EXPO\textit{NENTALLY CONVERGENT ALGORITHM TO GENERATE RANDOM POINTS IN A $d$–DIMENSIONAL BODY}

BENO\textsc{it} COLLINS, TERMEH KOUSHA, RAFA\textsc{L} KULIK, TOMASZ SZAREK, AND KAROL ŻYCZKOWSKI

\textsc{Abstract}. An algorithm to generate random points inside an arbitrary $d$–dimensional convex body $X$ with respect to the flat (Lebesgue) measure is proposed. It can be considered as an iterated functions system (IFS) with an infinite number of functions acting on $X$. We analyze the corresponding Markov operator which acts on the probability measures, and show that any initial measure converges exponentially to the uniform measure. Estimations for the convergence rate are derived in terms of the dimension $d$ and the ratio between the radius of the sphere inscribed in $X$ and the radius of the outscribed sphere. Concrete estimations are provided for the Birkhoff polytope containing bistochastic matrices, the set of quantum states acting on $N$–dimensional Hilbert space and its subset consisting of states with positive partial transpose.

1. Introduction

In this paper we consider the problem of generating a high dimensional random variable according to a given probability distribution. We focus on the case where the random variable lies in a vector space and the probability distribution is absolutely continuous and has smooth density with respect to the Lebesgue measure.

The simplest and fully solved case is the when all entries are independent. In this case, it is enough to sample each entry independently at random. But if the entries fail to be independent, it becomes difficult to sample from the given distribution efficiently.

Let us give some examples where sampling is still possible.

(i) In one dimension, it is possible to sample as soon as one is able to compute the cumulative function.

(ii) In dimension $n$, let $A$ be a Borel subset of $\mathbb{R}^n$, and $\mu$ be a measure on $\mathbb{R}$. If one wants to simulate the conditional probability $\mu_A(B) = \mu^\otimes(B \cap A)/\mu^\otimes(A)$, it is enough to sample according to $\mu^\otimes$ and keep the simulated value if an only if the random sample belongs to $A$. This technique is exact but it is efficiently implemented if and only if $\mu^\otimes(A)$ is large. In general, we are interested in a sequence of probability spaces indexed by $n$ with $n \to \infty$. In this setup, replacing $A$ by $A_n$, $\mu^\otimes(A_n)$ decays quickly to zero and this technique can not be applied.
(iii) In higher dimension too, there exists tricks to simulate interesting singular measures, such as for example the uniform measure on the unit sphere. In this case, it is enough to sample according to $\mathcal{N}(0,1)^{*n}$ and divide by the $l^2$. Although some interesting things can be said in a more general context of measures on manifolds, we do not discuss this problem here and plan to address it elsewhere.

Let us note also that the second item above can be generalized to the more general setup of a probability measure on $\mathbb{R}^n$ that is not uniform, but rather absolutely continuous with respect to the Lebesgue measure, and has a continuous density supported on a compact set. Denoting this density by $f$ and its support by $K$, due to the Fubini theorem it is enough to simulate the uniform measure on the compact set $\{(x, y), 0 \leq y \leq f(x)\} \subset K \times \mathbb{R}_+$ and take the first component. In principle this can be achieved through the conditional probability methods. This technique is sometimes known as the rejection method.

This is going to be the framework in which we are interested. Note however, that the measure of the set on which the necessary conditions are fulfilled can be too small to have any practical use in a large dimension. Therefore, instead of trying to achieve an explicit sampling, we analyze in this paper a scheme of an approximate sampling, that is close to the targeted sampling in total variation norm. Some techniques using SDES to achieve this have been developed but they are less systematic. Here we will focus on techniques based on Markov chains. Our approach may be considered as a variant of the slice sampling – see also [13, 15, 16, 17, 9].

In particular, our assumptions will be good enough to deal with the case of the uniform measure on convex bodies. This will lead us to a subsequent part of our paper where we deal with important applications of our results to Quantum Information Theory. There, we give theoretical convergence rates and also concrete simulations.

One important technical assumption that we will need to make is that the density of the measure to be simulated can be ‘easily’ computed at a given point. For example, if we are interested in the uniform measure on the convex body of separable states, the problem of finding out whether a state is separable or not is a ‘hard’ problem [6], and in particular it is ‘hard’ to tell whether the density is zero or not.

One important feature of our algorithm is that we don’t need to know the measure itself, but only a multiple of it. This is important because sometimes it is very difficult to compute partition functions.

The paper is organized as follows. In section 2 we formulate the problem and introduce necessary tools. In sections 3 and 4 we prove the convergence theorems for two types of sampling and provide general estimates for the convergence rates. These estimates are worked out in section 5 first for standard geometric objects, and it is shown that the convergence is faster for $d$-simplices than for $d$-cubes and $d$-balls. Later similar estimations are obtained for convex sets used in physical applications, including the Birkhoff polytope of bistochastic matrices, the set of quantum states and its subset consisting PPT states with positive partial transpose. The paper is concluded with some remarks in section 6.
2. CONVERGENCE THEOREM, GENERAL SETUP

Let \((X, \mathcal{A})\) and \((I, \mathcal{B})\) be two measurable spaces and let \(\mu, \nu\) be two probability measures on \(X\) and \(I\), respectively.

We shall assume that for any \(i \in I\) we have a transition function \(T_i : X \times \mathcal{A} \to [0, 1]\), i.e. \(T_i(x, \cdot)\) is a probability measure for any \(x \in X\) and for any \(A \in \mathcal{A}\) the function \(T_i(\cdot, A) : X \to [0, 1]\) is measurable. Additionally, we assume that \(\mu\) is invariant with respect to \(T_i\) for any \(i \in I\):

\[
\mu(A) = \int_X T_i(x, A)\mu(dx) \quad \text{for all } A \in \mathcal{A}.
\]

Now, if we assume that for any \(A \in \mathcal{A}\) the function \(T(\cdot, A) : I \times X \to [0, 1]\) is \(\mathcal{B} \otimes \sigma \mathcal{A}\)-measurable, then it follows from the Fubini theorem that the measure \(\mu\) is invariant with respect to the operator \(Q\) of the form

\[
Q\hat{\mu}(\cdot) = \int_X \int_I T_i(x, \cdot)\hat{\mu}(dx)\nu(di).
\]

By \(\mathcal{M}_1\) and \(\mathcal{M}\) we shall denote the set of all Borel measures and all probability Borel measures on \(X\), respectively.

By \(\| \cdot \|_{TV}\) we denote the total variation norm, i.e., if \(\hat{\mu} \in \mathcal{M} \setminus \mathcal{M}\), then \(\|\hat{\mu}\|_{TV} := \hat{\mu}^+(X) + \hat{\mu}^-(X)\), where \(\hat{\mu} = \hat{\mu}^+ - \hat{\mu}^-\) is the Jordan decomposition of the signed measure \(\hat{\mu}\).

We start with the following version of Doebelin’s theorem [5], which provides sufficient conditions for exponential convergence rates of the transition operator \(Q\).

**Proposition 2.1.** Assume that there exist \(\theta \in (0, 1), M \in \mathbb{N}\) and a measure \(\nu \in \mathcal{M}_1\) such that for any measurable set \(A\)

\[
Q^M(x, A) \geq \theta \nu(A) \quad \text{for any } x \in X.
\]

Then there exists a unique invariant measure \(\mu_* \in \mathcal{M}_1\) such that

\[
\|Q^n\mu - \mu_*\|_{TV} \leq C\alpha^n \quad \text{for all } \mu \in \mathcal{M}_1 \text{ and } n \geq 1,
\]

with the convergence rate \(\alpha = (1 - \theta)^{1/M}\) and prefactor \(C = 2(1 - \theta)^{-1}\).

Assume now that \(X\) is a bounded metric space and let \(\varphi : X \to \mathbb{R}\) be a Lipschitz function such that \(\int_X \varphi(x)\mu(dx) = 0\). Then we have (see Theorem 17.5.4 in [12]):

**Proposition 2.2.** Let \((X_n)\) be the Markov chain corresponding to the transition operator \(Q\). Under the hypothesis of Proposition 2.1, we have the Central Limit Theorem (CLT),

\[
\frac{\sum_{i=1}^n \varphi(X_i)}{\sqrt{n}} \Rightarrow W, \quad \text{as } n \to +\infty,
\]
where $W$ is a random variable with normal distribution $\mathcal{N}(0, D)$ for some $D \geq 0$ and the convergence is understood in law. Moreover, we have the Law of the Iterated Logarithm (LIL),

$$\limsup_{n \to +\infty} \frac{\sum_{i=1}^{n} \varphi(X_i)}{\sqrt{2n \log \log n}} = D$$

with probability 1. Of course the above implies that also

$$\liminf_{n \to +\infty} \frac{\sum_{i=1}^{n} \varphi(X_i)}{\sqrt{2n \log \log n}} = -D$$

with probability 1.

3. Convergence rate for the fixed basis algorithm

We first describe the fixed basis algorithm. This algorithm is conceptually the most closest one to slice sampling and can be seen as the first modification thereof. It also has the advantage that it is defined in a way that can be potentially generalized to manifolds (in contrast to algorithm 2).

3.1. Description of algorithm 1: fixed basis sampling. We consider a compact set $X \subset \mathbb{R}^d$. In a preliminary step let us choose a set of normed vectors $e = \{e_1, \ldots, e_l\}$ in $\mathbb{R}^d$ for $l \geq d$ such that $\text{Lin}\{e_1, \ldots, e_l\} = \mathbb{R}^d$. To generate a sequence of random points in $X$ repeat the following steps of the algorithm, illustrated in Fig. 1.

(i) Choose an arbitrary starting point $x_0 \in X$,

(ii) Draw randomly a vector $e_i$, where the direction $i$ is chosen with a uniform distribution among $(1, \ldots, l)$.

(iii) Find boundary points $x_1^{\min}, x_1^{\max} \in \partial X$ along the direction $e_i$: there exist positive numbers $a, b$ such that $x_1^{\min} = x_0 - ae_i$ and $x_1^{\max} = be_i$.

(iv) Select a point $x_1$ randomly with respect to the uniform measure in the interval $[x_1^{\min}, x_1^{\max}]$.

(v) Repeat the steps (ii)-(iv) to find subsequent random points $x_2, x_3, \ldots$.

This algorithm is very close to slice sampling. The main difference is that slice sampling proceeds recursively. It assumes one is able to simulate a uniform measure on an slice of codimension 1 (obtained by intersecting with an $n-1$ dimensional affine hyper plane).

3.2. Convergence rate with respect to the Lebesgue measure. Let $X$ be a compact subset $\mathbb{R}^d$ with a nonempty interior. Let a point $x_*$ in the interior of $X$ be given. We introduce two positive constants, $r$ and $R$, such that $B(x_*, r) \subset X \subset B(x_*, R)$, where $B(x_*, r)$ denotes a closed ball in $\mathbb{R}^d$. Let $e = \{e_1, \ldots, e_l\}$ such that $\text{Lin}\{e_1, \ldots, e_l\} = \mathbb{R}^d$
Figure 1. Algorithm to generate random points uniformly in a compact set $X$: one starts with any interior point $x_0 \in X$, picks randomly a direction (in this case $e_1$) finds two boundary points and draws $x_1$ randomly in the interval $[x_{1 \min}, x_{1 \max}]$. In the next step one chooses a next direction (here $e_2$), finds both boundary points and draws $x_2$ randomly in the interval $[x_{2 \min}, x_{2 \max}]$.

be given. We consider the Markov chain $\Phi$ corresponding to the algorithm illustrated in Fig. 1, which is described by the following transition function

\[
T(x, A) = \frac{1}{l} \sum_{i=1}^{l} \nu_{x, i}(A),
\]

where $\nu_{x, i}$ is the measure uniformly distributed over the set $\{y \in X : y-x = te_i \text{ for some } t \in \mathbb{R}\}$.

To characterize the set $X$ and the set of vectors $e = \{e_1, \ldots, e_l\}$, we will need the notion of $k$-accessibility illustrated in Fig. 2.

**Definition.** A compact set $X \subset \mathbb{R}^d$ is called $k$-accessible with respect to $e$, if there exists $x_* \in \text{int}X$ such that from any point $x \in X$ one can reach $x_*$ in not more than $k$ moves along the basis vectors. Thus there exist some sets $\{i_1, \ldots, i_k\} \subset \{1, \ldots, l\}$ and $\{\lambda_1, \ldots, \lambda_k\}, \lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that

\[
x + \sum_{j=1}^{m} \lambda_j e_{i_j} \in \text{int}X \quad \text{for any } m \leq k \quad \text{and} \quad x + \sum_{j=1}^{k} \lambda_j e_{i_j} = x_*.\]

**Theorem 3.1.** If $X \subset \mathbb{R}^d$ is $k$-accessible with respect to the basis $e = \{e_1, \ldots, e_l\}$, $x_* \in X$ and $B(x_*, r) \subset X \subset B(x_*, R)$ for some $r, R > 0$, then the chain $\Phi$ corresponding to the transition function $T$ given by (3.1) satisfies the hypothesis of Proposition 2.1 with

\[
M = k + d \quad \text{and} \quad \theta = b_d l^{-k-d} (r/R)^{k+d},
\]

where $b_d = \pi^{d/2} / \Gamma(d/2 + 1)$ denotes the volume of a unit ball in $\mathbb{R}^d$.

**Proof.** First, there is no restriction in assuming that $r \leq 1$. Fix $x \in X$ and let $\{i_1, \ldots, i_k\}$ and $\{\lambda_1, \ldots, \lambda_k\}$ be given according to $k$-accessibility of $X$. We are going to derive lower bounds for subdensities $f_m$ of $T^m(x, A)$ for $m = 1, \ldots, k$ defined on suitable spaces. Firstly,
Figure 2. Examples of $k$–accessibility in $2D$: a) an ellipse is $2$–accessible with respect to any orthogonal basis: any point $x$ can be transformed into a selected point $x_*$ in two moves along the basis vectors; b) the triangle $ABC$ is not accessible with respect to basis $(e_1, e_2)$, as starting from the corner $A$ one cannot move along the basis vectors, but this triangle is $2$–accessible with respect to the basis $(f_1, f_2)$.

Let $\tilde{X}_1 := \{x + t_1e_{i_1} : t_1 \in \mathbb{R}\} \cap X$. By the fact that $\tilde{X}_1$ contains a non-degenerate interval its one dimensional measure $\mathcal{L}_1$ is positive. Set $f_1(y) = 1/R$ for $y \in \tilde{X}_1$. Finally, we easily check that for any measurable set $A$ the following inequality holds

$$T(x, A) \geq (1/l) \int_{A \cap \tilde{X}_1} f_1 d\mathcal{L}_1.$$ 

Define $\tilde{X}_2 = \bigcup_{t_2 \in \mathbb{R}} (\tilde{X}_1 + t_2e_{i_2}) \cap X$. The set $\tilde{X}_2 \neq 0$ and its $t_2$–dimensional Lebesgue measure $\mathcal{L}_{t_2}$ is positive, where $t_2 = \dim \text{Lin}\{e_{i_1}, e_{i_2}\}$, by $k$–accessibility of $X$. Set $f_2(y) := (1/R)^2$ for $y \in \tilde{X}_2$ and observe that

$$T^2(x, A) \geq (1/l)^2 \int_{A \cap \tilde{X}_2} f_2 d\mathcal{L}_{t_2}.$$ 

By induction we define the sets $\tilde{X}_3, \ldots, \tilde{X}_m$. If we have done it for $m < k$, we may do it for $m + 1$ as well. Namely, we set $\tilde{X}_m := \bigcup_{t_m \in \mathbb{R}} (\tilde{X}_{m-1} + t_me_{i_m}) \cap X$. Then the set $\tilde{X}_m$ has positive $t_m$–dimensional Lebesgue measure, where $t_m = \dim \text{Lin}\{e_{i_1}, \ldots, e_{i_m}\}$. Moreover,

$$T^m(x, A) \geq (1/l)^m \int_{A \cap \tilde{X}_m} f_m d\mathcal{L}_{t_m},$$

where $f_m(y) := (1/R)^m$ for $y \in \tilde{X}_m$. In this way we obtain the constant $t_k = \dim \text{Lin}\{e_{i_1}, \ldots, e_{i_k}\}$ and the set $\tilde{X}_k \ni x_0$. Set

$$e_{i_{k+1}} = e_{q_1}, \ldots, e_{i_{k+d}} = e_{q_d},$$

where $q_1, \ldots, q_d \in \{1, \ldots, l\}$ are such that $\text{Lin}\{e_{q_1}, \ldots, e_{q_d}\} = \mathbb{R}^d$.

Now we repeat the procedure $d$ more times. Then $\tilde{X}_{k+d} \supset B(x_0, r)$ and $t_{k+d} = d$ and we finally obtain

$$T^{k+d}(x, A) \geq (1/R)^{k+d} \mathcal{L}_d(A \cap B(x_0, r)).$$
Since, \( B(x_0, r) \subset \tilde{X}_{k+d} \) we obtain
\[
T^{k+d}(x, A) \geq (1/Rl)^{k+d}b_d r^d \nu(A) \geq b_d l^{-k-d}(r/R)^{k+d} \nu(A),
\]
where \( \nu(A) = \mathcal{L}_d(A \cap B(x_0, r)) (\mathcal{L}_d(B(x_0, r)))^{-1} \).

This completes the proof. \( \square \)

From Proposition 2.2 it follows that

**Corollary 3.1.** Let \((X_n)_{n \geq 1}\) be the Markov chain corresponding to the transition function \(T\) and let \(\phi\) be an arbitrary Lipschitz function on \(X\) such that \(\int_X \phi d\mathcal{L}_d = 0\). Then \((\phi(X_n))_{n \geq 1}\) satisfies the CLT and LIL.

### 3.3. Case where the density is not uniform.

In this section we consider the case where the measure on a compact subset \(K\) of \(\mathbb{R}^n\) is not just the restriction of the Lebesgue measure, but a more general measure. We assume that the measure has a continuous strictly positive density with respect to the Lebesgue measure, and denote this density by \(f\).

In this case, we consider the subset \(\tilde{K}\) of \(\mathbb{R}^n \times \mathbb{R}^+\) given by
\[
\tilde{K} = \{(x, y) : x \in K, 0 < y \leq f(x)\}.
\]
This is clearly a compact set, and we can prove readily the following lemma, from which we derive our estimates.

**Proposition 3.1.** Assume that we are under the hypotheses of Theorem 3.1 for the set \(\tilde{K}\). Then the law of the first component \(x\) of \((x, y)\) converges towards the probability measure that we want to simulate \(f(x)dx\) at the speed given by Theorem 3.1.

**Proof.** This is a direct application of Fubini’s theorem. \( \square \)

### 4. Convergence rate for the random direction algorithm

Next, we describe algorithm 2, obtained from sampling in a random direction. As we will see, this algorithm converges faster in general, although it is conceptually less tied to slice sampling.

#### 4.1. Description of algorithm 2: random direction sampling.

We present here an alternative variant of the sampling algorithm, in which no fixed basis is selected, and the direction of a random walk is chosen randomly. As before we consider a compact set \(X \subset \mathbb{R}^d\).

(i) Choose an arbitrary starting point \(x_0 \in X\),

(ii) Draw randomly a unit vector \(e = [0, x]\), such that \(x \in S^d\) is a point taken according to the uniform distribution on the \(d\)-sphere,
(iii) Find boundary points $x_{1}^{\min}, x_{1}^{\max} \in \partial X$ along the direction $e$: there exist positive numbers $a, b$ such that $x_{1}^{\min} = x_{0} - ae$ and $x_{1}^{\max} + be$.

(iv) Select a point $x_{1}$ randomly with respect to the uniform measure in the interval $[x_{1}^{\min}, x_{1}^{\max}]$.

(v) Repeat the steps (ii)-(iv) to find subsequent random points $x_{2}, x_{3}, \ldots$

The only difference with respect to the previous version of the procedure is that each time the direction $e$ is taken randomly and that the direction used in step $i + 1$ does not depend on the direction used in step $i$.

4.2. Convergence rate with respect to the Lebesgue measure. For algorithm 2 based on a random choice of the direction one can obtain a convergence estimate analogous to Theorem 3.1.

Let $\Phi$ be a Markov chain described by the above algorithm and let $S$ be its transition kernel.

**Theorem 4.1.** Let $X \subset \mathbb{R}^{d}$ be a compact convex set. Assume that there exist $r, R > 0$ such that for some $x_{0} \in X$ one has $B(x_{0}, r) \subset X \subset B(x_{0}, R)$ for some $r, R > 0$. Then the chain $\Phi$ satisfies the hypothesis of Proposition 2.1 with

$$M = 1 \text{ and } \theta = (2/d) [(R/r + 1)^{d-1}(R/r)]^{-1}.$$  

**Proof.** We are going to evaluate $S(x, \cdot)$ for any $x \in X$. First assume that $X = B(x, R)$. It is easy to see that the transition function $S(x, \cdot)$ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}_{d}$. Its density is equal to

$$f_{x}(u) = (c_{d}\|u - x\|^{d-1}R)^{-1} \quad \text{for } u \in B(x, R) \setminus \{x\},$$

where $c_{d} = 2\pi^{d/2}/\Gamma(d/2)$ denotes the surface of the $d$-sphere of radius $1$. Indeed, by symmetry argument we see that $f_{x}(u) = \tilde{f}_{x}(\|u - x\|)$ for some function $\tilde{f}_{x} : [0, R] \to \mathbb{R}$. Then we have

$$\int_{0}^{u} \tilde{f}_{x}(v)c_{d}v^{d-1}dv = u/R \quad \text{for any } u \in [0, R].$$

Integrating both sides with respect to $u$ we obtain (4.2).

Now let $X$ be arbitrary. Since $X \subset B(x_{0}, R)$ we see that for any $x \in X$ the transition function $S(x, \cdot)$ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}_{d}$ and its density $f_{x}$ satisfies

$$f_{x}(u) \geq (c_{d}\|u - x\|^{d-1}2R)^{-1} \quad \text{for } u \in X.$$  

Thus

$$S(x, \cdot) \geq \int_{B(x_{0}, r)} f_{x}(u)\mathcal{L}_{d}(du) \geq (c_{d}(R + r)^{d-1}2R^{d-1})^{-1}\mathcal{L}_{d}(B(x, r))\nu(\cdot) \geq (b_{d}/c_{d})[(R/r + 1)^{d-1}(R/r)]^{-1}\nu(\cdot),$$
where \( \nu(\cdot) = \mathcal{L}_d(\cdot \cap B(x_0, r))/\mathcal{L}_d(B(x_0, r)) \) and \( b_d = \pi^{d/2}/\Gamma(d/2 + 1) \) denotes the volume of a unit ball in \( \mathbb{R}^d \). Since \( b_d/c_d = 2/d \), we finally obtain

\[
S(x, \cdot) \geq (2/d)[(R/r + 1)^{d-1}(R/r)]^{-1}\nu(\cdot)
\]

and the proof is complete. □

4.3. Historical remarks and a rough comparison of the convergence rates. In the previous sections we obtain two different convergence rates for the two different algorithms. Let us make here some general comments about the comparative speed of convergence of the two algorithms.

In principle, the two algorithms are not comparable. For example, if \( r/R \) is large in large dimension, then there is a possibility that algorithm 1 be better than algorithm 2. There definitely exists examples of this nature.

However, in all the examples we will consider, \( r/R \) becomes small as \( d \) becomes large, so \( 1/(R/r + 1) \) to a power of \( d \) appearing in algorithm 2 is larger than \( b_d(r/R)^{k+d} \) appearing in algorithm 1, all the more as the additional factor \( b_d \) for the volume of the unit \( l^2 \) ball decays fast to zero as \( d \) large.

The remainder of the paper is devoted to working out some examples in detail in order to understand the use of the two algorithms.

Let us mention that the theoretical speed of convergence (typically an exponential speed like \( 1 - (2/d)[(R/r + 1)^{d-1}(R/r)]^{-1}\) iteration steps for the generally faster algorithm 2) seems to be of slow speed for practical reasons when the dimension is large, of speed comparable to the conditional rejection algorithm described in item (ii) of the introduction. However, although the conditional rejection algorithm can be turned naively into a Markov chain with comparable convergence speed, algorithms (1) and (2) have the following theoretical properties: (A) they converge regardless of the starting measure, (B) they yield multiple random samples (the naive algorithm gives only one sample within a given time unit), (C) they come up with nice theoretical properties, such as a central limit theorem and a law of iterated logarithm.

Finally, let us note that from a numerical point of view, it seems that algorithm (1) and (2) seem to converge much faster than the bounds we are able to provide, at least if one restricts observables to eigenvalues (for example, algorithm one was run on 100 \( \times \) 100 self adjoint matrices with potential proportional to \( \exp(-NTrX^4)dX \), cf [11]) and it provides an eigenvalue distribution very close to that predicted by potential theory.

5. Applications to quantum information theory and statistical physics

In several problems of statistical and quantum physics one works with states defined on an \( N \) dimensional space. The corresponding sets of states form compact convex sets in \( \mathbb{R}^d \), were \( d = d(N) \). The same is true for the set of linear discrete transformations acting on
them. In the classical case one uses *stochastic* and *bistochastic matrices*, which send the set of probability simplex into itself, while in the quantum case one deals with *quantum operations*, formally defined as completely positive, trace preserving maps.

In several concrete applications, related e.g. to the theory of quantum information, one considers often various convex subsets of the above sets, and is interested to analyze properties of their typical elements. To this end it is important to develop an efficient algorithm to generate a sample of random points according to the flat measure in a given convex set $X$. In some cases there exist such algorithms dedicated to a given set: For instance, procedures to generate random quantum states were studied in [2, 7, 21], while other contributions deal with random subnormalized states [3], random bistochastic matrices [4] and random quantum operations [1].

However, the procedures mentioned above are dedicated to a particular problem and cannot be easily adopted to other convex sets. On the other hand, the sampling algorithm developed in this work is universal, as it allows one to generate sequences of random points distributed uniformly in an arbitrary convex body $X \subset \mathbb{R}^d$. We are going to characterize $X$ by the radius $R$ of the minimal outscribed sphere, the radius $r$ of the maximal inscribed sphere and by the barycenter $x_*$.

The rest of the paper is devoted to providing explicit estimates of the convergence rate of the algorithm to generate random points in $X$, which depends on the dimensionality $d$, the ratio $\mu = r/R$ and in algorithm 1 on the accessibility parameter $k$.

5.1. **Balls, cubes and simplices in $\mathbb{R}^d$**. For balls and cubes in $\mathbb{R}^d$ it is not difficult to generate random points according to the uniform measure, so we will not advocate to use the above algorithm for this purpose. However it is illuminating to compare estimations for the parameters determining the convergence rate according to Eq. (3.2).

5.1.1. **The Euclidean ball**. For a unit ball $B^d$ both radii coincide, $R = r$, so their ratio $\mu = r/R$ is equal to unity. Since for any choice of the basis $e$ the ball is $d$-accessible, estimation (3.2) gives $M = 2d$ and $\theta = b_d d^{-2d}$ where $b_d$ denotes the volume of a unit $d$-ball. This implies the convergence rate of algorithm 1 applied to a $d$–ball,

$$\alpha = (1 - \theta)^{1/M} = (1 - b_d d^{-2d})^{1/2d}.$$ 

On the other hand, in algorithm 2 we obtain the rate of convergence

$$\alpha = 1 - \theta = 1 - (2^d d)^{-1}.$$

5.1.2. **The unit cube**. For an unit cube $C^d$ the inscribed radius $r = 1/2$, and outscribed radius $R = \frac{1}{2}\sqrt{d}$ so the ratio reads $\mu = r/R = 1/\sqrt{d}$. If the basis $e$ is determined by the sides of the cube then $C^d$ is $d$–accessible. This implies $M = 2d$ and $\theta = b_d d^{-3d}$ and yields the convergence rate

$$\alpha = (1 - b_d d^{-3d})^{1/2d}.$$
Algorithm 2 gives the rate of convergence
\[ \alpha = 1 - \theta = 1 - (2/d)[(\sqrt{d} + 1)^{d-1} \sqrt{d}]^{-1}. \]

5.1.3. The simplex. For an \( N \)-simplex \( \Delta_N \) embedded in \( \mathbb{R}^d \) with \( d = N - 1 \) we have \( R = \sqrt{(N - 1)/N} \) and \( r = 1/\sqrt{N(N - 1)} \) so that \( \mu = 1/(N - 1) = d^{-1} \). Note that the simplex \( \Delta_N \) describes the set of classical states – \( N \)-point probability distributions.

For a \( d \)-simplex we can find a basis \( e \) such that the set \( \Delta_{d+1} \) is \( d \)-accessible. This is the case if the first vector \( e_1 \) is parallel to a side of the simplex, \( e_2 \) and \( e_1 \) span the plane parallel to a face of \( \Delta_{d+1} \), while adding an additional vector \( e_n \) spans a hyperplane containing an \( n \)-face of the simplex. For this choice of the basis one obtains therefore \( M = 2d \) and \( \theta = b_d d^{-4d} \), which implies
\[ \alpha = (1 - b_d d^{-4d})^{1/2d} \]
in algorithm 1. In algorithm 2, we have
\[ \alpha = 1 - \theta = 1 - (2/d)[(d + 1)^{d-1} d]^{-1}. \]

Thus algorithm 1 converges faster for a \( d \)-ball than for \( d \)-cube and \( d \)-simplex, as in the former case the ratio \( \mu = r/R \) is the largest.

5.2. Quantum states.

5.2.1. Quantum states. The set \( \Omega_N \) of density matrices (Hermitian and positive operators, \( \rho^* = \rho \geq 0 \), normalized by the trace condition \( \text{Tr} \rho = 1 \)) of size \( N \) has the dimension \( d = N^2 - 1 \). The radius of the out-sphere, equal to the Hilbert–Schmidt distance between a pure state \( \text{diag}(1, 0, \ldots, 0) \) and the maximally mixed state \( \rho_* = I/N \), reads \( R = \sqrt{(N - 1)/N} \). The radius of the inscribed sphere given by the distance between \( \rho_* \) and the center of a face, \( \text{diag}(0, 1, \ldots, 1)/(N - 1) \) is equal to \( r = 1/\sqrt{N(N - 1)} \) hence \( \mu = 1/(N - 1) = 1/(\sqrt{d + 1} - 1) \sim d^{-1/2} \).

Any quantum state \( \rho \in \Omega_N \) can be expressed in terms of the generalized Bloch vector \( \tau \),
\[ \rho = \frac{1}{N} I + \sum_{i=1}^{d} \tau_i \lambda_i. \]

Here \( \{\lambda_i\} \) is a set of \( d = N^2 - 1 \) traceless generators of the group \( SU(N) \), which form an orthonormal basis in the Hilbert–Schmidt space of operators of order \( N \). For \( N = 2 \) one usually takes three Pauli matrices \( \sigma_i \) while for \( N = 3 \) it is convenient to use eight Gell-Mann matrices \[13]. Since the state \( \rho \) is hermitian, the coordinates of the corresponding Bloch vector, \( \tau_i = \text{Tr} \lambda_i \rho \), are real. Thus the Bloch vector \( \tau = (\tau_1, \ldots, \tau_d) \) belongs to \( \mathbb{R}^d \) and the conditions for \( \tau \) to guarantee positivity of \( \rho \) are known \[10]. Setting some coefficients of \( \tau \) to zero corresponds to a projection onto a subspace and does not spoil positivity of \( \rho \).

Thus the \( d \)-dimensional convex set \( \Omega_N \) of quantum states is \( d \)-accessible with respect to the Bloch basis \( (\lambda_1, \ldots, \lambda_N) \). For this choice of the basis one obtains therefore \( M = 2d = \)
behave as analogously as a d–cube $C^d$ of dimension $d = N^2 - 1$.

In algorithm 2 we obtain in this case
$$\alpha = 1 - \theta \sim 1 - (2/d)[(\sqrt{d} + 1)^{d-1}\sqrt{d}]^{-1}.$$  

5.2.2. **PPT states.** Let $N = K^2$, so the Hilbert space has a tensor product structure, $\mathcal{H}_N = \mathcal{H}_A \otimes \mathcal{H}_B$. Then one defines a partial transposition, $T_2 = \mathbb{I} \otimes T$ and the set of PPT states (positive partial transpose) which satisfy $\rho^{T_2} \succeq 0$. This set forms a convex subset of $\Omega_N$ and it can be obtained as a joint part of $\Omega_N$ and its reflection $T_2(\Omega_N)$ [20]. The set of PPT states can be decomposed into cones of the same height $r = 1/\sqrt{N(N-1)}$ [19] hence $\mu = 1/(N-1) \sim d^{-1/2}$.

We were unable to show that the set of PPT states is $k$–accessible for some constant $k$. Therefore we may apply only algorithm 2 which gives the same rate of convergence as in the case of quantum states:
$$\alpha = 1 - \theta \sim 1 - (2/d)[(\sqrt{d} + 1)^{d-1}\sqrt{d}]^{-1}.$$  

5.3. **Stochastic matrices.** Stochastic matrices of order $N$ form a convex body of dimensionality $d = N(N-1)$ and play a role of classical maps, which send the simplex of $N$–point probability vectors into itself. Each column of a stochastic matrix $T$ consists of non-negative numbers which sum to unity, so it forms an $N$-simplex. Thus the set of stochastic matrices is equivalent to a Cartesian product of $N$ singlexes $\Delta_N$, so the estimates follow from section 5.1.3, as each column of $T$ can be generated independently.

5.4. **Bistochastic matrices.** The set $\mathcal{B}_N$ of bistochastic matrices of size $N$, called Birkhoff polytope and given by convex hull of all permutation matrices has dimensionality $d = (N-1)^2$. The radius of the the out-sphere of $\mathcal{B}_N$, equal to the Hilbert–Schmidt distance between identity and the uniform matrix $B$, containing all entries equal to $1/N$ reads $R = \sqrt{N - 1}$. The radius of the inscribed sphere given by the distance between $B$ and the matrix $B_0 = [NB_0 - \mathbb{I}]/(N-1)$ is equal to $r = 1/\sqrt{N-1}$, which implies $\mu = 1/(N-1)$.

Consider the set $\mathcal{C}$ of all matrices of the form
$$C_{i\alpha\beta\gamma} = \begin{bmatrix}
c_{11} = 0 & \cdots & 0 & \cdots & c_{1N} = 0 \\
\cdots & c_{i\alpha} = -1 & \cdots & c_{i\beta} = 1 & \cdots \\
0 & \cdots & 0 & \cdots & 0 \\
\cdots & c_{\gamma\alpha} = 1 & \cdots & c_{\gamma\beta} = -1 & \cdots \\
c_{N1} = 0 & \cdots & 0 & \cdots & c_{NN} = 0
\end{bmatrix}$$

for $i, \alpha, \beta, \gamma \in \{1, \ldots, N\}$. The set $\mathcal{C}$ will play the role of $e = \{e_1, \ldots, e_l\}$. Obviously $l = N^2(N-1)^2$. It may be verified that the set $\mathcal{B}_N$ is $(N-1)^3$–accessible with respect to $e$. To see it assume that $A = [a_{i,j}]_{1 \leq i,j \leq N}$ is a bistochastic matrix with $a_{i,\alpha} > 1/N$. Let
\[ \varepsilon = \min \{ a_{i,\alpha} - 1/N, 1/N \} \]. Observe that since \( a_{i,\alpha} > 1/N \) and the matrix is bistochastic, there exists \( \beta \) such that \( a_{i,\beta} < 1/N \). On the other hand, since \( a_{i,\beta} < 1/N \), there exists \( \gamma \) such that \( a_{\gamma,\beta} > 1/N \). Taking \( A - \varepsilon C_{i,\alpha,\beta,\gamma} \) we obtain the bistochastic matrix of the form

\[
\begin{pmatrix}
 a_{11} & \cdots & \cdots & a_{1N} \\
 \cdots & a_{i,\alpha} - \varepsilon = \frac{1}{N} & \cdots & a_{i,\beta} + \varepsilon > 0 & \cdots \\
 \cdots & \cdots & \cdots & \cdots & \cdots \\
 a_{N1} & \cdots & \cdots & a_{NN} \\
\end{pmatrix}.
\]

Repeating this procedure at most \( N - 1 \) times (possibly with different \( \beta \) and \( \gamma \)) we obtain a matrix with \( a_{i,\alpha} = 1/N \). To obtain the matrix with all entries equal to \( 1/N \) we have to apply this procedure to at most \( (N - 1)^2 \) entries and hence follows that \( k = (N - 1)^3 \).

Finally, we have \( M = N(N - 1)^2 \) and \( \theta = b(N-1)^2(N - 1)^{-4N(N-1)^2} \). Hence

\[ \alpha = (1 - \theta)^{1/M} = (1 - b(N-1)^2(N - 1)^{-4N(N-1)^2})^{N(N-1)^{-2}}. \]

In algorithm 2 we have

\[ \alpha = 1 - (2/d)[N^{d-1}(N - 1)]^{-1}. \]

6. Concluding Remarks

In this paper we proposed a universal algorithm to generate random points inside an arbitrary compact set \( X \) in \( \mathbb{R}^d \) according to the uniform measure. Any initial probability measure \( \mu \) transformed by the corresponding Markov operator converges exponentially to the invariant measure \( \mu^* \), uniform in \( X \). Explicit estimations for the convergence rate are derived in terms of the ratio \( \mu = r/R \) between the radii of the sphere inscribed inside \( X \) and the sphere outscribed on it and the number \( k \) determining the \( k \)-accessibility of the body with respect to a given orthogonal basis \( e \) in \( \mathbb{R}^d \).

Thus the algorithm presented here can be used in practice to generate, for instance, a sample of random quantum states. We reviewed in detail the case of a a composed quantum system, and of a sampling according to the uniform measure on the set of states with positive partial transpose. But our algorithm applies potentially to many other interesting collection of objects in quantum information theory, such as, for example, stochastic maps and bistochastic maps, trace non-increasing maps, etc.

Sampling random states satisfying a given condition and analyzing their statistical properties is relevant in the research on quantum entanglement and correlations in multi-partite quantum systems. A standard approach of generating random points from the entire set of quantum states with respect to the flat measure \([21]\) and checking a posteriori, whether the partial transpose of the state constructed is positive, becomes inefficient for large dimensions, as the relative volume of the set of PPT states becomes exponentially small \([22]\).
Note that the notion of $k$–accessibility plays a crucial role in obtaining our estimations. Running the algorithm for the triangle $ABC$ with the basis $e$ (see Fig. 2b), with respect to which it is not finitely accessible, one would cover an open subset of the triangle (with two corners excluded). Although this set has the full measure of the triangle it is an open set, so the convergence will not be exponential.

In general, for any $k$–accessible set, the lower parameter $k$ characterizing the accessibility is, the faster convergence of the Markov chain to the unique invariant measure $\mu_*$ one obtains.

Acknowledgements

BC, TK and RK were supported by NSERC discovery grants and Ontario’s ERA grants. The research of TS was supported by Polish Ministry of Science and Higher Education Grant no. IdP2011 000361 and EC grant RAQUEL. He was also supported by the National Science Centre of Poland, grant number DEC-2012/07/B/ST1/03320, while KZ acknowledges a support by the NSC grant DEC-2011/02/A/ST1/00119. The authors wish to thank David McDonald for many discussions. Part of this work emanates from TK’s PhD thesis. Substantial progress was achieved during several peer visits of the authors to their co–authors’ home institutions (and of TS and BC to Cambridge’s INI). The hospitality of the aforementioned institutions is warmly acknowledged. The authors are grateful to Konrad Szymanski for implementing the algorithm numerically.

References

[1] W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski, Random Quantum Operations, Phys. Lett. A 373, 320-324 (2009).
[2] S.L. Braunstein Phys. Lett. A 219, 169 (1996).
[3] V. Cappellini H.-J. Sommers and K. Życzkowski, "Subnormalized states and trace-non-increasing maps, J. Math. Phys. 48, 052110 (2007).
[4] V. Cappellini H.-J. Sommers, W. Bruzda and K. Życzkowski, Random bistochastic matrices, J. Phys. A 42 365209 (2009).
[5] W. Doeblin, Sur les propriétés asymptotiques de mouvement régis par certains types de chaînes simples, Bull. Math. Soc. Roum. Sci. 39, 57–115 (1937).
[6] L. Gurvits, Classical complexity and quantum entanglement, Journal of Computer and System Sciences 69, 448-484 (2004).
[7] M.J.W. Hall, Random quantum correlations and density operator distributions, Phys. Lett. A 242, 123-129 (1998).
[8] Harris, T. E. (1956) The existence of stationary measures for certain Markov processes. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 19541955, vol. II, pp. 113124. University of California Press, Berkeley and Los Angeles.
[9] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications Biometrika, 57, 97-109 (1970).
[10] G. Kimura, The Bloch vector for $n$–level systems, Phys. Lett. A 314, 339 (2003).
[11] T. Kousha, PhD thesis, university of Ottawa, 2012.
[12] S.P. Meyn and R.L. Tweedie, Markov chains and stochastic stability, Springer-Verlag, London (1993).
[13] Neal, Radford M. (2003) Slice sampling. Ann. Statist. 31 , no. 3, 705-767.
[14] Nechita, I. (2007) Asymptotics of random density matrices. Ann. Henri Poincare 8, 8, 15211538.
[15] Roberts, G. O.; Rosenthal, Jeffrey S. (1997) Geometric ergodicity and hybrid Markov chains. Electron. Comm. Probab. 2, no. 2, 1325 (electronic).
[16] Roberts, G. O.; Rosenthal, Jeffrey S. Harris recurrence of Metropolis- within-Gibbs and trans-dimensional Markov chains. Ann. Appl. Probab. 16, 2123-2139 (2006).
[17] Roberts, G. O.; Smith, A. F. M. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stochastic Process. Appl. 49, 207216 (1994).
[18] L. I. Schiff, Quantum Mechanics, New York, McGraw-Hill 1968.
[19] S. Szarek, I. Bengtsson and K. Życzkowski, On the structure of the body of states with positive partial transpose, J. Phys. A 39 L119-L126 (2006).
[20] K. Życzkowski and I. Bengtsson, On duality between quantum states and quantum maps, Open Syst. Inf. Dyn. 11, 3-42 (2004).
[21] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, J. Phys. A 34, 7111-7125 (2001).
[22] K. Życzkowski, P. Horodecki, A. Sanpera and M. Lewenstein, Volume of the set of separable states, Phys. Rev. A58, 883-892 (1998).

B.C: Département de Mathématique et Statistique, Université d’Ottawa, 585 King Edward, Ottawa, ON, K1N6N5 Canada, WPI AIMR, Tohoku, Sendai, 980-8577 Japan and CNRS, Institut Camille Jordan Université Lyon 1, France

E-mail address: bcollins@uottawa.ca

T.K: Département de Mathématique et Statistique, Université d’Ottawa, 585 King Edward, Ottawa, ON, K1N6N5 Canada

E-mail address: tkousha@uottawa.ca

R.K.: Département de Mathématique et Statistique, Université d’Ottawa, 585 King Edward, Ottawa, ON, K1N6N5 Canada

E-mail address: rkulik@uottawa.ca

T.S.: Department of Mathematics, University of Gdańsk, ul. Wita Stwosza 57, 80-952 Gdańsk, Poland

E-mail address: szarek@intertele.pl

K.Ż: Institute of Physics, Jagiellonian University, Cracow, and Center for Theoretical Physics, Polish Academy of Sciences, Warsaw

E-mail address: karol@tatry.if.uj.edu.pl