Conditional Quasi-Monte Carlo with Active Subspaces

Sifan Liu

Center for Computational Mathematics, Flatiron Institute

MCQMC 2024

Introduction

- Accurate estimation of expectations is a fundamental task across many fields.
- To evaluate $\mu = \mathbb{E}\left[f(x)\right]$, Monte Carlo methods sample $\mathbf{x}^{(i)} \sim \mathbb{P}$ and compute the sample average

$$\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i).$$

- The standard error of the Monte Carlo estimator is of order $O_p(n^{-1/2})$ for square-integrable functions.

Quasi-Monte Carlo

- QMC points are constructed deterministically to fill the unit hypercube $[0,1]^d$ more evenly than plain Monte Carlo samples.

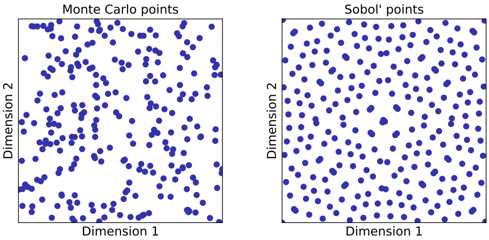


Figure: 256 Monte Carlo points (left) and Sobol' points (right, [Sob67]) in two dimensions.

$$D_n^* = \sup_{\mathbf{a}\in[0,1)^d} \left| \frac{\#\text{points in } [0,\mathbf{a})}{n} - \prod_{j=1}^{d} a_j \right|$$

Low-discrepancy sequence:

$$D_n^* = O(n^{-1}(\log n)^d)$$

Quasi-Monte Carlo

- QMC points are constructed deterministically to fill the unit hypercube $[0,1]^d$ more evenly than plain Monte Carlo samples.

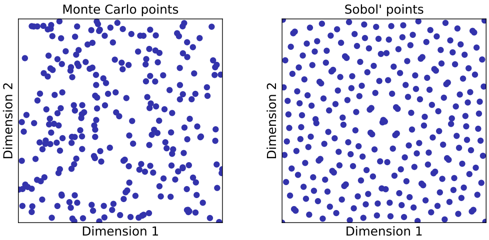


Figure: 256 Monte Carlo points (left) and Sobol' points (right, [Sob67]) in two dimensions.

$$D_n^* = \sup_{\mathbf{a}\in[0,1)^d} \left| \frac{\#\text{points in } [0,\mathbf{a})}{n} - \prod_{j=1}^d a_j \right|$$

Low-discrepancy sequence:

$$D_n^* = O(n^{-1}(\log n)^d)$$

Koksma-Hlawka inequality:

$$|\hat{\mu}_n - \mu| \leq \mathrm{V}(f) \cdot D_n^*$$

Quasi-Monte Carlo

- QMC points are constructed deterministically to fill the unit hypercube $[0,1]^d$ more evenly than plain Monte Carlo samples.
- QMC points can be randomized to get RQMC points.

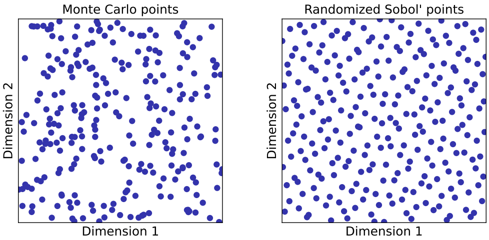


Figure: 256 Monte Carlo points (left) and Sobol' points (right, [Sob67]) in two dimensions.

$$D_n^* = \sup_{\mathbf{a}\in[0,1)^d} \left| \frac{\#\text{points in } [0,\mathbf{a})}{n} - \prod_{j=1}^d a_j \right|$$

Low-discrepancy sequence:

$$D_n^* = O(n^{-1}(\log n)^d)$$

Koksma-Hlawka inequality:

$$|\hat{\mu}_n - \mu| \le \mathrm{V}(f) \cdot D_n^*$$

Challenges

- Faster convergence rate requires higher-order smoothness.
- Error grows with dimension d.

We propose a generic method for reducing RQMC variance by

- improving the smoothness
- and reducing the "effective" dimension

Conditional Monte Carlo

- CMC is a variance reduction method for Monte Carlo estimators.
- Pre-integrating x_j: $\tilde{f}(\mathbf{x}_{-j}) = \mathbb{E}\left[f(\mathbf{x}) \mid \mathbf{x}_{-j}\right]$
  - $\mathbb{E}\left[\tilde{f}\right] = \mathbb{E}\left[f\right]$
  - $\mathrm{Var}(\tilde{f}) \leq \mathrm{Var}(f)$
- For RQMC, pre-integration can also improve smoothness.
  E.g. $f(x) = \mathbf{1}\left\{x_j \leq \phi(x_{-j})\right\}$ [GKLS18].

Which variable to pre-integrate

Pre-integrate x_1

$$\int f(x_1, x_2, \ldots, x_d) \mathrm{d}x_1$$

Which variable to pre-integrate

Pre-integrate x_2

$$\int f(x_1, x_2, \ldots, x_d)\mathrm{d}x_2$$

Which variable to pre-integrate

For integrals w.r.t. Gaussian density $\varphi(\mathbf{x})$, pre-integrate any linear combination of the variables:

$$\int f(U\mathbf{x})\varphi(\mathbf{x})\mathrm{d}x_1, \quad U \in \mathbb{R}^{d\times d} \text{ orthogonal}$$

Which variable to pre-integrate

The goal is to find a rotation U such that $\int f(U\mathbf{x})\varphi(\mathbf{x})\mathrm{d}x_1$

- has a tractable form
- achieves a large variance reduction
- improves smoothness
- reduces the effective dimension

Option pricing

- Suppose the asset price S_t follows the SDE

$$\frac{\mathrm{d}S(t)}{S(t)} = r\mathrm{d}t + \sqrt{V(t)}\mathrm{d}W^{(1)}(t),$$

  where $V(t)$ is the stochastic volatility satisfying

$$\mathrm{d}V(t) = a(V(t))\mathrm{d}t + b(V(t))\mathrm{d}W^{(2)}(t).$$

  where $W^{(1)}, W^{(2)}$ are two Brownian motions with correlation ρ.
- Asian call option: $\mathbb{E}\left(\frac{1}{T}\int_0^T S(t) - K\right)_+$
- Apply Euler-Maruyama discretization and Monte Carlo methods to simulate the path

$$\begin{aligned} \log S_{j+1} &= \log S_j + (r - V_j/2)\Delta t + \sqrt{V_j \Delta t}\left(\sqrt{1-\rho^2} z_{1,j+1} + \rho z_{2,j+1}\right), \\ V_{j+1} &= V_j + a(V_j)\Delta t + b(V_j)\sqrt{\Delta t} z_{2,j+1}, \quad \text{for } j = 0, \ldots, d-1. \end{aligned} \tag{1}$$

- The problem reduces to evaluate a Gaussian integral in $2d$ dimensions.

Pre-integration step

For a rotation matrix U, consider the integrand $f_U(\mathbf{x}) = f(U\mathbf{x})$.

Simplification of the integrand

If $U_{1:d,1} \geq 0$ and $U_{d+1:2d,1} = 0$, then the integrand takes the form

$$f_U(\mathbf{x}) = \left(\sum_{j=1}^{d} e^{a_j(\mathbf{x}_{-1}) + b_j x_1} - K \right)_+ ,$$

where $b_j > 0$ are constants, and $a_j(x_{-1})$ do not depend on x_1 for $1 \leq j \leq d$. In this case, the pre-integration step has a closed form

$$\int f_U(\mathbf{x}) \varphi(x_1) \mathrm{d}x_1 = \sum_{j=1}^{d} e^{a_j(\mathbf{x}_{-1}) + b_j^2/2} \bar{\Phi}(\gamma - b_j) - K \bar{\Phi}(\gamma),$$

where γ is such that $\sum_{j=1}^{d} e^{a_j(\mathbf{x}_{-1}) + b_j \gamma} = K$.

Pre-integration step

- The threshold γ can be found by a root-finding algorithm.
- Similar pre-integration steps for constant-volatility option pricing were used in [XW18, He19].
- We will choose U such that

$$U_{1:d,1} \geq 0, \quad U_{d+1:2d,1} = 0.$$

Error decomposition

- ANOVA decomposition: $f(\mathbf{x}) = \sum_{w\subseteq 1:d} f_w(\mathbf{x}_w)$, where $f_\emptyset \equiv \mu$; $\int f_w \mathrm{d}x = 0$ for $w \neq \emptyset$; and $\int f_w f_{w'} \mathrm{d}x = 0$ for $w \neq w'$.
- Monte Carlo variance:

$$\frac{1}{n}\mathrm{Var}(f) = \frac{1}{n}\sum_{w\subseteq 1:d} \sigma_w^2(f), \quad \text{where } \sigma_w^2(f) = \mathrm{Var}(f_w).$$

Error decomposition

- ANOVA decomposition: $f(\mathbf{x}) = \sum_{w \subseteq 1:d} f_w(\mathbf{x}_w)$, where $f_\emptyset \equiv \mu$; $\int f_w \mathrm{d}x = 0$ for $w \neq \emptyset$; and $\int f_w f_{w'} \mathrm{d}x = 0$ for $w \neq w'$.
- Monte Carlo variance:

$$\frac{1}{n}\mathrm{Var}(f) = \frac{1}{n}\sum_{w \subseteq 1:d} \sigma_w^2(f), \quad \text{where } \sigma_w^2(f) = \mathrm{Var}(f_w).$$

- Pre-integrating x_1

$$\frac{1}{n}\sum_{w:1\in w} 0 \cdot \sigma_w^2 + \frac{1}{n}\sum_{w:1\notin w} \sigma_w^2$$

pre-integrate x_1

Error decomposition

- ANOVA decomposition: $f(\mathbf{x}) = \sum_{w\subseteq 1:d} f_w(\mathbf{x}_w)$, where $f_\emptyset \equiv \mu$; $\int f_w \mathrm{d}x = 0$ for $w \neq \emptyset$; and $\int f_w f_{w'} \mathrm{d}x = 0$ for $w \neq w'$.
- Monte Carlo variance:

$$\frac{1}{n}\mathrm{Var}(f) = \frac{1}{n}\sum_{w\subseteq 1:d} \sigma_w^2(f), \quad \text{where } \sigma_w^2(f) = \mathrm{Var}(f_w).$$

- Pre-integrating x_1 and apply RQMC to the remaining variables:

$$\frac{1}{n}\sum_{w:1\in w} 0\cdot\sigma_w^2 + \frac{1}{n}\sum_{w:1\notin w} \Gamma_w\cdot\sigma_w^2$$

pre-integrate x_1 RQMC gain

Error decomposition

- ANOVA decomposition: $f(\mathbf{x}) = \sum_{w \subseteq 1:d} f_w(\mathbf{x}_w)$, where $f_\emptyset \equiv \mu$; $\int f_w \mathrm{d}x = 0$ for $w \neq \emptyset$; and $\int f_w f_{w'} \mathrm{d}x = 0$ for $w \neq w'$.
- Monte Carlo variance:

$$\frac{1}{n}\mathrm{Var}(f) = \frac{1}{n}\sum_{w \subseteq 1:d} \sigma_w^2(f), \quad \text{where } \sigma_w^2(f) = \mathrm{Var}(f_w).$$

- Pre-integrating x_1 and apply RQMC to the remaining variables:

$$\frac{1}{n}\sum_{w:1 \in w} 0 \cdot \sigma_w^2 + \frac{1}{n}\sum_{w:1 \notin w} \Gamma_w \cdot \sigma_w^2$$

pre-integrate x_1 RQMC gain

- We want to choose U such that large σ_w^2 values are multiplied by small factors.

Choosing the rotation: U_1

- Reduced variance:

$$\sum_{w:1\in w} \sigma_w^2 = \frac{1}{2}\mathbb{E}\left[(f_U(\mathbf{x}) - f_U(\tilde{x}_1, \mathbf{x}_{-1}))^2\right] \quad (x_1, \tilde{x}_1, \mathbf{x}_{-1}) \sim \mathcal{N}(0, I_{2d+1})$$
$$\approx U_1^\intercal \mathbb{E}\left[\nabla f(\mathbf{x}) \nabla f(\mathbf{x})^\intercal\right] U_1.$$

- To maximize the reduced variance, we could take U_1 to be the first eigenvector of

$$C := \mathbb{E}\left[\nabla f(\mathbf{x}) \nabla f(\mathbf{x})\right].$$

Choosing the rotation: U_1

- Reduced variance:

$$\sum_{w:1\in w} \sigma_w^2 = \frac{1}{2}\mathbb{E}\left[(f_U(\mathbf{x}) - f_U(\tilde{x}_1, \mathbf{x}_{-1}))^2\right] \quad (x_1, \tilde{x}_1, \mathbf{x}_{-1}) \sim \mathcal{N}(0, I_{2d+1})$$
$$\approx U_1^\intercal \mathbb{E}\left[\nabla f(\mathbf{x}) \nabla f(\mathbf{x})^\intercal\right] U_1.$$

- To maximize the reduced variance, we could take U_1 to be the first eigenvector of

$$C := \mathbb{E}\left[\nabla f(\mathbf{x}) \nabla f(\mathbf{x})\right].$$

- But we require $U_{1:d,1} \geq 0$ and $U_{d+1:2d,1} = 0$ for the pre-integration step to be tractable.
- So we select U_1 by solving the following problem:

$$\begin{aligned} \max_{\mathbf{v}\in\mathbb{R}^{2d}} \quad & \mathbf{v}^\intercal C \mathbf{v} \\ \text{s.t.} \quad & \mathbf{v}^\intercal \mathbf{v} = 1, \\ & \mathbf{v}_{1:d} \geq 0, \quad \mathbf{v}_{d+1:2d} = 0. \end{aligned}$$

Choosing the remaining columns $U_{2:2d}$

- To minimize the remaining variance $\sum_{w:1\notin w} \Gamma_w \sigma_w^2(f_U)$, we seek U such that large σ_w^2 values are multiplied by small Γ_w values.
- RQMC gain coefficients Γ_w are typically small for the first few projections.

Choosing the remaining columns $U_{2:2d}$

- To minimize the remaining variance $\sum_{w:1\notin w} \Gamma_w \sigma_w^2(f_U)$, we seek U such that large σ_w^2 values are multiplied by small Γ_w values.
- RQMC gain coefficients Γ_w are typically small for the first few projections.
- So we aim for the first few variables of f_U to explain most of the variance.
- For $k = 2, \ldots, 2d$,

$$\begin{aligned} \max_{\mathbf{v}} \quad & \mathbf{v}^\mathsf{T} C \mathbf{v} \\ \text{s.t.} \quad & \mathbf{v}^\mathsf{T}\mathbf{v} = 1, \quad \mathbf{v}^\mathsf{T} U_j = 0, \ \forall\, 1 \le j \le k-1. \end{aligned}$$

Choosing the remaining columns $U_{2:2d}$

- To minimize the remaining variance $\sum_{w:1\notin w} \Gamma_w \sigma_w^2(f_U)$, we seek U such that large σ_w^2 values are multiplied by small Γ_w values.
- RQMC gain coefficients Γ_w are typically small for the first few projections.
- So we aim for the first few variables of f_U to explain most of the variance.
- For $k = 2, \ldots, 2d$,

$$
\begin{aligned}
\max_{\mathbf{v}} \quad & \mathbf{v}^\intercal C \mathbf{v} \\
\text{s.t.} \quad & \mathbf{v}^\intercal \mathbf{v} = 1, \quad \mathbf{v}^\intercal U_j = 0, \ \forall\, 1 \le j \le k-1.
\end{aligned}
$$

- Given the first column U_1, $U_{2:2d}$ can be found by one eigendecomposition of $U_{1,\perp}^\intercal C U_{1,\perp}$.

Choosing the remaining columns $U_{2:2d}$

- To minimize the remaining variance $\sum_{w:1\notin w} \Gamma_w \sigma_w^2(f_U)$, we seek U such that large σ_w^2 values are multiplied by small Γ_w values.
- RQMC gain coefficients Γ_w are typically small for the first few projections.
- So we aim for the first few variables of f_U to explain most of the variance.
- For $k = 2, \ldots, 2d$,

$$\max_{\mathbf{v}} \quad \mathbf{v}^\intercal C \mathbf{v}$$
$$\text{s.t.} \quad \mathbf{v}^\intercal \mathbf{v} = 1, \quad \mathbf{v}^\intercal U_j = 0, \ \forall\, 1 \le j \le k-1.$$

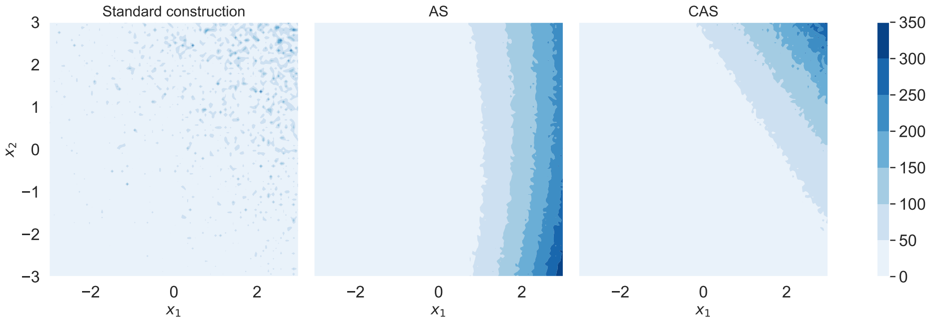
- Given the first column U_1, $U_{2:2d}$ can be found by one eigendecomposition of $U_{1,\perp}^\intercal C U_{1,\perp}$.
- Without the constraints on U_1, this procedure coincides with the active subspace method [Con15] and gradient PCA (GPCA) [XW19].

Heston model

- $\mathrm{d}S(t) = rS(t)\mathrm{d}t + \sqrt{V(t)}S(t)\mathrm{d}W^{(1)}t$.
- $\mathrm{d}V(t) = \kappa(\theta - V(t))\mathrm{d}t + \sigma\sqrt{V(t)}\mathrm{d}W^{(2)}(t)$.
- Number of time steps is $d = 32$.
- $\mathrm{Corr}(W^{(1)}, W^{(2)}) = 0.5$.
- Error reduction factors relative to plain Monte Carlo at sample size 2^{14}.

	RQMC			RQMC + preint		
K	STD	PCA	AS	STD	PCA	CAS
90	9.1	63.2	79.4	12.2	143.2	**352.1**
100	5.5	44.5	74.4	8.0	117.2	**347.1**
110	4.6	46.0	62.2	5.4	95.8	**254.0**

Visualization of the rotation



- Left: standard construction of Brownian motions
- Middle: Active subspace rotation
- Right: Constrained active subspace such that $U_{1:d,1} \geq 0$, $U_{d+1:2d,1} = 0$.

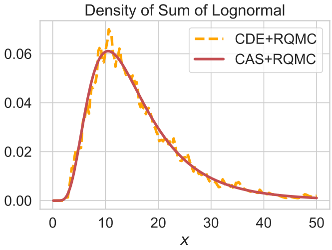
Spread option

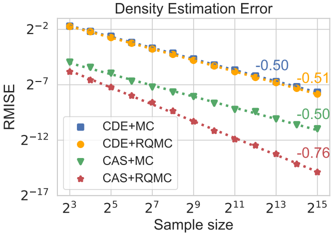
- $\mathrm{d}S^{(k)}(t) = rS^{(k)}(t)\mathrm{d}t + \sigma^{(k)}S^{(k)}(t)\mathrm{d}W^{(k)}(t)$
- $\mathrm{Corr}(W^{(1)}, W^{(2)}) = 0.5$
- $\mathbb{E}\left[(\bar{S}^{(1)} - \bar{S}^{(2)} - K)_+\right]$

	RQMC			RQMC + preint		
K	STD	PCA	AS	STD	PCA	CAS
-10	9.9	125.6	327.7	12.2	680.9	**4564.6**
0	5.1	70.0	223.0	6.4	352.4	**4400.2**
10	3.5	35.5	112.2	5.0	243.7	**4511.5**

Further applications

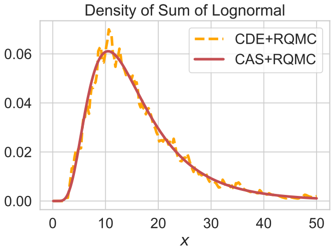
Conditional density estimation

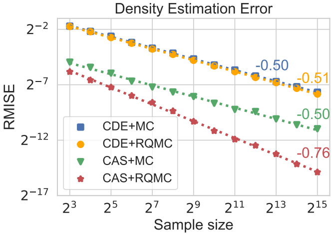




Further applications

Conditional density estimation





Simulating chemical reaction network

- $\mathrm{d}X_t = (\sum_{j=1}^J \nu_j a_j(X_t))\mathrm{d}t + \sum_{j=1}^J \nu_j \sqrt{a_j(X_t)}\mathrm{d}W_t^{(j)}$
- Error reduction factor for estimating $\mathbb{P}\left[X_d \leq K\right]$:

Threshold K	RQMC	**Proposed**
90	1.6	**1227.2**
100	2.2	**2103.6**
110	1.4	**1482.7**

Conclusions

- Conditional Monte Carlo is powerful if we pre-integrate an important variable.
- A generic approach is proposed for selecting the pre-integration variable while simutaneously reducing the effective dimension for RQMC.
- This method has broad applications, including option pricing under various models and beyond.

- Thank Prof. Art Owen for many helpful discussions.
- Based on:
  - Liu, S., & Owen, A. B. (2023). Preintegration via Active Subspaces. *SIAM Journal on Numerical Analysis.*
  - Liu, S. (2024+). Conditional Quasi-Monte Carlo with Constrained Active Subspaces. To appear *SIAM Journal on Scientific Computing.*

Thank you!

References I

P. G. Constantine.
Active subspaces: Emerging ideas for dimension reduction in parameter studies.
SIAM, Philadelphia, 2015.

Andreas Griewank, Frances Y Kuo, Hernan Leövey, and Ian H Sloan.
High dimensional integration of kinks and jumps—smoothing by preintegration.
Journal of Computational and Applied Mathematics, 344:259–274, 2018.

Zhijian He.
On the error rate of conditional quasi–Monte Carlo for discontinuous functions.
SIAM Journal on Numerical Analysis, 57(2):854–874, 2019.

I. M. Sobol'.
The distribution of points in a cube and the accurate evaluation of integrals.
USSR Computational Mathematics and Mathematical Physics, 7(4):86–112, 1967.

References II

Ye Xiao and Xiaoqun Wang.
Conditional quasi-monte carlo methods and dimension reduction for option pricing and hedging with discontinuous functions.
Journal of Computational and Applied Mathematics, 343:289–308, 2018.

Y. Xiao and X. Wang.
Enhancing quasi-Monte Carlo simulation by minimizing effective dimension for derivative pricing.
Computational Economics, 54(1):343–366, 2019.