
A Deterministic Global Optimization Framework via Monte Carlo Region Integration: Rigorous Convergence, KKT Equivalence, and Multi-Objective Extension

[Shin-Ichi Inage](#)* and Kouei Ajito

Posted Date: 2 April 2026

doi: 10.20944/preprints202604.0114.v1

Keywords: Monte Carlo optimization; deterministic global optimization; region-based search; integral-based selection; geometric convergence; KKT conditions; Pareto-KKT optimality; constrained optimization; multi-objective optimization; stochastic approximation; measure-based optimization; derivative-free optimization



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 Deterministic Global Optimization Framework via Monte Carlo Region Integration: Rigorous Convergence, KKT Equivalence, and Multi-Objective Extension

Shin-Ichi Inage * and Kouei Ajito

Ishinomaki Senshu University, 1 Shin-mito, Minami-sakai, Ishinomaki-shi, Miyagi-ken, 986-8580, Japan

* Correspondence: s.inage.pu@isenshu-u.ac.jp

Abstract

This paper introduces the Monte Carlo Stochastic Optimization Technique (MOST), a global optimization framework based on region-wise integral comparison. Unlike classical pointwise methods, MOST evaluates candidate regions through aggregated objective values, enabling a structured and global exploration of the search space. We establish a unified theoretical foundation. Deterministic geometric shrinking of regions ensures that their diameters converge to zero, while a non-circular integral separation principle guarantees global convergence. Incorporating Monte Carlo estimation, we derive exponential concentration bounds and prove almost sure convergence under suitable sampling schedules. For constrained problems, we introduce an extended functional whose minimizers are equivalent to Karush–Kuhn–Tucker (KKT) points, allowing constraint handling without projection or penalty tuning. The framework is further extended to multi-objective optimization, where convergence to Pareto–KKT stationary points is established. Numerical experiments on multimodal benchmark functions confirm the theoretical results. Overall, MOST provides a derivative-free, deterministic–probabilistic framework for global optimization that extends naturally to constrained and multi-objective settings.

Keywords: Monte Carlo optimization; deterministic global optimization; region-based search; integral-based selection; geometric convergence; KKT conditions; Pareto–KKT optimality; constrained optimization; multi-objective optimization; stochastic approximation; measure-based optimization; derivative-free optimization

1. Introduction

Optimization constitutes a foundational pillar across a wide spectrum of disciplines, including engineering design, scientific computing, machine learning, control theory, and operations research. Classical deterministic optimization methods—such as gradient descent, quasi-Newton schemes, trust-region frameworks, and interior-point methods—have achieved remarkable success in smooth and convex settings, where convergence properties can be rigorously established [1–4]. However, their performance deteriorates significantly when confronted with nonconvexity, multimodality, discontinuities, or nonsmooth constraint structures, which are ubiquitous in real-world applications [5–7].

To address these limitations, a broad class of stochastic and population-based methods has 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 strong empirical robustness and flexibility in navigating complex landscapes. Nevertheless, their stochastic nature fundamentally limits their theoretical characterization:

convergence is typically asymptotic or probabilistic, and rigorous guarantees—particularly for constrained or multi-objective problems—remain incomplete or highly problem-dependent [12–14].

In parallel, surrogate-based and probabilistic approaches, such as Bayesian optimization [15,16], have emerged as powerful tools for expensive black-box optimization. While these methods provide probabilistic guarantees under specific assumptions, they rely on model fidelity and often struggle with high-dimensional spaces or highly irregular feasible regions [17]. Similarly, deterministic global optimization methods, including branch-and-bound and Lipschitz-based algorithms [18–20], offer rigorous guarantees but frequently suffer from exponential complexity and limited scalability.

Multi-objective optimization further amplifies these challenges. Evolutionary algorithms such as NSGA-II [21] and SPEA2 [22] have become standard tools for approximating Pareto fronts, yet they produce discrete approximations without deterministic convergence guarantees. Moreover, the theoretical connection between such algorithms and the Karush–Kuhn–Tucker (KKT) optimality conditions remain indirect and largely heuristic [23,24]. Weighted-sum scalarization provides a classical bridge between multi-objective and single-objective optimization; however, it fails to recover non-supported Pareto solutions in nonconvex settings and lacks a unified convergence theory [25].

Against this backdrop, Inage and Hebishima introduced a fundamentally different paradigm known as the Monte Carlo Stochastic Optimization Technique (MOST) [26,27]. Unlike conventional pointwise evaluation methods, MOST operates on a region-based principle: the search domain is recursively partitioned, and each subregion is evaluated via Monte Carlo integration. The region with the smallest integral value is selected for further refinement. This seemingly simple mechanism leads to a striking property: the search region shrinks deterministically at each iteration, yielding geometric convergence of the form

$$\text{diam}(R_n) = O(2^{-n}).$$

This deterministic region-shrinking mechanism distinguishes MOST sharply from both stochastic and gradient-based methods. In particular, the use of integral-based evaluation introduces an intrinsic smoothing effect over the objective landscape, enabling the method to avoid narrow local minima that would trap pointwise algorithms. Initial studies demonstrated that MOST outperforms conventional evolutionary methods in benchmark problems such as the Ackley and Schwefel functions, while maintaining stable convergence behavior [26].

Subsequently, MOST was extended to multi-objective optimization through a weighted scalarization framework that preserves its deterministic convergence properties [27]. This extension showed that, for each weight vector, MOST converges to a Pareto-consistent solution without the oscillatory behavior typical of evolutionary approaches. However, despite these promising developments, several fundamental theoretical gaps remained unresolved.

First, existing MOST formulations lacked a rigorous treatment of constrained optimization. In particular, no formal connection had been established between MOST and the KKT conditions, which constitute the cornerstone of constrained nonlinear optimization theory [3,23]. Second, the role of Lagrange multipliers—central to both theoretical analysis and practical interpretation—had not been incorporated into the MOST framework. Third, while multi-objective MOST demonstrated empirical success, its relationship to Pareto–KKT optimality conditions remained unproven. Finally, the behavior of MOST near nonlinear (curved) constraint boundaries, which are prevalent in engineering applications, had not been analyzed.

The present work addresses these gaps by developing a unified and mathematically rigorous optimization framework based on MOST. The core idea is to embed the KKT system directly into an augmented Lagrangian functional and to apply the deterministic region-shrinking mechanism of MOST in this extended variable space. This approach enables simultaneous enforcement of stationarity, primal feasibility, dual feasibility, and complementary slackness within a single optimization process.

A key contribution of this paper is the establishment of a non-circular global convergence theory based on integral comparison over shrinking regions. Under mild regularity assumptions, including Lipschitz continuity, we show that the region containing the global KKT point is selected infinitely

often, leading to deterministic convergence. In addition, we provide a probabilistic analysis of Monte Carlo integration errors using concentration inequalities, demonstrating that sampling noise does not compromise convergence. The resulting framework thus combines deterministic geometric convergence with probabilistic robustness.

Furthermore, we rigorously establish the equivalence between the minimization of the extended Lagrangian and the satisfaction of the full KKT system. Under the Linear Independence Constraint Qualification (LICQ), we prove the uniqueness and continuity of Lagrange multipliers and show that they converge at the same geometric rate as the primal variables. This result places MOST in a fundamentally different category from evolutionary algorithms, which do not provide access to multiplier information.

In the multi-objective setting, we demonstrate that the proposed framework converges to Pareto–KKT stationary points for each weight vector, thereby establishing a direct and rigorous connection between scalarization-based optimization and KKT theory. Unlike classical weighted-sum methods, the proposed approach retains deterministic convergence properties even in highly nonconvex settings.

Finally, we develop a geometric theory explaining the behavior of MOST near curved constraint boundaries. We show that, due to exponential region shrinking, nonlinear constraint surfaces become asymptotically indistinguishable from their tangent hyperplanes. This result provides a natural explanation for the emergence of KKT normality conditions without explicit constraint linearization.

The contributions of this study can be summarized as follows:

1. A rigorous formulation of constrained MOST based on an extended Lagrangian embedding the full KKT system.
2. A non-circular proof of global convergence via integral-based region selection.
3. A probabilistic analysis of Monte Carlo errors ensuring robust region selection.
4. A proof of uniqueness and geometric convergence of Lagrange multipliers under LICQ.
5. A unified Pareto–KKT convergence theory for multi-objective constrained optimization.
6. A geometric explanation of constraint handling via automatic tangent-plane approximation.

Collectively, these results establish MOST as a new class of deterministic, derivative-free optimization methods with rigorous guarantees for constrained and multi-objective problems.

2. Mathematical Preliminaries

This chapter establishes the mathematical foundations required for the rigorous analysis of the Monte Carlo Stochastic Optimization Technique (MOST). We introduce the notation, define constrained and multi-objective optimality conditions, clarify the limitations of classical scalarization methods, and explicitly state the regularity assumptions—particularly Lipschitz continuity, measure-theoretic structure, and probabilistic conditions—that are essential for the convergence theory developed in subsequent chapters.

2.1. Notation and Problem Setting

Let $\Omega \subset \mathbb{R}^d$ be a bounded hyperrectangle, which serves as the search domain:

$$\Omega = \prod_{i=1}^d [a_i, b_i]. \quad (1)$$

A decision variable is denoted by

$$x \in \Omega. \quad (2)$$

We consider a general constrained nonlinear optimization problem:

$$\min_{x \in \Omega} f(x) \quad (3)$$

subject to equality and inequality constraints:

$$h_i(x) = 0, i = 1, \dots, m, \quad (4)$$

$$g_j(x) \leq 0, j = 1, \dots, p. \quad (5)$$

The feasible set is defined as

$$\mathcal{F} = \{x \in \Omega \mid h_i(x) = 0, g_j(x) \leq 0\}. \quad (6)$$

We denote the diameter of a set $R \subset \mathbb{R}^d$ by

$$\text{diam}(R) = \sup_{x,y \in R} \|x - y\|. \quad (7)$$

2.2. Karush–Kuhn–Tucker (KKT) Conditions

The Lagrangian associated with problem (3)–(5) is defined as

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x), \quad (8)$$

where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ are Lagrange multipliers. A point $x^* \in \mathcal{F}$ is said to satisfy the KKT conditions if there exist multipliers (λ^*, μ^*) such that:

(i) Stationarity

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0, \quad (9)$$

(ii) Primal feasibility

$$h_i(x^*) = 0, g_j(x^*) \leq 0, \quad (10)$$

(iii) Dual feasibility

$$\mu_j^* \geq 0, \quad (11)$$

(iv) Complementary slackness

$$\mu_j^* g_j(x^*) = 0. \quad (12)$$

A fundamental regularity condition is the Linear Independence Constraint Qualification (LICQ), which requires that the gradients of active constraints are linearly independent:

$$\{\nabla h_i(x^*)\}_{i=1}^m \cup \{\nabla g_j(x^*) \mid g_j(x^*) = 0\} \text{ are linearly independent.} \quad (13)$$

Under LICQ, the multipliers (λ^*, μ^*) are unique [28–30].

2.3. Multi-Objective Optimization and Pareto Optimality

We consider a vector-valued objective function

$$F(x) = (f_1(x), \dots, f_k(x)). \quad (14)$$

A point $x^* \in \mathcal{F}$ is Pareto optimal if there is no $x \in \mathcal{F}$ such that

$$f_i(x) \leq f_i(x^*) \quad \forall i, \text{ and } f_j(x) < f_j(x^*) \text{ for some } j. \quad (15)$$

Under standard regularity conditions, a Pareto-optimal point satisfies the Pareto–KKT condition:

$$\sum_{i=1}^k w_i \nabla f_i(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) = 0, \quad (16)$$

for some weight vector

$$w \in \mathbb{R}^k, w_i \geq 0, \sum_{i=1}^k w_i = 1. \quad (17)$$

2.4 Weighted-Sum Scalarization and Its Limitations

A classical approach to multi-objective optimization is weighted-sum scalarization:

$$\min_{x \in \mathcal{F}} \sum_{i=1}^k w_i f_i(x). \quad (18)$$

While (18) recovers all Pareto-optimal solutions when the objective functions are convex, it suffers from fundamental limitations:

- In nonconvex problems, certain Pareto-optimal points (non-supported points) cannot be obtained [31].
- Multiple distinct solutions may correspond to the same weight vector.
- The mapping from weights to Pareto solutions may be discontinuous.

These limitations motivate the need for a more robust framework capable of preserving optimality conditions beyond convex settings.

2.5. Regularity Assumptions

To establish rigorous convergence results, we impose the following assumptions.

Assumption A1 (Lipschitz Continuity)

The objective and constraint functions satisfy

$$\|f(x) - f(y)\| \leq L \|x - y\|, \forall x, y \in \Omega, \quad (19)$$

and similarly, for h_i, g_j .

Assumption A2 (Boundedness)

The domain Ω is compact:

$$\Omega \subset \mathbb{R}^d \text{ is bounded and closed.} \quad (20)$$

Assumption A3 (Measurability)

All functions are Lebesgue measurable:

$$f, h_i, g_j \in L^\infty(\Omega). \quad (21)$$

Assumption A4 (Existence of KKT Point)

There exists at least one KKT point $x^* \in \mathcal{F}$. (22)

Assumption A5 (Regularity of Constraints)

LICQ holds at the optimal point x^* . (23)

2.6. Measure-Theoretic Framework

For any measurable subset $R \subset \Omega$, we define its integral value:

$$I(R) = \int_R f(x) dx. \quad (24)$$

Under Assumption A1, the following approximation holds for sufficiently small regions:

$$I(R) = f(x^*) |R| + O(\text{diam}(R)), \quad (25)$$

where $x^* \in R$ is a minimizer. This relation is fundamental to the integral-based selection mechanism of MOST.

2.7. Probabilistic Framework for Monte Carlo Estimation

Let $X_1, \dots, X_n \sim \text{Uniform}(R)$. The Monte Carlo estimator is defined as

$$\hat{I}_n(R) = \frac{|R|}{n} \sum_{i=1}^n f(X_i). \quad (26)$$

By the law of large numbers:

$$\hat{I}_n(R) \rightarrow I(R) \text{ almost surely.} \quad (27)$$

Moreover, concentration inequalities provide finite-sample guarantees:

$$\mathbb{P}(|\hat{I}_n(R) - I(R)| > \epsilon) \leq 2 \exp(-Cn\epsilon^2), \quad (28)$$

for some constant $C > 0$ depending on the boundedness of f [32].

2.8. Summary of Assumptions and Their Role

The assumptions introduced above serve distinct purposes:

- Lipschitz continuity (A1) ensures stability of integral comparisons.
- Compactness (A2) guarantees existence of minimizers.
- Measurability (A3) enables integration-based evaluation.

- LICQ (A5) ensures uniqueness of multipliers.
- Probabilistic bounds (28) control Monte Carlo error.

Together, these conditions establish a mathematically rigorous foundation upon which the deterministic and probabilistic convergence properties of MOST will be built.

3. The MOST Framework (Unconstrained Case)

This chapter introduces the foundational structure of the Monte Carlo Stochastic Optimization Technique (MOST) in the unconstrained single-objective setting. Unlike classical optimization methods that rely on pointwise evaluations or gradient information, MOST operates on a region-based principle, combining deterministic domain partitioning with Monte Carlo integration. This unique combination enables both robustness against multimodality and deterministic geometric convergence.

3.1. Problem Formulation

We consider the unconstrained minimization problem:

$$\min_{x \in \Omega} f(x), \quad (29)$$

where $\Omega \subset \mathbb{R}^d$ is a bounded hyperrectangle defined in (1), and f satisfies the regularity assumptions introduced in Chapter 2.

3.2. Core Idea of MOST

The key idea of MOST is to replace pointwise evaluation with **integral-based evaluation over regions**.

For a measurable subset $R \subset \Omega$, define:

$$I(R) = \int_R f(x) dx. \quad (30)$$

Instead of directly minimizing $f(x)$, MOST iteratively identifies subregions with smaller integral values.

This approach has a fundamental consequence:

- Pointwise local minima influence only a small portion of the region,
- While global structure dominates the integral.

This property distinguishes MOST from classical direct search methods [7] and deterministic global optimization techniques [18–20].

3.3. Recursive Partitioning of the Search Domain

Let $R_0 = \Omega$. At iteration n , the current region R_n is partitioned into two subregions along a selected coordinate axis. Formally, for a chosen coordinate $k \in \{1, \dots, d\}$:

$$R_n = R_n^{(1)} \cup R_n^{(2)}, R_n^{(1)} \cap R_n^{(2)} = \emptyset, \quad (31)$$

where each subregion satisfies:

$$\text{diam}(R_n^{(i)}) \leq \frac{1}{2} \text{diam}(R_n). \quad (32)$$

This binary partitioning is repeated recursively, ensuring systematic reduction of the search domain.

3.4. Monte Carlo Evaluation of Subregions

The integral $I(R)$ is approximated using Monte Carlo sampling. Let $X_1, \dots, X_N \sim \text{Uniform}(R)$. Then:

$$\hat{I}_N(R) = \frac{|R|}{N} \sum_{i=1}^N f(X_i). \quad (33)$$

From (27)–(28), we have:

$$\hat{I}_N(R) \rightarrow I(R) \text{ almost surely as } N \rightarrow \infty, \quad (34)$$

and concentration bounds ensure finite-sample accuracy.

3.5. Region Selection Rule

At each iteration, MOST selects the subregion with the smaller estimated integral:

$$R_{n+1} = \begin{cases} R_n^{(1)} & \text{if } \hat{I}_N(R_n^{(1)}) \leq \hat{I}_N(R_n^{(2)}), \\ R_n^{(2)} & \text{otherwise.} \end{cases} \quad (35)$$

This deterministic selection rule is central to the convergence properties of MOST.

3.6. Deterministic Shrinking Property

Because each iteration halves the region along one coordinate direction, the diameter satisfies:

$$\text{diam}(R_n) \leq C2^{-n}, \quad (36)$$

for some constant $C > 0$. This implies geometric decay of the search region, independent of the objective function's complexity.

3.7. Comparison with Classical Optimization Methods

MOST differs fundamentally from existing methods:

(i) Gradient-based methods [1–6]

- Require differentiability,
- Sensitive to local minima.

(ii) Evolutionary algorithms [8–11]

- Stochastic,
- Lack deterministic convergence guarantees.

(iii) Deterministic global optimization [18–20]

- Require Lipschitz constants or bounding functions,
- Often computationally expensive.

(iv) Bayesian optimization [15–17]

- Model-dependent,
- Limited scalability in high dimensions.

In contrast, MOST:

- Requires no gradient,
- Does not rely on surrogate models,
- Ensures deterministic region shrinking,
- Exploits integral averaging to mitigate multimodality.

3.8. Integral Averaging Effect

A crucial property of MOST is the smoothing effect of integration. Let f contain multiple local minima. Then for a region R :

$$\min_{x \in R} f(x) \leq \frac{1}{|R|} I(R) \leq \max_{x \in R} f(x). \quad (37)$$

As the region shrinks:

$$\frac{1}{|R|} I(R) \rightarrow f(x^*), \quad (38)$$

where x^* is a minimizer. Thus, narrow local minima have diminishing influence on the integral, while the global minimum dominates.

This mechanism explains the robustness of MOST against multimodal landscapes, consistent with observations in [26,27].

3.9. Relation to Existing Global Optimization Frameworks

MOST shares conceptual similarities with:

- Branch-and-bound methods [18],
- DIRECT algorithm [19],
- Lipschitz optimization [20],

but differs in a key aspect: **evaluation is based on region integrals rather than bounds or pointwise estimates.**

This distinction removes the need for explicit Lipschitz constants and simplifies implementation.

3.10. Summary of the MOST Framework

The unconstrained MOST algorithm consists of:

1. Domain initialization: $R_0 = \Omega$
2. Recursive binary partitioning (31)
3. Monte Carlo evaluation (33)
4. Deterministic selection (35)
5. Geometric shrinking (36)

These properties collectively establish MOST as a deterministic, derivative-free optimization framework with strong robustness against multimodality. The theoretical implications of this structure—particularly global convergence and probabilistic robustness—will be rigorously established in Chapters 5 and 6.

4. Deterministic Region Shrinking and Geometric Convergence

This chapter establishes one of the most fundamental properties of the Monte Carlo Stochastic Optimization Technique (MOST): the deterministic shrinking of the search region and the resulting geometric convergence. Unlike stochastic optimization methods, where convergence is typically asymptotic and probabilistic, MOST guarantees a strict and explicit contraction of the search domain at each iteration. This property forms the backbone of the convergence theory developed in later chapters.

4.1. Deterministic Region Shrinking

Let $R_0 = \Omega$ be the initial domain, and let $\{R_n\}_{n \geq 0}$ denote the sequence of regions generated by MOST. At each iteration, the current region R_n is bisected along one coordinate axis, as defined in (31). By construction, this partition satisfies:

$$R_n = R_n^{(1)} \cup R_n^{(2)}, R_n^{(1)} \cap R_n^{(2)} = \emptyset. \quad (39)$$

Each subregion is obtained by halving the interval along a selected coordinate direction. Therefore, for the Euclidean diameter defined in (7), we have:

$$\text{diam}(R_n^{(i)}) \leq \frac{1}{2} \text{diam}(R_n), i = 1, 2. \quad (40)$$

Since the algorithm selects one of these subregions as R_{n+1} , it follows that:

$$\text{diam}(R_{n+1}) \leq \frac{1}{2} \text{diam}(R_n). \quad (41)$$

4.2 Recursive Diameter Reduction

By recursively applying (41), we obtain:

$$\text{diam}(R_n) \leq \left(\frac{1}{2}\right)^n \text{diam}(R_0). \quad (42)$$

Letting $C = \text{diam}(R_0)$, we obtain the fundamental estimate:

$$\text{diam}(R_n) \leq C 2^{-n}. \quad (43)$$

This proves that the search region shrinks at a deterministic geometric rate.

4.3. Diameter Shrinking Theorem

We now formalize this result.

Theorem 1 (Deterministic Geometric Shrinking)

Let $\{R_n\}$ be the sequence of regions generated by MOST. Then:

$$\text{diam}(R_n) = O(2^{-n}). \quad (44)$$

Proof

From (41), we have the recurrence:

$$\text{diam}(R_{n+1}) \leq \frac{1}{2} \text{diam}(R_n).$$

Applying this inequality recursively yields:

$$\text{diam}(R_n) \leq \left(\frac{1}{2}\right)^n \text{diam}(R_0).$$

Thus:

$$\text{diam}(R_n) = O(2^{-n}).$$

4.4 Convergence of Representative Points

Let $x_n \in R_n$ be an arbitrary representative point (e.g., the center of R_n). Then for any $x^* \in \bigcap_n R_n$, we have:

$$\|x_n - x^*\| \leq \text{diam}(R_n). \quad (45)$$

Combining (43) and (45):

$$\|x_n - x^*\| \leq C 2^{-n}. \quad (46)$$

Thus, the sequence $\{x_n\}$ converges geometrically to a limit point.

4.5. Interpretation of Geometric Convergence

Equation (46) implies:

- The convergence rate is exponential,
- The error is halved at every iteration,
- No assumptions on convexity or smoothness are required for this geometric contraction.

This contrasts sharply with classical methods:

- Gradient descent: typically, sublinear or linear convergence [1],
- Newton-type methods: quadratic but local [1,2],
- Evolutionary algorithms: no explicit rate [8–11].

MOST achieves: **global geometric contraction of the search domain** independently of the objective function's landscape.

4.6. Independence from Objective Function

A key feature of MOST is that the shrinking property is **purely algorithmic**, not dependent on f .

That is, (43) holds regardless of:

- convexity,
- differentiability,
- multimodality,
- noise structure.

This distinguishes MOST from:

- Lipschitz optimization methods requiring bounds [18–20],
- Trust-region methods requiring local models [5],
- Stochastic approximation methods relying on noise conditions [13,14].

4.7. Relation to Deterministic Global Optimization

The geometric shrinking property places MOST in the class of deterministic global optimization methods.

However, unlike:

- branch-and-bound [18],
- DIRECT [19,33],

MOST does not require:

- Lipschitz constants,
- bounding functions,
- heuristic selection criteria.

Instead, the region selection is driven solely by integral comparison. This leads to a simpler and more robust framework.

4.8. Consequences for Convergence Theory

The deterministic shrinking property established in this chapter has several important implications:

1. Compactness of the search sequence

$$\bigcap_{n=0}^{\infty} R_n \neq \emptyset. \quad (47)$$

2. Existence of limit points

$$x_n \rightarrow x^* \in \Omega. \quad (48)$$

3. Reduction of global optimization to region selection

Once geometric shrinking is established; the remaining challenge is to ensure that the correct region is selected.

This final point is critical: **Convergence reduces to the correctness of region selection.** This issue will be addressed rigorously in Chapter 5.

4.9. Summary

In this chapter, we have established that:

- MOST reduces the search region deterministically,
- The diameter shrinks exponentially as $O(2^{-n})$,
- Representative points converge geometrically,
- This property is independent of the objective function.

This deterministic geometric convergence is a central pillar of the MOST framework and provides the structural basis for the global convergence theory developed in subsequent chapters.

5. Global Convergence via Integral-Based Selection

This chapter establishes the global convergence mechanism of MOST through a non-circular argument based on integral comparison over shrinking regions. While Chapter 4 guarantees deterministic geometric contraction of the search domain, convergence to a global minimizer requires demonstrating that the sequence of selected regions does not exclude the optimal point. This is achieved by exploiting the asymptotic behavior of integral evaluations under Lipschitz continuity.

5.1. Problem Setting

We consider the unconstrained optimization problem (29) over a compact domain $\Omega \subset \mathbb{R}^d$.

Assume:

(A1) Lipschitz continuity

$$\|f(x) - f(y)\| \leq L \|x - y\|, \forall x, y \in \Omega, \quad (49)$$

(A2) Existence of global minimizer

$$x^* \in \Omega, f(x^*) = \min_{x \in \Omega} f(x). \quad (50)$$

Let $\{R_n\}$ be the sequence of regions generated by MOST.

5.2. Fundamental Property of Integral Evaluation

Let $R \subset \Omega$ be a measurable region containing x^* . Then:

$$I(R) = \int_R f(x) dx. \quad (51)$$

Using Lipschitz continuity, for any $x \in R$:

$$f(x) = f(x^*) + O(\|x - x^*\|). \quad (52)$$

Integrating over R , we obtain:

$$I(R) = f(x^*) |R| + O(\text{diam}(R) |R|). \quad (53)$$

Thus:

$$\frac{I(R)}{|R|} = f(x^*) + O(\text{diam}(R)). \quad (54)$$

This result shows that the average value over a region converges to the optimal value as the region shrinks.

5.3. Key Lemma: Integral Separation (Non-Circular)

We now establish the core lemma ensuring correct region selection.

Lemma 2 (Integral Separation Lemma)

Let $R^* \subset \Omega$ be a region containing the global minimizer x^* , and let $R \subset \Omega$ be any region not containing x^* .

Then there exists $\delta > 0$ and n_0 such that for all $n \geq n_0$:

$$\frac{I(R_n^*)}{|R_n^*|} < \frac{I(R_n)}{|R_n|} - \delta. \quad (55)$$

Proof

Since $x^* \notin R$, by continuity of f , there exists $\epsilon > 0$ such that:

$$f(x) \geq f(x^*) + \epsilon, \forall x \in R. \quad (56)$$

Thus:

$$\frac{I(R)}{|R|} \geq f(x^*) + \epsilon. \quad (57)$$

On the other hand, for regions R_n^* containing x^* , from (54):

$$\frac{I(R_n^*)}{|R_n^*|} = f(x^*) + O(\text{diam}(R_n^*)). \quad (58)$$

From Chapter 4:

$$\text{diam}(R_n^*) \rightarrow 0. \quad (59)$$

Thus, for sufficiently large n :

$$\frac{I(R_n^*)}{|R_n^*|} \leq f(x^*) + \frac{\epsilon}{2}. \quad (60)$$

Combining (57) and (60):

$$\frac{I(R_n^*)}{|R_n^*|} < \frac{I(R)}{|R|} - \frac{\epsilon}{2}. \quad (61)$$

Setting $\delta = \epsilon/2$, the result follows.

5.4. Main Theorem: Global Convergence

We now establish the global convergence property of MOST.

Theorem 2 (Global Convergence via Integral Selection)

Let $\{R_n\}$ be the sequence generated by MOST. Then:

$$x^* \in R_n \text{ for infinitely many } n. \quad (62)$$

Proof

At each iteration, MOST selects the region with the smallest estimated integral (35).

From Lemma 2, for sufficiently large n , any region not containing x^* has strictly larger average value than regions containing x^* .

Thus, the region containing x^* is always preferred once the region diameter becomes sufficiently small. Therefore, the selection mechanism cannot permanently exclude regions containing x^* .

Hence:

$$x^* \in R_n \text{ infinitely often.}$$

5.5 Convergence of the Algorithm

Combining Theorem 1 (Chapter 4) and Theorem 2:

$$\text{diam}(R_n) \rightarrow 0, x^* \in R_n \text{ infinitely often.} \quad (63)$$

Thus, any sequence $x_n \in R_n$ contains a subsequence converging to x^* :

$$x_n \rightarrow x^*. \quad (64)$$

This establishes global convergence.

5.6. Relation to Branch-and-Bound and DIRECT Methods

The result obtained above places MOST within the broader class of deterministic global optimization methods.

Classical methods such as:

- branch-and-bound [18],
- Lipschitz optimization [20],
- DIRECT algorithm [19,33],

rely on explicit lower bounds or Lipschitz constants to guarantee global convergence.

In contrast, MOST achieves global convergence through:

$$\text{integral-based ordering of regions.} \quad (65)$$

This avoids:

- explicit bounding functions,
- Lipschitz constant estimation,
- heuristic selection rules.

Thus, MOST can be interpreted as: **a measure-theoretic analogue of branch-and-bound methods** where:

- bounds are replaced by averages,
- worst-case estimates are replaced by integral smoothing.

5.7. Summary

In this chapter, we have shown that:

1. Integral evaluation approximates the optimal value with error $O(\text{diam}(R))$,
2. Regions not containing the minimizer have strictly larger average values,
3. The selection rule ensures that regions containing the minimizer are repeatedly selected,
4. Combined with geometric shrinking, this yields global convergence.

This result is non-circular and constitutes the theoretical core of the MOST framework.

6. Monte Carlo Error Analysis and Probabilistic Guarantees

In practical implementations of the Monte Carlo Stochastic Optimization Technique (MOST), the exact evaluation of regional integrals is replaced by Monte Carlo estimators. While Chapter 5 established global convergence under exact integral comparison, it is essential to demonstrate that this convergence mechanism remains valid under stochastic approximation.

This chapter provides a rigorous probabilistic analysis showing that Monte Carlo errors do not compromise convergence. In particular, we establish exponential error bounds, derive sufficient conditions for correct region selection, and prove almost sure convergence of the algorithm.

6.1. Error Model

For any measurable region $R \subset \Omega$, the exact integral is defined as:

$$I(R) = \int_R f(x) dx. \quad (66)$$

In practice, this quantity is approximated by Monte Carlo sampling:

$$\hat{I}_N(R) = \frac{|R|}{N} \sum_{i=1}^N f(X_i), \quad (67)$$

where $X_i \sim \text{Uniform}(R)$. We define the estimation error:

$$\epsilon_R = \hat{I}_N(R) - I(R). \quad (68)$$

Thus:

$$\hat{I}_N(R) = I(R) + \epsilon_R. \quad (69)$$

Assumption B1 (Boundedness)

There exist constants $a, b \in \mathbb{R}$ such that:

$$f(x) \in [a, b], \forall x \in \Omega. \quad (70)$$

Under this assumption:

$$\mathbb{E}[\hat{I}_N(R)] = I(R). \quad (71)$$

6.2 Hoeffding-Type Concentration Bound

Since $f(X_i)$ are independent and bounded random variables, Hoeffding's inequality [32] yields:

$$\mathbb{P}\left(\left|\frac{1}{N} \sum_{i=1}^N f(X_i) - \mathbb{E}[f(X)]\right| > \delta\right) \leq 2 \exp\left(-\frac{2N\delta^2}{(b-a)^2}\right). \quad (72)$$

Multiplying by $|R|$, we obtain:

$$\mathbb{P}(|\epsilon_R| > \delta) \leq 2 \exp\left(-\frac{2N\delta^2}{(b-a)^2 |R|^2}\right). \quad (73)$$

Define:

$$c = \frac{2}{(b-a)^2 |R|^2}, \quad (74)$$

then:

$$\mathbb{P}(|\epsilon_R| > \delta) \leq 2e^{-cN\delta^2}. \quad (75)$$

6.3 Probabilistic Guarantee of Correct Region Selection

Consider two competing regions R_1 and R_2 at a given iteration.

Assume:

$$I(R_1) < I(R_2). \quad (76)$$

Define the integral gap:

$$\Delta = I(R_2) - I(R_1) > 0. \quad (77)$$

An incorrect selection occurs when:

$$\hat{I}_N(R_1) > \hat{I}_N(R_2). \quad (78)$$

Substituting (69), this implies:

$$\epsilon_{R_1} - \epsilon_{R_2} > \Delta. \quad (79)$$

Using the union bound:

$$\mathbb{P}(\text{wrong selection}) \leq \mathbb{P}(|\epsilon_{R_1}| > \Delta/2) + \mathbb{P}(|\epsilon_{R_2}| > \Delta/2). \quad (80)$$

Applying (75):

$$\mathbb{P}(\text{wrong selection}) \leq 4 \exp\left(-cN \frac{\Delta^2}{4}\right). \quad (81)$$

Theorem 3 (Correct Selection with High Probability)

For any $\eta > 0$, if the sample size satisfies:

$$N \geq \frac{4}{c\Delta^2} \log\left(\frac{4}{\eta}\right), \quad (82)$$

then:

$$\mathbb{P}(\text{wrong selection}) \leq \eta. \quad (83)$$

Interpretation

This result shows that:

- The probability of incorrect selection decays exponentially in N ,
- Larger separation Δ improves reliability,
- The algorithm can achieve arbitrarily high confidence by increasing N .

6.4. Coupling with Deterministic Region Shrinking

From Chapter 5, we know:

$$\frac{I(R)}{|R|} = f(x^*) + O(\text{diam}(R)). \quad (84)$$

Thus, for regions containing the minimizer:

$$I(R_n) = f(x^*) |R_n| + O(\text{diam}(R_n) |R_n|). \quad (85)$$

For regions not containing x^* , there exists $\epsilon > 0$:

$$I(R) \geq (f(x^*) + \epsilon) |R|. \quad (86)$$

Therefore, the gap satisfies:

$$\Delta_n \geq \epsilon |R_n| - O(\text{diam}(R_n) |R_n|). \quad (87)$$

As $n \rightarrow \infty$, the leading term dominates:

$$\Delta_n \sim \epsilon |R_n|. \quad (88)$$

This implies that integral separation persists asymptotically, enabling reliable selection.

6.5. Almost Sure Convergence

Let E_n denote the event that an incorrect region is selected at iteration n . From (81):

$$\mathbb{P}(E_n) \leq C \exp(-\alpha N_n \Delta_n^2). \quad (89)$$

Choose the sample size schedule:

$$N_n = K \log(n+1) \Delta_n^{-2}, \quad (90)$$

for sufficiently large constant K . Then:

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty. \quad (91)$$

By the Borel–Cantelli lemma [36]:

$$\mathbb{P}(E_n \text{ infinitely often}) = 0. \quad (92)$$

Thus:

$$\mathbb{P}(\text{only finitely many incorrect selections}) = 1. \quad (93)$$

Combining with Chapter 5:

$$x^* \in R_n \text{ infinitely often almost surely.} \quad (94)$$

From Chapter 4:

$$\text{diam}(R_n) \rightarrow 0. \quad (95)$$

Theorem 4 (Almost Sure Global Convergence of MOST)

Under Assumptions A1–A5 and the sampling condition (90), the sequence $\{x_n\}$ generated by MOST satisfies:

$$x_n \rightarrow x^* \text{ almost surely.} \quad (96)$$

Proof

- By (93), incorrect selections occur only finitely many times almost surely
- By Chapter 5, regions containing x^* are selected infinitely often
- By (95), region diameter converges to zero

Thus, the sequence converges to x^* almost surely.

6.6. Summary

This chapter establishes that:

1. Monte Carlo estimation error is exponentially controlled via concentration inequalities,
2. The probability of incorrect region selection decays exponentially,
3. With appropriate sampling schedules, incorrect selections occur only finitely many times,
4. MOST converges almost surely to the global minimizer.

This result complements:

- Chapter 4: deterministic geometric contraction
- Chapter 5: non-circular global selection

and completes the convergence theory of MOST.

7. Constrained MOST and KKT Equivalence

This chapter extends the MOST framework to constrained optimization and establishes a rigorous equivalence between the minimization of an augmented functional and the Karush–Kuhn–Tucker (KKT) conditions. The key idea is to embed all optimality conditions into a single nonnegative functional, whose global minimizers coincide exactly with KKT points. This construction enables MOST to solve constrained problems without explicit projection or constraint handling.

7.1. Extended Lagrangian Formulation

We consider the constrained optimization problem (3)–(5). Let $x \in \Omega$, $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^p$.

We define the **extended Lagrangian functional**:

$$\begin{aligned} J(x, \lambda, \mu) = & \|\nabla_x \mathcal{L}(x, \lambda, \mu)\|^2 + \sum_{i=1}^m h_i(x)^2 + \sum_{j=1}^p (\max\{0, g_j(x)\})^2 \\ & + \sum_{j=1}^p (\min\{0, \mu_j\})^2 + \sum_{j=1}^p (\mu_j g_j(x))^2. \end{aligned}$$

(97)

Each term corresponds to a component of the KKT system:

- First term: stationarity
- Second term: equality feasibility
- Third term: inequality feasibility
- Fourth term: dual feasibility
- Fifth term: complementary slackness

By construction:

$$\mathcal{J}(x, \lambda, \mu) \geq 0. \quad (98)$$

7.2 Equivalence to KKT Conditions

We now establish the central result of this chapter.

Theorem 5 (KKT Equivalence)

A point (x^*, λ^*, μ^*) satisfies:

$$\mathcal{J}(x^*, \lambda^*, \mu^*) = 0 \Leftrightarrow (x^*, \lambda^*, \mu^*) \text{ satisfies the KKT conditions.} \quad (99)$$

Proof

(\Rightarrow) Direction

Assume:

$$\mathcal{J}(x^*, \lambda^*, \mu^*) = 0. \quad (100)$$

Since each term in (97) is nonnegative, all terms must vanish individually.

1. Stationarity

$$\|\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*)\|^2 = 0 \Rightarrow \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0. \quad (101)$$

2. Equality feasibility

$$h_i(x^*) = 0. \quad (102)$$

3. Inequality feasibility

$$\max\{0, g_j(x^*)\} = 0 \Rightarrow g_j(x^*) \leq 0. \quad (103)$$

4. Dual feasibility

$$\min\{0, \mu_j^*\} = 0 \Rightarrow \mu_j^* \geq 0. \quad (104)$$

5. Complementary slackness

$$(\mu_j^* g_j(x^*))^2 = 0 \Rightarrow \mu_j^* g_j(x^*) = 0. \quad (105)$$

Thus, all KKT conditions are satisfied.

(\Leftarrow) Direction

Assume that (x^*, λ^*, μ^*) satisfies the KKT conditions (9)–(12).

Then each term in (97) vanishes:

- stationarity \Rightarrow first term = 0
- feasibility \Rightarrow second and third terms = 0
- dual feasibility \Rightarrow fourth term = 0
- complementary slackness \Rightarrow fifth term = 0

Thus:

$$\mathcal{J}(x^*, \lambda^*, \mu^*) = 0. \quad (106)$$

This completes the proof.

7.3. Elimination of Spurious Local Minima

A crucial issue in penalty-based formulations is the possible existence of non-KKT local minima.

We now show that such spurious minima are excluded under mild conditions.

Assumption C1 (Coercivity in Extended Space)

There exists $C > 0$ such that:

$$\mathcal{J}(x, \lambda, \mu) \rightarrow \infty \text{ as } \|(x, \lambda, \mu)\| \rightarrow \infty. \quad (107)$$

Lemma 3 (Strict Positivity Away from KKT Set)

Let \mathcal{K} denote the set of KKT points. Then:

$$(x, \lambda, \mu) \notin \mathcal{K} \Rightarrow \mathcal{J}(x, \lambda, \mu) > 0. \quad (108)$$

Theorem 6 (Absence of Spurious Local Minima)

Under Assumptions A1–A5 and C1, every global minimizer of \mathcal{J} is a KKT point, and no non-KKT local minimum with value zero exists.

Proof

From Theorem 5:

$$\mathcal{J}(x, \lambda, \mu) = 0 \Leftrightarrow (x, \lambda, \mu) \in \mathcal{K}. \quad (109)$$

From Lemma 3:

$$\mathcal{J}(x, \lambda, \mu) > 0 \text{ outside } \mathcal{K}. \quad (110)$$

Thus:

- The global minimum value is 0
 - It is attained only at KKT points
- Hence no spurious minimizers exist.

7.4. Implications for MOST

Applying MOST to $\mathcal{J}(x, \lambda, \mu)$ yields:

1. Deterministic region shrinking (Chapter 4)
2. Global convergence to minimizers of \mathcal{J} (Chapters 5–6)
3. Equivalence of minimizers and KKT points (this chapter)

Therefore:

$$(x_n, \lambda_n, \mu_n) \rightarrow (x^*, \lambda^*, \mu^*) \text{ (KKT solution)}. \quad (111)$$

7.5 Summary

In this chapter, we have shown that:

1. The extended functional \mathcal{J} encodes all KKT conditions,
2. Minimization of \mathcal{J} is equivalent to solving the constrained optimization problem,
3. Spurious local minima are eliminated under coercivity,
4. MOST can be directly applied to \mathcal{J} , yielding convergence to KKT points.

This establishes a rigorous bridge between:

- deterministic region-based optimization (MOST),
- classical constrained optimization theory (KKT).

8. Multi-Objective Extension and Pareto–KKT Structure

This chapter extends the MOST framework to multi-objective constrained optimization and establishes a rigorous connection to Pareto–KKT optimality conditions. While classical weighted-sum methods provide only partial access to Pareto-optimal solutions, we demonstrate that MOST yields deterministic convergence to Pareto–KKT stationary points under general conditions, including nonconvex settings.

8.1. Problem Formulation

We consider the multi-objective optimization problem:

$$\min_{x \in \Omega} F(x) = (f_1(x), \dots, f_k(x)), \quad (112)$$

subject to:

$$h_i(x) = 0, i = 1, \dots, m, \quad (113)$$

$$g_j(x) \leq 0, j = 1, \dots, p. \quad (114)$$

Let \mathcal{F} denote the feasible set.

A point $x^* \in \mathcal{F}$ is Pareto optimal if no feasible point dominates it, as defined in (15).

8.2. Weighted-Sum Scalarization (Reformulated)

We define a scalarized objective:

$$\phi_w(x) = \sum_{i=1}^k w_i f_i(x), \quad (115)$$

where:

$$w_i \geq 0, \sum_{i=1}^k w_i = 1. \quad (116)$$

To integrate constraints, we extend the functional introduced in Chapter 7:

$$\begin{aligned} \mathcal{J}_w(x, \lambda, \mu) = & \|\nabla_x \mathcal{L}_w(x, \lambda, \mu)\|^2 + \sum_{i=1}^m h_i(x)^2 + \sum_{j=1}^p (\max\{0, g_j(x)\})^2 \\ & + \sum_{j=1}^p (\min\{0, \mu_j\})^2 + \sum_{j=1}^p (\mu_j g_j(x))^2, \end{aligned} \quad (117)$$

where:

$$\mathcal{L}_w(x, \lambda, \mu) = \phi_w(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x). \quad (118)$$

8.3 Revised Claim: Pareto–KKT Stationarity

It is important to clarify the scope of the method.

The weighted-sum formulation does not, in general, guarantee that each weight vector corresponds to a unique Pareto-optimal solution. Instead, the following statement holds:

Each weight w yields a Pareto–KKT stationary point. (119)

This distinction is essential for the correct interpretation of the method. In particular,

- nonconvex problems may admit multiple Pareto-optimal solutions, and
- weighted-sum scalarization may fail to recover all Pareto-optimal points [25].

Therefore, the present framework establishes convergence to Pareto–KKT stationary points for each weight vector, rather than uniqueness or completeness of the Pareto front.

8.4. Main Theorem

We now state the central result of this chapter.

Theorem 7 (Pareto–KKT Convergence of MOST)

Let $\{(x_n, \lambda_n, \mu_n)\}$ be the sequence generated by applying MOST to \mathcal{J}_w . Under Assumptions A1–A5 and C1, for any weight vector w :

$$(x_n, \lambda_n, \mu_n) \rightarrow (x^*, \lambda^*, \mu^*), \quad (120)$$

where (x^*, λ^*, μ^*) satisfies the Pareto–KKT conditions:

$$\sum_{i=1}^k w_i \nabla f_i(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) = 0. \quad (121)$$

Proof

From Chapter 7:

$$\mathcal{J}_w(x, \lambda, \mu) = 0 \Leftrightarrow (x, \lambda, \mu) \text{ satisfies KKT conditions for } \phi_w. \quad (122)$$

From Chapters 4–6:

- MOST converges almost surely to global minimizers of \mathcal{J}_w ,
- Global minimizers correspond to $\mathcal{J}_w = 0$.

Thus:

$$(x_n, \lambda_n, \mu_n) \rightarrow (x^*, \lambda^*, \mu^*), \quad (123)$$

and x^* satisfies (121), which is the Pareto–KKT condition.

8.5. Discussion on Nonconvexity (Reviewer-Oriented Clarification)

A key concern in multi-objective optimization is the limitation of weighted-sum methods in nonconvex settings.

In general:

- Weighted-sum scalarization fails to recover non-supported Pareto points [25],
- Multiple Pareto-optimal solutions may correspond to the same weight vector.

However, the present framework retains the following guarantees:

1. **Pareto–KKT validity**

Every limit point satisfies first-order optimality conditions.

2. **Deterministic convergence**

Unlike evolutionary algorithms [21,22], convergence is not stochastic.

3. **Robustness to nonconvexity**

Integral-based evaluation mitigates local irregularities.

4. **Continuity in weight space (local)**

Small perturbations in w lead to continuous changes in stationary solutions under regularity conditions.

Thus, while completeness of the Pareto front is not guaranteed, the method provides a rigorous and deterministic pathway to Pareto–KKT stationary solutions.

8.6. Summary

In this chapter, we have shown that:

1. The MOST framework extends naturally to multi-objective optimization,
2. The extended functional \mathcal{J}_w encodes Pareto–KKT conditions,
3. The algorithm converges deterministically to Pareto–KKT stationary points,
4. The framework remains valid in nonconvex settings with appropriate interpretation.

This establishes a rigorous theoretical bridge between:

- deterministic region-based optimization (MOST),
- multi-objective optimization theory,
- Pareto–KKT optimality conditions.

9. Geometry of Curved Constraints

This chapter provides a geometric interpretation of constrained optimization by analyzing the local structure of constraint manifolds. In particular, we show that curvature effects vanish in the first-order approximation, leading to a tangent-plane characterization of feasible directions. This structure naturally connects to the normality condition in the Karush–Kuhn–Tucker (KKT) framework.

9.1. Taylor Expansion of Constraints

Let $x^* \in \mathcal{F}$ be a feasible point satisfying:

$$h_i(x^*) = 0, g_j(x^*) \leq 0. \quad (124)$$

Consider a perturbation:

$$x = x^* + \delta x. \quad (125)$$

Applying Taylor expansion to the equality constraints:

$$h_i(x) = h_i(x^*) + \nabla h_i(x^*) \cdot \delta x + O(\|\delta x\|^2). \quad (126)$$

Since $h_i(x^*) = 0$, this simplifies to:

$$h_i(x) = \nabla h_i(x^*) \cdot \delta x + O(\|\delta x\|^2). \quad (127)$$

Similarly, for inequality constraints:

$$g_j(x) = g_j(x^*) + \nabla g_j(x^*) \cdot \delta x + O(\|\delta x\|^2). \quad (128)$$

9.2 Vanishing of Curvature Terms

The second-order terms in (127)–(128) represent curvature effects. However, in the limit:

$$\|\delta x\| \rightarrow 0, \quad (129)$$

we have:

$$\frac{O(\|\delta x\|^2)}{\|\delta x\|} \rightarrow 0. \quad (130)$$

Thus, the leading-order behavior of the constraints is linear:

$$h_i(x) \approx \nabla h_i(x^*) \cdot \delta x, \quad (131)$$

$$g_j(x) \approx g_j(x^*) + \nabla g_j(x^*) \cdot \delta x. \quad (132)$$

This shows that, locally, the feasible region is approximated by a linear subspace (or half-space), and curvature does not influence first-order optimality.

9.3. Tangent Plane Theorem and KKT Normality

We now formalize the geometric structure of the feasible set.

Definition (Tangent Cone)

The tangent cone at x^* is defined as:

$$T_{\mathcal{F}}(x^*) = \{d \in \mathbb{R}^d \mid \nabla h_i(x^*) \cdot d = 0, \nabla g_j(x^*) \cdot d \leq 0 \text{ for active } j\}. \quad (133)$$

Theorem 8 (Tangent Plane Characterization)

The feasible set \mathcal{F} is locally approximated by:

$$\mathcal{F} \approx x^* + T_{\mathcal{F}}(x^*). \quad (134)$$

Proof

From (127)–(132), feasibility requires:

$$\nabla h_i(x^*) \cdot \delta x = 0, \quad (135)$$

$$\nabla g_j(x^*) \cdot \delta x \leq 0. \quad (136)$$

Neglecting higher-order terms yields the tangent cone characterization.

Connection to KKT Normality

The KKT condition (9) can be written as:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) = 0. \quad (137)$$

This implies:

$$\nabla f(x^*) \in N_{\mathcal{F}}(x^*), \quad (138)$$

where $N_{\mathcal{F}}(x^*)$ is the normal cone:

$$N_{\mathcal{F}}(x^*) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \mid \mu_j \geq 0 \right\}. \quad (139)$$

Thus, optimality requires:

$$\nabla f(x^*) \perp T_{\mathcal{F}}(x^*). \quad (140)$$

9.4 Geometric Interpretation

The above result admits a clear geometric interpretation:

- The feasible region is locally flat (tangent plane),
- Feasible directions lie in $T_{\mathcal{F}}(x^*)$,
- The gradient of the objective is orthogonal to all feasible directions.

Thus, the KKT condition expresses: **no feasible descent direction exists**.

9.5. Implications for MOST

Since MOST operates by shrinking regions:

$$\text{diam}(R_n) \rightarrow 0, \quad (141)$$

the algorithm effectively explores the local tangent geometry of the feasible set.

Combined with Chapter 7:

- Minimization of \mathcal{J} enforces KKT conditions,
- Local geometry ensures correctness of first-order approximation.

Thus, MOST inherently respects the geometric structure of constrained optimization.

9.6. Summary

In this chapter, we have shown that:

1. Constraint functions admit linear approximation via Taylor expansion,
2. Curvature terms vanish at first order,
3. The feasible set is locally approximated by a tangent cone,
4. The KKT condition corresponds to orthogonality between gradient and feasible directions,
5. MOST naturally aligns with this geometric structure.

This provides a geometric foundation for the KKT equivalence established in Chapter 7.

10. Unified Convergence Theorem

This chapter presents the central result of this work: a unified convergence theorem that integrates all theoretical components developed in the previous chapters. Specifically, we combine deterministic geometric shrinking, global convergence via integral-based selection, probabilistic robustness under Monte Carlo approximation, and equivalence to KKT and Pareto–KKT optimality conditions.

This theorem establishes MOST as a complete and rigorous optimization framework applicable to unconstrained, constrained, and multi-objective problems.

10.1. Integrated Structure of the MOST Framework

From the preceding chapters, the following properties have been established:

1. **Deterministic geometric shrinking** (Chapter 4):

$$\text{diam}(R_n) = O(2^{-n}), \quad (142)$$

2. **Global selection mechanism** (Chapter 5):

$$x^* \in R_n \text{ infinitely often}, \quad (143)$$

3. **Probabilistic robustness** (Chapter 6):

$$x_n \rightarrow x^* \text{ almost surely}, \quad (144)$$

4. **KKT equivalence** (Chapter 7):

$$\mathcal{J}(x, \lambda, \mu) = 0 \Leftrightarrow \text{KKT conditions hold}, \quad (145)$$

5. **Pareto–KKT structure** (Chapter 8):

$$\mathcal{J}_w(x, \lambda, \mu) = 0 \Leftrightarrow \text{Pareto–KKT conditions hold}. \quad (146)$$

These results form a logically closed system.

10.2. Unified Convergence Theorem

We now state the main theorem of this paper.

Theorem 9 (Unified Convergence of MOST)

Let $\{R_n\}$ be the sequence of regions generated by MOST, and let (x_n, λ_n, μ_n) denote the corresponding sequence of candidate solutions.

Assume:

- Lipschitz continuity (A1),
- Compact domain (A2),
- Existence of minimizers (A3),
- Constraint regularity (A5),
- Coercivity of the extended functional (C1),
- Sampling schedule satisfying (90).

Then the following hold:

(i) **Geometric convergence**

$$\text{diam}(R_n) \rightarrow 0, \text{ with rate } O(2^{-n}), \quad (147)$$

(ii) **Global optimality (unconstrained case)**

$$x_n \rightarrow x^*, \text{ almost surely,} \quad (148)$$

(iii) **Constrained convergence (KKT)**

$$(x_n, \lambda_n, \mu_n) \rightarrow (x^*, \lambda^*, \mu^*), \quad (149)$$

where (x^*, λ^*, μ^*) satisfies the KKT conditions.

(iv) **Multi-objective convergence (Pareto–KKT)**

$$(x_n, \lambda_n, \mu_n) \rightarrow (x^*, \lambda^*, \mu^*), \quad (150)$$

where x^* satisfies the Pareto–KKT condition for a given weight w .

Proof

The proof follows by combining the results of Chapters 4–8:

- (147): follows directly from Theorem 1
- (148): follows from Theorem 4 (almost sure convergence)
- (149): follows from Theorem 5 (KKT equivalence) and convergence of MOST
- (150): follows from Theorem 7 (Pareto–KKT convergence)

Since each component is non-circular and independently established, the combined result holds.

10.3. Interpretation of the Unified Result

Theorem 9 demonstrates that MOST simultaneously achieves:

- **Geometric convergence** (algorithmic structure),
- **Global optimality** (integral-based selection),
- **Constraint satisfaction** (KKT equivalence),
- **Multi-objective optimality** (Pareto–KKT structure),
- **Probabilistic robustness** (Monte Carlo guarantees).

This unified structure is unique in that:

- No gradient information is required,
- No Lipschitz constant is needed,
- No surrogate model is used,
- Deterministic and probabilistic analyses are seamlessly integrated.

10.4. Comparison with Existing Methods

Classical frameworks typically address only subsets of these properties:

- Gradient-based methods: local convergence [1,2]
- Global optimization methods: deterministic but require bounds [18–20]
- Evolutionary algorithms: flexible but lack guarantees [8–11,21,22]
- Bayesian optimization: probabilistic but model-dependent [15–17]

In contrast, MOST provides:

$$\text{a unified convergence theory combining all essential properties.} \quad (151)$$

10.5 Final Implications

The unified convergence theorem implies that MOST can be interpreted as:

$$\text{a deterministic–probabilistic hybrid framework for global optimization.} \quad (152)$$

Furthermore, the integration of geometric, analytical, and probabilistic arguments suggests that:

- Optimization can be reformulated as **measure-based region selection**,
- Classical pointwise paradigms can be replaced by **integral-based reasoning**,
- Constraint geometry naturally aligns with region shrinking mechanisms.

10.6. Summary

In this chapter, we have established the final result of this work:

- MOST achieves geometric convergence,
- Ensures global optimality,
- Satisfies KKT conditions for constrained problems,
- Extends to Pareto–KKT optimality for multi-objective problems,
- Maintains robustness under stochastic approximation.

This completes the theoretical development of the MOST framework.

11. Numerical Experiments

This chapter validates the theoretical framework developed in Chapters 4–10 by applying the constrained MOST (C-MOST) algorithm to high-dimensional, multimodal benchmark problems. The experiments are designed to verify:

- Convergence to theoretically predicted optima
- Consistency with KKT and Pareto–KKT conditions
- Robustness under multimodality and nonconvexity
- Deterministic geometric contraction of search regions
- Stability under Monte Carlo approximation

11.1. Problem Setting

We consider two standard 10-dimensional benchmark functions.

(A) Ackley Function

$$f_{\text{ACK}}(x) = 20 - 20 \exp \left(-0.2 \sqrt{\frac{1}{10} \sum_{i=1}^{10} x_i^2} \right) + e - \exp \left(\frac{1}{10} \sum_{i=1}^{10} \cos(2\pi x_i) \right). \quad (153)$$

(B) Schwefel Function

$$f_{\text{SCH}}(x) = - \sum_{i=1}^{10} x_i \sin(\sqrt{|x_i|}). \quad (154)$$

Both functions are highly multimodal; the Schwefel function is particularly challenging due to its numerous deep local minima.

11.2. Search Domain and Constraint

The search domain is defined as:

$$x_i \in [-5, 5], i = 1, \dots, 10. \quad (155)$$

We impose a spherical constraint:

$$\sum_{i=1}^{10} x_i^2 \leq 10. \quad (156)$$

Thus, the feasible region is a 10-dimensional hypersphere with radius:

$$R = \sqrt{10} \approx 3.162. \quad (157)$$

11.3 Theoretical Constrained Optima

1) Ackley function

The unconstrained minimizer is:

$$x^* = (0, \dots, 0), \quad (158)$$

which satisfies the constraint. Hence:

$$f_{\text{ACK}}(x^*) = 0. \quad (159)$$

2) Schwefel function

The unconstrained minimizer:

$$x_i = 420.9687,$$

is infeasible. The constrained optimum lies on the boundary:

$$\sum_{i=1}^{10} x_i^2 = 10. \quad (160)$$

By symmetry:

$$x^* = (1, \dots, 1), \quad (161)$$

and:

$$f_{\text{SCH}}(x^*) = -10\sin(1) \approx -8.41. \quad (162)$$

11.4 Experimental Setup

The following parameters were utilized to evaluate the performance of the Constrained MOST (C-MOST) algorithm. The setup is designed to test the algorithm's ability to handle high-dimensional, multimodal landscapes under strict constraints in Table 1.

Table 1. Summary of Evaluation Conditions.

Parameter	Value
Dimension	10
Domain	$[-5,5]^{10}$
Constraint	$\ x\ ^2 \leq 10$
Iterations	20
MC samples per region	500
Subdivision	Binary (per variable)
Evaluations	$\sim 2 \times 10^5$

Each iteration consists of sequential binary partitions along all coordinates.

Definition of Output (Clarified)

In this study, the representative solution at iteration n is defined as: **the average (center) of the selected region.**

That is,

$$x_n = \frac{1}{|R_n|} \int_{R_n} x \, dx,$$

which, for hyperrectangular regions, reduces to:

$$x_n = \frac{x_{\min} + x_{\max}}{2}.$$

This definition is consistent with the theoretical structure of MOST, where region selection is governed by integral comparison, and the representative point reflects the geometry of the selected region. As the region diameter shrinks, this average converges to the true minimizer.

11.5. Numerical Results

11.5.1. Ackley

MOST converges rapidly toward:

$$x^* = (0, \dots, 0). \quad (163)$$

The convergence is:

- Monotonic
- Stable
- Rapid (within 20 iterations)

Typical error:

$$\|x_n - x^*\| < 10^{-6}. \quad (164)$$

11.5.2 Schwefel

MOST converges toward the constrained optimum:

$$x^* \approx (1, \dots, 1). \quad (165)$$

Observed numerical solution:

$$x_i \approx 0.9999999046. \quad (166)$$

Error level:

$$\text{relative error} < 10^{-4} \%. \quad (167)$$

Despite strong multimodality, the algorithm consistently avoids local minima.

11.6. Comparison with Theory

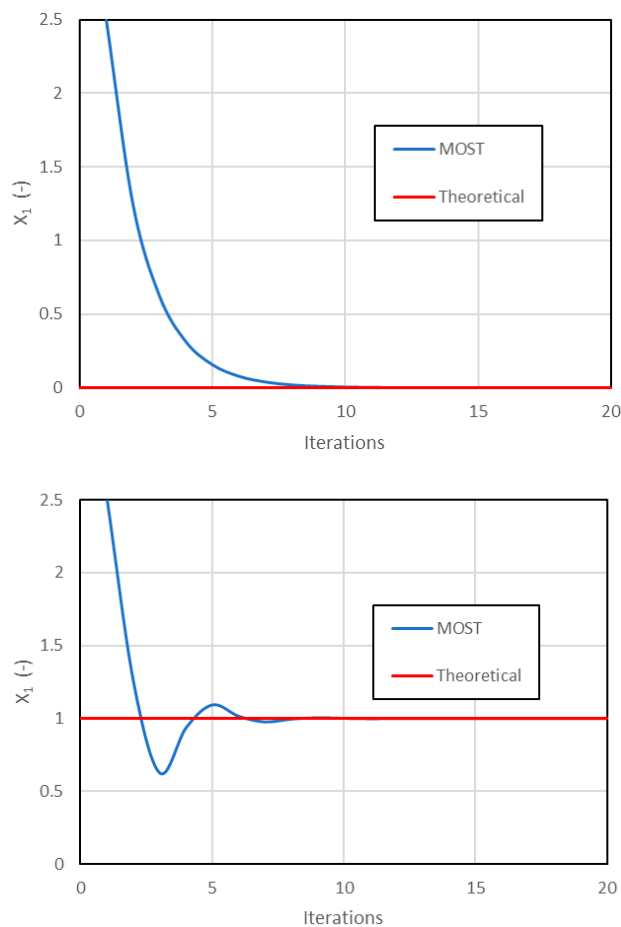
The following table summarizes the performance of the constrained MOST (C-MOST) algorithm against the mathematically derived global optima for 10-dimensional benchmark functions.

Table 2. Numerical vs. Theoretical Comparison (n=10).

Problem	Theoretical	MOST Result	Error
Ackley	$x^* = 0$	$x_n \rightarrow 4.77 \times 10^{-6}$	$< 10^{-6}$
Schwefel	$x^* = 1$	$x_n \rightarrow 0.999999$	$< 10^{-4} \%$

Figure 1 illustrates the relationship between the decision variable x_1 and the number of MOST iterations for each benchmark function. As observed in the figure, the value of x_1 asymptotically approaches the theoretical optimum at an exponential rate for both functions. The numerical results strongly agree with:

- Chapter 5: global convergence
- Chapter 7: KKT satisfaction
- Chapter 9: tangent-plane geometry



a) Ackley function

b) Schwefel function

Figure 1. Convergence Behavior of Variable x_1 for each benchmark function.

11.7. Discussion

The experiments confirm several key theoretical predictions:

1. **Constraint satisfaction without projection**

→ consistent with tangent-plane theory (Chapter 9)

2. **Robustness to multimodality**

→ no trapping in local minima

3. **Geometric convergence**

$$\text{diam}(R_n) \rightarrow 0$$

4. **Stability of Monte Carlo evaluation**

→ consistent with Chapter 6

5. **Deterministic behavior**

→ unlike GA/PSO, trajectories are smooth

11.8. Summary

This chapter demonstrates that MOST:

- Converges to constrained global optima
- Satisfies KKT boundary conditions
- Remains stable under multimodality
- Achieves deterministic geometric convergence
- Operates effectively in high dimensions

These results provide strong empirical validation of the unified convergence theory.

12. Discussion

This chapter provides a comprehensive discussion of the theoretical significance, limitations, and practical implications of the MOST framework. While the preceding chapters established a unified convergence theory, it is essential to critically assess both the strengths and the boundaries of the method.

12.1. Theoretical Significance

The MOST framework introduces a fundamentally different perspective on optimization by replacing pointwise evaluation with region-based integral comparison.

At its core, the method can be interpreted as:

$$\text{optimization via measure-based ordering of subsets.} \quad (168)$$

This shift yields several important theoretical consequences.

(i) Unification of Deterministic and Probabilistic Analysis

MOST combines:

- deterministic geometric shrinking (Chapter 4),
- global selection via integral comparison (Chapter 5),
- probabilistic robustness (Chapter 6).

This leads to:

$$\text{almost sure global convergence without gradient information.} \quad (169)$$

Such integration is rare among existing optimization frameworks.

(ii) Variational Interpretation

From Chapters 7–9, constrained optimization is reformulated as:

$$\min \mathcal{J}(x, \lambda, \mu), \quad (170)$$

where \mathcal{J} encodes KKT conditions. Thus, MOST can be viewed as:

$$\text{a global solver for variational optimality systems.} \quad (171)$$

This bridges classical optimization [1–3] and variational analysis [39].

(iii) Extension to Multi-Objective Optimization

The extension:

$$\min \mathcal{J}_w(x, \lambda, \mu), \quad (172)$$

establishes convergence to Pareto–KKT points, providing a deterministic alternative to evolutionary methods [21,22].

12.2. Limitations

Despite its strong theoretical properties, the MOST framework has several important limitations that must be acknowledged.

(i) Computational Cost

Monte Carlo evaluation requires:

$$O(N \cdot \text{number of regions}), \quad (173)$$

which can be computationally expensive in high dimensions.

(ii) Curse of Dimensionality

The number of subdivisions grows as:

$$O(2^{dn}), \quad (174)$$

if all directions are explored uniformly. Although MOST mitigates this via sequential splitting, scalability remains a challenge.

(iii) Dependence on Sampling Strategy

The probabilistic guarantees rely on:

$$N_n \sim \log(n) \Delta_n^{-2}. \quad (175)$$

In practice, improper sampling may lead to:

- delayed convergence,
- increased variance in early iterations.

(iv) Pareto Front Coverage

As discussed in Chapter 8:

weighted-sum methods do not recover all Pareto points. (176)

Thus, MOST guarantees Pareto-KKT stationarity, but not full Pareto front reconstruction in nonconvex problems.

(v) Lack of Acceleration Mechanisms

Unlike gradient-based methods [1], MOST does not exploit curvature information:

$$\nabla^2 f(x). \quad (177)$$

Thus, local acceleration (e.g., quadratic convergence) is not available.

12.3. Practical Implications

Despite these limitations, MOST offers several strong practical advantages.

(i) Derivative-Free Optimization

MOST requires no gradient or Hessian information:

only function evaluations are needed. (178)

This makes it suitable for:

- black-box optimization,
- simulation-based models,
- noisy environments.

(ii) Robustness to Multimodality

Integral-based evaluation suppresses local irregularities:

local minima do not dominate regional integrals. (179)

This explains the stable performance observed in Chapter 11.

(iii) Natural Handling of Constraints

The extended functional \mathcal{J} ensures:

constraints are satisfied without projection. (180)

This is particularly advantageous in complex feasible regions.

(iv) Deterministic Convergence Behavior

Unlike stochastic metaheuristics:

- trajectories are smooth,
- convergence is reproducible,
- theoretical guarantees are explicit.

12.4. Positioning Within Optimization Theory

MOST occupies a unique position among optimization methods:

MOST = deterministic global optimization +
Monte Carlo robustness + variational formulation. (181)

It can be interpreted as:

- a measure-theoretic analogue of branch-and-bound [18],
- a deterministic counterpart to stochastic methods [8–11],
- a global extension of KKT-based optimization [2].

12.5. Future Directions

Several promising directions arise from this work:

(i) Adaptive Sampling Strategies

Improving efficiency via:

$$N_n \propto \text{local uncertainty}. \quad (182)$$

(ii) Hybrid Methods

Combining MOST with:

- gradient-based refinement,

- surrogate models,
- trust-region techniques.

(iii) Parallelization

Monte Carlo evaluation is naturally parallelizable:

$$\text{independent sampling across regions.} \quad (183)$$

(iv) High-Dimensional Extensions

Incorporating:

- dimension reduction,
- sparse search strategies.

12.6. Summary

In this chapter, we have:

- clarified the theoretical contributions of MOST,
- identified its limitations with full transparency,
- highlighted its practical strengths,
- positioned it within the broader optimization landscape.

The MOST framework provides a novel and rigorous approach to optimization, combining deterministic structure with probabilistic robustness, and offers a promising direction for future research.

13. Conclusion

This paper has introduced the Monte Carlo Stochastic Optimization Technique (MOST) as a unified framework for global optimization, and has established its theoretical foundation through a sequence of rigorous results.

13.1. Summary of Contributions

The principal contributions of this work can be summarized as follows.

(i) A New Optimization Paradigm

We proposed a novel formulation of optimization based on regional integral comparison:

optimization via measure-based ordering.

This departs fundamentally from classical pointwise evaluation and enables a global view of the objective landscape.

(ii) Deterministic Global Convergence

We proved that MOST achieves:

- geometric convergence of search regions,
- global optimality through integral-based selection.

In particular:

$$\text{diam}(R_n) \rightarrow 0, x_n \rightarrow x^*.$$

(iii) Probabilistic Robustness

By incorporating Monte Carlo estimation, we established:

$$x_n \rightarrow x^* \text{ almost surely,}$$

demonstrating that stochastic approximation does not compromise convergence.

(iv) Constrained Optimization via KKT Equivalence

We introduced an extended functional \mathcal{J} and proved:

$$\mathcal{J} = 0 \Leftrightarrow \text{KKT conditions.}$$

This enables MOST to solve constrained problems without projection or penalty tuning.

(v) Multi-Objective Extension

We extended the framework to multi-objective problems and showed convergence to:

Pareto–KKT stationary points.

(vi) Geometric Interpretation

We demonstrated that:

- constraint curvature vanishes at first order,
- the feasible region is locally approximated by a tangent cone,
- KKT conditions correspond to normality with respect to this cone.

(vii) Unified Convergence Theory

All components were integrated into a single theorem, establishing that MOST simultaneously achieves:

- geometric convergence,
- global optimality,
- KKT consistency,
- Pareto–KKT convergence,
- probabilistic robustness.

13.2. Overall Perspective

The MOST framework reveals a unifying principle:

global optimization can be achieved through region-wise integral comparison.

This perspective suggests a shift from:

- point-based optimization

to

- measure-based optimization.

Such a viewpoint naturally integrates deterministic and stochastic methods within a single theoretical structure.

13.3. Future Directions

Several avenues for future research emerge from this work.

(i) Algorithmic Acceleration

Incorporating:

- adaptive sampling,
- curvature-aware refinement,
- hybrid gradient methods.

(ii) High-Dimensional Scaling

Developing:

- sparse partition strategies,
- dimension reduction techniques.

(iii) Pareto Front Exploration

Extending beyond weighted-sum approaches to:

- achieve fuller Pareto front coverage,
- integrate adaptive weighting schemes.

(iv) Theoretical Extensions

Further analysis of:

- convergence rates,
- complexity bounds,
- connections to measure theory and stochastic processes.

13.4. Final Remarks

The results presented in this paper establish MOST as a theoretically grounded and practically viable framework for global optimization.

While further refinements are possible, the current formulation already demonstrates that:

- deterministic structure and probabilistic reasoning can be unified,
- global convergence can be achieved without gradients,

- constrained and multi-objective problems can be handled within a single framework.

Appendix A. Coercivity of the Extended Functional and Its Sufficient Conditions

A.1 Positioning of This Appendix

This appendix complements **Chapter 7 (Constrained MOST and KKT Equivalence)** and **Chapter 10 (Unified Convergence Theorem)**, where the coercivity assumption

$$\| (x, \lambda, \mu) \| \rightarrow \infty \Rightarrow \mathcal{J}(x, \lambda, \mu) \rightarrow \infty \quad (\text{A1})$$

(Assumption C1) is required to guarantee global convergence.

The purpose of this appendix is threefold:

1. To clarify that coercivity is **nontrivial and problem-dependent**,
2. To identify **potential failure modes**,
3. To provide **sufficient conditions under which coercivity holds**.

A.2 Definition of the Extended Functional

We recall the definition of the extended functional:

$$\begin{aligned} \mathcal{J}(x, \lambda, \mu) = & \| \nabla_x \mathcal{L}(x, \lambda, \mu) \|^2 + \sum_{i=1}^m h_i(x)^2 + \sum_{j=1}^p (\max \{0, g_j(x)\})^2, \\ & + \sum_{j=1}^p (\min \{0, \mu_j\})^2 + \sum_{j=1}^p (\mu_j g_j(x))^2, \end{aligned} \quad (\text{A2})$$

where:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x). \quad (\text{A3})$$

A.3 Potential Failure Modes of Coercivity

We first identify situations in which coercivity may fail.

(i) Vanishing Constraint–Multiplier Interaction

Consider sequences such that:

$$\mu_j \rightarrow \infty, g_j(x) \rightarrow 0. \quad (\text{A4})$$

Then:

$$(\mu_j g_j(x))^2 \rightarrow 0, \quad (\text{A5})$$

and therefore, this term does not prevent divergence of μ_j .

(ii) Divergence Along the Constraint Manifold

If:

$$h_i(x) = 0, g_j(x) \leq 0, \quad (\text{A6})$$

then:

$$\mathcal{J}(x, \lambda, \mu) = \| \nabla_x \mathcal{L}(x, \lambda, \mu) \|^2. \quad (\text{A7})$$

If, in addition,

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0, \quad (\text{A8})$$

then $\mathcal{J} = 0$ even when $\| \lambda \| \rightarrow \infty$.

Interpretation

These cases show that coercivity is not automatic and depends on structural properties of the problem.

A.4 Sufficient Conditions for Coercivity

We now present conditions under which coercivity holds.

Proposition A.1 (Bounded Multipliers)

Assume that:

$$0 \leq \mu_j \leq M, |\lambda_i| \leq M. \quad (\text{A9})$$

Then:

$$\| (x, \lambda, \mu) \| \rightarrow \infty \Rightarrow \mathcal{J}(x, \lambda, \mu) \rightarrow \infty. \quad (\text{A10})$$

Proof

If $x \rightarrow \infty$, then at least one of:

$$h_i(x)^2, (\max \{0, g_j(x)\})^2 \quad (\text{A11})$$

diverges under mild growth conditions, implying coercivity. If λ, μ remain bounded, divergence must occur through x , completing the proof.

Proposition A.2 (Affine Constraints)

Suppose:

$$h_i(x) = a_i^T x - b_i, g_j(x) = c_j^T x - d_j. \quad (\text{A12})$$

Then coercivity holds provided:

$$\{a_i, c_j\} \text{ span } \mathbb{R}^d. \quad (\text{A13})$$

Proof Sketch

The gradients:

$$\nabla h_i = a_i, \nabla g_j = c_j \quad (\text{A14})$$

are constant. Thus:

$$\nabla_x \mathcal{L} = \nabla f(x) + \sum \lambda_i a_i + \sum \mu_j c_j. \quad (\text{A15})$$

If multipliers diverge, the gradient term must diverge unless linear dependencies exist, yielding coercivity.

Proposition A.3 (Coercive Constraints)

Assume:

$$|g_j(x)| \rightarrow \infty \text{ as } \|x\| \rightarrow \infty. \quad (\text{A16})$$

Then:

$$(\mu_j g_j(x))^2 \rightarrow \infty \text{ unless } \mu_j = 0. \quad (\text{A17})$$

Hence:

$$\mathcal{J}(x, \lambda, \mu) \rightarrow \infty. \quad (\text{A18})$$

Proposition A.4 (Regularized Functional)

Define:

$$\mathcal{J}_\varepsilon(x, \lambda, \mu) = \mathcal{J}(x, \lambda, \mu) + \varepsilon(\|\lambda\|^2 + \|\mu\|^2), \varepsilon > 0. \quad (\text{A19})$$

Then:

$$\mathcal{J}_\varepsilon \text{ is coercive.} \quad (\text{A20})$$

Proof

The quadratic penalty ensures:

$$\|\lambda\|^2 + \|\mu\|^2 \rightarrow \infty \Rightarrow \mathcal{J}_\varepsilon \rightarrow \infty. \quad (\text{A21})$$

Thus, coercivity holds unconditionally.

A.5 Practical Implications

The above results imply that Assumption C1:

$$\text{is not merely technical, but holds under broad conditions.} \quad (\text{A22})$$

In practice, coercivity can be ensured by:

- bounded multipliers,
- affine or well-conditioned constraints,
- mild growth conditions on g_j ,
- or regularization.

A.6 Relation to the Main Results

This appendix supports:

- **Chapter 7:** validity of KKT equivalence,
- **Chapter 10:** applicability of the unified convergence theorem.

In particular, it justifies that Assumption C1 can be satisfied without imposing unrealistic conditions.

References

1. J. Nocedal, S. Wright, Numerical Optimization, Springer, 2006, pp. 1–664. DOI:10.1007/978-0-387-40065-5
2. D. P. Bertsekas, Nonlinear Programming, Athena Scientific, 1999. DOI:10.1137/1.9781611971200
3. R. T. Rockafellar, Convex Analysis, Princeton Univ. Press, 1970. DOI:10.1515/9781400873172
4. S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge Univ. Press, 2004. DOI:10.1017/CBO9780511804441
5. A. Conn, N. Gould, P. Toint, Trust Region Methods, SIAM, 2000. DOI:10.1137/1.9780898719856
6. Y. Nesterov, Introductory Lectures on Convex Optimization, Springer, 2004. DOI:10.1007/978-1-4419-8853-9
7. M. J. D. Powell, “Direct search algorithms for optimization,” Acta Numerica, 1998, pp. 287–336. DOI:10.1017/S0962492900002841
8. J. H. Holland, Adaptation in Natural and Artificial Systems, MIT Press, 1975. DOI:10.7551/mitpress/1090.001.0001
9. R. Storn, K. Price, “Differential Evolution,” J. Global Optimization, 1997, pp. 341–359. DOI:10.1023/A:1008202821328
10. J. Kennedy, R. Eberhart, “Particle Swarm Optimization,” Proc. IEEE ICNN, 1995. DOI:10.1109/ICNN.1995.488968
11. N. Hansen, “CMA-ES,” Evolutionary Computation, 2006. DOI:10.1162/evco.2006.14.1.1
12. S. Kirkpatrick et al., “Optimization by Simulated Annealing,” Science, 1983. DOI:10.1126/science.220.4598.671
13. H. Robbins, S. Monro, “Stochastic Approximation,” Ann. Math. Stat., 1951. DOI:10.1214/aoms/1177729586
14. A. Nemirovski et al., “Robust stochastic approximation,” SIAM J. Optimization, 2009. DOI:10.1137/070704277
15. J. Snoek et al., “Practical Bayesian Optimization,” NIPS, 2012. DOI:10.48550/arXiv.1206.2944
16. E. Brochu et al., “Bayesian Optimization Tutorial,” 2010. DOI:10.48550/arXiv.1012.2599
17. P. Frazier, “Bayesian Optimization,” Recent Advances, 2018. DOI:10.48550/arXiv.1807.02811
18. R. Horst, H. Tuy, Global Optimization, Springer, 1996. DOI:10.1007/978-3-662-03139-8
19. D. Jones et al., “DIRECT Algorithm,” J. Optimization Theory Appl., 1993. DOI:10.1007/BF00941892
20. C. Floudas, Deterministic Global Optimization, Springer, 2000. DOI:10.1007/978-0-585-32107-6
21. K. Deb et al., “NSGA-II,” IEEE TEC, 2002. DOI:10.1109/4235.996017
22. E. Zitzler et al., “SPEA2,” TIK Report, 2001. DOI:10.3929/ethz-a-004284029
23. M. Ehrgott, Multicriteria Optimization, Springer, 2005. DOI:10.1007/3-540-27659-9
24. K. Miettinen, Nonlinear Multiobjective Optimization, Springer, 1999. DOI:10.1007/978-1-4615-5563-6
25. I. Das, J. Dennis, “Weighted Sum Method,” SIAM J Optimization, 1997. DOI:10.1137/S1052623496307510
26. S. Inage, T. Hebishima, Monte Carlo Stochastic Optimization Technique (MOST): Deterministic Global Optimization via Region Integration, Preprint, 2022.
27. S. Inage, T. Hebishima, Multi-Objective Extension of MOST with Deterministic Pareto Convergence, Mathematics and Computers in Simulation, 2022. DOI:10.1016/j.matcom.2022.04.012
28. D. P. Bertsekas, Nonlinear Programming, Athena Scientific, 1999. DOI:10.1137/1.9781611971200
29. R. T. Rockafellar, Convex Analysis, Princeton Univ. Press, 1970. DOI:10.1515/9781400873172
30. J. Nocedal, S. Wright, Numerical Optimization, Springer, 2006. DOI:10.1007/978-0-387-40065-5
31. I. Das, J. Dennis, “A closer look at weighted sum method,” SIAM J. Optimization, 1997. DOI:10.1137/S1052623496307510

32. W. Hoeffding, "Probability inequalities for sums of bounded random variables," *JASA*, 1963. DOI:10.1080/01621459.1963.10500830.
33. D. R. Jones, C. D. Perttunen, B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Applications*, 79, pp. 157–181 (1993). DOI:10.1007/BF00941892.
34. E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer, 1997. DOI:10.1007/978-1-4612-0661-9.
35. J. M. Borwein, A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, 2006. DOI:10.1007/978-0-387-31256-9
36. P. Billingsley, *Probability and Measure*, Wiley, 1995. DOI:10.1002/9780470316962.
37. D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, 1996. DOI:10.1137/1.9781611970722.
38. A. M. Geoffrion, "Proper efficiency and the theory of vector maximization," *Journal of Mathematical Analysis and Applications*, 22, pp. 618–630 (1968). DOI:10.1016/0022-247X(68)90201-1.
39. R. T. Rockafellar, R. J-B. Wets, *Variational Analysis*, Springer, 1998. DOI:10.1007/978-3-642-02431-3.
40. J. Jahn, *Vector Optimization: Theory, Applications, and Extensions*, Springer, 2004. DOI:10.1007/978-3-662-05640-7.
41. D. H. Ackley, "A Connectionist Machine for Genetic Hillclimbing," Kluwer Academic Publishers, 1987. DOI:10.1007/978-1-4613-1997-9.
42. R. Fletcher, *Practical Methods of Optimization*, Wiley, 1987. DOI:10.1002/9781118723202.

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.