On Integration Methods Based on Scrambled Nets of Arbitrary Size

Mathieu Gerber*

University of Lausanne and CREST

We consider the problem of evaluating $I(\varphi) = \int_{[0,1]^s} \varphi(x)dx$ for a function $\varphi \in L_2[0,1]^s$. In situations where $I(\varphi)$ can be approximated by an estimate of the form $N^{-1} \sum_{n=0}^{N-1} \varphi(x^n)$, with $\{x^n\}_{n=0}^{N-1}$ a point set in $[0,1]^s$, Owen (1995, 1997a, b, 1998) shows that the $O_P(N^{-1/2})$ Monte Carlo convergence rate can be improved by taking for $\{x^n\}_{n=0}^{N-1}$ the first $N = \lambda b^m$ points, $1 \leq \lambda < b$, of a scrambled $(t,s)$-sequence in base $b \geq 2$. For more complex integration problems, Gerber and Chopin (2014) have recently developed a sequential quasi-Monte Carlo (SQMC) algorithm which has an error of size $O_P(N^{-1/2})$ for bounded and continuous functions $\varphi$, when it uses the first $N = \lambda b^m$ points of scrambled $(t,s)$-sequences as inputs. In this paper we extend these results by relaxing the constraint on $N$. Our main contribution is to provide a bound for the variance of scrambled net quadratures which is of order $O(N^{-1})$, without any restriction on $N$. This bound allows us to provide simple conditions to get an integration error of size $O_P(N^{-1/2})$ for functions that depend on the quadrature size $N$ and, as a corollary, to establish that SQMC reaches the $O_P(N^{-1/2})$ convergence rate for any patterns of $N$. Finally, we show in a numerical study that for scrambled net quadrature rules we can relax the constraint on $N$ without any loss of efficiency when the integrand $\varphi$ is a discontinuous function while, for the univariate version of SQMC, taking $N = \lambda b^m$ may only provide moderate gains.

Key-words: Randomized quasi-Monte Carlo; Scrambling; Sequential quasi-Monte Carlo

*mathieu.gerber@ensae.fr
1. Introduction

We consider the problem of evaluating

\[ I(\varphi) = \int_{[0,1]^s} \varphi(x)dx \]

for a function \( \varphi \in L_2[0,1)^s \). Focussing first on unweighed quadrature rules,

\[ I(P_N, \varphi) = \frac{1}{N} \sum_{n=0}^{N-1} \varphi(x^n), \]

where \( P_N = \{x^n\}_{n=0}^{N-1} \) is a set of \( N \) points in \([0,1)^s\), the simplest way to approximate \( I(\varphi) \) is to use the Monte Carlo estimator which selects for \( P_N \) a set of \( N \) independent uniform random variates on \([0,1)^s\). The central limit theorem then ensures that the variance of the approximation error \( I(P_N, \varphi) - I(\varphi) \) is of order \( O(N^{-1}) \). However, it is now well known that this rate can be improved by taking for \( P_N \) a randomized quasi-Monte Carlo (RQMC) point set. In particular, Owen (1995) proposes a randomization scheme for \((t,s)\)-sequences in base \( b \geq 2 \), known as nested scrambling, such that the variance of the quadrature rule \( I(P_N, \varphi) \) decreases faster than \( N^{-1} \) when \( P_N \) is the set made of the first \( N \) points of the resulting randomized sequence Owen (1997a, 1998). Owen (1997a) also establishes that, in this case, \( \text{Var}(I(P_N, \varphi)) \leq c_0 N^{-1} \sigma^2 \) for a constant \( c_0 > 0 \) independent of \( \varphi \) and where \( N^{-1} \sigma^2 = N^{-1} \int_{[0,1)^s} (\varphi(x) - I(\varphi))^2dx \) is the variance of a Monte Carlo quadrature rule of the same size. Interestingly, Owen (1997a) shows that the constant \( c_0 \) has the additional property to be independent of the dimension \( s \).

In some complicated settings, the function \( \varphi \) can not be computed explicitly and/or the dimension \( s \) is too large for a simple unweighted quadrature rule \( I(P_N, \varphi) \) to be efficient. Important examples where such a problem arise are parameter and state inference in state-space models. Recently, Gerber and Chopin (2014) have developed a sequential quasi-Monte Carlo (SQMC) algorithm to carry out sequential inference in this class of models. When this algorithm uses points taken from scrambled \((t,s)\)-sequences as inputs, it outperforms Monte Carlo methods with an error of size \( O(N^{-1/2}) \) for continuous and bounded functions \( \varphi \) (Gerber and Chopin 2014, Theorem 7).

One limitation of all these results is that they only apply for \( N = \lambda b^m, 1 \leq \lambda < b \), this restriction arising because \((t,s)\)-sequences in base \( b \) are characterized by their equidistribution properties on sets of \( b^m \) consecutive points, \( m \geq t \) (see section 2 for a review on \((t,s)\)-sequences). From a practical point of view, this means that a variance reduction can only be obtained at the price of a sharply increasing running time, which may reduce the attractiveness of scrambled net integration methods when one is interested e.g. to reach a given level of precision at the lowest computational effort.

The objective of this paper is to study quadrature rules and SQMC based on scrambled nets of arbitrary size. Our main theoretical contribution is to provide a bound for the variance of the scrambled net quadrature rule \( I(P_N, \varphi) \) which shows that the \( c(N^{-1}) \) convergence rate obtained by Owen (1997a, 1998) under the restriction \( N = \lambda b^m \) in
fact holds for any $N$. This bound also provides conditions to have an error of size $O_P(N^{-1/2})$ for the integral of functions $\varphi_N$ which depend on the quadrature size $N$, as it typically happens in sequential estimation methods. The main consequence of this result is to establish the asymptotic superiority of SQMC over sequential Monte Carlo algorithms without any restriction of $N$. In addition to this bound, we show two interesting properties of scrambled net quadratures of arbitrary size. First, when points of a scrambled $(0,s)$-sequence are used, the variance of the quadrature rule admits a bound of the form $c_0^*\sigma^2N^{-1}$ for a constant $c_0^* > 0$ which is independent of the integrand $\varphi$ and of the dimension $s$. Second, Yue and Mao (1999) establish that for smooth integrands the integration error of quadratures based on scrambled $(\lambda,t,m,s)$-sequences is of order $O_P(N^{-1}(\log N)^{s-1}$. We note in this work that their computations in fact imply an error of size $O_P(N^{-1})$. In a recent paper, Owen (2014) shows that this rate is the best we can achieve uniformly in $N$ for equally weighted quadrature rules and therefore, on this class of functions, quadratures based on scrambled sequences have the optimal worst case behaviour.

The rest of this paper is organized as follows. Section 2 gives the notations and the background material used in this work. The announced results for a quadrature rule $I(P^N, \varphi)$ based on scrambled nets are formally stated in section 3. In section 4 we provide conditions to get the $O_P(N^{-1/2})$ convergence rate for integrands that depend on $N$ and discuss their application in the context of SQMC. In section 5 the question of the impact of $N$ on the convergence rate for both scrambled nets quadrature rules and for SQMC is analysed in a numerical study while section 6 concludes.

2. Background

In this section we provide the background material on $(t,s)$-sequences, scrambled sequences and on the Haar-like decomposition of $L^2[0,1]^s$ introduced by Owen (1997a). Only the concepts and the results used in this paper are presented. For a complete exposition of these notions we refer the reader, respectively, to Dick and Pillichshammer (2010, chapter 4), Owen (1995) and Owen (1997a, 1998).

For integers $s \geq 1$ and $b \geq 2$, let

$$C^b = \left\{ E = \prod_{j=1}^{s} \left[ a_j b^{-d_j}, (a_j + 1)b^{-d_j} \right] \subseteq [0,1]^s, a_j, d_j \in \mathbb{N}, a_j < b^{d_j}, j = 1, \ldots, s \right\}$$

be the set of all $b$-ary boxes.

Let $t$ and $m$ be two positive integers such that $m \geq t$. Then, the point set $\{x^n\}_{n=0}^{b^m-1}$ is called a $(t,m,s)$-net in base $b$ if every $b$-ary box of volume $b^{t-m}$ contains exactly $b^t$ points, while the point set $\{x^n\}_{n=0}^{b^m-1}$, $1 \leq \lambda < b$, is called a $(\lambda,t,m,s)$-net if every $b$-ary box of volume $b^{t-m}$ contains exactly $\lambda b^t$ points and no $b$-ary box of volume $b^{t-m-1}$ contains more than $b^t$ points. A sequence $(x^n)_{n \geq 0}$ of points in $[0,1]^s$ is called a $(t,s)$-sequence in base $b \geq 2$ if for any integers $a \geq 0$ and $m \geq t$ the point set $\{x^n\}_{n=ab^m}^{(a+1)b^m-1}$ is a
(t, m, s)-net in base b. Finally, note that if \((x^n)_{n\geq 0}\) is a \((t, s)\)-sequence in base b, then, for \(1 \leq \lambda < b\), \(\{x^n\}_{n=0}^{\lambda b^m+(\lambda b)^m-1}\) is a \((\lambda, t, m, s)\)-net for any integers \(\lambda, t, m\geq 0\) and \(m \geq t\).

To introduce the Haar-like decomposition of \(L^2[0,1]^s\) developed by Owen (1997a), let \(u \subseteq S := \{1, \ldots, s\}\), \(\kappa\) a vector of \(|u|\) non negative integers \(k(u,j), j \in \{1, \ldots, |u|\}\), \(|\kappa| = \sum_{j=1}^{|u|} k(u,j)\), and

\[ \mathcal{E}^b_{u,\kappa} = \left\{ \prod_{j=1}^{|u|} \left[ a_j b^{-d_j}, (a_j + 1) b^{-d_j} \right) \in \mathcal{E}^b : d_j = k(u,j) + 1 \text{ if } j \in u \text{ and } d_j = 0 \text{ if } j \notin u \right\}. \]

Then, Owen (1997a) shows that

\[ \varphi(x) = \sum_{u \subseteq S} \sum_{\kappa} \nu_{u,\kappa}(x) \]

where, for any \(u \subseteq S\), \(\kappa\), \(\sum_{\kappa} = \sum_{k(u,1)=0}^{\infty} \cdots \sum_{k(u,|u|)=0}^{\infty}\) and \(\nu_{u,\kappa}\) is a step function, constant over each of the \(b^{|u|+|\kappa|}\) sets \(E \in \mathcal{E}^b_{u,\kappa}\) and which integrates to zero over any \(b\)-ary box that strictly contains a set \(E \in \mathcal{E}^b_{u,\kappa}\). These step functions are mutually orthogonal and \(\nu_{\emptyset,()}\) is constant over \([0,1]^s\). The resulting ANOVA decomposition of \(\varphi\) is given by

\[ \sigma^2 = \sum_{|u|>0} \sum_{\kappa} \sigma^2_{u,\kappa} \]  

(1)

with \(\sigma^2_{u,\kappa} = \int \nu_{u,\kappa}^2(x)dx\).

Let \(P^N = \{x^n\}_{n=0}^{N-1}\), \(x^n = (x^n_1, \ldots, x^n_s)\), be the first \(N\) points of a \((t, s)\)-sequence in base \(b\) where, for \(j = 1, \ldots, s\), \(x^n_j = \sum_{i=1}^{N} a_{j,ni} b^{-i}\) with \(a_{j,ni} \in \{0, \ldots, b-1\}\) for all \(n\) and \(i\). Owen (1995) proposes a method to randomly permute the digits \(a_{j,ni}\) such that the scrambled point set \(\tilde{P}^N = \{\tilde{x}^n\}_{n=0}^{N-1}\) preserves the equidistribution properties of the original net \(P^N\). In addition, under this randomization scheme, each \(\tilde{x}^n\) is marginally uniformly distributed on \([0,1]^s\) and Owen (1997a) shows that

\[ \text{Var}(I(\tilde{P}^N, \varphi)) = 1 \cdot \frac{1}{N} \sum_{|u|>0} \sum_{\kappa} \Gamma_{u,\kappa} \sigma^2_{u,\kappa} \]  

(2)

where \(\Gamma_{u,\kappa}\) depends on the properties of the non scrambled point set \(\{x^n\}_{n=0}^{N-1}\). In particular, for an arbitrary value of \(N \in \mathbb{N}^s\), the gain factors \(\Gamma_{u,\kappa}\) are bounded by Hickernell and Yue (2001) Lemma 11

\[ \Gamma_{u,\kappa} \leq \gamma^{t+1} \left( \frac{b+1}{b-1} \right)^{s+1}. \]

(3)

When the point set \(P^N\) is a \((\lambda, t, m, s)\)-net, the gain factors can be more precisely controlled. Notably, Owen (1998) Lemma 2) obtains

\[ \text{Var}(I(\tilde{P}^N, \varphi)) = 1 \cdot \frac{1}{N} \sum_{|u|>0} \sum_{|\kappa|>m-t-|u|} \Gamma_{u,\kappa} \sigma^2_{u,\kappa} \]  

(4)
where \( \Gamma_{u,\kappa} \leq \Gamma_t \) with \( \Gamma_0 = e \) if \( b \geq s \) (Owen 1997b, Theorem 1; Hickernell and Yue 2001, Lemma 6) and, for \( t > 0 \), \( \Gamma_t = b'(b+1)^2/(b-1)^s \) (Owen 1998, Lemma 4). Together with equation (4), these bounds \( \Gamma_t \) for the gain factors imply that
\[
\text{Var} \left( I(\tilde{P}^N, \varphi) \right) = o(N^{-1}), \quad \text{Var} \left( I(\hat{P}^N, \varphi) \right) \leq \Gamma_t \frac{\sigma^2}{N}
\]
where we recall that \( \tilde{P}^N \) contains the first \( N = \lambda b^m \) points of a scrambled \((t, s)\)-sequence in base \( b \).

We conclude this section by noting that all the results presented in this work also hold for the computationally cheaper scrambling method proposed by Matoušek (1998), although in what follows we will only refer to the scrambling technique developed by Owen (1995) for ease of presentation. In addition, even if it is not always explicitly mentioned, all the scrambled nets we consider in this work are made of the first \( N \) points of a scrambled \((t, s)\)-sequence.

### 3. Quadratures based on scrambled nets of arbitrary size

#### 3.1. Error bounds

A first result concerning the error bound of quadratures based on scrambled nets of an arbitrary size \( N \) can be directly deduced from (2) and (3). Indeed, if \( \tilde{P}^N \) contains the first \( N \in \mathbb{N}^* \) points of a scrambled \((t, s)\)-sequence in base \( b \geq 2 \), these two bounds imply that
\[
\text{Var} \left( I(\tilde{P}^N, \varphi) \right) \leq \frac{\sigma^2 b^{t+1} \left( \frac{b+1}{b-1} \right)^s}{N}
\]
so that the variance of scrambled nets quadrature is never larger than a constant time the Monte Carlo variance.

The following Theorem is the main result of this work and provides a sharper bound for the integration error (see Appendix A.1 for a proof).

**Theorem 1.** Let \( \varphi \in L_2[0, 1]^s \), \( \sigma^2 = \int_{[0,1]^s} \varphi^2(x)dx - \left( \int_{[0,1]^s} \varphi(x)dx \right)^2 \) and \( \hat{P}^N = \{\tilde{\mathbf{x}}^{n}\}_{n=0}^{N-1} \) be the first \( N \) points of a \((t, s)\)-sequence in base \( b \geq 2 \) scrambled as in Owen (1995). Let \( N \in \mathbb{N}^* \) be such that \( b^k \leq N < b^{k+1} \) for an integer \( k \geq t + s + 1 \). Then,
\[
\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{\mathbf{x}}^n) \right) \leq b^t \left( \frac{b+1}{b-1} \right)^s \frac{2}{N} \left( 1 + c_b B^{(k)}_t \right) + c_b \left( b^{t+1} \left( \frac{b+1}{b-1} \right)^{s+1} \right)
\]
where \( c_b = \frac{(b-1)^{1/2}}{b-1} \) and, for \( c \in \{0, \ldots, k - s - 1\} \),
\[
B^{(k)}_c = \sum_{|u|>0} \sum_{|\kappa|>k-c-|u|} \sigma^2_{u,\kappa} + \sum_{|u|>0} \sum_{|\kappa|\leq k-c-|u|} \frac{1}{b^{k-c-|u|}} \sum_{|\kappa|\leq k-c-|u|} \sigma^2_{u,\kappa} b^{c|\kappa|}.
\]
Thanks to Kronecker’s Lemma (see e.g. Shiryaev[1996] Lemma 2, p.390), this Theorem implies that for any square integrable function the error is of size $o_P(N^{-1/2})$ without any restriction on $N$. When $t = 0$, we note from the proof of this result that the variance of quadratures based on points taken from a scrambled $(0,s)$-sequences is never larger than a universal constant $c^*$ time the Monte Carlo variance. These two results are collected in the following Corollary.

**Corollary 1.** Consider the set-up of Theorem 1. Then, for any $\varphi \in L^2[0,1]^s$, we have

$$\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{x}^n) \right) = o(N^{-1}).$$

In addition, for $t = 0$,

$$\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{x}^n) \right) \leq \frac{\sigma^2}{N} (3 + 2\sqrt{2}) \approx 15.84 \frac{\sigma^2}{N}.$$

**Proof.** To prove the error rate, let $\tilde{\sigma}_{u,l}^2 = \sum_{\kappa : |\kappa| = l} \sigma_{u,\kappa}^2$ for $l \in \mathbb{N}$ and note that, for any fixed integers $a > 0$, $0 \leq c < k - s$ and $u \subseteq S$,

$$\frac{1}{b^a(k-c-|u|)} \sum_{|\kappa| \leq k-c-|u|} \sigma_{u,\kappa}^2 b^{|\kappa|} = \sum_{|u| > 0} \frac{1}{b^a(k-c-|u|)} \sum_{l=0}^{k-c-|u|} b^{al} \tilde{\sigma}_{u,l}^2 = \sum_{|u| > 0} \frac{1}{b^a(k-c-|u|+1)} \sum_{l=1}^{k-c-|u|+1} b^{al} \tilde{\sigma}_{u,l-1}^2$$

which converges to zero by Kronecker’s Lemma (see e.g. Shiryaev[1996] Lemma 2, p.390).

Also, because $\sum_{|u| > 0} \sum_{\kappa} \sigma_{u,\kappa}^2 = \sigma^2$, this shows that, as $N \to +\infty$, $B_t^{(k)} \to 0$ and $B_t^{(k)}_{t+1} \to 0$ and therefore, using Theorem 1,

$$N \text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{x}^n) \right) \to 0$$

as $N \to +\infty$. The proof of the bound for $t = 0$ is postponed to Appendix A.2.

At this point it worth mentioning that the $o_P(N^{-1/2})$ convergence rate for quadratures based on scrambled nets of arbitrary size was simultaneously established by Art Owen (personal communication) using a more direct proof. Nevertheless, the bound given in Theorem 1 also allows to study situations where the integrand depends on the size of the quadrature rule $N$, as explained in section 4.
\[ 3.2. \text{Error rate for smooth integrands} \]

For a hyperrectangle \( J = [a, b] \subset [0, 1]^s \) and a set \( u \subseteq S \), let \( J^u = [a^u, b^u] \) where we write \( x^u \) the projection of \( x \in [0, 1]^s \) onto \([0, 1]^u\), the \(|u|\)-dimensional unit hypercube with coordinates in \( u \), and

\[
\Delta_u(\varphi, J) = \sum_{v \subseteq u} (-1)^{|v|} \varphi_v(a)
\]

where \( \varphi_v(a) = \varphi(x) \) with \( x_i = a_i \) if \( i \in v \) and \( x_i = b_i \) otherwise.

Yue and Mao (1999) analyse the integration error for the class of real valued functions \( \varphi \) that satisfy the following general Lipschitz condition:

\[
|\Delta_u(\varphi, J)| \leq C \lambda_u(J^u), \quad \forall u \subseteq S, \quad \forall J = [a, b] \subset [0, 1]^s
\]

for a constant \( C > 0 \) and where \( \lambda_u \) is the Lebesgue measure on \([0, 1]^u\). For these smooth integrands, Yue and Mao (1999) establish that

\[
\forall \alpha > 0 \quad \max_{1 \leq \alpha \leq 2} \sum_{|u| > \alpha} \sum_{|v| > m} \sigma_{u,v}^2 = O(b^{-2m}s^{-1})
\]

and therefore, according to Owen (1998, Theorem 2), we have, for \( m \geq t + s - 1 \),

\[
\forall \alpha > 0 \quad \max_{1 \leq \alpha \leq 2} \sum_{|u| > \alpha} \sum_{|v| > m-t-|u|} \sigma_{u,v}^2 = O(b^{-2m}s^{-1})
\]

showing that the error is of size \( O_P(N^{-1.5}(\log N)^{-\frac{s-1}{2}}) \) for quadratures based on scrambled \((\lambda, t, m, s)\)-nets. From this results, Yue and Mao (1999, Theorem 4) deduce that quadratures based on nets of arbitrary size have an error rate of order \( O_P(N^{-1}(\log N)^{-\frac{s}{2}}) \) but as we will see in the next few lines the right rate is in fact \( O_P(N^{-1}) \).

As previously, \( \hat{P}_N = \{ \hat{x}_n \}_{n=0}^{N-1} \) denotes the set containing the first \( b^{k+1} > N \geq b^k \) points of a scrambled \((t, s)\)-sequence in base \( b \). The standard way to analyse the variance of a scrambled net quadrature rule of arbitrary size is to decompose \( \hat{P}_N \) into scrambled \((a_m, t, m, s)\)-nets \( \hat{P}_m \), \( m = t, \ldots, (k - t + 1) \), and a remaining set \( \hat{P} \) that contains \( \hat{n} < b^k \) points (see the proof of Theorem 1 for more details). Let \( \hat{P}' = \hat{P} \cup_{m=t}^{t+s-2} \hat{P}_m \). Then, using trivial inequalities, we have

\[
\text{Var} \left( I(\hat{P}_N, \varphi) \right) \leq \frac{1}{N^2} \left( \text{Var} \left( \sum_{n \in \hat{P}_N} \varphi(\hat{x}_n) \right)^{1/2} + \sum_{m=t+s-1}^{k} \text{Var} \left( \sum_{n \in \hat{P}_m} \varphi(\hat{x}_n) \right)^{1/2} \right)^2.
\]

From (6), \( \text{Var} \left( \sum_{n \in \hat{P}_N} \varphi(\hat{x}_n) \right)^{1/2} < c_1 \) for a constant \( c_1 \) while, using (8),

\[
\sum_{m=t+s-1}^{k} \text{Var} \left( \sum_{n \in \hat{P}_m} \varphi(\hat{x}_n) \right)^{1/2} = O \left( \sum_{m=t+s-1}^{k} b^{-m/2}m^{s-1} \right) = O(1)
\]

7
since $\sum_{m=t+1}^{\infty} b^{-m/2} m^{-\frac{1}{2}} < +\infty$. We therefore conclude that $N^2 \text{Var} \left( I(\hat{P}^N, \varphi) \right) < \bar{c}$ for $N$ large enough and for a constant $\bar{c}$. Recently, Owen (2014, Theorem 2) has shown that the best possible error rate we can have uniformly on $N$ is of order $O_P(N^{-1})$ and therefore there exists a constant $\underline{c} < \bar{c}$ such that, for $N$ large enough,

$$\underline{c} < N^2 \text{Var} \left( I(\hat{P}^N, \varphi) \right) < \bar{c}.$$ 

Hence, on the class of function satisfying the general Lipschitz condition (7), scrambled net quadrature rules have the optimal worst case behaviour.

4. Error rate for integrands that depend on the quadrature size

We now analyse the behaviour of the quadrature $I(\hat{P}^N, \varphi_N)$ where $(\varphi_N)_{N \geq 1}$ is a sequence of real valued functions. Using Theorem 1, we can deduce the following result concerning the error size of $I(\hat{P}^N, \varphi_N)$.

**Corollary 2.** Consider the set-up of Theorem 1. Let $(\varphi_N)_{N \geq 1}$ be a sequence of functions such that, for any $u \subseteq \{1, \ldots, s\}$ and for any $\kappa(u)$, we have, as $N \to +\infty$,

$$\sigma^2_N \to \sigma^2 < +\infty, \quad \sigma^2_{N,u} \to \sigma^2_{u},$$

where $\sigma^2_N = \sum_{|u| > 0} \sum_{\kappa} \sigma^2_{N,u,\kappa}$ and $\sigma^2 = \sum_{|u| > 0} \sum_{\kappa} \sigma^2_{u,\kappa}$. Then,

$$\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi_N(\tilde{x}^n) \right) = o(N^{-1}).$$

**Proof.** Let $k$ be the largest power of $b$ such that $b^k \leq N$. Then, by Theorem 1 to prove the result we first need to show that, for $a \in \{1/2, 1\}$ and $c \in \{t, t+1\}$, we have

$$\sum_{|u| > 0} \sum_{|\kappa| \leq k-c-|u|} \sigma^2_{N,u,\kappa} b^{a|\kappa|} = o(1).$$

To establish this result, let $a$ and $c$ be as above, $k' = k - c - |u| + 1$ (for $k$ large enough), $\tilde{k} = \lfloor k'/2 \rfloor$ and $S_{u,p} = \sum_{l=1}^{p+1} \sigma^2_{N,u,1-l}$ where $\sigma^2_{N,u,1}$ is defined as in the proof of Corollary 1. Note that the positive and increasing sequence $(S_{u,p})_{p \geq 1}$ converges to $\sigma^2_{N,u} = \sum_{|\kappa| > 0} \sigma^2_{N,u,\kappa}$ as $p \to +\infty$. Then, using summation by part and similar computations as in the proof of Kronecker’s Lemma (see e.g. Shiryaev 1996, Lemma 2, p.390), we have
the proof is complete.

that, as $N \to \infty$, as required. To conclude the proof note that these computations also imply
so that

Then, using Fatou’s Lemma,

because each $\sigma^2_{N,u,l}$ is a finite sum of some $\sigma^2_{N,u,\kappa}$’s and, by assumption, $\sigma^2_{N,u,\kappa} \to \sigma^2_{u,\kappa}$ for any $u$ and $\kappa$. This shows that the second term of (9) converges to zero as $N \to +\infty$. The above computations also show that, for any $u \subseteq S$, $\sigma^2_{N,u} \to 0$ as $N \to +\infty$. Hence, the right-hand side of (9) goes to zero as $N$ increases, as required. To conclude the proof note that these computations also imply that, as $N \to +\infty$,

and the proof is complete.

In practice the sequence of functions $(\varphi_N)_{N \geq 1}$ is often such that $\varphi_N \to \varphi$ where $I(\varphi)$ is the quantity of interest. The classical situation where this set-up occurs is when we are
estimating $I(\varphi)$ using a sequential method such as the array-RQMC algorithm developed by [L’Ecuyer et al., 2006] or the SQMC algorithm proposed by [Gerber and Chopin, 2014]. For this latter, a direct consequence of Corollary 2 is to relax the constraint on $N$ in Theorem 7 of Gerber and Chopin (2014), showing that on the class of continuous and bounded functions SQMC asymptotically outperforms standard sequential Monte Carlo algorithms without any restriction on how the number of “particles” $N$ grows.

Providing a complete description of SQMC is behind the scope of this work (see section 5.2 for an example of SQMC algorithm). Nevertheless, to get some insight about how Corollary 2 applies to this class of algorithms, we describe now how this results can be used to study a scrambled net version of the sampling importance resampling (SIR) algorithm proposed by [Rubin, 1988], which is iteratively used in SQMC.

To keep the presentation simple we consider a (toy) SIR algorithm designed to estimate the univariate expectation $\pi(f) := \int_{X} f(x) \pi(x) dx$, with $\pi$ a density function on $X \subseteq \mathbb{R}$.

Let $q(x) dx$ be a proposal distribution on $[0, 1)$, $P_N^1 = \{x_n^1\}_{n=0}^{N-1}$ be a (deterministic) QMC point set in $[0, 1)$ and $P_N^2 = \{\tilde{x}_n^2\}_{n=0}^{N-1}$ be a scrambled net. Then, noting $F^{-1}_{\mu}$ the (generalized) inverse of $F_\mu$, the CDF corresponding to the probability measure $\mu$ on $\mathbb{R}$, one QMC version of the SIR algorithm may be follows:

1. Compute $z_n^1 = F^{-1}_q(x_n^1)$ and $w_n = \pi(z_n^1)/q(z_n^1)$ for $n = 1, \ldots, N$;

2. Compute $I(P_N^2, \varphi_N)$ where $\varphi_N = f \circ F^{-1}_N$ with $F^{-1}_N$ the (generalized) inverse of the empirical CDF $F_N(z) = \sum_{n=1}^{N} w_n^m \mathbb{I}(z_n^1 \leq x)$.

Assume that the functions $f(z)$ and $\pi(z)/q(z)$ are continuous and bounded and that $F^{-1}_q$ is continuous. Then, Gerber and Chopin (2014, Theorem 1) shows that $\varphi_N \to \varphi = f \circ F^{-1}_\pi$, with $I(\varphi) = \pi(f)$ the quantity we want to estimate. In addition, it can be shown that, under these assumptions, step 2. of the above SIR algorithm verifies the assumptions of Corollary 2 (see the proof of Gerber and Chopin, 2014, Theorem 7) so that the quadrature rule $I(P_N^2, \varphi_N)$ reaches, without any constraint on $N$, the $\mathcal{O}(N^{-1/2})$ convergence rate. Note that for sake of simplicity we assume here that $P_N^1$ is a deterministic QMC point set but this analysis extends trivially to the case where we also use a scrambled net in step 1. of this QMC SIR algorithm.

5. Numerical Study

In this section we illustrate the main findings of this paper. All the simulations rely on a Sobol’ sequence that is scrambled using the method proposed by Owen (1995). We recall that $b = 2$ for the Sobol’ sequence. All means square errors (MSEs) presented below are estimated from 100 independent repetitions.
5.1. Scrambled net quadratures

As in He and Owen (2014), our objective is to estimate the s-dimensional integral $I(\varphi_i)$, $i = 1, \ldots, 3$, where

$$\varphi_1(x) = \sum_{i=1}^s x_i, \quad \varphi_2(x) = \sum_{i=1}^s \max \left( \sum_{i=1}^s x_i - \frac{s}{2}, 0 \right), \quad \varphi_3(x) = \mathbb{I}(\sum_{i=1}^s x_i > \frac{s}{2})(x).$$

Note that the integrands $\varphi_1$ and $\varphi_2$ are both Lipschitz continuous but $\varphi_2$ is not everywhere differentiable. For $i = 1, \ldots, 3$ we estimate the integral $I(\varphi_i)$ using the quadrature rule $I(\tilde{P}^N, \varphi_i)$ where, as mentioned above, $\tilde{P}^N$ is the set containing the first $N$ points of a scrambled Sobol’ sequence.

Figure 1 shows the evolution of the mean square error $I(\tilde{P}^N, \varphi_i)$, $i = 1, \ldots, 3$, for $N$ ranging from 1 to $2^{16}$ and for $s = 3$. In addition to the Mses we have reported the Monte Carlo $\mathcal{O}(N^{-1})$ reference line to illustrate the result of Corollary 1, namely that the convergence rate is faster than $N^{-1}$ for any pattern of $N$. To compare quadrature rules based on nets of arbitrary size with those based on $(\lambda, t, m, s)$-nets, Figure 1 also shows the evolution of the Mses along the subsequence $N = 2^m$. The interesting point to note here is that the advantage of using $(\lambda, t, m, s)$-nets over nets of arbitrary size decreases as the integrand becomes “less smooth”. Indeed, for the smooth and Lipschitz function $\varphi_1$ it is clear from Figure 1 that there is no gain to take $N \neq 2^m$ since we can observe that the cheapest way to reach any given level of MSE is to take a power of the base of the Sobol’ sequence. For the function $\varphi_2$ the same observation hold only for $N \geq 2^5$ and the gain of using $(\lambda, t, m, s)$-nets is clearly smaller than for the estimation of $I(\varphi_1)$. Finally, the advantage of taking a powers of 2 for the quadrature size has completely disappeared for the discontinuous function $\varphi_3$.

To understand these observations recall that the variance of the quadrature rule $I(\tilde{P}^N, \varphi)$ is bounded above by (see section 3.2)

$$\text{Var} \left( I(\tilde{P}^N, \varphi) \right) \leq \frac{C}{N^2} \left( \mathbb{I}(N = 2^k) + \sum_{j=t+s-1}^k \text{Var} \left( \sum_{n \in \tilde{P}_j} \varphi(\tilde{x}^n) \right) \right)^{1/2}$$

for a constant $C$ and where, in our setting, the $\tilde{P}_j$’s are scrambled $(t, j, s)$-nets in base $b = 2$ and $k$ is the largest integer such that $2^k \leq N$. In addition, note that the cost of choosing an arbitrary value for $N$ is to impose a lower bound of order $N^{-2}$ for the variance of the quadrature rule since it is the best rate we can achieve uniformly on $N$ (Owen, 2014, Theorem 2). Therefore, together with inequality (10), this shows that $\text{Var}(I(\tilde{P}^N, \varphi)) = \mathcal{O}(N^{-2})$ for integrands $\varphi$ which are such that $\text{Var}(I(\tilde{P}^{b^m}, \varphi)) = \mathcal{O}(b^{-\alpha m})$ for a $\alpha \geq 2$. On the contrary, for a non smooth integrand $\varphi$ which verifies $\text{Var}(I(\varphi), \tilde{P}^{b^m})) = \mathcal{O}(b^{-\alpha m})$ for a $\alpha \in (0, 2)$, the upper bound (10) suggests that we will obtain the same convergence rate regardless of the pattern for $N$.

As illustrated in Figure 1, the error size of quadratures based on scrambled $(\lambda, t, m, s)$-nets depends positively on the smoothness of the integrand (for theoretical results, see...
Owen 1997b, 1998; Yue and Mao, 1999; Hickernell and Yue, 2001). Hence, taking $N = \lambda b^k$ is the best choice for the “smooth” integrands $\varphi_1$ and $\varphi_2$ since then the MSE goes to zero much faster than $N^{-2}$. Note that for $\varphi_2$ the MSE obtained by taking $N = \lambda b^k$ decreases slower than for $\varphi_1$ so that, as a result, $N$ should be larger to rule out the choice $N \neq \lambda b^k$. Finally, for the discontinuous function $\varphi_3$ the convergence rate of the MSE when using $(\lambda, t, m, s)$-nets is too slow for the choice of $N$ to influence that of the MSE.

5.2. Likelihood function estimation in state-space models

We now study the problem of estimating the likelihood function of the following generic univariate state-space model

$$
\begin{align*}
(y_k | z_k &\sim \mathcal{N}(\mu_y(z_k), \sigma_y^2(z_k)), \quad k \geq 0 \\
z_k | z_{k-1} &\sim \mathcal{N}(\mu_z(z_{k-1}), \sigma_z^2(z_{k-1})), \quad k \geq 1 \\
z_0 &\sim \mathcal{N}(\mu_0, \sigma_0^2)
\end{align*}
$$

(11)

where $(y_k)_{k \geq 0}$ is the observation process, $(z_k)_{k \geq 0}$ is the hidden Markov process and where $\mu_q : \mathbb{R} \to \mathbb{R}$ and $\sigma_q^2 : \mathbb{R} \to \mathbb{R}^+$, $q \in \{z, y\}$, are known functions.

Given a set of $T$ observations $\{y_k\}_{k=0}^{T-1}$ we denote by $p(y_{0:T-1})$ the likelihood function of the model (11), which can not be computed explicitly. Indeed, writing $f(\cdot, \mu, \sigma^2)$ the density function of the $\mathcal{N}(\mu, \sigma^2)$ distribution, it is easy to see that $p(y_{0:T-1}) = I(\varphi_T)$ where $\varphi_T : [0,1)^T \to \mathbb{R}$ is given by

$$
\varphi_T(x_0, \ldots, x_{T-1}) = \tilde{\varphi}_T \circ F_T^{-1}(x_0, \ldots, x_{T-1})
$$

with $\tilde{\varphi}_T(z_0, \ldots, z_{T-1}) = \prod_{k=0}^{T-1} f(y_k, \mu_y(z_k), \sigma_y^2(z_k))$ and $F_T$ the Rosenblatt transformation (see Rosenblatt 1952 for a definition) of the probability measure

$$
f(z_0, \mu_0, \sigma_0^2)dz_0 \prod_{k=1}^{T-1} f(z_k, m_z(z_{k-1}), \sigma_z^2(z_{k-1}))dz_k.
$$

In practical scenarios, the time horizon $T$ is large (at least several dozen) and the function $\varphi_T$ is concentrated in a tiny region of the integration domain. Consequently, simple unweighted quadrature rules require a huge amount of points to provide a precise estimate of $p(y_{0:T-1})$. An efficient way to get an unbiased estimate $p^N(y_{0:T-1})$ of $p(y_{0:T-1})$ is to use a SQMC algorithm (Gerber and Chopin 2014), that is a QMC version of sequential Monte Carlo methods which are now standard tools to handle this kind of problems (see e.g. Doucet et al. 2001). The suitable SQMC algorithm for the generic state-space model (11) is presented in Algorithm 1 where we use the standard notation $\Phi(\cdot)$ for the CDF of the $\mathcal{N}(0,1)$ distribution.

In this simulations study we analyse the MSE of $\log p^N(y_{0:T-1})$ when at steps 1 and 10 of Algorithm 1 the RQMC point sets are the first $N$ points of independent scrambled Sobol’ sequences, where $N = 4i$ for $i = 3, \ldots, 11$. Note that in Algorithm 1 it is clear that $\sum_{n=1}^{N} w_n = \sum_{n=1}^{N} \varphi_{N,k}(x_n^i)$ where, for $x = (x_1, x_2) \in [0,1]^2$,

$$
\varphi_{N,k}(x) = f(y_k, \mu_y \circ g(x), \sigma_y^2 \circ g(x)), \quad g(x) = \mu_z \circ F_{N,k}^{-1}(x_1) + \sigma_z \circ F_{N,k-1}^{-1}(x_1) \Phi^{-1}(x_2).
$$
Figure 1: Mean square error of $I(\tilde{P}^N, \varphi_1)$ (top-left), $I(\tilde{P}^N, \varphi_2)$ (top-right) and $I(\tilde{P}^N, \varphi_3)$ where $\tilde{P}^N$ contains the first $N$ points of a scrambled Sobol’ sequence. The dashed lines are the Monte Carlo $O(N^{-1})$ reference lines, the dotted lines present the results along the subsequence $N = 2^m$ for $m = 0, \ldots, 16$ and the solid lines the MSEs for any $N \in \{1, \ldots, 2^{16}\}$. The results are obtained from 100 independent repetitions.
Algorithm 1 SQMC Algorithm to estimate $p(y_{0:T-1})$ in the state-space model $(11)$

1: Generate a RQMC point set $\{\tilde{x}^n_0\}_{n=1}^N$ in $[0, 1)$.
2: for $n = 1 \rightarrow N$ do
3: Compute $z^n_0 = \mu_0 + \sigma_0 \Phi^{-1}(\tilde{x}^n_0)$ and $w^n_0 = f(y_0; \mu_y(z^n_0), \sigma^2_y(z^n_0))$.
4: end for
5: for $n = 1 \rightarrow N$ do
6: Normalize the weights: $W^n_0 = w^n_0 / \sum_{m=1}^N w^n_m$.
7: end for
8: Compute $p^N(y_0) = N^{-1} \sum_{n=1}^N w^n_0$.
9: for $k = 1 \rightarrow T - 1$ do
10: Generate a RQMC point set $\{\tilde{x}^n_k\}_{n=1}^N$ in $[0, 1)^2$. let $\tilde{x}^n_k = (\tilde{x}^n_k, \tilde{v}^n_k)$.
11: for $n = 1 \rightarrow N$ do
12: Compute $\tilde{z}^{n}_{k-1} = F_{N,k-1}^{-1}(\tilde{x}^n_k)$ where $F_{N,k-1}(z) = \sum_{m=0}^N W_{k-1}^m \mathbb{I}(z_{k-1}^m \leq z)$.
13: Compute $z^n_k = \mu_z(\tilde{z}^n_{k-1}) + \sigma_z(\tilde{z}^n_{k-1}) \Phi^{-1}(\tilde{v}^n_k)$.
14: Compute $w^n_k = f(y_k; \mu_y(z^n_k), \sigma^2_y(z^n_k))$.
15: end for
16: for $n = 1 \rightarrow N$ do
17: Normalize the weights: $W^n_k = w^n_k / \sum_{m=1}^N w^n_m$.
18: end for
19: Compute $p^N(y_{0:k}) = p^N(y_{0:k-1}) N^{-1} \sum_{n=1}^N w^n_k$.
20: end for
21: return $p^N(y_{0:T-1})$. 

14
Figure 2: Mean square error for the estimation of \( \log p(y_{0:T-1}) \) in the SV model (12) (left plot) and in the toy example (13) (right plot). The dotted line is the Monte Carlo \( O(N^{-1}) \) reference line, the dashed line presents the results for SQMC for \( N = 2^m, m = 3, \ldots, 13 \), while the solid are for SQMC with \( N = 4i, i = 3, \ldots, 211 \). The results are obtained from 100 independent runs of the SQMC algorithm.

Since the function \( F_{N,k-1}^{-1} \) is discontinuous, the results of the previous section therefore suggest that the gain of restricting \( N \) to be powers of the Sobol' sequence is small in the context of SQMC. In addition, it is wroth remarking that this gain will also depend on the regularity of the functions \( \mu_z \) and \( \sigma_z \).

**Stochastic volatility (SV) model.** We first consider the following simple univariate SV model

\[
\begin{aligned}
    y_k | z_k & \sim \mathcal{N}(0, e^{z_k}), & k \geq 0 \\
    z_k | z_{k-1} & \sim \mathcal{N}(0.9z_{k-1}, 0.1), & k \geq 1 \\
    z_0 & \sim \mathcal{N}(0, \frac{0.1}{1-0.9^2})
\end{aligned}
\]  

(12)

from which a set of 100 observations is generated. Figure 2 (left plot) presents the MSE of the estimator \( p^N(y_{0:T-1}) \) and the \( O(N^{-1}) \) Monte Carlo reference line. As expected, we see that the \( o_P(N^{-1}) \) convergence rate for the SQMC algorithm holds uniformly on \( N \). However, we observe in this example that selecting \( N = 2^m \) is optimal as soon as \( N \geq 2^7 \) in the sense that this choice guarantees the smallest MSE for a given computational budget. The fact that very quickly it is desirable to restrict \( N \) to be a power of 2 is not surprising because the SV model (12) is an example of state-space model (11) where the functions \( \mu_z \) and \( \sigma^2_z \) are both very smooth and Lipschitz.
A non-linear and non-stationary model  We now consider the following non-linear and non-stationary well known toy example in the particle filtering literature (see e.g. Gordon et al. 1993)

\[
\begin{align*}
  y_k | z_k & \sim N \left( \frac{z_k^2}{20}, 1 \right), \quad k \geq 0 \\
  z_k | z_{k-1} & \sim N \left( 0.5z_{k-1} + 25 \frac{z_{k-1}}{1+z_{k-1}^2} + 8 \cos(1.2k), 10 \right), \quad k \geq 1 \\
  z_0 & \sim N(0, 2)
\end{align*}
\]

(13)

from which we again simulate a set of 100 observations. Note that, as for the SV model (12), the function \( \mu_z \) is everywhere differentiable but now it is no longer Lipschitz. Consequently, we expect that using \((t, m, s)\)-nets to estimate the likelihood function \( p(y_{0:T-1}) \) is less profitable than for the SV model. This point is confirmed in the Figure 2 (right plot) where we show the evolution of the MSE as a function of \( N \). We indeed observe from this plot that the gain of using a number of particles which is a power of two becomes now apparent only for \( N \) larger than \( 10^{13} \approx 1778 \). In addition, it is only for \( m \geq 12 \) that taking \( N = 2^m \) is the fastest way to achieve any further improvement in term of MSE. Finally, the gain of SQMC compared to Monte Carlo techniques can be assessed from Figure 2 where we have also represented the Monte Carlo \( O(N^{-1}) \) reference line.

To conclude this section it is worth mentioning that to keep the presentation of SQMC simple we have only shown simulations for univariate models. In the multivariate version of SQMC, the resampling step of Algorithm 1 (step 12) requires to sort the particles along a Hilbert space filling curve. Since the Hilbert curve is \((1/d)\)-Hölder continuous, with \( d \) the dimension of the state variable, the estimation problem becomes less smooth as \( d \) increases. In light of the observations of this simulation study, this suggests that the gain of restricting \( N \) to be powers of the base of the underlying \((t, s)\)-sequence is smaller than for univariate models. This point was confirmed in non reported simulations study conducted for the bivariate version of the SV model (12), where the gain of using \((t, m, s)\)-nets as input of SQMC has completely disappeared.

6. Conclusion

Together with the works of Yue and Mao (1999) and Hickernell and Yue (2001), the present analysis concludes to show that the results of Owen (1997a,b, 1998) obtained for quadrature rules based on \((\lambda, t, s, m)\)-nets are in fact true for quadrature rules based on the first \( N \) points of scrambled \((t, s)\)-sequences without any restriction on the pattern of \( N \), namely, to sum-up:

1. For any square integrable functions the integration error goes to zero faster than for the classical Monte Carlo estimator;

2. For any square integrable functions the variance of scrambled quadrature rules is bounded by the Monte Carlo variance multiplied by a constant independent of the integrand;
3. The constant in 2. is uniform with respect to the dimension for scrambled \((0,s)\)-sequences;

4. For smooth integrands an explicit convergence rate (better than \(N^{-1/2}\)) can be computed (see Yue and Mao, 1999; Hickernell and Yue, 2001).

In a simulation study, we show that quadratures based on scrambled \((\lambda, t, m, s)\)-nets outperform those based on nets of arbitrary size when the integrand \(\varphi\) of interest is smooth. More precisely, using scrambled \((\lambda, t, m, s)\)-nets is for such functions the fastest way to reach any given level of \(\text{MSE}\). Nevertheless, as the integrand becomes less smooth, this gain decreases and completely disappears for discontinuous functions.

The second important result proved in this paper is the asymptotic superiority of the sequential quasi-Monte Carlo algorithm proposed by Gerber and Chopin (2014) over standard sequential Monte Carlo methods without any restriction on how the number of particles grows. Since SQMC involves integration of discontinuous functions the behaviour of the \(\text{MSE}\) when the algorithm takes scrambled \((\lambda, t, m, s)\)-nets as inputs should not be too different compared to what we would get when scrambled nets of arbitrary size are used. This point is illustrated in a simulation study based in two univariate state-space models and we argue that for multivariate models it is very unlikely to expect any gain of using as input for SQMC only points of scrambled sequences that form \((\lambda, t, m, s)\)-nets.

**Acknowledgements**

I thank Nicolas Chopin and Florian Pelgrin for useful remarks, and Art Owen for having shared with me his shorter proof for the first part of Corollary he derived when I was writing this manuscript.

**References**

Dick, J. and Pillichshammer, F. (2010). *Digital nets and sequences: discrepancy theory and quasi-Monte Carlo integration*. Cambridge University Press.

Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Gerber, M. and Chopin, N. (2014). Sequential Quasi-Monte Carlo. *arXiv:1402.4039*.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Comm., Radar, Signal Proc.*, 140(2):107–113.

He, Z. and Owen, A. B. (2014). Extensible grids: uniform sampling on a space-filling curve. *arXiv:1406.4549*.
Hickernell, F. J. and Yue, R.-X. (2001). The mean square discrepancy of scrambled $(t, s)$-sequences. *SIAM Journal of Numerical Analysis*, 38:1089–1112.

L’Ecuyer, P., Lécot, C., and Tuffin, B. (2006). A randomized quasi-Monte Carlo simulation method for Markov chains. In *Monte Carlo and quasi-Monte Carlo methods 2004*, pages 331–342. Springer Berlin Heidelberg.

Matoušek, J. (1998). On the $L_2$-discrepancy for anchored boxes. *Journal of Complexity*, 14:527–556.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional conference series in applied mathematics.

Owen, A. B. (1995). Randomly permuted $(t, m, s)$-nets and $(t, s)$-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statistics*, volume 106, pages 299–317. Springer, New York.

Owen, A. B. (1997a). Monte Carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910.

Owen, A. B. (1997b). Scramble net variance for integrals of smooth functions. *Ann. Stat.*, 25(4):1541–1562.

Owen, A. B. (1998). Scrambling Sobol’ and Niederreiter-Xing points. *Journal of complexity*, 14(4):466–489.

Owen, A. B. (2014). A constraint on extensible quadrature rules. *arXiv:1404.5363*.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23(3):470–472.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In Bernardo, M., DeGroot, K., Lindley, D., and Smith, A., editors, *Bayesian Statistics 3*. Oxford University Press.

Shiryaev, A. N. (1996). *Probability*. Springer.

Yue, R.-X. and Mao, S.-S. (1999). On the variance of quadrature over scrambled nets and sequences. *Statist. Prob. Letters*, 44:267–280.

### A. Proofs

A.1. **Proof of Theorem**

We first prove the following Lemma that plays a key role in the proof of Theorem.
Lemma 2. Let \( k \) and \( t \) be two integers such that \( k \geq t \geq 0 \) and \( v_m \in [0, b - 1] \), \( m = 0, \ldots, k \). Then,

\[
\sum_{m=t}^{k} v_m b^m \sum_{|u|>0} \sum_{|\varphi|>m-t-|u|} \sigma^2_{u,\varphi} \leq \left( \sum_{m=t}^{k} v_m b^m \right) \sum_{|u|>0} \sum_{|\varphi|>k-t-|u|} \sigma^2_{u,\varphi} + b^k \left( \sum_{|u|>0} \frac{1}{b^{k-t-|u|}} \sum_{|\varphi|\leq k-t-|u|} \sigma^2_{u,\varphi} b^{k|\varphi|} \right) \quad (14)
\]

where we use the convention that empty sums are null.

Proof. For \( u \subseteq S \) and for \( l \in \mathbb{Z} \), let \( \tilde{\sigma}^2_{u,l} = \sum_{|\varphi|=l} \sigma^2_{u,\varphi} \) if \( l \geq 0 \) and \( \tilde{\sigma}^2_{u,l} = 0 \) otherwise.

To simplify the notations, let \( k_t = k - t \) and \( v'_m = v_{m+t} \). Then,

\[
\sum_{m=t}^{k} v_m b^m \sum_{|u|>0} \sum_{|\varphi|>m-t-|u|} \sigma^2_{u,\varphi} = b^t \sum_{m=0}^{k_t} v'_m b^m \sum_{|u|>0} \sum_{l>m-|u|} \tilde{\sigma}^2_{u,l}. \quad (15)
\]

In order to study the second term of (15), let \( u \subseteq S \) be such that \( k_t \geq |u| \). Then,

\[
\sum_{m=0}^{k_t} v'_m b^m \sum_{l\leq m-|u|} \tilde{\sigma}^2_{u,l} = \sum_{m=|u|}^{k_t} v'_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l} = \sum_{m=|u|}^{k_t-|u|} \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l} + \sum_{m=|u|}^{k_t-|u|} v'_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l}.
\]

Since

\[
\sum_{m=0}^{k_t-|u|} v'_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l} = \sum_{l=0}^{k_t-|u|} \tilde{\sigma}^2_{u,l} - \sum_{m=0}^{l+|u|-1} v_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l},
\]

with \( b^t \sum_{m=l+|u|-1}^{k_t-|u|} v'_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l} = N_l \sum_{m=0}^{l+|u|-1} v_m b^m \), we obtain

\[
b^t \sum_{m=0}^{k_t-|u|} v'_m b^m \sum_{l=0}^{m-|u|} \tilde{\sigma}^2_{u,l} = \sum_{l=0}^{k_t-|u|} \left( N_l - \sum_{m=0}^{l+|u|-1} v_m b^m \right) \tilde{\sigma}^2_{u,l}.
\]

Therefore, using (15) and the convention that empty sums are null,

\[
\sum_{m=t}^{k} v_m b^m \sum_{|u|>0} \sum_{|\varphi|>m-t-|u|} \sigma^2_{u,\varphi} = N_l \sigma^2 - \sum_{|u|>0} \sum_{l=0}^{k_t-|u|-1} \left( N_l - \sum_{m=0}^{l+|u|-1} v_m b^m \right) \tilde{\sigma}^2_{u,l} = N_l \sum_{|u|>0} \sum_{l>k_t-|u|} \tilde{\sigma}^2_{u,l} + \sum_{|u|>0} \sum_{l=0}^{k_t-|u|-1} \sum_{m=0}^{l+|u|-1} v_m b^m.
\]
Finally, since \( v_m \leq b - 1 \), we have, for \( u \subseteq S \) such that \( k_t \geq |u| \),

\[
\sum_{l=0}^{k_t-|u|} \tilde{\sigma}_{u,l}^2 \sum_{m=0}^{l+|u|+t-1} v_m b^m \leq (b-1) \sum_{l=0}^{k_t-|u|} \tilde{\sigma}_{u,l}^2 \frac{b^l+(l+|u|+t-1)(b-t)-1}{b-1} \leq b^k \left( \frac{1}{b^{k-t-|u|}} \sum_{l=0}^{k_t-|u|} \tilde{\sigma}_{u,l}^2 b^l \right).
\]

This shows that

\[
\sum_{m=t}^{k} v_m b^m \sum_{|u|>0} \sum_{|\kappa|>m-t-|u|} \sigma_{u,\kappa}^2 \leq \left( \sum_{m=t}^{k} v_m b^m \right) \sum_{|u|>0} \sum_{|\kappa|>k-t-|u|} \sigma_{u,\kappa}^2 + b^k \left( \sum_{|u|>0} \frac{1}{b^{k-t-|u|}} \sum_{l \leq k-t-|u|} \tilde{\sigma}_{u,l}^2 b^l \right) \tag{16}
\]

and the proof of the Lemma is complete.

\[\square\]

To prove the Theorem, and following the proof of [Niederreiter, 1992, Lemma 4.11, p.56], we decompose \( \{\tilde{x}^n\}_{n=0}^{N-1}, N \geq 1, \) into scrambled \((\lambda_m, t, m, s)\)-nets \( \tilde{P}_m, m = t, \ldots, (k - t + 1) \), and a remaining set \( \tilde{P} \) that contains strictly less than \( b^t \) points. We recall that \( k \) is the largest power of \( b \) such that \( b^k \leq N \).

To construct this partition of \( \{\tilde{x}^n\}_{n=0}^{N-1} \), let \( N = \sum_{m=0}^{k} a_m b^m \) be the expansion of \( N \) in base \( b \), with \( a_m \in \{0, \ldots, b-1\} \) and \( a_k \neq 0 \). Then, let \( \tilde{P}_k = \{\tilde{x}^n\}_{n=0}^{a_k b^k - 1} \) and, for \( 0 \leq m \leq k-1 \), let \( \tilde{P}_m \) be the point set made of the \( \tilde{x}^n \)'s with \( \sum_{h=m+1}^{k} a_h b^h \leq n < \sum_{h=m}^{k} a_h b^h \).

By definition of a \((t, s)\)-sequence, \( \tilde{P}_m \) is a scrambled \((a_m, t, m, s)\)-nets in base \( b \geq 2 \) for \( m = t, \ldots, k \) while \( \tilde{P} = \bigcup_{m=0}^{k-1} \tilde{P}_m \) has cardinality strictly smaller than \( b^t \).

Using this decomposition of \( \{\tilde{x}^n\}_{n=0}^{N-1} \) we have, using the convention that empty sums
are equal to zero,

\[
\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{x}^n) \right) = \text{Var} \left( \frac{1}{N} \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) + \frac{1}{N} \sum_{m=t}^{k} \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \right)
\]

\[
\leq \left\{ \frac{1}{N} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) \right)^{1/2} + \frac{1}{N} \sum_{m=t}^{k} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \right)^{1/2} \right\}^2
\]

\[
= \frac{1}{N^2} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) \right) + \left\{ \frac{1}{N} \sum_{m=t}^{k} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \right)^{1/2} \right\}^2
\]

\[
+ 2 \left\{ \frac{1}{N} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) \right)^{1/2} \right\} \left\{ \frac{1}{N} \sum_{m=t}^{k} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \right)^{1/2} \right\}.
\]

(17)

For the first term (and in the case where \( \tilde{P} \) is not empty) let \( \tilde{n} = |\tilde{P}| < b^t \) and note that, using (11)-3,

\[
\text{Var} \left( \frac{1}{\tilde{n}} \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) \right) = \frac{1}{\tilde{n}} \sum_{|u|>0} \sum_{\kappa} \Gamma_u,\kappa \sigma^2_{u,\kappa} \leq \frac{1}{\tilde{n}} b^{t+1} \left( \frac{b+1}{b-1} \right)^{s+1} \sigma^2
\]

and therefore,

\[
\frac{1}{N^2} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}} \varphi(\tilde{x}) \right) \leq \frac{\tilde{n}}{N} \left( b^{t+1} \left( \frac{b+1}{b-1} \right)^{s+1} \right) \frac{\sigma^2}{N} \leq \frac{b^t}{N} \left( b^{t+1} \left( \frac{b+1}{b-1} \right)^{s+1} \right) \frac{\sigma^2}{N}.
\]

(18)

To bound the second term of (17) define, for \( m \geq t \), \( \tilde{I}_m = (a_m b^m)^{-1} \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \) if \( m \) is such that \( a_m \neq 0 \) and set \( \tilde{I}_m = 0 \) otherwise. Then,

\[
\left\{ \frac{1}{N} \sum_{m=t}^{k} \text{Var} \left( \sum_{\tilde{x} \in \tilde{P}_m} \varphi(\tilde{x}) \right)^{1/2} \right\}^2 = \left\{ \frac{1}{N} \sum_{m=t}^{k} \text{Var}(a_m b^m \tilde{I}_m)^{1/2} \right\}^2
\]

\[
= \frac{1}{N^2} \sum_{m=t}^{k} \text{Var}(a_m b^m \tilde{I}_m) + \frac{2}{N^2} \sum_{k \geq m \geq n \geq t} \text{Var}(a_m b^m \tilde{I}_m)^{1/2} \text{Var}(a_n b^n \tilde{I}_n)^{1/2}.
\]

(19)

Using (14) we have, for \( m \geq t \) such that \( a_m \neq 0 \), \( \text{Var}(\tilde{I}_m) \leq \Gamma_t (a_m b^m)^{-1} \sum_{|u|>0} \sum_{|\kappa|>m-t-|u|} \sigma^2_{u,\kappa} \) and therefore, using Lemma 2

\[
\sum_{m=t}^{k} \text{Var}(a_m b^m \tilde{I}_m) \leq \Gamma_t \sum_{m=t}^{k} a_m b^m \sum_{|u|>0} \sum_{|\kappa|>m-t-|u|} \sigma^2_{u,\kappa} \leq \Gamma_t B^{(k)}_t
\]
where \( B_t^{(k)} \) is as in the statement of the Theorem. Hence, since \( b^k \leq N \),

\[
\frac{1}{N} \sum_{m=1}^{k} \text{Var}(a_m b^m \hat{I}_m) \leq \Gamma_t \left( \sum_{|u|>0} \sum_{|\kappa|>k-u} \sigma_{u,\kappa}^2 + \sum_{|u|>0} \sum_{|\kappa| \leq k-u} \frac{1}{|k-u|} \sum_{\kappa} \sigma_{u,\kappa}^2 b^{(|\kappa|)} \right).
\]

To study the second term of (19), let \( m > n \geq t \). Then, easy computations show that

\[
\frac{a_n a_m b^{n+m}}{\Gamma_t} \text{Var}(\hat{I}_m)^{1/2} \text{Var}(\hat{I}_n)^{1/2} \leq \left( \sum_{|u|>0} \sum_{|\kappa|>m-t-u} \sigma_{u,\kappa}^2 \right)^{1/2} \left( \sum_{|u|>0} \sum_{|\kappa|>n-t-u} \sigma_{u,\kappa}^2 \right)^{1/2} \leq \frac{1}{2} \left( \sum_{|u|>0} \sum_{|\kappa|>m-t-u} \sigma_{u,\kappa}^2 + \sum_{|u|>0} \sum_{|\kappa|>n-t-u} \sigma_{u,\kappa}^2 \right).
\]

Therefore,

\[
\frac{2}{\Gamma_t} \sum_{m=t+1}^{k} \sum_{n=t}^{m-1} \text{Var}(a_m b^m \hat{I}_m)^{1/2} \text{Var}(a_n b^n \hat{I}_n)^{1/2} \leq \sum_{m=t+1}^{k} \sum_{n=t}^{m-1} \left\{ (a_m a_n)^{1/2} b^{n+m/2} \sum_{|u|>0} \sum_{|\kappa|>m-t-u} \sigma_{u,\kappa}^2 + (a_n a_m)^{1/2} b^{m+n/2} \sum_{|u|>0} \sum_{|\kappa|>n-t-u} \sigma_{u,\kappa}^2 \right\} \leq (b-1)^{1/2} \sum_{m=t+1}^{k} (a_m b^{m/2}) \sum_{|u|>0} \sum_{|\kappa|>m-t-u} \sigma_{u,\kappa}^2 \sum_{n=t}^{m-1} b^{n/2} \leq c_b \left( \sum_{m=t+1}^{k} a_m b^m \sum_{|u|>0} \sum_{|\kappa|>m-t-u} \sigma_{u,\kappa}^2 \right) + \sum_{m=t+1}^{k} a_m^{1/2} b^{m/2} \sum_{n=t}^{m-1} a_n^{1/2} b^{n/2} \sum_{|u|>0} \sum_{|\kappa|>n-t-u} \sigma_{u,\kappa}^2 \right) \quad (20)
\]

with \( c_b = \frac{(b-1)^{1/2}}{b^{1/2}} \) and where the first term after the last inequality sign is bounded by \( B_t^{(k)} \) by Lemma 2. For the second term, we have, using Lemma 2 (where \( k \) is replaced
by \( m - 1 \) and \( b \) by \( b^{1/2} \),

\[
\sum_{m=t+1}^{k} a_m b^{m/2} \left( \sum_{n=t}^{m-1} a_n^{1/2} b^{n/2} \sum_{|u|>0} \sum_{|x|>|x|-|u|} \sigma_{u,x}^2 \right) 
\leq \sum_{m=t+1}^{k} a_m^{1/2} b^{m/2} \left( \sum_{n=t}^{m-1} a_n^{1/2} b^{n/2} \sum_{|u|>0} \sum_{|x|>|x|-|u|} \sigma_{u,x}^2 
+ b \frac{m+1}{2} \sum_{|u|>0} b^{m-1-|u|} \sum_{l \leq m-1-|u|} \sigma_{u,l}^2 b^{l/2} \right)
\]

where the right hand side is bounded by

\[
c_b \left( \sum_{m=t+1}^{k} a_m b^{m} \sum_{|u|>0} \sum_{|x|>|x|-|u|} \sigma_{u,x}^2 \right) 
+ b^{l+1/2} \sum_{|u|>0} b^{m} \sum_{m=t+1}^{k} a_m^{1/2} b^{m/2} \sum_{l \leq m-1-|u|} b^{l+1/2} \sigma_{u,l}^2. \quad (21)
\]

Note that by Lemma 2 the term in bracket in the expression (21) is bounded by \( B^{(k)}_{t+1} \). In addition, in the same spirit as for the derivation of the upper bound in equation (16), the second term of (21) can be rewritten as

\[
b^{l+1/2} \sum_{m=l}^{k} (a_{m+1} b^m)^{1/2} \sum_{l \geq 0} b^{l+1/2} \sigma_{u,l}^2 \]

with \( b^{l+1/2} \sum_{m=l}^{k} (a_{m+1} b^m)^{1/2} \leq c_b b^{k+1/2} \). Therefore,

\[
b^{l+1/2} \sum_{|u|>0} \sum_{m=t+1}^{k} a_m b^{m/2} \sum_{l \leq m-1-|u|} b^{l+1/2} \sigma_{u,l}^2 \leq c_b b^{k+1/2} \sum_{|u|>0} \sum_{l=0}^{k-1} b^{l+1/2} \sigma_{u,l}^2 
= c_b b^{k+1/2} \sum_{|u|>0} \sum_{l=0}^{k-1} b^{l+1/2} \sigma_{u,l}^2 
\leq c_b b^k \sum_{|u|>0} \sum_{l=0}^{k-1} b^{l+1/2} \sigma_{u,l}^2.
\]
We conclude the proof using the fact that $(c_1^{1/2} + c_2^{1/2})^2 \leq 2(c_1 + c_2)$.

### A.2. Proof of the bound for $t = 0$

To prove the bound for $t = 0$, first note that, in this case, $\tilde{P} = \emptyset$. In addition, a $(0, s)$-sequence in base $b$ exists only if $b \geq s$ (see [Dick and Pillichshammer, 2010, Corollary 4.36, p.141]) and therefore the gain factors $\Gamma_{u, \kappa}$ are bounded by $\Gamma_0 = e$. Hence,

$$\sum_{m=0}^{k} \text{Var} \left( a_m b^m \hat{I}_m \right) \leq \Gamma_0 N \sigma^2$$

and, from (20),

$$2 \sum_{m > n \geq t} \frac{\text{Var}(a_m b^m \hat{I}_m)^{1/2} \text{Var}(a_n b^n \hat{I}_n)^{1/2}}{\Gamma_0} \leq c_b N \sigma^2 + \sum_{m=1}^{k} a_m^{1/2} b^{m/2} \left( \sum_{n=0}^{m-1} a_n^{1/2} b^{n/2} \sum_{|u| > 0} \sum_{|\kappa| > n - |u|} \sigma_{u, \kappa}^2 \right)$$

$$\leq c_b N \sigma^2 + \sigma^2 \sum_{m=t+1}^{k} a_m^{1/2} b^{m/2} \sum_{n=0}^{m-1} a_n^{1/2} b^{n/2}$$

$$\leq 2 c_b N \sigma^2.$$

Using (19), we conclude that

$$\text{Var} \left( \frac{1}{N} \sum_{n=0}^{N-1} \varphi(\tilde{x}^n) \right) \leq \frac{\sigma^2}{N} e (1 + 2c_b) \leq \frac{\sigma^2}{N} e (3 + 2\sqrt{2})$$

because $c_b \leq 1 + \sqrt{2}$, $\forall b \geq 2$. 

24