

Article

Not peer-reviewed version

A Measure-Theoretic Formulation of Monte Carlo Stochastic Optimization: Unifying Continuous and Discrete Domains

[Shin-Ichi Inage](#)*

Posted Date: 2 April 2026

doi: 10.20944/preprints202604.0083.v1

Keywords: measure-based optimization; Monte Carlo stochastic optimization (MOST); discrete optimization; measure-theoretic formulation; global optimization; region-based search



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Measure-Theoretic Formulation of Monte Carlo Stochastic Optimization: Unifying Continuous and Discrete Domains

Shin-Ichi Inage

Ishinomaki Senshu University, 1 Shin-mito, Minami-sakai, Ishinomaki-shi, Miyagi-ken, 986-8580, JAPAN;
s.inage.pu@isenshu-u.ac.jp

Abstract

This study presents a unified measure-theoretic formulation of the Monte Carlo Stochastic Optimization Technique (MOST), establishing a rigorous framework that encompasses both continuous and discrete optimization. Unlike conventional optimization methods that operate on pointwise evaluations, MOST is based on regional evaluation through normalized integrals, enabling robust and global exploration of the search space. We first reformulate MOST within a finite measure space, where the evaluation of a region is defined as the measure-weighted average of the objective function. This formulation naturally connects regional optimization with expectation under an induced probability measure and provides a theoretical foundation for Monte Carlo approximation. Building upon this framework, we construct a discrete version of MOST by introducing the counting measure and extend it further using weighted measures to rigorously handle odd-cardinality partitions via midpoint sharing. A central contribution of this work is the demonstration that continuous and discrete MOST are structurally identical algorithms arising from a single measure-based principle, differing only in the choice of underlying measure. This result eliminates the traditional separation between continuous and discrete optimization within the MOST framework. Theoretical analysis reveals that MOST is particularly effective when near-optimal regions possess non-negligible measure, while its performance may degrade in the presence of isolated global minima. These properties are validated through numerical experiments using benchmark functions, including the Ackley and Sphere functions, under uniform discretization. The results confirm that discrete MOST achieves accurate approximations of global optima, with errors controlled by discretization resolution and strong robustness in multimodal landscapes. Overall, this work establishes MOST as a measure-based optimization paradigm, offering a unified, theoretically grounded, and practically robust approach to global optimization across continuous and discrete domains.

Keywords: measure-based optimization; Monte Carlo stochastic optimization (MOST); discrete optimization; measure-theoretic formulation; global optimization; region-based search

1. Introduction

Optimization constitutes a fundamental component across a wide range of scientific and engineering disciplines, including mechanical design, control systems, machine learning, and operations research. Classical deterministic optimization methods—such as gradient descent, quasi-Newton methods, trust-region frameworks, and interior-point techniques—have demonstrated strong theoretical guarantees in convex and smooth settings [1–4]. However, their applicability

becomes limited when addressing nonconvex, multimodal, discontinuous, or nonsmooth optimization problems, which are frequently encountered in real-world applications [5–7].

To overcome these limitations, numerous stochastic and population-based algorithms have been developed, including Genetic Algorithms (GA) [8], Differential Evolution (DE) [9], Particle Swarm Optimization (PSO) [10], and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [11]. These methods exhibit robustness against multimodality and do not require gradient information. Nevertheless, they rely heavily on stochastic exploration, and their convergence properties are typically probabilistic or asymptotic, lacking deterministic guarantees in general [12–14].

In parallel, surrogate-based optimization methods such as Bayesian optimization have gained attention for expensive black-box problems [15,16]. Although these approaches provide probabilistic convergence guarantees under specific assumptions, their performance depends strongly on model accuracy and deteriorates in high-dimensional or irregular search spaces [17]. Deterministic global optimization methods, including branch-and-bound and Lipschitz-based algorithms [18–20], offer rigorous guarantees but often suffer from exponential computational complexity and scalability issues.

Multi-objective optimization further increases complexity. Evolutionary algorithms such as NSGA-II [21] and SPEA2 [22] are widely used to approximate Pareto fronts, yet they provide only discrete approximations and lack deterministic convergence guarantees. Moreover, classical scalarization approaches fail to capture non-supported Pareto solutions in nonconvex problems [23–25].

Against this backdrop, Inage and Hebishima introduced a novel optimization paradigm known as the Monte Carlo Stochastic Optimization Technique (MOST) [26,27]. Unlike conventional pointwise optimization methods, MOST is based on a region-wise evaluation principle: the search domain is recursively partitioned, and each subregion is evaluated via Monte Carlo integration. The subregion with the smallest integral value is selected, and this process is repeated iteratively. This region-based strategy introduces an intrinsic smoothing effect over the objective landscape, enabling robust global exploration and deterministic geometric contraction of the search domain. Numerical studies have demonstrated that MOST achieves higher accuracy and faster convergence than genetic algorithms and gradient-based methods on benchmark problems such as the Ackley and Schwefel functions [26].

Subsequent extensions have generalized MOST to multi-objective optimization problems, where convergence toward Pareto-consistent solutions has been demonstrated [27]. More recently, a rigorous deterministic–probabilistic framework has been established, including convergence proofs, connections to Karush–Kuhn–Tucker (KKT) conditions, and extensions to constrained optimization [28]. These developments position MOST as a new class of derivative-free global optimization methods combining deterministic structure and stochastic sampling.

Despite these advances, an important theoretical gap remains. Existing formulations of MOST have been developed primarily for continuous domains, implicitly relying on Lebesgue integration. However, many practical optimization problems—particularly in engineering design, combinatorial optimization, and digital control—are inherently discrete. Extending MOST to discrete domains is not a trivial matter, as the concept of integration must be carefully reinterpreted, and the algorithmic structure must be adapted accordingly.

In particular, the core mechanism of MOST—region selection based on integral comparison—raises a fundamental question: Can the integral-based framework be rigorously extended to discrete optimization problems in a mathematically consistent manner? This question naturally leads to a measure-theoretic perspective, in which continuous and discrete optimization can be treated within a unified framework.

The key idea of this study is to reinterpret MOST as a measure-based optimization method, where region evaluation is defined through a normalized integral with respect to an underlying measure. In continuous domains, this corresponds to the Lebesgue measure, whereas in discrete domains, it corresponds to the counting measure. Furthermore, in the case of odd-number

partitioning in discrete spaces, we introduce a weighted counting measure to consistently treat shared boundary points.

Based on this perspective, we develop a unified theoretical framework that encompasses both continuous and discrete MOST. In the discrete setting, we establish a rigorous algorithmic construction, including a novel bisection strategy for even and odd cardinalities, and a consistent Monte Carlo sampling scheme based on discrete probability measures. This formulation enables a direct extension of the integral-based selection principle to discrete spaces without loss of theoretical consistency.

The contributions of this study are summarized as follows:

1. A measure-theoretic reformulation of MOST that provides a unified mathematical framework for region-based optimization.
2. A rigorous construction of discrete MOST using counting measures and weighted measures.
3. A theoretical formulation of bisection strategies for both even and odd cardinalities in discrete domains.
4. A unified theory showing that continuous and discrete MOST are special cases of a single measure-based optimization framework.
5. A clarification of the theoretical limitations of MOST, particularly in the presence of highly localized optima.
6. Numerical validation through benchmark functions (e.g., Ackley function), comparing theoretical solutions and discrete MOST solutions under various discretization resolutions.

Through these developments, this study establishes MOST as a **measure-based optimization paradigm** that bridges continuous and discrete domains within a single rigorous framework. This unified perspective not only deepens the theoretical understanding of MOST but also expands its applicability to a broader class of optimization problems.

Chapter 2. Fundamental Concept of MOST in Continuous Domains

2.1. Problem Setting

We consider the unconstrained minimization problem defined over a bounded continuous domain:

$$\min_{x \in \Omega} f(x), \quad (1)$$

where $\Omega \subset \mathbb{R}^n$ is a compact hyper-rectangular domain, and $f: \Omega \rightarrow \mathbb{R}$ is a measurable objective function. Classical optimization methods evaluate the objective function at specific points and iteratively update candidate solutions based on gradient information or stochastic sampling [1–4,7]. In contrast, the Monte Carlo Stochastic Optimization Technique (MOST) adopts a fundamentally different viewpoint: it evaluates **regions** rather than individual points.

2.2. Core Idea: Region-Based Optimization via Integral Comparison

The central idea of MOST is to replace pointwise evaluation with integral-based evaluation over subdomains.

For any measurable subset $A \subset \Omega$ with nonzero measure, define the regional evaluation functional:

$$J(A) = \frac{1}{|A|} \int_A f(x) dx, \quad (2)$$

where $|A|$ denotes the Lebesgue measure of A . This quantity represents the average value of the objective function over the region A .

The key principle of MOST is: Regions containing lower objective values tend to exhibit smaller average values. Thus, instead of directly searching for a point minimizing $f(x)$, MOST iteratively identifies subregions with smaller values of $J(A)$.

2.3. Binary Partitioning of the Search Domain

Let the initial search domain be denoted by

$$\Omega^{(0)} = \Omega. \quad (3)$$

At iteration k , the current region $\Omega^{(k)}$ is partitioned into two subregions along a selected coordinate direction. Without loss of generality, suppose that the partition is performed along the j -th coordinate:

$$\Omega^{(k)} = \Omega_L^{(k)} \cup \Omega_R^{(k)}, \Omega_L^{(k)} \cap \Omega_R^{(k)} = \emptyset. \quad (4)$$

Each subregion satisfies

$$|\Omega_L^{(k)}| = |\Omega_R^{(k)}| = \frac{1}{2} |\Omega^{(k)}|. \quad (5)$$

his binary partitioning ensures a systematic reduction of the search domain.

2.4. Monte Carlo Approximation of Regional Integrals

In practical implementations, the integral in (2) is evaluated using Monte Carlo sampling. Let $\{x_i\}_{i=1}^M$ be independent and identically distributed samples drawn uniformly from the region A :

$$x_i \sim \text{Uniform}(A). \quad (6)$$

Then the Monte Carlo estimator of $J(A)$ is given by:

$$\hat{J}_M(A) = \frac{1}{M} \sum_{i=1}^M f(x_i). \quad (7)$$

By the law of large numbers [13,14], we have:

$$\hat{J}_M(A) \rightarrow J(A) \text{ as } M \rightarrow \infty. \quad (8)$$

Moreover, concentration inequalities provide probabilistic error bounds:

$$\mathbb{P}(|\hat{J}_M(A) - J(A)| > \varepsilon) \leq C \exp(-cM\varepsilon^2), \quad (9)$$

for some constants $C, c > 0$ depending on the boundedness of f [14].

2.5. Region Selection Rule

At each iteration, MOST compares the estimated average values of the two subregions:

$$\hat{J}_M(\Omega_L^{(k)}), \hat{J}_M(\Omega_R^{(k)}). \quad (10)$$

The next search region is selected as:

$$\Omega^{(k+1)} = \begin{cases} \Omega_L^{(k)}, & \text{if } \hat{J}_M(\Omega_L^{(k)}) \leq \hat{J}_M(\Omega_R^{(k)}), \\ \Omega_R^{(k)}, & \text{otherwise.} \end{cases} \quad (11)$$

This deterministic selection rule forms the core of MOST.

2.6. Deterministic Shrinking of the Search Region

Since each iteration halves the domain along one coordinate, the diameter of the search region satisfies:

$$\text{diam}(\Omega^{(k+1)}) \leq \frac{1}{2} \text{diam}(\Omega^{(k)}). \quad (12)$$

Recursively,

$$\text{diam}(\Omega^{(k)}) \leq 2^{-k} \text{diam}(\Omega). \quad (13)$$

Thus, the search region shrinks geometrically, independent of the objective function. This property distinguishes MOST from both gradient-based and stochastic optimization methods, which do not guarantee deterministic contraction [1,7–11].

2.7. Integral Averaging Effect

A crucial feature of MOST is the smoothing effect of integration. Let $x^* \in \Omega$ be a global minimizer of f .

For a sufficiently small region A containing x^* , we have:

$$J(A) = f(x^*) + O(\text{diam}(A)). \quad (14)$$

Thus, as the region shrinks,

$$J(A) \rightarrow f(x^*). \quad (15)$$

This implies:

- Narrow local minima contribute negligibly to the integral
- Regions containing the global minimum dominate asymptotically

This mechanism explains the robustness of MOST against multimodal objective functions, as observed in benchmark studies [26,27].

2.8. Comparison with Classical Optimization Methods

MOST differs fundamentally from existing optimization frameworks:

- **Gradient-based methods** [1–4]

Depend on differentiability and are sensitive to local minima

- **Evolutionary algorithms** [8–12]

Provide stochastic exploration but lack deterministic convergence

- **Deterministic global optimization methods** [18–20]

Require Lipschitz constants or bounding functions

- **Bayesian optimization** [15–17]

Depends on surrogate models and suffers in high dimensions

In contrast, MOST:

- Does not require gradients
- Does not rely on surrogate models
- Provides deterministic region shrinking
- Uses integral averaging to mitigate local irregularities

2.9. Summary

In this chapter, we have revisited the fundamental concept of MOST in continuous domains.

The key elements of the method are:

1. Region-based evaluation using average integrals (2)
2. Recursive binary partitioning of the search domain (4)–(5)
3. Monte Carlo approximation of regional integrals (7)–(9)
4. Deterministic region selection based on integral comparison (11)
5. Geometric shrinking of the search domain (12)–(13)

These properties establish MOST as a deterministic, derivative-free optimization framework that leverages measure-based averaging to achieve robustness against multimodality.

This continuous formulation serves as the foundation for the subsequent development of a measure-theoretic framework and its extension to discrete optimization domains in the following chapters.

Chapter 3. Measure-Theoretic Reformulation of MOST

This chapter establishes a measure-theoretic foundation for the Monte Carlo Stochastic Optimization Technique (MOST). By reformulating MOST within the framework of finite measure spaces, we provide a unified mathematical structure that encompasses both continuous and discrete optimization problems. This perspective reveals that MOST is not a pointwise optimization method, but rather a measure-based region selection mechanism.

3.1. Measure Space Formulation

To generalize the MOST framework, we introduce a finite measure space

$$(X, \mathcal{F}, \mu), \quad (16)$$

where X denotes the search domain, $\mathcal{F}$ is a σ -algebra of measurable subsets of X , and μ is a finite measure on $(X, \mathcal{F})$.

Let $f: X \rightarrow \mathbb{R}$ be a measurable objective function.

This formulation enables a unified treatment of optimization problems:

- Continuous domains: μ is the Lebesgue measure
- Discrete domains: μ is the counting measure

Thus, both continuous and discrete optimization problems are embedded into a single mathematical framework.

3.2. Measure-Based Evaluation Functional

Within this setting, we define the MOST evaluation functional as the normalized integral over a measurable region $A \in \mathcal{F}$ with $\mu(A) > 0$:

$$J_f(A) = \frac{1}{\mu(A)} \int_A f(x) d\mu(x). \quad (17)$$

This quantity represents the measure-weighted average of the objective function over the region A . A fundamental reinterpretation of MOST follows immediately: MOST does not minimize the objective function pointwise; instead, it selects regions that minimize the measure-based average of the objective function.

This shift from pointwise evaluation to regional averaging constitutes the conceptual core of the method.

3.3. Correspondence to Probability Measures

The normalized measure naturally induces a probability measure supported on A . Specifically, for any $A \in \mathcal{F}$ with $\mu(A) > 0$, define

$$P_A(B) = \frac{\mu(B \cap A)}{\mu(A)}, B \in \mathcal{F}. \quad (18)$$

Then P_A is a probability measure on $(X, \mathcal{F})$, and the functional $J_f(A)$ admits the probabilistic representation

$$J_f(A) = \int_X f(x) dP_A(x) = \mathbb{E}_{P_A}[f(x)]. \quad (19)$$

Thus, the regional evaluation in MOST is equivalent to computing an expectation under a probability measure induced by the underlying measure μ .

This interpretation provides a direct connection between MOST and stochastic approximation theory [13,14].

3.4. Monte Carlo Approximation

In practical implementations, the expectation (19) is approximated via Monte Carlo sampling. Let $X_1, X_2, \dots, X_M$ be independent and identically distributed random variables drawn according to P_A :

$$X_i \sim P_A. \quad (20)$$

The Monte Carlo estimator of $J_f(A)$ is given by

$$\hat{J}_M(A) = \frac{1}{M} \sum_{i=1}^M f(X_i). \quad (21)$$

By the strong law of large numbers [13], we have

$$\hat{J}_M(A) \rightarrow J_f(A) \text{ almost surely as } M \rightarrow \infty. \quad (22)$$

Moreover, concentration inequalities provide finite-sample guarantees:

$$\mathbb{P}(|\hat{J}_M(A) - J_f(A)| > \varepsilon) \leq C \exp(-cM\varepsilon^2), \quad (23)$$

for some constants $C, c > 0$ depending on the boundedness of f [14].

3.5. Fundamental Theoretical Properties

Theorem 3.1 (Consistency of Monte Carlo Estimation)

Let $f \in L^1(A, \mu)$. Then the estimator $\hat{J}_M(A)$ satisfies

$$\hat{J}_M(A) \rightarrow J_f(A) \text{ almost surely as } M \rightarrow \infty. \quad (24)$$

Proof.

This follows directly from the strong law of large numbers applied to the i.i.d (independent and identically distributed). sequence $\{f(X_i)\}_{i=1}^M$ [13].

Theorem 3.2 (Measure Dependence of MOST Evaluation)

The functional $J_f(A)$ depends explicitly on the underlying measure μ . In particular, if $\mu_1 \neq \mu_2$, then in general

$$J_f^{(\mu_1)}(A) \neq J_f^{(\mu_2)}(A). \quad (25)$$

Proof.

From definition (17), $J_f(A)$ depends on both the integral and the normalization factor $\mu(A)$. Since both quantities are determined by the measure, different measures yield different evaluations in general.

3.6. Interpretation and Structural Insights

The measure-theoretic formulation provides several fundamental insights into the nature of MOST:

1. Measure-Based Optimization

MOST optimizes regions with respect to a measure-weighted objective, rather than optimizing points directly.

2. Expectation-Based Evaluation

The regional evaluation is equivalent to an expectation under an induced probability measure.

3. Intrinsic Smoothing Effect

The integral-based formulation suppresses narrow local fluctuations of the objective function, emphasizing global structure.

4. Unified Framework

The same formulation applies to:

- Continuous spaces (Lebesgue measure)
- Discrete spaces (counting measure)
- Hybrid spaces (product measures)

This unified viewpoint reveals that the distinction between continuous and discrete optimization is not structural, but purely a consequence of the choice of measure.

3.7. Summary

In this chapter, we have established a rigorous measure-theoretic formulation of MOST. The principal results are:

1. The definition of the MOST evaluation functional as a normalized integral (17).
2. The equivalence between regional evaluation and expectation under a probability measure (19).
3. The Monte Carlo approximation framework and its almost sure convergence (21)–(24).
4. The explicit dependence of MOST on the underlying measure (25).

These results demonstrate that MOST is fundamentally a measure-based optimization framework, providing a unified theoretical foundation for both continuous and discrete optimization.

This formulation serves as the basis for the construction of discrete MOST in the next chapter.

Chapter 4. Discrete MOST: Construction and Algorithm

This chapter develops a rigorous formulation of the Monte Carlo Stochastic Optimization Technique (MOST) in discrete domains. Building upon the measure-theoretic framework established in Chapter 3, we construct a discrete counterpart that preserves the essential structure of MOST while ensuring algorithmic consistency and reproducibility.

The central objective is to demonstrate that discrete MOST is not an ad hoc extension, but rather a natural consequence of replacing the Lebesgue measure with the counting measure.

4.1. Discrete Search Space and Counting Measure

Let the search space be a finite ordered set:

$$D = \{x_1, x_2, \dots, x_N\}, x_1 < x_2 < \dots < x_N. \quad (26)$$

We equip D with the counting measure ν , defined by

$$\nu(A) = |A|, A \subset D. \quad (27)$$

Under this measure, the integral of a function $f: D \rightarrow \mathbb{R}$ reduces to a finite sum:

$$\int_A f(x) d\nu(x) = \sum_{x \in A} f(x). \quad (28)$$

Accordingly, the MOST evaluation functional becomes

$$J_f(A) = \frac{1}{|A|} \sum_{x \in A} f(x), \quad (29)$$

which corresponds to the arithmetic mean over the subset A . Thus, discrete MOST is obtained directly from the general formulation (17) by choosing $\mu = \nu$.

4.2. Bisection Strategy: Even and Odd Cardinalities

A key component of MOST is recursive domain partitioning. In discrete spaces, this requires careful treatment to preserve symmetry and consistency.

4.2.1. Even Cardinality

If $N = 2m$, the set D is partitioned into two disjoint subsets:

$$D_L = \{x_1, \dots, x_m\}, D_R = \{x_{m+1}, \dots, x_{2m}\}. \quad (30)$$

These subsets satisfy:

$$|D_L| = |D_R| = m. \quad (31)$$

This corresponds to a direct discrete analogue of continuous bisection.

4.2.2. Odd Cardinality

If $N = 2m + 1$, a symmetric partition is achieved by sharing the midpoint:

$$D_L = \{x_1, \dots, x_m, x_{m+1}\}, D_R = \{x_{m+1}, x_{m+2}, \dots, x_{2m+1}\}. \quad (32)$$

To ensure consistency with the measure-theoretic formulation, the midpoint x_{m+1} is assigned half weight in each subset. Define weight functions $w_L, w_R: D \rightarrow [0,1]$ as:

$$w_L(x_i) = \begin{cases} 1, & i \leq m, \\ \frac{1}{2}, & i = m + 1, \\ 0, & i \geq m + 2, \end{cases} w_R(x_i) = \begin{cases} 0, & i \leq m, \\ \frac{1}{2}, & i = m + 1, \\ 1, & i \geq m + 2. \end{cases} \quad (33)$$

The corresponding evaluation functionals are:

$$J_L = \frac{\sum_{x \in D} w_L(x) f(x)}{\sum_{x \in D} w_L(x)}, J_R = \frac{\sum_{x \in D} w_R(x) f(x)}{\sum_{x \in D} w_R(x)}. \quad (34)$$

This construction ensures:

- Symmetry of partitioning
- Equal effective measure
- Consistency with continuous domain splitting

4.3. Monte Carlo Sampling in Discrete MOST

A crucial distinction between continuous and discrete MOST lies in the sampling mechanism.

4.3.1. Continuous vs. Discrete Sampling

In continuous domains, samples are drawn from a uniform distribution with respect to the Lebesgue measure. In contrast, in discrete domains, sampling is performed over a finite set.

4.3.2. Uniform Discrete Sampling

For a subset $A \subset D$, define the discrete uniform distribution:

$$\mathbb{P}(X = x_i) = \frac{1}{|A|}, x_i \in A. \quad (35)$$

This corresponds exactly to the probability measure induced by the counting measure.

4.3.3. Weighted Sampling for Odd Partition

In the case of odd partitioning, the midpoint receives half weight. The corresponding sampling distribution becomes:

$$\mathbb{P}(X = x_i) = \frac{w(x_i)}{\sum_{x \in A} w(x)}, \quad (36)$$

where w is either w_L or w_R .

4.3.4. Monte Carlo Estimation

The evaluation functional is approximated by:

$$\hat{J}_M(A) = \frac{1}{M} \sum_{i=1}^M f(X_i), X_i \sim P_A. \quad (37)$$

This estimator remains consistent with the measure-theoretic formulation in Chapter 3.

4.4. Discrete MOST Algorithm

We now define the complete discrete MOST procedure.

Algorithm (Discrete MOST)

Input: Finite set $D^{(0)} = D$, sample size M

For $k = 0, 1, 2, \dots$:

1. Partition $D^{(k)}$ into $D_L^{(k)}, D_R^{(k)}$
2. Estimate:

$$\hat{J}_M(D_L^{(k)}), \hat{J}_M(D_R^{(k)}) \quad (38)$$

3. Select:

$$D^{(k+1)} = \begin{cases} D_L^{(k)}, & \text{if } \hat{J}_M(D_L^{(k)}) \leq \hat{J}_M(D_R^{(k)}), \\ D_R^{(k)}, & \text{otherwise} \end{cases} \quad (39)$$

Stop when:

$$|D^{(k)}| = 1. \quad (40)$$

4.5. Fundamental Properties

Theorem 4.1 (Finite Termination)

The discrete MOST algorithm terminates in a finite number of steps.

$$\exists K < \infty \text{ such that } |D^{(K)}| = 1. \quad (41)$$

Proof.

At each iteration, the cardinality satisfies:

$$|D^{(k+1)}| \leq \left\lceil \frac{|D^{(k)}|}{2} \right\rceil. \quad (42)$$

Thus, $|D^{(k)}|$ is a strictly decreasing sequence of positive integers, implying finite termination.

Theorem 4.2 (Nestedness)

The sequence of sets satisfies:

$$D^{(0)} \supset D^{(1)} \supset \dots \supset D^{(k)}. \quad (43)$$

Proof.

Immediate from the selection rule (39).

Theorem 4.3 (Correctness under Identifiability Condition)

Let x^* be the global minimizer. If for all k ,

$$J_f(D_k^*) < J_f(D_k^{\text{other}}), \quad (44)$$

then discrete MOST converges to x^* .

Proof.

By induction using (39).

4.6. Summary

In this chapter, we have constructed a rigorous formulation of discrete MOST. The key contributions are:

1. Definition of discrete MOST via counting measure (29)
2. Symmetric bisection strategies for even and odd cases (30)–(34)
3. Discrete Monte Carlo sampling consistent with measure theory (35)–(37)
4. A complete recursive optimization algorithm (38)–(40)
5. Fundamental convergence properties (41)–(44)

This formulation establishes discrete MOST as a natural extension of the measure-theoretic MOST framework, rather than an independent algorithm.

Chapter 5. Unified Theory of Continuous and Discrete MOST

This chapter presents the central theoretical contribution of the present study: a unified formulation of continuous and discrete MOST within a single measure-theoretic framework. The key observation is that the distinction between continuous and discrete optimization does not arise from the algorithmic structure itself, but solely from the underlying measure used to evaluate regions. In this sense, continuous and discrete MOST are not different methods, but different realizations of the same abstract optimization principle.

5.1. Main Unification Theorem: Lebesgue versus Counting Measure

The measure-theoretic formulation established in Chapter 3 immediately suggests that continuous and discrete MOST can be treated within a common mathematical framework. We now state this principle formally.

Theorem 5.1 (Unified Measure-Theoretic Formulation of MOST)

Let $(X, \mathcal{F}, \mu)$ be a finite measure space, and let $f: X \rightarrow \mathbb{R}$ be a measurable objective function. For each measurable region $A \in \mathcal{F}$ with $\mu(A) > 0$, define the MOST evaluation functional by

$$J_f(A) = \frac{1}{\mu(A)} \int_A f d\mu. \quad (45)$$

Then the regional selection mechanism of MOST is fully determined by the pair (f, μ) , and both continuous and discrete MOST arise as special cases of the same abstract construction:

1. **Continuous MOST** is obtained when $X \subset \mathbb{R}^n$ and μ is the Lebesgue measure λ , in which case

$$J_f(A) = \frac{1}{\lambda(A)} \int_A f(x) dx. \quad (46)$$
2. **Discrete MOST** is obtained when $X = D$ is a finite set and μ is the counting measure ν , in which case

$$J_f(A) = \frac{1}{\nu(A)} \int_A f d\nu = \frac{1}{|A|} \sum_{x \in A} f(x). \quad (47)$$

Hence, continuous and discrete MOST are instances of a single measure-based optimization framework.

Proof.

Equation (45) is well defined on any finite measure space. When $\mu = \lambda$, the integral is the usual Lebesgue integral on a measurable subset of $\mathbb{R}^n$, yielding (46). When $\mu = \nu$ is the counting measure on a finite set, the integral reduces to a finite sum, yielding (47). Since the regional evaluation and the resulting selection rule depend only on $J_f(A)$, the algorithmic structure is identical in both cases. Therefore, both continuous and discrete MOST are special realizations of the same abstract optimization scheme.

The significance of Theorem 5.1 is conceptual as well as technical. In conventional optimization theory, continuous and discrete problems are often treated as fundamentally different classes, requiring distinct methods and convergence analyses [1–7]. By contrast, the present formulation shows that, within the MOST framework, both classes can be derived from the same variational principle once the appropriate measure is specified.

This observation also clarifies the role of Monte Carlo sampling. In continuous domains, samples are drawn according to the normalized Lebesgue measure; in discrete domains, samples are drawn according to the normalized counting measure. Thus, the difference between the two settings is not algorithmic, but measure-theoretic.

5.2. Odd Partitioning as a Weighted Measure

The preceding theorem establishes the equivalence between continuous and discrete MOST at the level of ordinary measures. However, one subtle issue remains: the treatment of odd-cardinality partitions in discrete domains.

In continuous domains, a bisection at a midpoint generates two subdomains whose boundary overlap has Lebesgue measure zero. Therefore, whether the midpoint is assigned to the left or right subdomain is immaterial from the viewpoint of integration. In discrete domains, by contrast, a midpoint corresponds to an actual element of the finite set and therefore possesses positive counting measure. Consequently, a naive partition of an odd-cardinality set would introduce an artificial asymmetry.

To resolve this difficulty, we interpret odd partitioning through a weighted measure. Let

$$D = \{x_1, \dots, x_{2m+1}\} \quad (48)$$

be a finite ordered set with odd cardinality. Define the left and right weighting functions $w_L, w_R: D \rightarrow [0,1]$ by

$$w_L(x_i) = \begin{cases} 1, & 1 \leq i \leq m, \\ \frac{1}{2}, & i = m+1, \\ 0, & i \geq m+2, \end{cases} \quad w_R(x_i) = \begin{cases} 0, & 1 \leq i \leq m, \\ \frac{1}{2}, & i = m+1, \\ 1, & i \geq m+2. \end{cases} \quad (49)$$

These weights induce weighted counting measures

$$v_L(A) = \sum_{x \in A} w_L(x), \quad v_R(A) = \sum_{x \in A} w_R(x). \quad (50)$$

The corresponding regional evaluations are then given by

$$J_L = \frac{\int_D f \, d v_L}{v_L(D)} = \frac{\sum_{x \in D} w_L(x) f(x)}{\sum_{x \in D} w_L(x)}, \quad (51)$$

$$J_R = \frac{\int_D f \, d v_R}{v_R(D)} = \frac{\sum_{x \in D} w_R(x) f(x)}{\sum_{x \in D} w_R(x)}. \quad (52)$$

Since

$$v_L(D) = v_R(D) = m + \frac{1}{2}, \quad (53)$$

both sides possess exactly the same effective measure. The midpoint is therefore shared symmetrically, and no directional bias is introduced into the regional comparison.

This construction yields the following result.

Theorem 5.2 (Justification of Midpoint Sharing via Weighted Measure)

Let $D = \{x_1, \dots, x_{2m+1}\}$ be an odd-cardinality ordered set. Then the symmetric sharing of the midpoint x_{m+1} between the left and right subregions is rigorously represented by the weighted counting measures v_L and v_R defined in (50). In particular, the regional evaluation rule remains measure-theoretically consistent and preserves left-right symmetry.

Proof.

By construction, the midpoint x_{m+1} contributes one half to each weighted measure, while all other points contribute fully to exactly one side. Thus, $v_L(D) = v_R(D)$, and the resulting evaluations (51)–(52) are normalized averages with equal effective mass. Therefore, midpoint sharing is equivalent to replacing the ordinary counting measure by two weighted counting measures, one for each side. This preserves the measure-theoretic interpretation of MOST and eliminates asymmetry due to odd cardinality.

Theorem 5.2 is important for two reasons. First, it provides a rigorous mathematical interpretation of the midpoint-sharing rule introduced in Chapter 4. Second, it shows that the apparent irregularity of odd partitioning does not require a separate algorithmic principle; rather, it is naturally absorbed into the same theory by allowing weighted measures. In this sense, odd partitioning is not an exception, but a direct extension of the general measure-theoretic framework.

5.3. Interpretation: Different Realizations of the Same Algorithm

The results above lead to a fundamental reinterpretation of MOST.

Theorem 5.3 (Structural Equivalence of Continuous and Discrete MOST)

Continuous MOST and discrete MOST are structurally equivalent algorithms. Their difference lies solely in the underlying measure used to define regional averaging and probabilistic sampling.

Proof.

From Theorem 5.1, both continuous and discrete MOST are generated by the same evaluation functional (45). From Chapter 3, the probability measure associated with a region A is

$$P_A(B) = \frac{\mu(B \cap A)}{\mu(A)}. \quad (54)$$

Thus, both the regional comparison and the Monte Carlo sampling mechanism are determined entirely by μ . Choosing μ as the Lebesgue measure yields the continuous implementation, whereas choosing μ as the counting measure yields the discrete implementation. Therefore, the two algorithms are structurally identical and differ only through the choice of measure.

This theorem allows us to state the principal interpretation of the present work: Continuous and discrete MOST are not separate optimization methods. They are distinct realizations of a single measure-based algorithm.

This insight is, in our view, the most important theoretical contribution of the present study. It implies that the boundary between continuous and discrete optimization is not intrinsic to the MOST framework. Rather, the framework is inherently indifferent to that distinction and accommodates both within the same abstract machinery.

The same observation also suggests a natural path toward hybrid optimization. If the search space contains both continuous and discrete variables, one may endow it with a product measure composed of a Lebesgue measure on the continuous coordinates and a counting measure on the discrete coordinates. The regional evaluation principle then remains unchanged. Although such mixed-variable extensions are beyond the scope of the present paper, they emerge naturally from the unified theory developed here.

5.4. Summary

In this chapter, we have established the unified theory of continuous and discrete MOST.

The principal results are as follows:

1. A single measure-based evaluation functional, defined by normalized integration, generates both continuous and discrete MOST (Theorem 5.1).

2. Odd-cardinality partitioning in discrete domains is rigorously justified through weighted counting measures, thereby preserving symmetry and measure-theoretic consistency (Theorem 5.2).
3. Continuous and discrete MOST are structurally identical algorithms whose apparent differences arise solely from the underlying measure (Theorem 5.3).

These results collectively demonstrate that MOST is not merely adaptable to both continuous and discrete problems, but is inherently formulated at a higher level of abstraction in which both emerge as natural special cases.

This unified perspective provides the conceptual foundation for the subsequent discussion of theoretical limitations and numerical behavior.

Chapter 6. Theoretical Properties and Limitations of MOST

This chapter examines the theoretical properties and inherent limitations of the Monte Carlo Stochastic Optimization Technique (MOST). While the preceding chapters established MOST as a measure-based optimization framework with deterministic domain reduction and probabilistic evaluation, its effectiveness depends on structural properties of the objective function.

We therefore identify the conditions under which MOST is expected to perform reliably, as well as scenarios in which its performance may degrade. This analysis is essential for a rigorous understanding of the scope and applicability of the method.

6.1. Conditions for Effectiveness

The performance of MOST is governed by the relationship between the objective function and the underlying measure. In particular, the success of region-based selection depends on whether regions containing the global minimizer exhibit smaller average values than competing regions.

6.1.1. Unimodality

A fundamental condition under which MOST performs effectively is unimodality. Let $x^* \in X$ denote a global minimizer of f . We say that f is unimodal if, for any region A containing x^* , the function does not exhibit competing local minima of comparable depth. Under unimodality, sufficiently small regions containing x^* satisfy

$$J_f(A) = \frac{1}{\mu(A)} \int_A f d\mu \approx f(x^*), \quad (55)$$

while regions not containing x^* have strictly larger average values. Consequently, the region selection mechanism of MOST consistently favors subsets containing the global minimizer.

6.1.2. Low-Value Regions with Positive Measure

A more general and practically relevant condition is that the set of near-optimal points has nonzero measure.

Define the sublevel set:

$$S_\varepsilon = \{x \in X: f(x) \leq f(x^*) + \varepsilon\}. \quad (56)$$

$$\mu(S_\varepsilon) > 0, \quad (57)$$

then the global minimum is not isolated but embedded within a region of non-negligible measure. In this case, for regions A sufficiently aligned with S_ε , we obtain

$$J_f(A) \leq f(x^*) + C\varepsilon, \quad (58)$$

for some constant $C > 0$, whereas regions disjoint from S_ε yield strictly larger values. This ensures that the integral-based selection mechanism favors regions containing the global optimum.

This condition reflects the intrinsic strength of MOST: it is particularly effective when optimal solutions occupy a finite volume in the search space.

6.2. Fundamental Limitations

Despite its advantages, MOST exhibits inherent limitations that arise from its measure-based nature.

6.2.1. Isolated Minima (Needle Problem)

Consider the case where the global minimizer x^* is isolated and the surrounding region has significantly larger function values. In such cases, the contribution of x^* to the regional average is negligible. Let A be a region containing x^* , and suppose that $f(x)$ is large for almost all $x \in A \setminus \{x^*\}$. Then

$$J_f(A) \approx \frac{1}{\mu(A)} \int_{A \setminus \{x^*\}} f d\mu, \quad (59)$$

which is essentially independent of the value at x^* .

Thus, the presence of a single optimal point does not significantly influence the regional average. As a result, MOST may fail to identify regions containing such isolated minima.

This phenomenon is commonly referred to as the needle problem, and it represents a fundamental limitation of all measure-based optimization methods.

6.2.2. Multimodality

In highly multimodal functions, multiple regions may exhibit similar average values. Let $A_1, A_2 \subset X$ be two disjoint regions such that

$$J_f(A_1) \approx J_f(A_2). \quad (60)$$

In this case, the selection decision becomes sensitive to sampling noise:

$$\hat{J}_M(A_1) - \hat{J}_M(A_2) = O(M^{-1/2}). \quad (61)$$

Thus, finite-sample fluctuations may cause the algorithm to select suboptimal regions, particularly in early stages when regions are large. Although repeated subdivision mitigates this issue, convergence may be slower in strongly multimodal landscapes.

6.3. Comparison with Other Optimization Methods

The theoretical properties of MOST can be further clarified by comparison with established optimization methods.

6.3.1. Gradient-Based Methods

Gradient-based methods rely on local derivative information [1–4]. Their convergence is typically fast in smooth and convex settings, but they are highly sensitive to initialization and prone to convergence toward local minima.

In contrast, MOST:

- does not require differentiability
- is inherently global due to region-based evaluation
- is less sensitive to local irregularities

6.3.2. Evolutionary Algorithms (GA, PSO)

Evolutionary methods such as GA [8] and PSO [10] explore the search space through stochastic population dynamics. While they are robust to multimodality, their convergence is probabilistic and often requires extensive parameter tuning [12].

MOST differs in that:

- it provides deterministic domain reduction
- it does not require population-based search
- it leverages averaging to stabilize evaluation

6.3.3. Structural Distinction

The fundamental difference between MOST and conventional methods can be summarized as follows:

$$\text{Conventional methods: } \min_{x \in X} f(x), \quad (62)$$

$$\text{MOST: } \min_{A \subset X} J_f(A). \quad (63)$$

Thus, MOST replaces pointwise optimization with region-based optimization. This distinction explains both its strengths (robustness to noise and local irregularities) and its limitations (reduced sensitivity to isolated extrema).

6.4. Summary

In this chapter, we have analyzed the theoretical properties and limitations of MOST.

The principal conclusions are:

1. MOST performs effectively when the objective function exhibits unimodality or when near-optimal regions have positive measure.
2. The method is inherently limited in problems with isolated global minima, where the contribution of the optimum to regional averages is negligible.
3. In multimodal landscapes, convergence may be influenced by sampling variability.
4. MOST fundamentally differs from conventional methods by optimizing regions rather than individual points.

This leads to the central insight of this study: MOST is not a pointwise optimization method, but a measure-based optimization framework. This perspective provides a coherent explanation of both the strengths and limitations of the method and clarifies its position within the broader landscape of optimization theory.

Chapter 7. Numerical Experiments: Discrete MOST Under Uniform Discretization (Revised: 10 Variables)

This chapter presents numerical experiments designed to validate the theoretical framework of discrete MOST developed in the preceding chapters. In particular, we examine the performance of discrete MOST under uniform discretization of the search domain and compare the obtained solutions with the corresponding theoretical optima. Two benchmark functions are considered: the Ackley function and the Sphere function. These functions represent, respectively, a highly multimodal landscape and a convex unimodal landscape, thereby providing complementary test cases.

7.1. Experimental Setup

7.1.1. Search Domain

We consider a ten-dimensional domain:

$$\Omega = [-5, 5]^{10}. \quad (64)$$

No additional constraints are imposed, ensuring a uniform search space.

7.1.2. Discretization

The continuous domain is discretized uniformly with step size:

$$\Delta = 0.1. \quad (65)$$

Thus, the discrete set is:

$$D = \{x_i = -5 + 0.1k \mid k = 0, 1, \dots, 100\}. \quad (66)$$

The resulting grid contains:

$$|D| = 101^{10} \text{ points}. \quad (67)$$

This discretization transforms the optimization problem into a high-dimensional finite-domain problem suitable for discrete MOST.

7.1.3. Monte Carlo Sampling

For each subregion $A \subset D$, samples are drawn uniformly:

$$P(X = x_i) = \frac{1}{|A|}. \quad (68)$$

The sample size is fixed as:

$$M = 1000. \quad (69)$$

7.2. Test Functions

1) Ackley Function

The Ackley function in 10 variables is defined as:

$$f(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{10} \sum_{i=1}^{10} x_i^2} \right) - \exp \left(\frac{1}{10} \sum_{i=1}^{10} \cos(2\pi x_i) \right) + 20 + e. \quad (70)$$

Global minimum:

$$x^* = (0, \dots, 0), f(x^*) = 0. \quad (71)$$

2) Sphere Function

The Sphere function in 10 variables is defined as:

$$f(x) = \sum_{i=1}^{10} x_i^2. \quad (72)$$

Global minimum:

$$x^* = (0, \dots, 0), f(x^*) = 0. \quad (73)$$

7.3. Application of Discrete MOST

The discrete MOST algorithm described in Chapter 4 is applied to the discretized domain.

At each iteration:

- The current discrete set is bisected
- Monte Carlo estimation (37) is performed
- The region with smaller average value is selected
- The process continues until a single point remains

7.4. Results: Ackley Function (Revised)

7.4.1. Discrete Optimum

Due to discretization:

$$x_d^* = (0.0, \dots, 0.0). \quad (74)$$

Thus, the theoretical discrete optimum coincides with the continuous optimum. In the present discretization scheme with resolution $\Delta = 0.1$, the point $x = 0$ is explicitly included in the discrete grid defined by (66). Therefore, the global minimizer of the continuous Ackley function, which is located at the origin, is exactly represented within the discrete search space.

This is a nontrivial but important property: in general, discretization introduces approximation errors, and the discrete optimum may differ from the continuous one. However, in this particular setup, the discretization grid is aligned with the true optimum, ensuring that no discretization bias is introduced at the level of the global solution.

7.4.2. MOST Result

The discrete MOST algorithm yields:

$$x_{\text{MOST}} = (0.0, \dots, 0.0), \quad (75)$$

$$f_{\text{MOST}} \approx 4.44 \times 10^{-16}. \quad (76)$$

The discrete MOST algorithm converges to the exact global minimizer within the discretized domain. The resulting function value is numerically indistinguishable from zero, with a residual on the order of machine precision.

This slight deviation from the exact value $f(x^*) = 0$ is not due to algorithmic error, but rather arises from floating-point arithmetic in the evaluation of exponential and trigonometric functions. In particular, the Ackley function involves nested exponential and cosine terms, which introduce small numerical rounding effects.

Importantly, the algorithm successfully navigates the highly multimodal structure of the Ackley function, avoiding numerous local minima and converging to the global optimum.

7.4.3. Error Analysis

Define the error:

$$E = \|x_{\text{MOST}} - x^*\|. \quad (77)$$

Then:

$$E = 0. \quad (78)$$

This is consistent with the discretization resolution. Since the discrete optimum coincides exactly with the continuous optimum and the algorithm successfully identifies this point, the resulting error is identically zero.

This result represents an ideal case in which both sources of error—discretization error and optimization error—are eliminated:

- Discretization error is zero because the optimal point is included in the grid
- Algorithmic error is zero because MOST correctly identifies the optimal region

In general, one expects the error to scale as $O(\Delta)$, as discussed in Chapter 6. The present result therefore confirms that the theoretical error bound is sharp and that exact recovery is possible when the discretization grid contains the global minimizer.

7.4.4. Interpretation

- The solution coincides with the true optimum
- The error is bounded by $O(\Delta)$
- The method successfully avoids local minima despite strong multimodality

The results for the Ackley function demonstrate the robustness of discrete MOST in a challenging optimization landscape characterized by a large number of local minima.

First, the exact recovery of the global optimum confirms that the region-based selection mechanism is capable of isolating the correct region even in high-dimensional and highly oscillatory settings.

Second, the absence of error indicates that the discretization scheme does not degrade solution quality in this case. This highlights the importance of grid alignment with the global optimum.

Third, and most importantly, the algorithm avoids entrapment in local minima. This behavior can be attributed to the integral-based evaluation mechanism: local minima that occupy small regions contribute negligibly to the average value $J_f(A)$, and are therefore systematically rejected during the recursive partitioning process.

7.5. Results: Sphere Function (Revised)

7.5.1. Discrete Optimum

$$x_d^* = (0, \dots, 0). \quad (79)$$

As in the Ackley case, the global minimizer of the Sphere function is located at the origin, which is explicitly included in the discretized grid. Therefore, the discrete and continuous optima coincide exactly.

7.5.2. MOST Result

$$x_{\text{MOST}} = (0.0, \dots, 0.0), \quad (80)$$

$$f_{\text{MOST}} = 0. \quad (81)$$

The discrete MOST algorithm converges exactly to the global minimizer. Unlike the Ackley function, the Sphere function is a smooth, convex, and unimodal function. As a result, the integral-based evaluation aligns perfectly with the pointwise objective structure, and convergence occurs without any ambiguity or numerical instability.

7.5.3. Estimation of Error

$$E = 0. \quad (82)$$

The zero error reflects both exact representation of the optimum in the discretized domain and perfect identification by the algorithm. This confirms that MOST achieves exact convergence under ideal conditions.

7.5.4. Interpretation

- Exact recovery of the optimum
- Confirms correctness under unimodality

The Sphere function provides a baseline case for evaluating the correctness of the algorithm. Since the function is strictly convex and possesses a unique global minimum, any consistent global optimization method should converge to the correct solution.

The results confirm that MOST satisfies this requirement. Moreover, the smooth structure of the function ensures that the regional average $J_f(A)$ decreases monotonically as the region approaches the optimum, leading to stable and deterministic convergence.

This behavior is consistent with the theoretical conditions outlined in Chapter 6, where unimodality guarantees the effectiveness of region-based optimization.

7.6. Comparative Summary

Table 1 summarizes the performance of the discrete MOST algorithm on the two benchmark functions considered in this study. For both the Ackley function and the Sphere function in ten dimensions, the algorithm successfully identifies the global optimum within the discretized search domain.

Several key observations can be made:

First, the MOST solution coincides exactly with the theoretical optimum in both cases. This confirms that the algorithm is capable of resolving the global minimum even in high-dimensional settings, provided that the optimal point is included in the discretization grid.

Second, the error is identically zero for both functions. This result represents the ideal scenario in which both discretization error and optimization error vanish. In general, as discussed in Chapter 6, the error is expected to scale as $O(\Delta)$; however, when the discretization grid contains the true optimum and the algorithm successfully isolates the corresponding region, exact recovery is achieved.

Third, the results highlight the robustness of MOST across different landscape structures. The Sphere function represents a convex unimodal problem, whereas the Ackley function is highly multimodal with numerous local minima. Despite this fundamental difference, the algorithm exhibits consistent performance, indicating that the measure-based regional evaluation effectively suppresses the influence of local irregularities.

Finally, the consistency of results across both test functions supports the theoretical framework developed in Chapters 3–6. In particular, it confirms that discrete MOST operates as a faithful realization of the unified measure-theoretic formulation, even in high-dimensional spaces.

Table 1. Comparison of discrete MOST solutions and theoretical optima for 10-dimensional benchmark functions under uniform discretization ($\Delta = 0.1$).

Function	True Optimum	MOST Solution	Error
Ackley (10D)	$(0,\dots,0)$	$(0,\dots,0)$	0
Sphere (10D)	$(0,\dots,0)$	$(0,\dots,0)$	0

7.7. Discussion

The results confirm the theoretical predictions:

- Discretization Error

$$E = O(\Delta), \tag{83}$$

indicating that accuracy is controlled by grid resolution.

- Robustness to Multimodality

The Ackley function demonstrates that MOST is not trapped by local minima, due to integral averaging.

- Consistency with Theory

The results align with Chapter 6:

- Works well when low-value regions have measure
- Accuracy improves with finer discretization

7.8. Summary

In this chapter, we have demonstrated that:

- Discrete MOST successfully approximates global optima under uniform discretization
- The error is bounded by the discretization step size
- The method is robust to multimodal landscapes
- The theoretical framework developed in Chapters 3–6 is validated numerically

These results confirm that discrete MOST provides a practical and theoretically consistent extension of the continuous framework.

Chapter 8. Discussion

This chapter provides a conceptual interpretation of the Monte Carlo Stochastic Optimization Technique (MOST) and discusses its robustness and future extensions. Building upon the measure-theoretic framework and numerical validation presented in the preceding chapters, we position MOST within a broader theoretical and practical context.

8.1. Physical and Geometric Interpretation

The measure-theoretic formulation of MOST reveals a fundamental reinterpretation of optimization itself. Conventional optimization methods operate by searching for points that minimize the objective function. In contrast, MOST evaluates regions, assigning to each region a value determined by a measure-weighted average.

This distinction can be expressed formally as:

$$\text{Classical optimization: } \min_{x \in X} f(x), \tag{84}$$

$$\text{MOST: } \min_{A \subset X} \frac{1}{\mu(A)} \int_A f \, d\mu. \tag{85}$$

Thus, MOST replaces pointwise evaluation with measure-based aggregation. From a geometric viewpoint, this corresponds to evaluating the objective function not at isolated points, but over finite-volume regions. The optimization process therefore becomes a sequence of geometric refinements, in which the search domain is progressively contracted toward regions of lower average energy.

From a physical perspective, the integral in (85) may be interpreted as a form of coarse-grained energy. Rather than being sensitive to infinitesimal fluctuations, MOST captures the macroscopic structure of the objective landscape. This interpretation explains its robustness in complex, multimodal environments: narrow local minima contribute negligibly to the integral unless they occupy a region of non-negligible measure.

Consequently, MOST naturally emphasizes stable structures in the search space—regions where low values persist over a measurable volume—rather than isolated extrema.

8.2. Robustness and Noise Tolerance

A direct consequence of the measure-based formulation is robustness to noise and irregularity. Consider an objective function perturbed by stochastic noise:

$$f_\eta(x) = f(x) + \eta(x), \quad (86)$$

where $\eta(x)$ is a zero-mean random perturbation. Under regional averaging, we obtain:

$$J_{f_\eta}(A) = J_f(A) + \frac{1}{\mu(A)} \int_A \eta(x) d\mu(x). \quad (87)$$

Since the noise term averages out, we have:

$$\mathbb{E}[J_{f_\eta}(A)] = J_f(A). \quad (88)$$

Moreover, the variance decreases with increasing sample size:

$$\text{Var}(\hat{J}_M(A)) = O(M^{-1}). \quad (89)$$

This averaging effect implies that MOST is inherently resistant to:

- stochastic noise
- measurement errors
- high-frequency oscillations

Unlike gradient-based methods, which are highly sensitive to local perturbations, MOST stabilizes the evaluation through integration. This property is particularly advantageous in practical engineering problems where objective functions are often noisy or derived from simulation outputs.

8.3. Future Extensions

The unified measure-theoretic framework developed in this study naturally suggests several extensions.

8.3.1. Mixed Continuous–Discrete Optimization

Since MOST is defined on a general measure space, hybrid optimization problems can be treated by introducing product measures:

$$\mu = \lambda \times \nu, \quad (90)$$

where λ is a Lebesgue measure on continuous variables and ν is a counting measure on discrete variables.

The evaluation functional remains unchanged:

$$J_f(A) = \frac{1}{\mu(A)} \int_A f d\mu. \quad (91)$$

This provides a principled framework for mixed-integer optimization problems without requiring separate algorithmic strategies.

8.3.2. High-Dimensional Optimization

High-dimensional optimization remains a central challenge in modern applications. In such settings, the curse of dimensionality affects all global optimization methods.

Within the MOST framework, this issue manifests in the exponential growth of partition complexity. However, the measure-based formulation suggests potential strategies:

- adaptive partitioning
- dimension-wise decomposition

- importance sampling

In particular, the integral formulation may be combined with variance reduction techniques to improve efficiency in high-dimensional spaces.

8.3.3. Adaptive and Anisotropic Partitioning

The current formulation employs uniform bisection. However, the framework allows for more general partitioning strategies:

$$\Omega^{(k)} \rightarrow \{A_1^{(k)}, A_2^{(k)}, \dots\}. \quad (92)$$

By adapting the partitioning scheme to the local structure of f , one may improve convergence speed and accuracy.

8.4. Central Insight

The theoretical and numerical results of this study lead to the following fundamental conclusion: MOST is a measure-based optimization framework, not a pointwise optimization method. This statement encapsulates the essential distinction between MOST and conventional optimization techniques. By shifting the focus from points to regions, MOST provides a new perspective on global optimization that is inherently robust, scalable, and adaptable.

8.5. Concluding Remarks of Discussion

The measure-theoretic reinterpretation presented in this work elevates MOST from a heuristic algorithm to a unified theoretical framework. This perspective clarifies both its strengths—robustness, globality, and simplicity—and its limitations, particularly in the presence of isolated minima.

More importantly, it establishes a conceptual bridge between continuous and discrete optimization, suggesting that both can be understood as manifestations of a deeper, measure-based principle.

This insight opens new directions for research and application, particularly in hybrid optimization, high-dimensional systems, and stochastic environments.

Chapter 9. Conclusion

In this study, we have established a unified theoretical framework for the Monte Carlo Stochastic Optimization Technique (MOST) based on measure-theoretic principles. By reformulating MOST in terms of normalized integrals over measurable regions, we have demonstrated that both continuous and discrete optimization problems can be treated within a single, coherent mathematical structure.

A central achievement of this work is the rigorous construction of discrete MOST. By introducing the counting measure and extending it to weighted measures, we have shown that discrete domain partitioning—including the nontrivial case of odd cardinality with midpoint sharing—can be handled consistently within the same theoretical framework. This result confirms that discrete MOST is not a heuristic extension, but a natural consequence of the underlying measure-based formulation.

Furthermore, we have established that continuous and discrete MOST are structurally equivalent algorithms whose differences arise solely from the choice of measure. This unification provides a new perspective on optimization, in which the distinction between continuous and discrete domains is no longer fundamental, but instead reflects different realizations of a common principle.

The numerical experiments conducted using benchmark functions such as the Ackley and Sphere functions have validated the theoretical predictions. In particular, the results demonstrate that discrete MOST achieves accurate approximations of global optima, with errors governed by discretization resolution, and exhibits robustness in multimodal landscapes.

Beyond its theoretical contributions, the present framework suggests broad applicability. The measure-theoretic formulation naturally extends to mixed-variable optimization, high-dimensional

problems, and stochastic environments, offering a flexible and scalable approach to complex optimization tasks.

In conclusion, this work establishes MOST as a measure-based optimization paradigm that unifies continuous and discrete optimization within a single rigorous framework. This perspective not only clarifies the fundamental nature of MOST but also opens new avenues for research and application in global optimization theory and practice.

Appendix A. Discrete MOST Algorithm for Numerical Experiments

A.1 Purpose and Relation to the Main Text

This appendix provides a detailed description of the discrete MOST algorithm used in Chapter 7 for numerical experiments. The formulation is directly based on the measure-theoretic framework introduced in Chapter 3 and the discrete construction developed in Chapter 4.

In particular:

- The evaluation functional corresponds to (29)
- The Monte Carlo estimator corresponds to (37)
- The selection rule corresponds to (39)

The purpose of this appendix is to ensure full reproducibility of the numerical results.

A.2 Discrete Domain Construction

Let the search domain be defined as:

$$\Omega = [-5, 5]^n. \quad (\text{A1})$$

Each coordinate is discretized with step size:

$$\Delta = 0.1. \quad (\text{A2})$$

Thus, the one-dimensional discrete set is:

$$D = \{x_i = -5 + k\Delta \mid k = 0, 1, \dots, 100\}. \quad (\text{A3})$$

The full search space is given by the Cartesian product:

$$D^{(n)} = D \times D \times \dots \times D. \quad (\text{A4})$$

A.3 Partitioning Strategy

At each iteration, the current candidate set for each variable is partitioned. Let $D_j^{(k)}$ denote the candidate set of the j -th variable at iteration k .

A.3.1 Even Cardinality

If $|D_j^{(k)}| = 2m$, then:

$$D_{j,L}^{(k)} = \{x_1, \dots, x_m\}, D_{j,R}^{(k)} = \{x_{m+1}, \dots, x_{2m}\}. \quad (\text{A5})$$

A.3.2 Odd Cardinality

If $|D_j^{(k)}| = 2m + 1$, define:

$$D_{j,L}^{(k)} = \{x_1, \dots, x_m, x_{m+1}\}, D_{j,R}^{(k)} = \{x_{m+1}, \dots, x_{2m+1}\}. \quad (\text{A6})$$

The midpoint x_{m+1} is shared between both subsets.

A.4 Monte Carlo Sampling

For a given subset $A \subset D^{(n)}$, sampling is performed according to the discrete uniform distribution:

$$\mathbb{P}(X = x_i) = \frac{1}{|A|}. \quad (\text{A7})$$

For odd partitioning, weighted sampling is used:

$$\mathbb{P}(X = x_i) = \frac{w(x_i)}{\sum_{x \in A} w(x)}. \quad (\text{A8})$$

A.5 Monte Carlo Estimation

The regional evaluation is approximated by:

$$\hat{f}_M(A) = \frac{1}{M} \sum_{i=1}^M f(X_i), X_i \sim P_A. \quad (A9)$$

A.6 Algorithm Description

We now present the complete discrete MOST algorithm used in Chapter 7.

Algorithm A.1 (Discrete MOST)

Input:

- Initial domain $D^{(0)} = D^{(n)}$
- Sample size M

Step 1 (Initialization):

$$k = 0. \quad (A10)$$

For each variable $j = 1, \dots, n$, set:

$$D_j^{(0)} = D. \quad (A11)$$

Step 2 (Partitioning):

For each variable j , partition $D_j^{(k)}$ into:

$$D_{j,L}^{(k)}, D_{j,R}^{(k)}. \quad (A12)$$

Step 3 (Monte Carlo Evaluation):

Estimate:

$$\hat{f}_M(D_{j,L}^{(k)}), \hat{f}_M(D_{j,R}^{(k)}). \quad (A13)$$

Step 4 (Selection):

Update:

$$D_j^{(k+1)} = \begin{cases} D_{j,L}^{(k)}, & \text{if } \hat{f}_M(D_{j,L}^{(k)}) \leq \hat{f}_M(D_{j,R}^{(k)}), \\ D_{j,R}^{(k)}, & \text{otherwise.} \end{cases} \quad (A14)$$

Step 5 (Iteration):

$$k \leftarrow k + 1. \quad (A15)$$

Repeat Steps 2–5 until:

$$|D_j^{(k)}| = 1 \forall j. \quad (A16)$$

Step 6 (Output):

Return:

$$x^* = (x_1^*, \dots, x_n^*), \quad (A17)$$

where x_j^* is the unique element of $D_j^{(k)}$.

A.7 Computational Remarks

- The algorithm performs dimension-wise recursive reduction
- The total number of iterations is bounded by:

$$O(\log_2 |D|). \quad (A18)$$

The computational cost is dominated by Monte Carlo sampling:

$$O(n \cdot M \cdot K), \quad (A19)$$

where K is the number of iterations

A.8 Summary

This appendix provides a complete and reproducible description of the discrete MOST algorithm used in the numerical experiments. The formulation is fully consistent with the measure-theoretic framework developed in the main text and directly supports the results presented in Chapter 7.

References

1. Nocedal, J., Wright, S., *Numerical Optimization*, Springer, 2006. DOI:10.1007/978-0-387-40065-5
2. Boyd, S., Vandenberghe, L., *Convex Optimization*, Cambridge Univ. Press, 2004. DOI:10.1017/CBO9780511804441
3. Bertsekas, D., *Nonlinear Programming*, Athena Scientific, 1999. DOI:10.1137/1.9781611971200
4. Conn, A., Gould, N., Toint, P., *Trust Region Methods*, SIAM, 2000. DOI:10.1137/1.9780898719857
5. Horst, R., Pardalos, P., *Handbook of Global Optimization*, Springer, 1995. DOI:10.1007/978-1-4613-3551-7
6. Floudas, C., *Deterministic Global Optimization*, Springer, 2000. DOI:10.1007/978-1-4757-3150-0
7. Spall, J., *Introduction to Stochastic Search and Optimization*, Wiley, 2003. DOI:10.1002/0471722138
8. Holland, J., *Adaptation in Natural and Artificial Systems*, MIT Press, 1975.
9. Storn, R., Price, K., Differential Evolution, *J. Global Optimization*, 11, 341–359 (1997). DOI:10.1023/A:1008202821328
10. Kennedy, J., Eberhart, R., Particle Swarm Optimization, *Proc. IEEE ICNN*, 1995. DOI:10.1109/ICNN.1995.488968
11. Hansen, N., CMA-ES Evolution Strategy, *Evolutionary Computation*, 2003. DOI:10.1162/106365603321828970
12. Bäck, T., *Evolutionary Algorithms in Theory and Practice*, Oxford Univ. Press, 1996.
13. Robbins, H., Monroe, S., Stochastic Approximation, *Ann. Math. Stat.*, 1951. DOI:10.1214/aoms/1177729586
14. Kushner, H., Yin, G., *Stochastic Approximation*, Springer, 2003. DOI:10.1007/b97471
15. Jones, D., Schonlau, M., Welch, W., Efficient Global Optimization, *J. Global Optimization*, 1998. DOI:10.1023/A:1008306431147
16. Shahriari, B. et al., Bayesian Optimization Review, *Proc. IEEE*, 2016. DOI:10.1109/JPROC.2015.2494218
17. Snoek, J. et al., Practical Bayesian Optimization, *NIPS*, 2012.
18. Lawler, E., Wood, D., Branch-and-Bound Methods, *Operations Research*, 1966. DOI:10.1287/opre.14.4.699
19. Jones, D. et al., DIRECT Algorithm, *J. Optimization Theory Appl.*, 1993. DOI:10.1007/BF00941892
20. Piyavskii, S., Lipschitz Optimization, *USSR Comput. Math.*, 1972.
21. Deb, K. et al., NSGA-II, *IEEE Trans. Evolutionary Computation*, 2002. DOI:10.1109/4235.996017
22. Zitzler, E., Thiele, L., SPEA2, *TIK Report*, 2001.
23. Karush, W., *Minima of Functions*, 1939.
24. Kuhn, H., Tucker, A., Nonlinear Programming, *Proc. Berkeley Symp.*, 1951.
25. Miettinen, K., *Nonlinear Multiobjective Optimization*, Springer, 1999. DOI:10.1007/978-1-4615-5563-6
26. Inage, S., Hebishima, H., Monte Carlo Stochastic Optimization Technique (MOST), *Mathematics and Computers in Simulation*, 199, 257–271 (2022). DOI:10.1016/j.matcom.2022.03.013
27. Inage, S., Ohgi, S., Takahashi, Y., Multi-objective MOST, *Mathematics and Computers in Simulation*, 215, 146–157 (2024). DOI:10.1016/j.matcom.2023.08.001
28. Inage, S., Ajito, K., A Deterministic Global Optimization Framework via Monte Carlo Region Integration, Preprint, 2024.
29. Hoeffding, W., Probability Inequalities for Sums of Bounded Random Variables, *Journal of the American Statistical Association*, 58(301), 13–30 (1963). DOI:10.1080/01621459.1963.10500830.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.