Lecture 10. Simulated Annealing and QMC

Tiejun Li[1,2]

[1]School of Mathematical Sciences (SMS),
&
[2]Center for Machine Learning Research (CMLR),
Peking University,
Beijing 100871,
P.R. China
tieli@pku.edu.cn

Office: No. 1 Science Building, Room 1376E

Table of Contents

Simulated Annealing

- We already have very efficient algorithms for traditional convex programming.

Simulated Annealing

- We already have very efficient algorithms for traditional convex programming.
- But how about the non-convex programming problems, such as the following combinatorial optimization problem?

Traveling Salesman Problem

Suppose there are N cities and there exists one path $(l_{ij} = l_{ji})$ for each two. Try to find a minimal path passing all the cities such that each city is passed and only passed one time.

$$\min_{x \in X} H(x) = \sum_{i=1}^{N} l_{x_i x_{i+1}}, \quad x_{N+1} := x_1.$$

$X = \{(x_1, \ldots, x_N),\ x_1, \ldots, x_N$ is a permutation of $1, 2, \ldots, N\}$

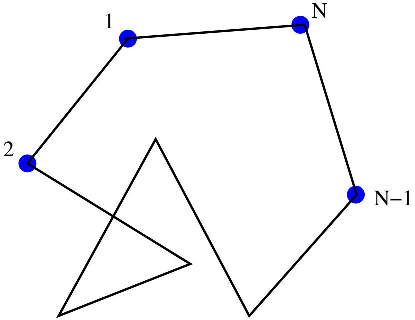

Figure: Traveling Salesman Problem

Traveling Salesman Problem

- The number of all the possible paths is $O(N!)$. It is a typical combinatorial explosion problem (NP-hard problem).

Traveling Salesman Problem

- The number of all the possible paths is $O(N!)$. It is a typical combinatorial explosion problem (NP-hard problem).
- This number increases exponentially with N, and there is not any simple rules for the function $H(x)$.

Traveling Salesman Problem

- The number of all the possible paths is $O(N!)$. It is a typical combinatorial explosion problem (NP-hard problem).
- This number increases exponentially with N, and there is not any simple rules for the function $H(x)$.
- The traditional algorithms are inapplicable here.

Image restoration problem

Reference Image Noisy Image Denoised Image

Figure: Image denoising problem

Image restoration problem

Suppose there are J pixels for an image, and there are 256 colors for each pixel.

- Any image can be represented as one element in

$$X = \{(x_1, \ldots, x_J) \;:\; x_i \in \{0, 1, \ldots, 255\}\}.$$

The smoothness of an image is defined as

$$H(x) = \alpha \sum_{\langle s,t \rangle} (x_s - x_t)^2, \qquad \alpha > 0,$$

where $\langle s, t \rangle$ means the neighboring pixels in the lattice among $x = (x_1, \ldots, x_J)$.

Image restoration problem

Suppose there are J pixels for an image, and there are 256 colors for each pixel.

- Any image can be represented as one element in

$$X = \{(x_1, \ldots, x_J) \ : \ x_i \in \{0, 1, \ldots, 255\}\}.$$

  The smoothness of an image is defined as

$$H(x) = \alpha \sum_{\langle s,t \rangle} (x_s - x_t)^2, \qquad \alpha > 0,$$

  where $\langle s, t \rangle$ means the neighboring pixels in the lattice among $x = (x_1, \ldots, x_J)$.

- Then define the comparison function for images x and y where y is the reference image

$$H(x|y) = \alpha \sum_{\langle s,t \rangle} (x_s - x_t)^2 + \frac{1}{2\sigma^2} \sum_s (x_s - y_s)^2.$$

Image restoration problem

- An image recovering problem for polluted y may be proposed as minimizing the following function:

$$\min_{x \in X} H(x|y).$$

Image restoration problem

- An image recovering problem for polluted y may be proposed as minimizing the following function:

$$\min_{x \in X} H(x|y).$$

- The number of all the possible states is 256^J! Traditional algorithms are still inapplicable here!

Image restoration problem

- An image recovering problem for polluted y may be proposed as minimizing the following function:

$$\min_{x \in X} H(x|y).$$

- The number of all the possible states is 256^J! Traditional algorithms are still inapplicable here!
- Simulated annealing algorithm is one of the framework to handle this kind of non-convex global optimization problem from stochastics viewpoint.

Image restoration problem

- An image recovering problem for polluted y may be proposed as minimizing the following function:

$$\min_{x \in X} H(x|y).$$

- The number of all the possible states is 256^J! Traditional algorithms are still inapplicable here!
- Simulated annealing algorithm is one of the framework to handle this kind of non-convex global optimization problem from stochastics viewpoint.
- While the effectivity is still under discussion.

Simulated Annealing: Basic Framework

Otimization problem:

- For optimization problem

$$\min_{x \in X} H(x),$$

Define the global minimizers of $H(x)$

$$M = \{x_0 \;:\; H(x_0) = \min_{x \in X} H(x)\},$$

and introduce the parameter $\beta > 0$, define

$$\Pi^\beta(x) = \frac{1}{Z_\beta} e^{-\beta H(x)}, \qquad Z_\beta = \sum_{x \in X} \exp(-\beta H(x)),$$

then $\Pi^\beta(x)$ is a probability distribution on X.

Simulated Annealing: Theorem

Theorem $\Pi^\beta(x)$ has the property

$$\lim_{\beta\to+\infty}\Pi^\beta(x)=\begin{cases}\frac{1}{|M|} & \text{if } x\in M,\\ 0 & \text{else.}\end{cases}$$

and if β is sufficiently large, then $\Pi^\beta(x)$ is monotonely increasing as a function of β for any $x\in M$, and $\Pi^\beta(x)$ is monotonely decreasing as a function of β for any $x\notin M$.

- Proof. Rewrite

$$\Pi^\beta(x)=\frac{e^{-\beta(H(x)-m)}}{\sum_{z:H(z)=m}e^{-\beta(H(z)-m)}+\sum_{z:H(z)>m}e^{-\beta(H(z)-m)}}$$

$$\xrightarrow{\beta\to+\infty}\begin{cases}\frac{1}{|M|}, & x\in M,\\ 0, & x\notin M,\end{cases}$$

where $m=\min_x H(x)$.

Simulated Annealing: Theorem

- If $x \in M$, we have

$$\Pi^\beta(x) = \frac{1}{|M| + \sum_{z:H(z)>m} e^{-\beta(H(z)-m)}},$$

then $\Pi^\beta(x)$ monotonely increases with β increasing.

Simulated Annealing: Theorem

- If $x \in M$, we have

$$\Pi^{\beta}(x) = \frac{1}{|M| + \sum_{z:H(z)>m} e^{-\beta(H(z)-m)}},$$

  then $\Pi^{\beta}(x)$ monotonely increases with β increasing.
- If $x \notin M$, we have

$$\begin{aligned}
\frac{\partial \Pi^{\beta}(x)}{\partial \beta} &= \frac{1}{\tilde{Z}_{\beta}^{2}} \Big(e^{-\beta(H(x)-m)}(m - H(x))\tilde{Z}_{\beta} - e^{-\beta(H(x)-m)} \sum_{z \in X} e^{-\beta(H(z)-m)}(m - H(z)) \Big) \\
&= \frac{1}{\tilde{Z}_{\beta}^{2}} \Big(e^{-\beta(H(x)-m)} [(m - H(x))\tilde{Z}_{\beta} - \sum_{z \in X} e^{-\beta(H(z)-m)}(m - H(z))] \Big),
\end{aligned}$$

  where

$$\tilde{Z}_{\beta} \triangleq \sum_{z \in X} \exp(-\beta(H(z) - m)).$$

  Pay attention that

$$\lim_{\beta \to +\infty} [(m - H(x))\tilde{Z}_{\beta} - \sum_{z \in X} e^{-\beta(H(z)-m)}(m - H(z))] = |M|(m - H(x)) < 0,$$

  The proof is completed.

Remark

- The construction of $\Pi^{\beta}(x)$ opens a new way to optimize $H(x)$ via stochastics.

Remark

- The construction of $\Pi^{\beta}(x)$ opens a new way to optimize $H(x)$ via stochastics.
- The theorem shows that if we can generate the random sequence with distribution $\Pi^{\beta}(x)$, then the random numbers will finally jump among the minimizers when $\beta = +\infty$. This procedure is called annealing.

Remark

- The construction of $\Pi^\beta(x)$ opens a new way to optimize $H(x)$ via stochastics.
- The theorem shows that if we can generate the random sequence with distribution $\Pi^\beta(x)$, then the random numbers will finally jump among the minimizers when $\beta = +\infty$. This procedure is called annealing.
- It corresponds to the physical crystallization. In physics, β corresponds to $1/T$, where T is temperature. Global energy minimization means a perfect crystal without defects.The observed crystals with defects in nature can be understood as the local minimum state.

Remark

- The construction of $\Pi^{\beta}(x)$ opens a new way to optimize $H(x)$ via stochastics.
- The theorem shows that if we can generate the random sequence with distribution $\Pi^{\beta}(x)$, then the random numbers will finally jump among the minimizers when $\beta = +\infty$. This procedure is called annealing.
- It corresponds to the physical crystallization. In physics, β corresponds to $1/T$, where T is temperature. Global energy minimization means a perfect crystal without defects.The observed crystals with defects in nature can be understood as the local minimum state.
- In order to obtain a perfect crystal, one may image the following process: The crystals will take the form of liquids in the high temperature, then one decreases the temperature very slowly until the perfect crystal forms at the zero temperature. This is the basic idea of simulated annealing.

Remark

- The construction of $\Pi^\beta(x)$ opens a new way to optimize $H(x)$ via stochastics.
- The theorem shows that if we can generate the random sequence with distribution $\Pi^\beta(x)$, then the random numbers will finally jump among the minimizers when $\beta = +\infty$. This procedure is called annealing.
- It corresponds to the physical crystallization. In physics, β corresponds to $1/T$, where T is temperature. Global energy minimization means a perfect crystal without defects.The observed crystals with defects in nature can be understood as the local minimum state.
- In order to obtain a perfect crystal, one may image the following process: The crystals will take the form of liquids in the high temperature, then one decreases the temperature very slowly until the perfect crystal forms at the zero temperature. This is the basic idea of simulated annealing.
- The random number generation with distribution $\Pi^\beta(x)$ can be created by Metropolis algorithm.

Theoretical results: Formulation

- Assuming the Metropolis sampler for simulated annealing is

$$P^\beta(x,y) = \begin{cases} G(x,y)\frac{\pi^\beta(y)}{\pi^\beta(x)}, & \pi^\beta(y) < \pi^\beta(x) \text{ and } x \neq y, \\ G(x,y), & \pi^\beta(y) \geq \pi^\beta(x) \text{ and } x \neq y, \\ 1 - \sum_{z \neq x} P^\beta(x,z) & x = y. \end{cases}$$

  where $G(x,y)$ is the proposal matrix. It is symmetric as before.

Theoretical results: Formulation

- Assuming the Metropolis sampler for simulated annealing is

$$P^\beta(x,y) = \begin{cases} G(x,y)\frac{\pi^\beta(y)}{\pi^\beta(x)}, & \pi^\beta(y) < \pi^\beta(x) \text{ and } x \neq y, \\ G(x,y), & \pi^\beta(y) \geq \pi^\beta(x) \text{ and } x \neq y, \\ 1 - \sum_{z\neq x} P^\beta(x,z) & x = y. \end{cases}$$

  where $G(x,y)$ is the proposal matrix. It is symmetric as before.
- In order to state the fundamental theorem of simulated annealing, we make the following definitions.

Theoretical results: Some Definitions

Definition (Neighborhood system)

The neighborhood system of x is defined as $N(x) = \{y \in X | x \neq y, G(x, y) > 0\}$.

Theoretical results: Some Definitions

Definition (Neighborhood system)

The neighborhood system of x is defined as $N(x) = \{y \in X | x \neq y, G(x, y) > 0\}$.

Definition (Communication length)

Given x and y, if there exists sequence $x = u_0, u_1, \ldots, u_{\sigma(x,y)} = y$ such that $u_{j+1} \in N(u_j)$ for any $j = 0, 1, \ldots, \sigma(x, y) - 1$, then we say that the states x and y communicate, where $\sigma(x, y)$ is the length of the shortest path along which x and y communicate.

Theoretical results: Some Definitions

Definition (Neighborhood system)

The neighborhood system of x is defined as $N(x) = \{y \in X | x \neq y, G(x,y) > 0\}$.

Definition (Communication length)

Given x and y, if there exists sequence $x = u_0, u_1, \ldots, u_{\sigma(x,y)} = y$ such that $u_{j+1} \in N(u_j)$ for any $j = 0, 1, \ldots, \sigma(x,y) - 1$, then we say that the states x and y communicate, where $\sigma(x,y)$ is the length of the shortest path along which x and y communicate.

Definition (Maximal local increase of energy)

The maximal local increase of energy is defined as

$$\Delta = \max\{H(y) - H(x) : x \in X, y \in N(x)\}.$$

Theoretical results

Theorem (Fundamental theorem of simulated annealing)

Suppose that X is a finite set, $H(x)$ is a nonconstant function, $G(x, y)$ is a symmetric irreducible proposal matrix,

$$\tau = \max\{\sigma(x, y) : x, y \in X\}.$$

If the annealing procedure is chosen such that $\beta(n) \leq \frac{1}{\tau\Delta} \ln n$, then for any initial distribution ν, we have

$$\lim_{n\to+\infty} \|\nu P^{\beta(1)} \cdots P^{\beta(n)} - \Pi^{\infty}\| = 0.$$

Remark

- The theorem shows that the annealing rate must be slow enough such that $\beta(n) \leq \frac{1}{\tau\Delta} \ln n$. It is a very very slow rate because $n \geq \exp(\tau\Delta\beta(n))$, we need $n \sim \exp(N_0)$ if $\beta(n) = N_0 \gg 1$.

Remark

- The theorem shows that the annealing rate must be slow enough such that $\beta(n) \leq \frac{1}{\tau\Delta} \ln n$. It is a very very slow rate because $n \geq \exp(\tau\Delta\beta(n))$, we need $n \sim \exp(N_0)$ if $\beta(n) = N_0 \gg 1$.
- This means high accuracy needs exponential computing time, which is impossible for realistic computation.

Remark

- The theorem shows that the annealing rate must be slow enough such that $\beta(n) \leq \frac{1}{\tau\Delta} \ln n$. It is a very very slow rate because $n \geq \exp(\tau\Delta\beta(n))$, we need $n \sim \exp(N_0)$ if $\beta(n) = N_0 \gg 1$.
- This means high accuracy needs exponential computing time, which is impossible for realistic computation.
- We should take more rapid annealing rates such as $\beta(n) \sim p^{-n}$ $(p \lesssim 1)$ or others. Of course, it has no theoretical foundations.

Table of Contents

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices
  - Take N very large — Huge computational effort;

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices
  - Take N very large — Huge computational effort;
  - Variance reduction techniques.

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices
  - Take N very large — Huge computational effort;
  - Variance reduction techniques.
- In the follows we will introduce the QMC to replace the pseudo-random sequence with quasi-random sequence.

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices
  - Take N very large — Huge computational effort;
  - Variance reduction techniques.
- In the follows we will introduce the QMC to replace the pseudo-random sequence with quasi-random sequence.
- It improves the convergence rate to $O((\ln N)^k N^{-1})$, where k depends on the space dimension.

Quasi-Monte Carlo Method

- The standard MC is of $O(\frac{\sigma}{\sqrt{N}})$. In order to improve the accuracy, one has two choices
  - Take N very large — Huge computational effort;
  - Variance reduction techniques.
- In the follows we will introduce the QMC to replace the pseudo-random sequence with quasi-random sequence.
- It improves the convergence rate to $O((\ln N)^k N^{-1})$, where k depends on the space dimension.
- Finally we will find that QMC is essentially a deterministic method which is very similar with MC.

Discrepancy

The concept of discrepancy is an estimate of the uniformity of the points.

- For N points $\{x_n\}_{n=1}^{N}$ belonging to the unit d-cube $I^d = [0,1]^d$, define

$$R_N(J) = \frac{1}{N}\#\{x_n \in J\} - m(J)$$

for any set $J \subset I^d$, where $\#\{x_n \in J\}$ means the number of the points in set J, and $m(J)$ is the measure of J.

Discrepancy

The concept of discrepancy is an estimate of the uniformity of the points.

- For N points $\{x_n\}_{n=1}^{N}$ belonging to the unit d-cube $I^d = [0,1]^d$, define

$$R_N(J) = \frac{1}{N}\#\{x_n \in J\} - m(J)$$

  for any set $J \subset I^d$, where $\#\{x_n \in J\}$ means the number of the points in set J, and $m(J)$ is the measure of J.
- Intuitively $R_N(J)$ is the difference between the exact volume and the random sampling estimate of the volume.

Rectangles

Definition (Rectangles)

Define the whole set of rectangles in I^d as

$$E = \{J(x, y) \ : \ (0, 0, \ldots, 0) \leq x \leq y \leq (1, 1, \ldots, 1)\},$$

where $x \leq y$ means $x_i \leq y_i$,$i = 1, \ldots, d$, $J(x, y)$ means the set of rectangles with the lower left node x and the upper right node y.
Define

$$E^* = \{J(0, y) \ : \ (0, 0, \ldots, 0) \leq y \leq (1, 1, \ldots, 1)\}.$$

Discrepancy: Definition

Definition (Discrepancy)

The L^∞-discrepancy of a sequence $\{x_n\}_{n=1}^N$ is defined as

$$D_N = \sup_{J \in E} |R_N(J)|;$$

Discrepancy: Definition

Definition (Discrepancy)

The L^∞-discrepancy of a sequence $\{x_n\}_{n=1}^N$ is defined as

$$D_N = \sup_{J \in E} |R_N(J)|;$$

and the L^2-discrepancy

$$T_N = \left(\int_{(x,y) \in I^{2d}, x \leq y} R_N(J(x,y))^2 dx dy \right)^{\frac{1}{2}}.$$

Discrepancy: Definition

Definition (Discrepancy)

The L^∞-discrepancy of a sequence $\{x_n\}_{n=1}^N$ is defined as

$$D_N = \sup_{J \in E} |R_N(J)|;$$

and the L^2-discrepancy

$$T_N = \Big(\int_{(x,y)\in I^{2d}, x \le y} R_N(J(x,y))^2 dx dy \Big)^{\frac{1}{2}}.$$

The L^p-discrepancy can be defined similarly. Specially we define the discrepancy

$$D_N^* = \sup_{J \in E^*} |R_N(J)|,$$

$$T_N^* = \left(\int_{I^d} R_N(J(0,x))^2 dx \right)^{\frac{1}{2}}.$$

Total variation

The total variation of function

- In 1D case, the total variation of a function is defined as the sum of the jumps:

  $$V[f] = \sup_{\tau} \sum_i |f(x_{i+1}) - f(x_i)|,$$

  where τ is taken to all the possible partitions of the domain. If f is differentiable, then

  $$V[f] = \int_0^1 |df| = \int_0^1 |f'(x)| dx.$$

Total variation

The total variation of function

- In 1D case, the total variation of a function is defined as the sum of the jumps:

$$V[f] = \sup_{\tau} \sum_i |f(x_{i+1}) - f(x_i)|,$$

where τ is taken to all the possible partitions of the domain. If f is differentiable, then

$$V[f] = \int_0^1 |df| = \int_0^1 |f'(x)| dx.$$

- The total variation of function f in unit d-cube $[0,1]^d$ is defined as

$$V[f] = \int_{I^d} \left| \frac{\partial^d f}{\partial x_1 \cdots \partial x_d} \right| dx_1 \cdots dx_d + \sum_{i=1}^{d} V[f_1^{(i)}],$$

where $f_1^{(i)}$ is the restriction of f on the boundary $x_i = 1$. It is a recursive definition of total variation.

Koksma-Hlawka Theorem

Theorem (Koksma-Hlawka Theorem)

For any sequence $\{x_n\}_{n=1}^N \subset I^d$ and the function f with bounded variation in I^d, the integration error $\mathcal{E}$ obeys the following inequality

$$\mathcal{E}[f] \leq V[f] D_N^*,$$

where $\mathcal{E}[f] \triangleq |I[f] - I_N[f]| = |\int_{I^d} f(x)dx - \frac{1}{N}\sum_{i=1}^N f(x_i)|$.

Koksma-Hlawka Theorem: Intuitive Proof

Intuitive Proof. For the function $f(x)$ which takes value 0 on the boundary of I^d, define

$$R(x) = R_N(J(0, x)).$$

Koksma-Hlawka Theorem: Intuitive Proof

Intuitive Proof. For the function $f(x)$ which takes value 0 on the boundary of I^d, define

$$R(x) = R_N(J(0,x)).$$

Then

$$dR(x) = \{\frac{1}{N}\sum_{i=1}^{N}\delta(x - x_i) - 1\}dx,$$

where $dR = \frac{\partial^d R}{\partial x_1 \cdots \partial x_d}, \ dx = dx_1 \cdots dx_d$.

Koksma-Hlawka Theorem: Intuitive Proof

So we have

$$
\begin{aligned}
\mathcal{E}[f] &= \left| \int_{I^d} f(x)dx - \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right| \\
&= \left| \int_{I^d} \{1 - \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i)\} f(x) dx \right| \\
&= \left| \int_{I^d} R(x) df(x) \right| \\
&\leq (\sup_x R(x)) \int_{I^d} |df(x)| = D_N^* V[f].
\end{aligned}
$$

$\square$

Koksma-Hlawka Theorem: Implication

- Koksma-Hlawka theorem shows that the discretization error can be described by the total variation $V[f]$ and the discrepancy for the sample points.

Koksma-Hlawka Theorem: Implication

- Koksma-Hlawka theorem shows that the discretization error can be described by the total variation $V[f]$ and the discrepancy for the sample points.
- QMC gives some special quasi random sequences which have good discrepancy properties. It is a pure number theoretic result.

Quasi-random Sequences

Definition

A sequence $\{x_n\}_{n=1}^{N} \subset I^d$ is called quasi-random if

$$D_N \leq C(\ln N)^k N^{-1},$$

in which c and k are constants that are independent of N, but may depend on the dimension d.

Quasi-random Sequences

Definition

A sequence $\{x_n\}_{n=1}^{N} \subset I^d$ is called quasi-random if

$$D_N \leq C(\ln N)^k N^{-1},$$

in which c and k are constants that are independent of N, but may depend on the dimension d.

- Van der Corput sequence $(d = 1)$:
  The generation of sequence $\{x_i\}_{i=1}^{N}$ is composed of two steps:
  Step1. Write out n in base 2:

  $$n = \Big(a_m a_{m-1} \cdots a_1 a_0\Big)_2,$$

  where $(\cdot)_2$ means in base 2, $a_i \in \{0, 1\}$ is the i-th bit of n;
  Step2. Generate x_n in base 2

  $$x_n = \Big(0.\, a_0 a_1 \cdots a_m\Big)_2.$$

Quasi-random Sequences

- Halton sequence ($d > 1$):
  Denote $x_n = (x_n^1, x_n^2, \ldots, x_n^d)$, where the k-th component x_n^k is obtained by two steps.
  Step1. Write out n in base p_k. (where p_k is the k-th prime number, e.g. $p_1 = 2, p_2 = 3$)

$$n = \left(a_{m_k}^k a_{m_k-1}^k \cdots a_1^k a_0^k\right)_{p_k};$$

  Step2. Generate x_n^k in base p_k:

$$x_n^k = \left(0.\, a_0^k a_1^k \cdots a_{m_k}^k\right)_{p_k}.$$

The number theorists has proved

$$D_N(\text{Halton}) \leq C_d (\ln N)^d N^{-1}.$$

Limitations of QMC

QMC has the following limitations:

- QMC are designed for integration and are not directly applicable to simulations. This is because of the correlations between the points of a quasi-random sequence.

Limitations of QMC

QMC has the following limitations:

- QMC are designed for integration and are not directly applicable to simulations. This is because of the correlations between the points of a quasi-random sequence.
- Because the theoretical basis of QMC is from Koksma-Hlawka theorem, and the generation style of quasi-random numbers is very special, it is commonly applied to the integral in rectangle with the form $\int_{I^d} f(x)dx$. For the powerful MCMC method, how to design the corresponding QMC version is not clear.

Limitations of QMC

QMC has the following limitations:

- QMC are designed for integration and are not directly applicable to simulations. This is because of the correlations between the points of a quasi-random sequence.
- Because the theoretical basis of QMC is from Koksma-Hlawka theorem, and the generation style of quasi-random numbers is very special, it is commonly applied to the integral in rectangle with the form $\int_{I^d} f(x)dx$. For the powerful MCMC method, how to design the corresponding QMC version is not clear.
- QMC is found to lose its effectiveness when the dimension of the integral becomes large. This can be anticipated from the bound $(\ln N)^d N^{-1}$ on discrepancy. For large dimension d, this bound is dominated by the $(\ln N)^d$ term unless $N > e^d$.

Limitations of QMC

QMC has the following limitations:

- QMC are designed for integration and are not directly applicable to simulations. This is because of the correlations between the points of a quasi-random sequence.
- Because the theoretical basis of QMC is from Koksma-Hlawka theorem, and the generation style of quasi-random numbers is very special, it is commonly applied to the integral in rectangle with the form $\int_{I^d} f(x)dx$. For the powerful MCMC method, how to design the corresponding QMC version is not clear.
- QMC is found to lose its effectiveness when the dimension of the integral becomes large. This can be anticipated from the bound $(\ln N)^d N^{-1}$ on discrepancy. For large dimension d, this bound is dominated by the $(\ln N)^d$ term unless $N > e^d$.
- QMC is found to lose its effectiveness if the integrand f is not smooth. The factor $V[f]$ in the Koksma-Hlawka inequality is an indicator of this dependence.

Limitations of QMC

All in all:

- QMC is suitable for the integration in which the space dimension is not so big, the integrand f is relatively smooth.

Limitations of QMC

All in all:

- QMC is suitable for the integration in which the space dimension is not so big, the integrand f is relatively smooth.
- Though it has better convergence rate than Monte Carlo method, its applicability is limited.