Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

AN ITERATIVE ALGORITHM FOR SAMPLING FROM MANIFOLDS

Chang-han Rhee

Industrial and Systems Engineering
& Biomedical Engineering
Georgia Institute of Technology
313 Ferst Drive
Atlanta, GA, 30324, USA

Enlu Zhou

Industrial and Systems Engineering
Georgia Institute of Technology
755 Ferst Drive
Atlanta, GA, 30324, USA

Peng Qiu

Biomedical Engineering
Georgia Institute of Technology
313 Ferst Drive
Atlanta, GA, 30324, USA

ABSTRACT

We develop an algorithm that generates samples from a given probability distribution on a manifold embedded in a Euclidean space based only on the ability to evaluate the mapping defined by the parametrization of the manifold. In particular, we do not assume the ability to evaluate the derivatives of the mapping and the ability to tell whether a given point in the ambient space belongs to the manifold or not. The new approach is useful when the manifold is analytically intractable and highly nonlinear—for example, in studying complex regulatory networks in systems biology where the mapping is typically defined by the solution of a system of ordinary differential equations.

1 INTRODUCTION

Suppose that we are interested in studying a manifold $\mathscr{M}$ parametrized by a mapping $f : A \subseteq \mathbb{R}^m \to \mathbb{R}^n$ where $m \leq n$, i.e., $\mathscr{M} = f(A) \subset \mathbb{R}^n$. We are interested in the case where we only have limited information regarding the behavior of f: we assume that we can only evaluate f and the evaluation is computationally expensive. More specifically, we do not have the ability to determine whether a given point y in the ambient space $\mathbb{R}^n$ belongs to $\mathscr{M}$, and we cannot evaluate the derivatives of f exactly. A naive approach would be evaluating f on randomly sampled points $\{x_1, x_2, \ldots\}$ in the parameter space $A \subseteq \mathbb{R}^m$, and study the manifold with the points $\{f(x_1), f(x_2), \ldots\} \subset \mathscr{M}$. This approach can be inefficient if f is a non-linear function which maps the vast majority of the parameters in A into a small region of the manifold, and most of the interesting behavior of the manifold is explained by a small region of A. In such cases, it would be useful if one can generate samples from a given density (with respect to the natural area measure inherited from $\mathbb{R}^n$) on the manifold $\mathscr{M}$ instead of the parameter space. Figure 1 illustrates this point. Consider, for instance, the manifold

$$\mathscr{M} = \{(e^{-\theta_1 t_1} + e^{-\theta_2 t_1}, e^{-\theta_1 t_2} + e^{-\theta_2 t_2}, e^{-\theta_1 t_3} + e^{-\theta_2 t_3}) : 0 \leq \theta_2 < \theta_1 \leq 100\} \tag{1}$$

with $t_1 = 1, t_2 = 2, t_3 = 4$. The left plot of Figure 1 shows the samples obtained by first generating θ_1 and θ_2 from the uniform distribution on the parameter space $[0,100] \times [0,100] \cap \{(\theta_1, \theta_2) : \theta_1 > \theta_2\}$, and then mapping into the manifold $\mathscr{M}$. The right plot was generated from the uniform distribution on $\mathscr{M}$. (The exact meaning of the uniform distribution on a manifold will be discussed in Section 2.) The left plot shows that most of the points in the parameter space are mapped into a very small region in $\mathscr{M}$, and hence, one's understanding of the manifold based on such samples can be limited and even misleading. One can get much better understanding of the manifold by studying uniform samples as we can see from the right plot of Figure 1. We will get back to this example in Section 3.

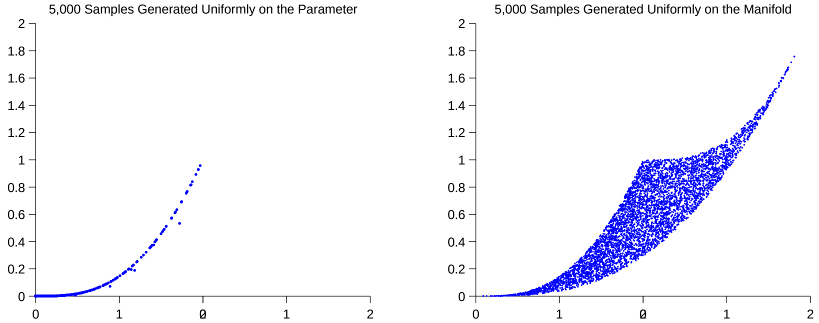

Figure 1: Uniform in parameter vs. uniform on manifold.

In general, uniform samples from a given set can be obtained by constructing a random walk that is confined to the set and simulating the Markov chain associated with the random walk for a sufficiently long time to produce samples from its equilibrium distribution. Designing such a random walk, however, typically requires extra information beyond the ability to evaluate f—such as membership inquiry whether a point of interest belongs to the manifold. Hence, it is not obvious how to carry out such strategies in our context. Another approach is to take advantage of the analytic information of the parametrization. The area formula in geometric measure theory tells us that one can generate uniform samples from a manifold by generating samples from the density proportional to the m-dimensional Jacobian of f in the parameter space and then mapping the samples to the manifold. However, to generate samples directly from such a density, one needs to get a handle on the global properties of the derivatives of f which is not available in many contexts. An obvious approach is, again, to resort on Markov chain Monte Carlo, but it requires multiple evaluation of the derivatives of f to generate a single approximate sample. In this paper, we introduce an iterative algorithm that can generate samples from a given density on a manifold based only on the ability to evaluate f and the target density.

The motivation behind the developments in this paper was to devise an efficient tool for studying biological networks. For example, enzymatic regulatory networks are often modeled with systems of ordinary differential equations. Given a network topology of the regulatory network, the associated differential equations can be considered as a mapping from a set of coefficients (of the system of differential equations) to the solution of the differential equations. Systems biologists can learn about the possible behaviors that can be acheived by the regulatory network with a given network topology by studying the manifold defined by the associated mapping; see, for example, Ma et al. (2009). However, since the analytic solutions of such differential equations are rarely available, one has to resort on (computationally expensive) numerical solutions to evaluate such mappings, and the derivatives are hard to compute accurately. Moreover, such mappings are often highly non-linear so that uniform sampling in the parameter space results in the samples that are concentrated in a small region on the manifold. Because of these difficulties, traditional sampling methods can be inefficient in this context. On the other hand, the new algorithm suggested in this paper is much easier to apply in studying such manifolds because it only requires the ability to evaluate the mapping.

The remainder of this paper is organized as follows: Section 2 discusses the main idea and introduces an iterative algorithm. Section 3 illustrates how the algorithm works with several examples.

2 SAMPLING FROM MANIFOLDS WITHOUT DERIVATIVES

In this section, we introduce an algorithm that generates uniform samples from a manifold $\mathscr{M} \triangleq f(A) \subset \mathbb{R}^n$ based only on the ability to evaluate f. (The algorithm can be easily generalized to sample from a given (non-uniform) probability density on the manifold; we will discuss such an extension later in this section.) More specifically, we assume that f is a one-to-one mapping with sufficient regularity, so that the image $f(A)$ is an m-dimensional manifold embedded in $\mathbb{R}^n$. We wish to generate uniform samples from the manifold: that is, we aim for sampling from the uniform density with respect to the m-dimensional Hausdorff measure

$$\mathscr{H}^m(B) = \lim_{\delta \to 0} \inf_{\substack{B \subseteq \cup S_i, \\ \mathrm{diam}(S_i) \leq \delta}} \sum \alpha_m \left(\frac{\mathrm{diam}(S_i)}{2} \right)^m \tag{2}$$

where the infimum is over all countable coverings $\{S_i \subseteq \mathbb{R}^n : i \in I\}$ of B, $\mathrm{diam}(S_i) \triangleq \sup\{|x-y| : x, y \in S_i\}$, and α_m is the volume of the m-dimensional unit ball. This is a natural extension of Lebesgue measure, and if $m = n$, $\mathscr{H}^m$ coincides with the m-dimensional Lebesgue measure.

As Diaconis, Holmes, and Shahshahani (2013) pointed out, the area formula (see, for example, Federer 1996, Section 3.2.5) dictates how one should sample from the uniform distribution with respect to the Hausdorff measure.

Result (Area Formula) If $f : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz and $m \leq n$,

$$\int_A g(f(x)) J_m f(x) \lambda^m(dx) = \int_{\mathbb{R}^n} g(y)(\#\{A \cap f^{-1}(y)\}) \; \mathscr{H}^m(dy) \tag{3}$$

where λ^m is the m-dimensional Lebesgue measure, A is λ^m-measurable, $g : \mathbb{R}^n \to \mathbb{R}$ is Borel, $\#S$ denotes the cardinality of the set S, and $J_k f$ is the k-dimensional Jacobian of f. In this special case where $k = m \leq n$, the k-dimensional Jacobian is equal to

$$J_m f(x) = \sqrt{\det\big(Df(x)^T Df(x)\big)},$$

where $Df(x)$ is the differential of f at x.

It should be noted that the Lipschitz continuity of f implies that $J_m f$ is well-defined almost everywhere with respect to λ^m by the Rademacher theorem (see, for example, Federer 1996, Section 3.1.6). In theory, formula (3) gives a complete answer to the question of how to generate uniform samples from the manifolds: generate samples $x_1, x_2, \ldots$in parameter space from the density proportional to $J_m f$ and then apply f to $x_1, x_2, \ldots$ to get the uniform samples from the manifold. While this strategy is not viable as such in our context (since it requires explicit knowledge of the derivatives of f), one might consider a procedure that evaluates f at low discrepancy points such as Latin hypercube design points in parameter space, and then approximates the Jacobian via finite difference and interpolating those points. Once a good approximation of the Jacobian is obtained in this way, one can proceed with standard random number generation techniques such as acceptance-rejection or Markov chain Monte Carlo to generate approximate samples from $\mathscr{H}^m$. However, this procedure suffers from the same difficulty as naive uniform sampling in parameter space: it can be inefficient when most of the parameter space is mapped into a small region of the manifold and the rest of the manifold is covered by a small region of the parameter space.

Before introducing our approach, we would like to point out that our setting is different from the traditional setting under which algorithms based on random walks—such as hit-and-run (Boneh and Golan 1979, Smith 1984) and shake-and-bake (Boender et al. 1991)—were developed. In such settings, the manifold is typically a convex set or the boundary of a convex set and the set is specified by a set of constraints (eg. a polytope given by a system of linear inequalities), and hence it is easy to tell whether a given point in $\mathbb{R}^n$ (the ambient space in which the manifold is embedded) is an element of the manifold. In our setting, on the other hand, the manifold is not necessarily a convex set or the boundary of a convex set and we don't have a direct way to answer the membership inquiry. In fact, such question—whether a certain region in the ambient space is covered by the manifold or not—is one of the very questions we would like to address by sampling from the manifold.

The main idea of our approach is to approximate the uniform distribution by starting with a number of random samples and repeatedly perturbing the pre-image of the samples in the parameter space and resampling the perturbed samples on the manifold $\mathscr{M}$. This approach is based on two observations: first, given a set of points on the manifold, one can discard samples that are in concentrated regions and duplicate ones that are in sparse regions so that the remaining samples are distributed more uniformly, and such a procedure can be implemented based only on the Euclidean distances between the points in the ambient space $\mathbb{R}^n$; second, while constructing a perturbation directly on the manifold is hard, we can perturb the pre-images of the samples in the parameter space based on the information we have learned about the mapping f so far. Our algorithm starts with N initial samples $x_1, x_2, \cdots, x_N$ generated in the parameter space and iterates the resampling step and the perturbation step to update the samples. The algorithm returns the image $f(x_1), f(x_2), \cdots, f(x_N)$ of the current samples—or equivalently, the empirical distribution $\eta_N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{f(x_i)}$ of the samples—when the procedure gets close enough to the equilibrium. For large enough N, the returned empirical measure η_N is expected to be a good approximation of the uniform distribution.

In the resampling step, we resample (with replacement) from the current samples with higher probability from the sparse region and with lower probability from the dense region so that the resulting samples cover the support of the original samples (approximately) uniformly. The precise choice of the probability of resampling each point can be determined by viewing this step as applying Boltzmann-Gibbs transformation Ψ_G associated with the potential G to the empirical distribution of the current samples, where G is inversely proportional to the density of the current samples. Let $\mathscr{P}$ be the space of probability measures on $\mathbb{R}^n$. The Boltzmann-Gibbs transformation $\Psi_G : \mathscr{P} \to \mathscr{P}$ associated with the potential $G : \mathbb{R}^n \to \mathbb{R}_+$ is defined as follows:

$$\Psi_G(\eta)(dy) \triangleq \frac{1}{\eta G} G(y)\eta(dy),$$

where ηG denotes $\int_{\mathbb{R}^n} G(y)\eta(dy)$ as usual. Note that if $\eta(dy) = p(y)\mathscr{H}^m(dy)$ and $G(y) = 1/p(y)$, then $\Psi_G(\eta)(dy) = \mathscr{H}^m(dy)$. That is, if the distribution η has density p with respect to the Hausdorff measure, then the Boltzmann-Gibbs transformation associated with the potential G which is inversely proportional to the density p gives the uniform distribution with respect to the Hausdorff measure. In our setting, we of course do not have an exact expression for the density $p(y)$ of the current samples. Instead, we can use the empirical distribution $\tilde{\eta}(dy) = \frac{1}{N}\sum_{i=1}^N \delta_{f(x_i)}$ (where δ_y denotes the Dirac measure) as an approximation of η, and the estimate $\tilde{p}$ of the density p constructed from the samples $f(x_1), \cdots, f(x_N)$ as an approximation of p. Then, for $\widetilde{G}(y) = 1/\tilde{p}(y)$, we expect $\Psi_{\widetilde{G}}(\tilde{\eta})(dy) \approx \mathscr{H}^m(dy)$ on the regions where $\mathscr{M}$ is "well-supported" by the current samples. In this case, the Boltzmann-Gibbs transformation of the empirical distribution can be approximated by resampling from the samples $f(x_1), \cdots, f(x_N)$ with probability proportional to $\big(1/\tilde{p}(f(x_1)), \cdots, 1/\tilde{p}(f(x_N))\big)$.

Note that it is straightforward to extend this idea to non-uniform target distributions. Given a density $r(y)$, one can set $G(y) \triangleq r(y)/p(y)$ instead of $1/p(y)$ so that $\Psi_G(\eta)(dy) = r(y)\mathscr{H}^m(dy)$. For empirical distribution $\tilde{\eta}$, we can approximate $\Psi_{\tilde{G}}(\tilde{\eta})(dy) \approx r(y)\mathscr{H}^m(dy)$ by resampling from $f(x_1), \cdots, f(x_N)$ with probability proportional to $\big(r(f(x_1))/\tilde{p}(f(x_1)), \cdots, r(f(x_N))/\tilde{p}(f(x_N))\big)$.

In the perturbation step we attempt to explore new points $\{x'_1, \cdots, x'_N\}$ in the manifold by perturbing the pre-images $\{x_1, \cdots, x_N\}$ of the current samples $\{f(x_1), \cdots, f(x_N)\}$ in the parameter space. In view of the traditional random walk type approaches, one may consider choosing a line l through the current point x, and picking a point y uniformly from the intersection $\mathscr{M} \cap l$ of the manifold as the next step. However, this strategy is not operable in our context because we don't have the ability to tell whether a point in the ambient space $\mathbb{R}^n$ belongs to the manifold $\mathscr{M}$, and hence we can not even tell if a point belongs to such an intersection. Instead, we can perturb the samples in the parameter space. However, a naive choice of the perturbation such as uniform or normal distribution centered at the current location in the parameter space may lead to a distribution far from uniform when mapped into the manifold $\mathscr{M}$ because of the non-linearity of the mapping f. Our aim in this step is to perturb the current samples and explore the parameter space in a way that the new samples are from a smoothed distribution of the current empirical distribution, so that if $\tilde{\eta}$ is close to uniform, then the empirical distribution of the perturbed points are still close to the uniform distribution. If we can identify the density $q(x)$ from which $x_1, \cdots, x_N$ are generated, we can perturb each sample x_i using diffusion; see Botev, Grotowski, and Kroese (2010) for detailed exposition of the idea of using diffusion in density estimation problem. Note that $q(x)$ solves the Fokker-Planck equation associated with the following diffusion process

$$\begin{aligned} dX(t) &= \sqrt{1/q(X(t))}dW(t); \\ X(0) &= x_i. \end{aligned}$$

Therefore, if x_i is distributed according to q then $X(t) = x_i + \int_0^t \sqrt{1/q(X(s))}dW(s)$ is also expected to be from q for $t > 0$. To approximate such a diffusion, we first approximate q with current samples $\{x_1, \cdots, x_N\}$, and discretize the SDE, for example, with Euler scheme:

$$\begin{aligned} \tilde{X}(t+h) - \tilde{X}(t) &= \sqrt{1/\tilde{q}(\tilde{X}(t))}(W(t+h) - W(t)); \\ \tilde{X}(0) &= x_i \end{aligned}$$

where $\tilde{q}$ is the approximation of q. Then we can simulate the discrete process $\tilde{X}$ up to time T to get $x_i' = \tilde{X}(T)$, whose distribution is close to x_i.

In both resampling and perturbation steps, we need to approximate the density from samples. We suggest computing the approximate density at each sample x_i as follows:

1. Find the k-th nearest neighbor of x_i. Let r be the distance from the current sample to the k-th nearest neighbor;
2. Set the density at the current sample equal to $\frac{k/N}{\alpha_m r^m}$,

where α_m is the volume of the m-dimensional unit ball. Note that due to the resampling, two or more samples can share the same location before the perturbation. If the k-nearest neighbor distance is 0, we first find the sample x_j' whose distance r' from x_i is strictly positive, then count the number k' of samples within the distance r', and set $\tilde{q}$ equal to $\frac{k'/N}{\alpha_m (r')^m}$. Although unlikely for large N, there is a strictly positive probability that all N samples get concentrated in one point after resampling. To deal with this, we introduce a maximum $M_N \gg 1$ which is sufficiently large and $M_N \to \infty$ as $N \to \infty$, and set $\tilde{q}(x) = M_N$ if the estimate is greater than M_N. For $x \notin \{x_1, \ldots, x_N\}$, we set $\tilde{q}(x) = \tilde{q}(x_j)$ where $j = \arg\min_{k=1,\ldots,N} \|x_k - x\|_2$. This is a variant of k-nearest neighbor density estimation (see, for example, Terrell and Scott 1992.)

The resampling and perturbation steps are repeated until the algorithm converges to the uniform distribution. To determine such convergence, one can monitor the distribution of the density estimate $\tilde{p}(f(x_j))$ computed from the k-nearest neighbor distances for reasonably large k. For example, one can stop when $\tilde{p}(f(x_j))/\sum_{i=0}^N \tilde{p}(f(x_i)) - 1/N$ is concentrated around 0 and the average k-nearest neighbor distance stops expanding. The whole procedure described so far in this section is summarized in Algorithm 1.

We expect that this procedure is convergent (i.e., stable) under general conditions; but in case the parameter space A is a nice compact set such as a hypercube, it is easy to see that the algorithm has to be convergent by viewing the procedure as a Markov chain evolving on the state space $A^N \subseteq \mathbb{R}^{m \times N}$. Let $\mathbf{x}(k) \triangleq (x_1(k), x_2(k), \cdots, x_N(k)) \in A^N$ denote the N samples generated in k^{th} iteration. Then, by construction, the density estimate $\tilde{q}$ in the k^{th} perturbation step is bounded from above and below uniformly w.r.t. the configuration $\mathbf{x}(k-1)$ of the previous samples, and hence, the conditional density of $\mathbf{x}(k)$ (given $\mathbf{x}(k-1)$)

should be bounded from below uniformly w.r.t. $\mathbf{x}(k-1)$. This implies that the Markov chain $\mathbf{x}$ is uniformly ergodic, and hence, the algorithm converges at an exponential rate for any fixed N. We also expect that the Algorithm 1 generates samples from the correct distribution (for large N). Consider an idealized version of the algorithm. Let η_0 be an initial distribution with density p_0 on the manifold, and suppose that the support of η_0 covers the entire manifold. Now, recursively define $\eta_{k+1} \triangleq \Psi_{1/p_k}(\eta_k)$ where p_k is the density of η_k with respect to the Hausdorff measure $\mathscr{H}^m$. Then, p_k's are constant for $k \geq 1$; that is, η_k's are uniform distributions with respect to the Hausdorff measure. For large N, the resampling step is a good approximation of the Boltzmann-Gibbs transformation $\Psi_{1/p}$, and the perturbation step changes the resulting distribution little. Therefore, the empirical measure η_k^N of samples generated in the k^{th} iteration becomes good approximation of η_k, which is the uniform distribution. Note that the idealized version achieves the uniform distribution instantly (in a single iteration). This suggests that we can expect a very quick convergence in case η_0^N supports the manifold reasonably well. Such quick convergence behaviors can be observed in Example 1 and 2 in Section 3. On the other hand, Example 3 in Section 3 illustrates that even in the unfortunate cases where the initial samples do not cover the manifold well, our algorithm can explore and discover the hidden part of the manifold efficiently. The theoretical analysis of the convergence of the algorithm and the identification of general conditions for the convergence and the reliable stopping criterion are the major directions for future research.

Algorithm 1 Resample-Perturb Iterative Algorithm

```
Generate $N$ samples $x_1,\cdots,x_N \in A$ from an initial distribution $p_0$
while $\eta_N$ is not uniform or changes significantly do
    $\{x'_1,\cdots,x'_N\} \leftarrow \textsc{Resample}(\{x_1,\cdots,x_N\})$
    $\{x_1,\cdots,x_N\} \leftarrow \textsc{Perturb}(\{x'_1,\cdots,x'_N\})$
    $\eta_N \leftarrow \frac{1}{N}\sum_{i=1}^N \delta_{f(x_i)}$
end while
return $\eta_N$

function $\textsc{Resample}(\{x_1,\cdots,x_N\})$
    $G_i \leftarrow \frac{1/\textsc{ApproxDensity}(f(x_i);f(x_1),\cdots,f(x_N))}{\sum_{j=1}^N 1/\textsc{ApproxDensity}(f(x_j);f(x_1),\cdots,f(x_N))}, \qquad i=1,\cdots,N$
    for $i = 1:N$ do
        Sample $x'_i$ from $\mathbf{P} = (G_1,\cdots,G_N)$    (i.e, $\mathbf{P}(x'_i = x_j) = G_j$)
    end for
    return $\{x'_1,\cdots,x'_N\}$
end function

function $\textsc{Perturb}(\{x'_1,\cdots,x'_N\})$
    $h \leftarrow T/M$
    for $i = 1:N$ do
        $x_i \leftarrow x'_i$
        for $j = 1:M$ do
            $Z \sim N(0,I)$
            $x_i \leftarrow x_i + \sqrt{h/\textsc{ApproxDensity}(x_i;x'_1,\cdots,x'_N)}Z$
        end for
    end for
    return $\{x_1,\cdots,x_N\}$
end function
```

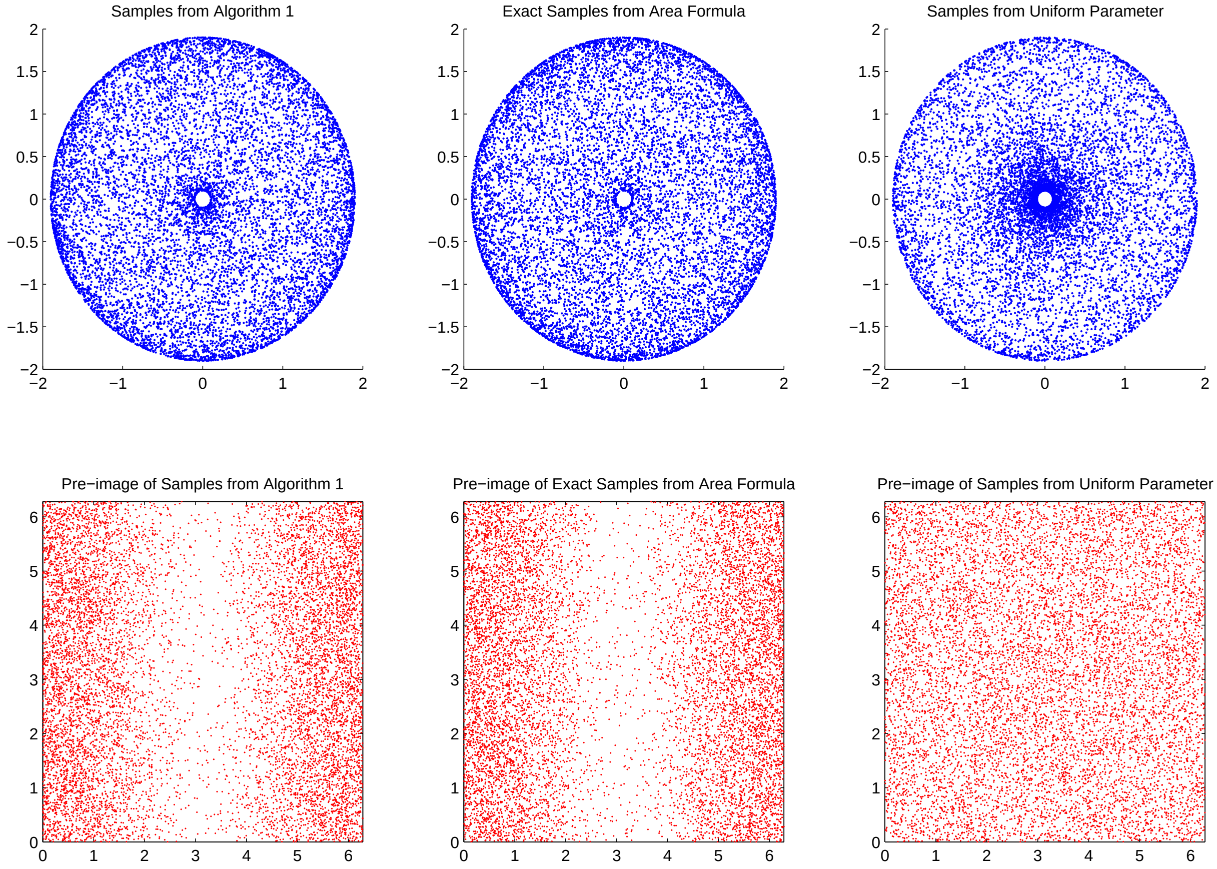

Figure 2: (Uniform Samples From Torus) Comparison of the 10,000 samples from Algorithm 1, area formula, and uniform distribution in parameter space.

3 EXAMPLES

In this section, we examine our algorithm with a few simple examples. For all three examples, we have used $k = 5$ for number of nearest neighbors, $T = 0.05$ for the perturbation size, $n = 30$ for the number of time discretization steps for the diffusion approximation.

Example 1 (Uniform Samples from Torus) Diaconis, Holmes, and Shahshahani (2013) show how to sample from a torus using the area formula (3). Consider a torus

$$\mathscr{M} = \{((R + r\cos\theta)\cos\psi, (R + r\cos\theta)\sin\psi, r\sin\theta) : 0 \le \theta, \psi < 2\pi\}$$

where $0 < r < R$. The major radius R is the distance from the center of the tube to the center of the torus, and the minor radius r is the radius of the tube. An obvious parametrization of $\mathscr{M}$ and its 2-dimensional Jacobian are

$$f(\theta, \psi) = ((R + r\cos\theta)\cos\psi, (R + r\cos\theta)\sin\psi, r\sin\theta), \qquad J_2 f(\theta, \psi) = r(R + r\cos\theta).$$

In view of the area formula (3), one can generate (exactly) uniform samples on $\mathscr{M}$ by generating samples on $[0, 2\pi] \times [0, 2\pi]$ from the density $g(\theta, \psi) \propto R + r\cos\theta$. This can be easily done for example, by generating ψ

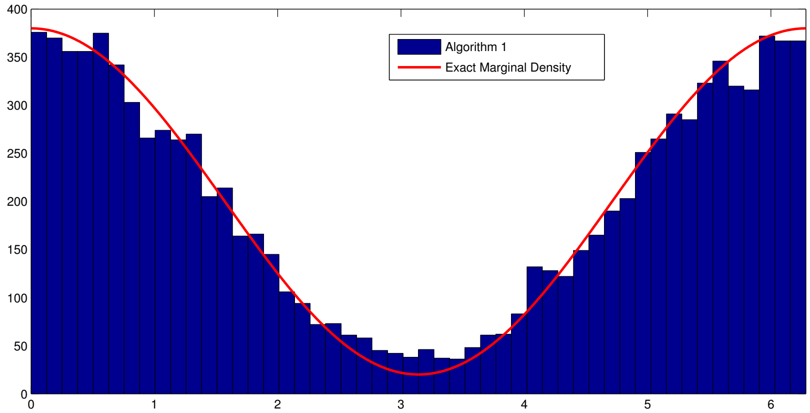

Figure 3: (Uniform Samples From Torus) Histogram from 10,000 samples of θ's generated by Algorithm 1. Red line shows the exact target marginal density computed from the area formula.

from the uniform distribution on $[0,2\pi]$, and (independently) generating θ from the density $\frac{1}{2\pi R}(R+r\cos\theta)$, via acceptance rejection; see Diaconis, Holmes, and Shahshahani (2013) for more detail. Figure 2 compares the samples produced in three different ways for $R=1$ and $r=0.9$. The upper plot shows the samples projected on $x-y$ plane, and the lower plot shows the pre-image of the samples in the parameter space. The two plots on the left show the samples from Algorithm 1 after two iterations. Similarly, the plots in the middle were produced with the 10,000 samples generated by the area formula (as described above), and the right plots show 10,000 samples generated by uniformly sampling in the parameter space, i.e., $\theta\sim U[0,2\pi]$ and $\psi\sim U[0,2\pi]$. We can see that the samples in the upper right plot are more concentrated as they get closer to the center of the torus when compared to the other plots. Figure 3 shows the histogram of θ generated by Algorithm 1, and the exact marginal density of $g(\theta,\psi)=\frac{1}{4\pi^2 R}(R+r\cos\theta)$.

Example 2 (Non-uniform Density on Torus) In this example, we consider the same manifold $\mathscr{M}$ as in Example 1, but we sample from a non-uniform distribution. Suppose that we are interested in the shape of $\mathscr{M}$ in the proximity of a given point, say $(0,1,0)$, and hence, we want to sample more frequently from the closer parts of the manifold to the point, and less frequently from farther parts of the manifold. For the purpose, we choose a density proportional to the reciprocal of the squared distance from $(0,1,0)$. That is, we sample from $P(d\mathbf{x})=r(\mathbf{x})\mathscr{H}^2(d\mathbf{x})$ where $r(x,y,z)\propto 1/(x^2+(y-1)^2+z^2)$. Again, in view of the area formula, one can sample from P by sampling from a density proportional to

$$r(f(\theta,\psi))g(\theta,\psi)\propto\frac{R+r\cos\theta}{((R+r\cos\theta)\cos\psi)^2+((R+r\cos\theta)\sin\psi-1)^2+(r\sin\theta-2)^2}.$$

Figure 4 and 5 compare the samples produced by Algorithm 1 (after two iterations), exact samples generated by the area formula, and the samples generated uniformly in the parameter space.

Example 3 (Exponential Model) Here we consider a manifold

$$\mathscr{M}=\{(e^{-\theta_1 t_1}+e^{-\theta_2 t_1},e^{-\theta_1 t_2}+e^{-\theta_2 t_2},e^{-\theta_1 t_3}+e^{-\theta_2 t_3}):0\le\theta_2<\theta_1\le 100\}$$

where $0<t_1<t_2<t_3$. The first derivative

$$Df(\theta_1,\theta_2)=\begin{pmatrix}-t_1\exp(-\theta_1 t_1) & -t_1\exp(-\theta_2 t_1)\\ -t_2\exp(-\theta_1 t_2) & -t_2\exp(-\theta_2 t_2)\\ -t_3\exp(-\theta_1 t_3) & -t_3\exp(-\theta_2 t_3)\end{pmatrix}, \tag{4}$$

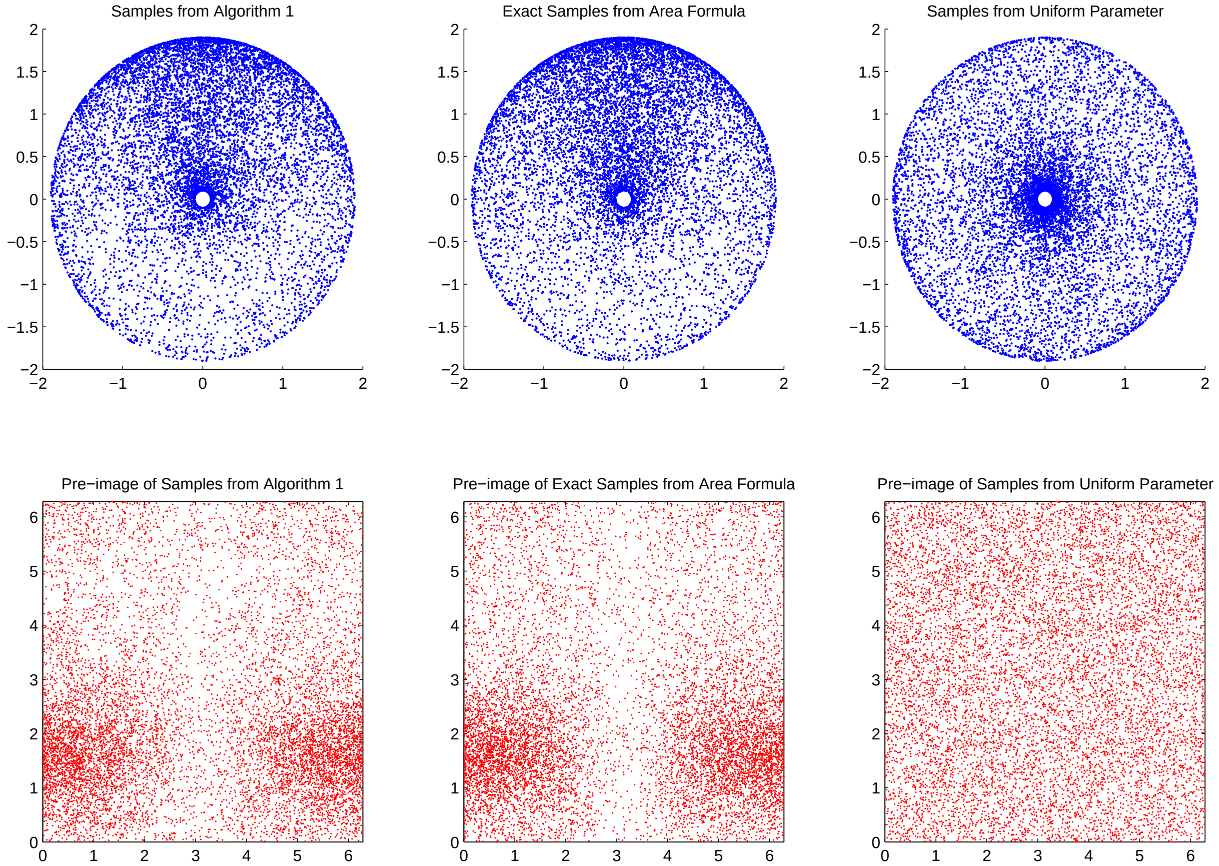

Figure 4: (Non-Uniform Density on Torus) Comparison of the 10,000 samples from Algorithm 1, area formula, and uniform distribution in parameter space.

and the 2-dimensional Jacobian

$$J_2 f(\theta_1,\theta_2) = \sqrt{\alpha_{12}+\alpha_{13}+\alpha_{23}} \tag{5}$$

where

$$\alpha_{ij} = t_i^2 t_j^2 \exp\{-2(\theta_1 t_j + \theta_2 t_i)\}\{\exp(\theta_1-\theta_2)(t_j-t_i)-1\}^2. \tag{6}$$

Figure 6 shows the result for $t_1 = 1$, $t_2 = 2$, $t_3 = 4$ with $5,000$ samples. The left plot was produced by Algorithm 1, (the upper plot shows the samples projected on a plane perpendicular to the vector $(1,-1,0)$, and the lower plot shows the pre-image of the samples in the parameter space); the plots in the middle show the 5,000 samples generated by the area formula; the right plots show 5,000 samples generated by uniformly sampling in the parameter space, i.e., $\theta_1,\theta_2 \sim \text{Unif}\big([0,100]\times[0,100]\cap\{\theta_1>\theta_2\}\big)$. The samples generated uniformly in the parameter space are concentrated in a small part of the boundary of the manifold. The Algorithm 1 was started with initial samples distributed uniformly in the parameter space, and the final samples were obtained after 10 iterations. The progression of the algorithm is illustrated in Figure 8.

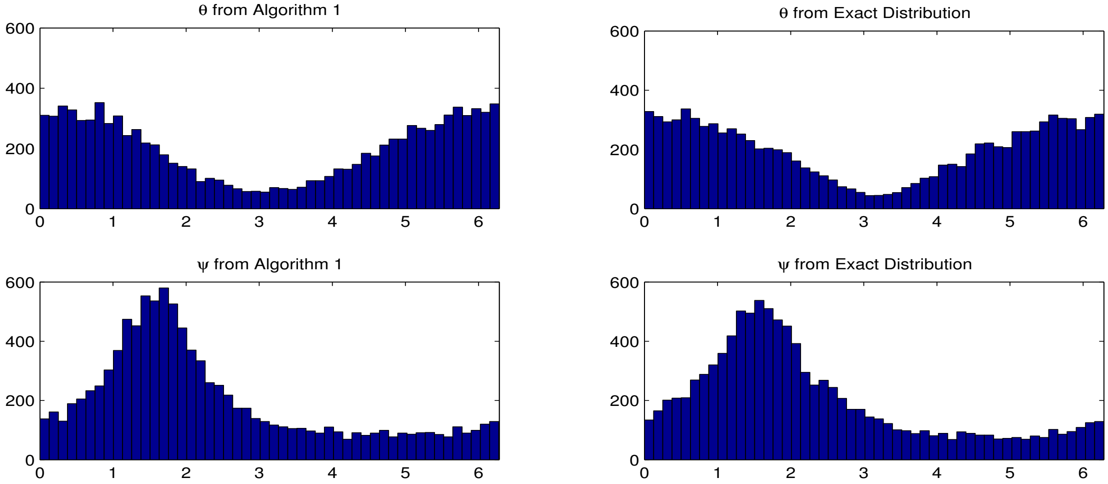

Figure 5: (Non-Uniform Density on Torus) Histograms from 10,000 samples generated by Algorithm 1, and the exact area formula.

REFERENCES

Boender, C., R. Caron, A. Rinnooy Kan, J. McDonald, H. Edwin Romeijin, R. L. Smith, J. Telgen, and A. Vorst. 1991. “Shake-and-bake algorithms for generating uniform points on the boundary of bounded polyhedra”. *Operations Research* 39 (6): 945–954.

Boneh, A., and A. Golan. 1979. “Constraints redundancy and feasible region boundedness by random feasible point generator (RFPG)”. In *Proceedings of the 3rd European Congress on Operations Research*. Amsterdam.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese. 2010. “Kernel density estimation via diffusion”. *Annals of Statistics* 38 (5): 2916–2957.

Diaconis, P., S. Holmes, and M. Shahshahani. 2013. “Sampling from a manifold”. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*, edited by G. Jones and X. Shen, 102–125.

Federer, H. 1996. *Geometric Measure Theory*. Berlin: Springer.

Ma, W., A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. 2009. “Defining Network Topologies that Can Achieve Biochemical Adaptation”. *Cell* 138 (4): 760–773.

Smith, R. L. 1984. “Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions”. *Operations Research* 32 (6): 1296–1308.

Terrell, G. R., and D. W. Scott. 1992. “Varaible kernel density estimation”. *Annals of Statistics* 20 (3): 1236–1265.

AUTHOR BIOGRAPHIES

CHANG-HAN RHEE is currently a postdoctoral fellow in the School of Industrial & Systems Engineering and Department of Biomedical Engineering at Georgia Institute of Technology. He holds a M.S. and Ph.D. in Computational and Mathematical Engineering from Stanford University. He graduated with B.Sc. in the Departments of Mathematics and Computer Science at Seoul National University, South Korea. His research interests include stochastic simulation, sensitivity analysis, experimental design, and stochastic control. His email address is rhee@gatech.edu and his web page is http://rhee.gatech.edu/.

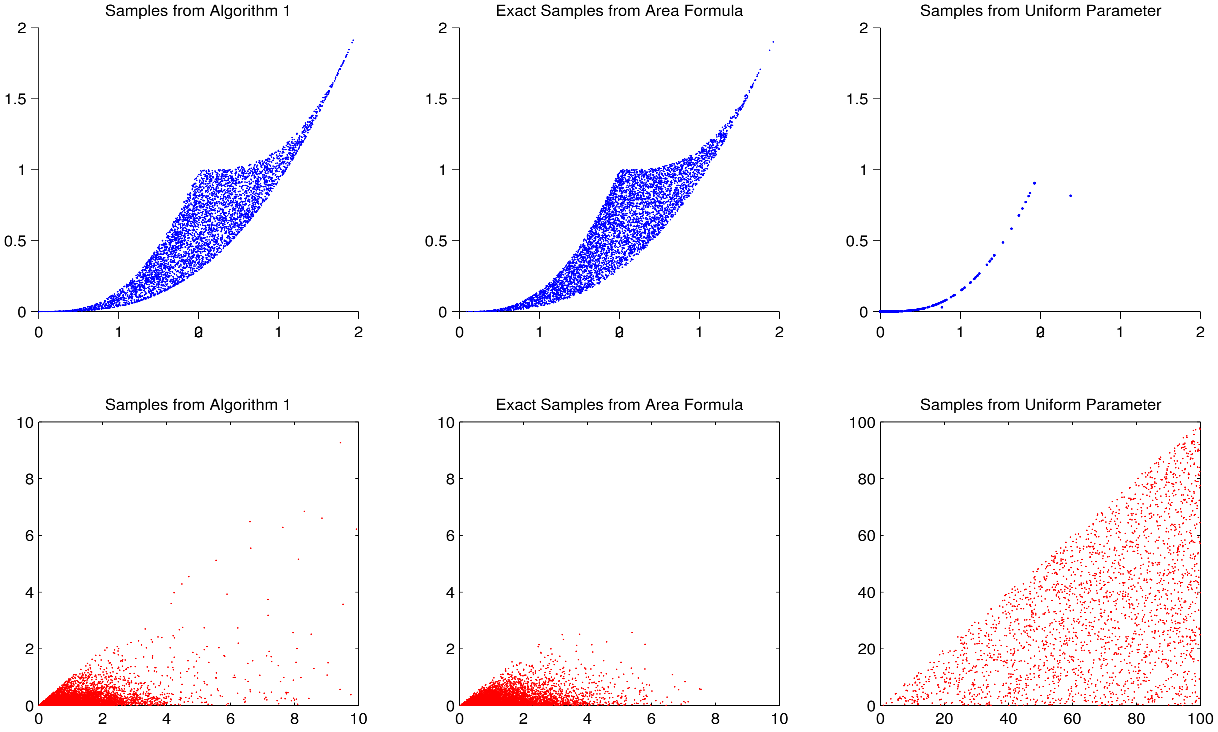

Figure 6: (Exponential Model) Comparison of the 5,000 samples from Algorithm 1, area formula, and uniform distribution in parameter space.

ENLU ZHOU is an Assistant Professor in the H. Milton Stewart School of Industrial & Systems Engineering at Georgia Institute of Technology. She received the B.S. degree with highest honors in electrical engineering from Zhejiang University, China, in 2004, and the Ph.D. degree in Electrical Engineering from the University of Maryland, College Park, in 2009. She then joined the faculty of the University of Illinois at Urbana-Champaign in the Department of Industrial and Enterprise Systems Engineering. In 2013, she joined the H. Milton Stewart School of Industrial & Systems Engineering at Georgia Tech. Her research interests include stochastic control and simulation optimization, with applications towards financial engineering. Her email address is enlu.zhou@isye.gatech.edu and her web page is http://enluzhou.gatech.edu/.

PENG QIU is an Assistant Professor in the Department of Biomedical Engineering at Georgia Institute of Technology. He received the B.S. degree in Electronic Science and Technology from University of Science and Technology of China, China, in 2003, and the Ph.D. degree in Electrical Engineering from the University of Maryland, College Park, in 2007. He was a postdoc fellow at the Departement of Radiology at Stanford University from 2007 to 2010. He then joined the faculty of the University of Texas M.D. Anderson Cancer Center in the Department of Bioinformatics and Computational Biology. In 2013, she joined the Department of Biomedical Engineering at Georgia Tech. His research interests are in bioinformatics and computational biology, focusing on statistical signal processing, machine learning, control systems and optimization. His email address is peng.qiu@bme.gatech.com and his web page is http://pengqiu.gatech.edu/.

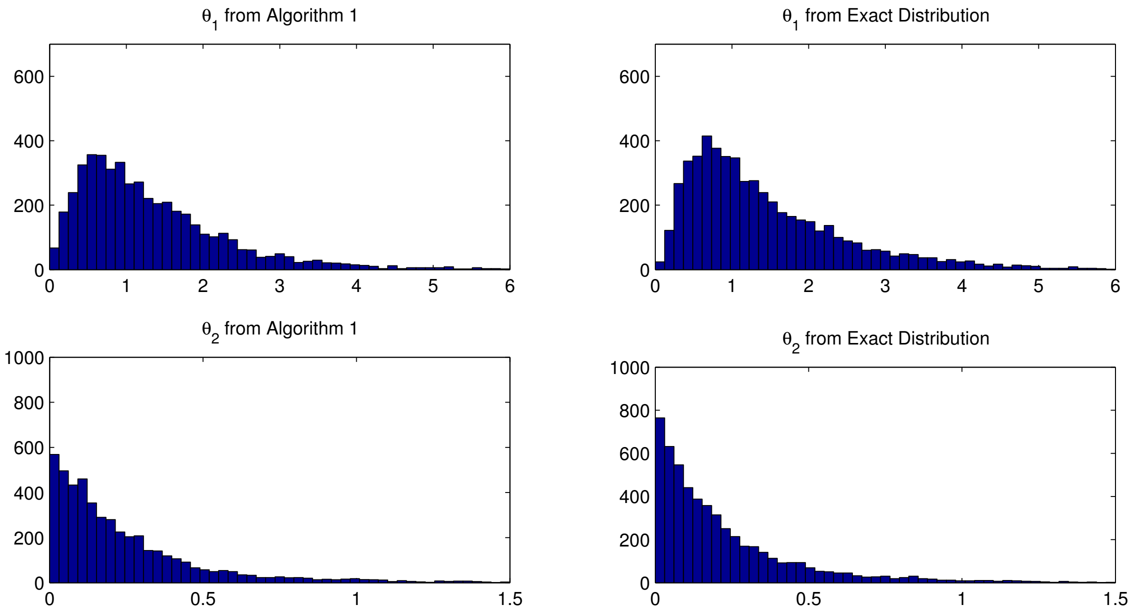

Figure 7: (Exponential Model) Histograms from 5,000 samples generated by Algorithm 1, and the exact area formula.

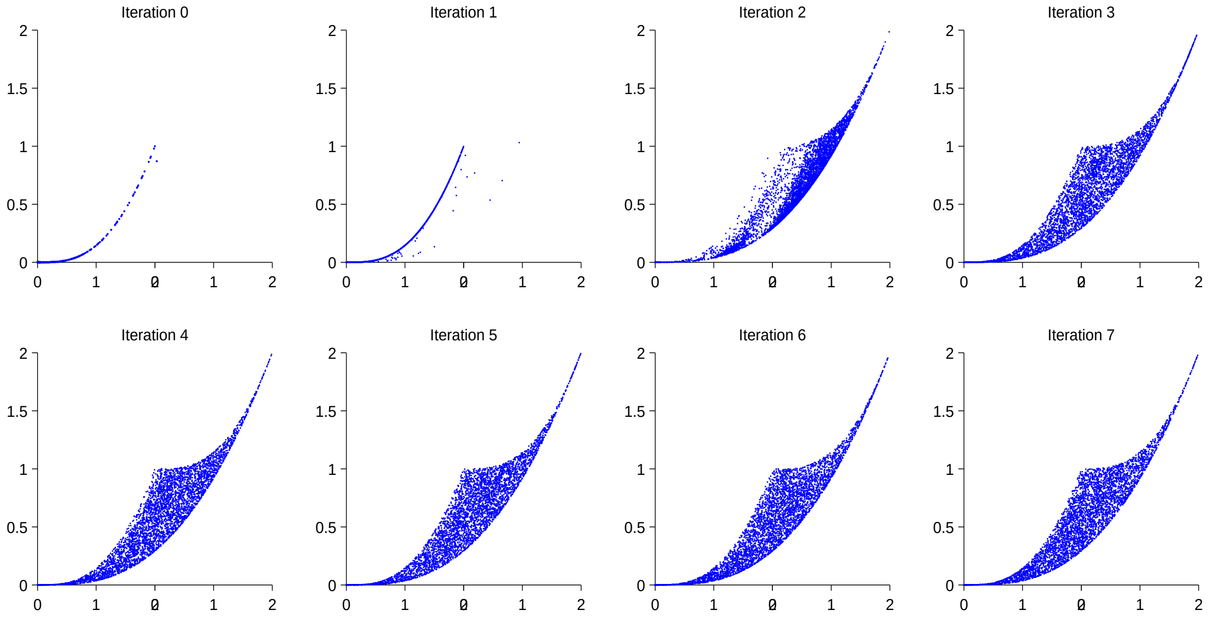

Figure 8: (Exponential Model) The progression of Algorithm 1.