# QMC Methods for Integrals and Integral Equations

Emanouil Atanassov

emanouil@parallel.bas.bg

Department of Grid Technologies and Applications

Institute for Parallel Processing

Bulgarian Academy of Sciences

# Outline

- Motivation
- History
- Advantages
- Application to integral equations
- First steps
- Conclusion

# Purpose of QMC methods

- Improve the convergence of a wide range of modeling and simulation algorithms, running on modern distributed and HPC computing resources
- Achieve better convergence rates but retain some advantages of MC methods
  - parallelizable
  - usable in high dimensional settings
  - provide error bounds
  - low requirements for smoothness
- Unknown quantities are expressed as answers to integration problems, possibly of infinite dimension

# Motivation

- More than 70% of computing power of HPC and supercomputing resources is used for Monte Carlo simulations‘
- Typical convergence of MC methods is $\mathrm{O}\left(N^{-\frac{1}{2}}\right)$
- Advantages of MC methods:
  - inherently parallel
  - provide error estimates due to Central Limit Theorem
  - low smoothness requirements
  - rate does not depend on dimension
  - applicable in infinite dimensional settings

# Requirements on QMC methods

- Replace MC methods with little change in the codes and algorithms. Ideally only the pseudorandom number generator is replaced with low-discrepancy sequence generators
- Faster convergence rate for a wide range of functions
- Provide statistical aposteriori error estimates
- Run efficiently in distributed environments
  - high parallel efficiency
  - reproducibility of results
  - failover capability
- Usability in high-dimension problems

# History of QMC methods

- Uniform distribution of sequences $\mod 1$

  **Definition 0.1** *A sequence $\sigma = \{x_j\}_{j=1}^{\infty}$ is uniformly distributed in $\mathbb{E}^s$ if*

$$\lim_{N\to\infty} \frac{A_N(J)}{N} = \mu(J)$$

  *for each interval $J \subset \mathbb{E}^s$. This is equivalent to*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} f(x_j) = \int_{\mathbb{E}^s} f(x)dx$$

  *for each continuous function.*

# History of QMC methods - constructions fron

- **Theorem 0.1** *(H. Weyl, 1916) The sequence* $\{n\alpha\}$ *is uniformly distributed in* $\mathbb{E}^s$*, if* $1, \alpha_1, \ldots, \alpha_s$ *are linearly independent over the field of rationals.*
- Korobov showed how a net of numbers $\left\{\frac{na}{p}\right\}$ can give better convergence than Monte Carlo for a class of periodic functions with bounded derivatives. $p$ must be prime. The numbers $a = (a_1, \ldots, a_s)$ are called *optimal coefficients.*

# Discrepancy

- **Definition 0.2** *For a sequence* $\sigma = (x_i)_{i=1}^N \subset \mathbb{E}^s$*, the discrepancy is defined as*

$$D_N(\sigma) = sup_{J \subset \mathbb{E}^s} \frac{A_N(J)}{N} - \mu(J)$$

*where* $J$ *is an* $s$*-dimensional interval.*

- Star discrepancy - if the intervals start with $0$. $L^p$ discrepancy, diaphony are other measures of irregularity of distribution
- Best possible rate (conjecture) - $\mathrm{O}\left(N^{-1} \log^s N\right)$

# Theorem of Koksma - Hlawka

- **Theorem 0.2** *For a function $f$ of bounded variation in the sense of Hardy and Krause*

$$\left| \frac{1}{N} \sum_{j=1}^{N} f(x_j) - \int_{\mathbb{E}^s} f(x) dx \right| < V(f) D_N(\sigma)$$

- This is largely theoretical result, since computing discrepancy is hard, and theoretical estimates largely overstate the discrepancy.

# History of QMC methods - digital constructio

- **Definition 0.3** *The sequence of Van der Corput - Halton in base* $p$ *is obtained by representing*

$$n = \sum_{j=0}^{\infty} a_j p^j$$

*and then defining*

$$\phi_p(n) = \sum_{j=0}^{\infty} a_j p^{-j-1}$$

*. If* $p_1, \ldots, p_s$ *are pairwise relatively prime, the sequence* $(\phi_{p_1}(n)), \ldots, \phi_{p_s}(n))$ *, is uniformly distributed in* $\mathbb{E}^s$*. Typically* $p_1, \ldots, p_s$ *are the first* $s$ *primes.*

- Faure has shown how permutting the digits can give better convergence.

# Digital $(t, m, s)$ nets

- **Definition 0.4** *For a given dimension $s \geq 1$ and integers $b \geq 2$ and $0 \leq t \leq m$, a $(t, m, s)$-net in base $b$ is a point set $P$ consisting of $b^m$ points in $\mathbb{E}^s$ such that every subinterval $J$ of $\mathbb{E}^s$ of the form*

$$J = \prod_{i=1}^{s} \left[ a_i b^{-d_i}, (a_i + 1) b^{-d_i}) \right)$$

*with integers $d_i \geq 0$ and $0 \geq a_i \geq b^{d_i}$ for $1 \leq i \leq s$ and with $\mu(J) = b^{t-m}$ contains exactly $b^t$ points of $P$.*

- The sequences of Sobol, Faure, Niederreiter, are of this type and have best possible order of discrepancy.

# The notion of effective dimension

- If we can represent a Monte Carlo algorithm as an integration problem in $\mathbb{E}^s$, we call $s$ the effective dimension of the algorithm.
- Decreasing the effective dimension makes QMC methods more efficient.
- Infinite dimensional uniformly distributed sequences do exist.
- For a given problem, there is a competition between Sobol, Faure, Niederreiter, Halton and other constructions.
- If the functions are periodic, Korobov' sequences are to be preferred (but they are not infinite).

# ANOVA decomposition

- A function $f$ can be decomposed as:

$$f(x) = f_0 + f_1(x_1) + \cdots + f_s(x_s) + \sum_{i,j} f_{i,j}(x_i, x_j) + \cdots + f_{1,\ldots,s}(x_1, \ldots, x_s)$$

where $f_0$ is constant and all the functions are orthogonal. Then the variance of $f$ can be expressed as

$$D = D_1 + \cdots + D_s + D_{1,2} + \cdots + D_{1,\ldots,s}$$

where

$$D = \int_{\mathbb{E}^s} f^2(x) dx$$

and

$$D_{i_1,\ldots,i_k} = \int_0^1 \cdots \int_0^1 f^2_{i_1,\ldots,i_k}(x_{i_1}, \ldots, x_{i_k})\, dx_1 \ldots dx_k$$

- In many integration problems the low-dimensional interactions contribute most of the total variance

# The Brownian bridge construction with QMC

- How to sample a Wiener process $W(t)$ with QMC? Suppose $t \in [0, 1]$.
- Let $s$ be the number of intermediate points. Consider low-discrepancy sequence $\sigma = (x_j^1, \ldots, x_j^s)$ $W(1)$ is sampled with $invnorm(x_j^1)$. Then $W(\frac{1}{2})$ is sampled with mean $W(0) + W(1)$ and standard deviation $\frac{1}{2}$, using $x_j^2$. After that we sample $W(\frac{1}{4})$ and then $W(\frac{3}{4})$, etc.
- The advantage is that the variance is concentrated towards the first coordinates, which are usually better distributed (Sobol, Halton).

# The use of scrambling

- Obtaining error estimates in QMC is not possible in the same way as in MC - the samples are correlated.
- Introduce a source of randomness into the algorithm
- Basic scrambling
  - Add the same random number to all samples ($mod 1$).
  - **xor** with the same random binary number
  - Perform the same permutation of $p$ digits
- If several runs of the QMC algorithm are performed, with different scramblings, then statistical error estimate can be obtain. It will not be as accurate as in MC, because number of (super-)samples is rather small, even though number of terms of the sequence used may be large.

# Owen's construction

- Consider one-dimensional case. Fix the base $b$ and select a tree of random permutations, all of them independently choosen.
- Permute the first digit, using the first permutation $\tau_1$. Depending on the result, permute the second digit, using the permutation from the corresponding node in the tree. Continute until all digits are permuted (instead of continuing the process, at some point one can simply choose a random number, since no other point can arrive at the same branch).
- Advantage of Owen's construction - expected error is

$$\mathrm{O}\left(N^{-\frac{3}{2}}\log N\right)$$

  for functions in $L^2$.
- These methods are called randomized QMC and are effectively the norm.

# Transforming MC into QMC

- If MC is sampling trajectories, we consider one trajectory to result from one point in $\mathbb{E}^s$. $s$ is the constructive dimensionality. From one QMC point we obtain one trajectory.
- If $s$ is not limited, we can use hybrid methods, where we use QMC for part of the variables and MC for the rest.
- Always use scrambling, even without making error estimation, otherwise you can hit a singularity.
- Example - integral equations that are resolved by Neumann series expansion
- It is also advantageous to skip the first $N$ elements of a sequence, where $N$ is large. This is not equivalent to scrambling.

# Transforming MC into QMC

- Not all QMC methods result from simple replacement of pseudorandom numbers with low-discrepancy sequences. Example - Runge-Kutta QMC
- QMC has advantage if constructive dimensionality is low (10, max 20), or if the ANOVA decomposition is such that variance is concentrated in low-dimensional interactions.
- Example - in finance Sobol sequences with several thousand coordinates are being used in production.

# MC vs QMC

- Not all QMC methods result from simple replacement of pseudorandom numbers with low-discrepancy sequences. Example - Runge-Kutta QMC
- QMC has advantage if constructive dimensionality is low (10, max 20), or if the ANOVA decomposition is such that variance is concentrated in low-dimen sional interactions.
- Example - in finance Sobol sequences with several thousand coordinates are being used in production.
- In most cases direct replacement is not worse, and can gain because of speed of generation. With increase of number of trajectories pseudorandom number generators start to exhibit strange correlations.

# Parallel implementation

- QMC can not discard or lose partial results
- Blocking or leap-frogging is used for parallelisation.
- Not as flexible as MC, but reasonably efficient (close to 100%).
- Offers reproducibility of results - small initial state gives all the information needed.

# Cost analysis

- QMC can be plugged in with little effort
- Standard software implementations exist, but load-balancing is difficult. Aborted computations should re-submitted.
- Most integration methods gain from using QMC
- In integral equations, the improvement depends on the practical problem, and some re-ordering or re-design may be necessary. The smoothness of the sub-integral function is important.

# Conclusion

- Changing the algorithm from MC to QMC can be done in a few lines of code.
- Standard software implementations exist, even for Matlab.
- Strong improvement is to be expected in most integration problems. QMC combines with other techniques for variance reduction.
- Usable in high-dimensional problems
- Have been used in ensemble Monte Carlo methods
- Ray tracing in computer graphics has been done with QMC