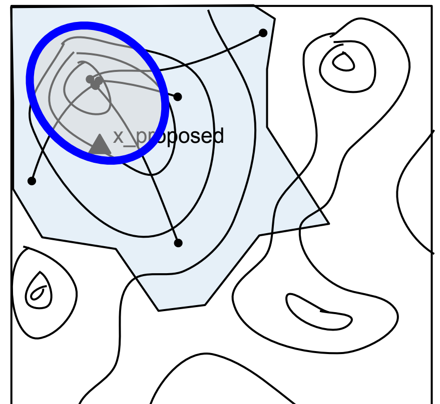
Contour plot of unnormalized multi-modal density showing trajectories of an L-BFGS optimization program; and the basin of attraction for one of the modes.

3

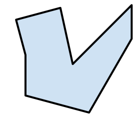


To estimate the proposal density for use in meta-inference MH, the regenerator guesses initial point(s) x_init that could have led to a given x_final. Initially, we guess x_init from a poor importance distribution. The proposal density estimates will have high variance resulting in a low acceptance rate.

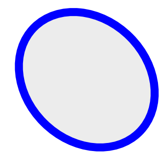
4



We adapt the regeneration distribution by modeling the sampled x_init, x_proposed pairs using a Dirichlet process mixture model. Each time we perform a proposal we add the result to the DPMM. Inference in the DPMM is performed continuously in parallel to our MH chain. We samle x_init from the approximate DPMM conditional distribution given x_proposed; which approaches the optimal distribution (the basin of attraction) with more data.

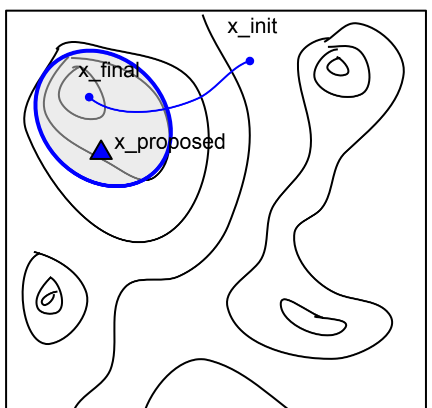


In 3, 4: the importance distribution q(x_init | x_proposed) used to estimate the proposal density.



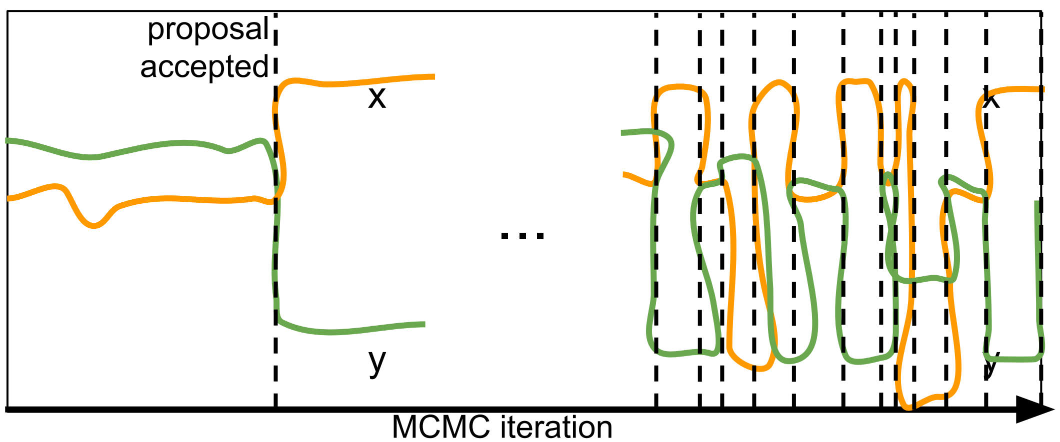
In 2, 3, 4: The distribution p(x_proposed | x_final) representing the noise added to the proposal value after the deterministic optimization. The distribution can be a multivariate Gaussian fit to the mode by using local 2nd order information, or some other simple noise model.

2



A proposal is made by first choosing an initialization point x_init uniformly, then running the optimization procedure to convergence at x_final, and sampling from a ball around x_final.

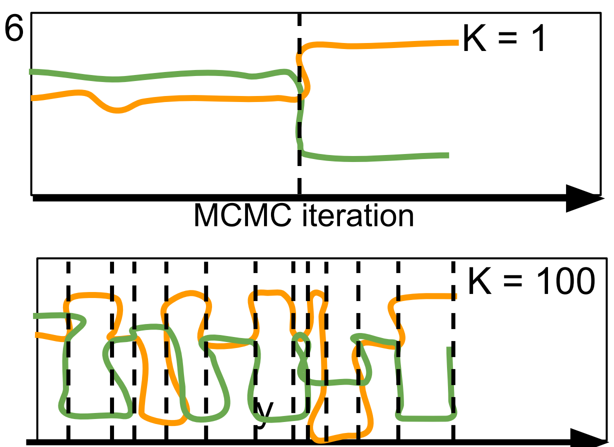
5



We use a cycle of a local random walk proposal and the optimization-based mode-jumping proposal. The rejection rate for the mode-jumping proposal is high when the regenerator is initially poor without adaptation.

After adaptation,the accept rate for the mode-jumping proposal improves. **Note that the proposal distribution itself does not change, only the ability to estimate the proposal density is responsible for the improved acceptance rate.**

6



By sampling multiple regeneration importance particles we can reduce the variance of our proposal density estimates and improve the acceptance rate for a fixed importance distribution without adaptation. K is the number of importance samples.