“There is a real need for this book ... There are numerous books on generalized convexity with applications, but there are only a few books concerned with pseudolinear functions. There are many books on nonconvex optimization with applications, but they do not contain pseudolinear functions. There are also several books on vector optimization, but they do not contain pseudolinear functions.”

—Nan-Jing Huang, Sichuan University

Pseudolinear Functions and Optimization is the first book to focus exclusively on pseudolinear functions, a class of generalized convex functions. It discusses the properties, characterizations, and applications of pseudolinear functions in nonlinear optimization problems.

The book describes the characterizations of solution sets of various optimization problems. It examines multiobjective pseudolinear, multiobjective fractional pseudolinear, static minmax pseudolinear, and static minmax fractional pseudolinear optimization problems and their results. The authors extend these results to locally Lipschitz functions using Clarke subdifferentials. They also present optimality and duality results for \( \eta \)-pseudolinear and semi-infinite pseudolinear optimization problems.

The authors go on to explore the relationships between vector variational inequalities and vector optimization problems involving pseudolinear functions. They present characterizations of solution sets of pseudolinear optimization problems on Riemannian manifolds as well as results on pseudolinearity of quadratic fractional functions. The book also extends \( \eta \)-pseudolinear functions to pseudolinear and \( \eta \)-pseudolinear fuzzy mappings and characterizations of solution sets of pseudolinear fuzzy optimization problems and \( \eta \)-pseudolinear fuzzy optimization problems. The text concludes with some applications of pseudolinear optimization problems to hospital management and economics.

This book encompasses nearly all the published literature on the subject along with new results on semi-infinite nonlinear programming problems. It will be useful to readers from mathematical programming, industrial engineering, and operations management.
Pseudolinear Functions and Optimization
Pseudolinear Functions and Optimization

Shashi Kant Mishra
Balendu Bhooshan Upadhyay
To
Our Beloved Parents
Smt. Shyama Mishra, Shri Gauri Shankar Mishra
and
Smt. Urmila Upadhyay, Shri Yadvendra D. Upadhyay
## Contents

Foreword xiii  
Preface xv  
List of Figures xvii  
List of Tables xix  
Symbol Description xxi

1 Basic Concepts in Convex Analysis 1  
  1.1 Introduction 1  
  1.2 Basic Definitions and Preliminaries 2  
  1.3 Matrices 6  
  1.4 Derivatives and Hessians 17  
  1.5 Convex Sets and Properties 20  
  1.6 Relative Interiors 29  
  1.7 Convex Functions and Properties 34  
  1.8 Generalized Convex Functions 43  
  1.9 Separations and Alternative Theorems 47  
  1.10 Subdifferential Calculus 52  
      1.10.1 Convex Subdifferentials 52  
      1.10.2 Dini and Dini-Hadamard Derivatives 57  
      1.10.3 Clarke Subdifferentials 64  
  1.11 Optimality Criteria 70  
      1.11.1 Unconstrained Minimization Problem 70  
      1.11.2 Constrained Minimization Problems 74  
      1.11.3 Scalar Optimization Problems and Optimality Criteria 76  
      1.11.4 Multiobjective Optimization and Pareto Optimality 78  
      1.11.5 Necessary and Sufficient Conditions for Pareto Optimality 80  
      1.11.6 Nondifferentiable Optimality Conditions 83  
  1.12 Duality 85
2 Pseudolinear Functions: Characterizations and Properties 87

2.1 Introduction .................................................. 87
2.2 Differentiable Pseudolinear Functions ....................... 88
2.3 Local Characterizations of Pseudolinear Functions ......... 96
2.4 Second Order Characterizations of Pseudolinear Functions .... 100
2.5 Semilocally Pseudolinear Functions ..................... 105
2.6 Dini Derivatives and Pseudolinear Functions ............... 110
2.7 Pseudolinearity in Terms of Bifunctions ................. 113
2.8 Directionally Differentiable Pseudolinear Functions ...... 119
2.9 Locally Lipschitz Pseudolinear Functions and Clarke Subdifferentials ................. 123
2.10 Weakly Pseudolinear Functions ............................. 127
2.11 Properties of the Functional Class of Differentiable at a Point Functions ................. 129

3 Constrained Pseudolinear Optimization: Characterizations of Solution Sets 133

3.1 Introduction .................................................. 133
3.2 Characterizations of the Solution Set of (P) Using Differentiable Pseudolinear Functions ................. 134
3.3 Solution Sets of Linear Fractional Programs ............... 138
3.4 Characterizations of the Solutions Set of (P) Using $\ast$-Pseudolinear Functions ................. 139
3.5 Characterization of the Solution Set of (P) Using $h$-Pseudolinear Functions ................ 141
3.6 Characterization of the Solution Set of (P) Using Locally Lipschitz Pseudolinear Functions and Clarke Subdifferential .... 145

4 Characterizations of Solution Sets in Terms of Lagrange Multipliers 149

4.1 Introduction .................................................. 149
4.2 Definitions and Preliminaries .............................. 150
4.3 Pseudolinear Optimization Problems ..................... 150
4.4 Linear Fractional Programming Problems .................. 154
4.5 Vector Linear Fractional Programming Problem .......... 155
4.6 Nonsmooth Pseudolinear Optimization Problems with Linear Inequality Constraints ................. 157
4.7 Nonsmooth Vector Pseudolinear Optimization Problems with Linear Inequality Constraints .......... 162
4.8 Necessary Optimality Conditions for a Mathematical Program in Terms of Bifunctions ................. 163
<table>
<thead>
<tr>
<th>11.1 Introduction</th>
<th>287</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.2 Definitions and Preliminaries</td>
<td>288</td>
</tr>
<tr>
<td>11.2.1 Necessary and Sufficient Optimality Conditions</td>
<td>290</td>
</tr>
<tr>
<td>11.3 Nonsmooth Vector Variational Inequality and Vector Optimization Problems</td>
<td>292</td>
</tr>
<tr>
<td>11.3.1 Necessary and Sufficient Optimality Conditions</td>
<td>293</td>
</tr>
<tr>
<td>11.4 Pseudoaffine Variational Inequality Problems</td>
<td>296</td>
</tr>
</tbody>
</table>

**12 An Extension of Pseudolinear Functions and Variational Inequalities**

<table>
<thead>
<tr>
<th>12.1 Introduction</th>
<th>301</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.2 Definitions and Preliminaries</td>
<td>302</td>
</tr>
<tr>
<td>12.3 Characterizations of PPM-Maps</td>
<td>302</td>
</tr>
<tr>
<td>12.4 Affine PPM-Maps</td>
<td>307</td>
</tr>
<tr>
<td>12.5 Solution Set of a PPM Variational Inequality Problem</td>
<td>308</td>
</tr>
<tr>
<td>12.6 General Form of Pseudoaffine Maps Defined on the Whole Space</td>
<td>312</td>
</tr>
<tr>
<td>12.7 Reduction to the Continuous Case</td>
<td>315</td>
</tr>
<tr>
<td>12.8 Elementary Properties of Pseudoaffine Maps Defined on the Whole Space</td>
<td>317</td>
</tr>
<tr>
<td>12.9 The Case $n = 2$</td>
<td>320</td>
</tr>
<tr>
<td>12.10 Line Properties of Pseudoaffine Maps</td>
<td>321</td>
</tr>
</tbody>
</table>
Contents

12.11 The Case Where $T$ Has at Least One Zero 324
12.12 The Case $n = 3$ 327
12.13 The Case Where $T$ Has No Zeros 331

13 $\eta$-Pseudolinear Functions: Characterizations of Solution Sets 337

13.1 Introduction 337
13.2 Characterization of Differentiable $\eta$-Pseudolinear Functions 338
13.3 Characterization of Solution Set of (P) Using Differentiable $\eta$-Pseudolinear Functions 346
13.4 $\eta$-Pseudolinear Functions and Vector Optimization 348
13.5 Characterizations of Locally Lipschitz $\eta$-Pseudolinear Functions 354
13.6 Characterization of Solution Set of (P) Using Locally Lipschitz $\eta$-Pseudolinear Functions 360

14 Smooth Pseudolinear Functions on Riemannian Manifolds 365

14.1 Introduction 365
14.2 Characterizations of Pseudolinear Functions 366
14.3 Construction of Pseudolinear Functions 372
14.4 Characterizations of $\eta$-Pseudolinearity on Differentiable Manifolds 373
14.5 $\eta$-Pseudolinear Programming Problem on Differentiable Manifolds 377
14.6 New Examples of $\eta$-Pseudolinear Functions 378

15 On the Pseudolinearity of Quadratic Fractional Functions 381

15.1 Introduction 381
15.2 Definitions and Preliminaries 382
15.3 Generalized Linearity of Quadratic Fractional Functions 385
15.4 Affinity of Quadratic Fractional Functions 403
15.5 Gradient of Pseudolinear Quadratic Fractional Functions 406
15.6 A Larger Class of Quadratic Fractional Functions 408
15.7 Pseudolinear Quadratic Fractional Functions in Optimization 411
15.8 Local Analysis of Pseudolinearity through the Implicit Function Approach 413
15.9 Pseudolinearity of Generalized Quadratic Fractional Functions 421
15.10 The Carosi-Martein Theorem 424
Contents

16 Pseudolinear Fuzzy Mapping 427

16.1 Introduction .................................................. 427
16.2 Definition and Preliminaries ................................. 428
16.3 Characterizations of Pseudolinear Fuzzy Mappings ........ 433
16.4 Characterization of Solution Sets ............................ 436

17 Pseudolinear Functions and Their Applications 439

17.1 Introduction .................................................. 439
17.2 A Hospital Management Problem and Algorithm .......... 440
17.2.1 Mathematical Background of the Algorithm and Pseudo-
linear Functions .................................................. 441
17.2.2 A Less Nonlinear Program ................................ 442
17.2.3 Characterization of a Global Solution .................... 444
17.2.4 A Pseudolinear Simplex Algorithm ......................... 445
17.2.5 Degeneracy .................................................. 450
17.2.6 Example ..................................................... 451
17.3 Pseudolinear Quadratic Fractional Functions and
Simplex Algorithm .................................................. 452
17.3.1 Description of the Algorithms ........................... 452
17.3.2 Basic Definitions and Preliminaries ....................... 453
17.3.3 The Convex Simplex Algorithm ............................ 457
17.4 The Le Chatelier Principle in
Pseudolinear Programming ....................................... 458
17.4.1 Karush-Kuhn-Tucker Optimality Conditions ............. 459
17.4.2 The Le Chatelier Principle ................................ 460

Bibliography 463

Index 485
Optimization is a discipline that plays a key role in modeling and solving problems in most areas of engineering and sciences. Optimization tools are used to understand the dynamics of information networks, financial markets, solve logistics and supply chain problems, design new drugs, solve biomedical problems, design renewable and sustainable energy systems, reduce pollution, and improve health care systems. In general, we have convex optimization models, which, at least in theory, can be solved efficiently, and nonconvex optimization models that are computationally very hard to solve. Designing efficient algorithms and heuristics for solving optimization problems requires a very good understanding of the mathematics of optimization. For example, the development of optimality conditions, the theory of convexity, and computational complexity theory have been instrumental for the development of computational optimization algorithms.

The book *Pseudolinear Functions and Optimization* by Shashi Kant Mishra and Balendu Bhooshan Upadhyay, sets the mathematical foundations for a class of optimization problems. Although convexity theory plays the most important role in optimization, several attempts have been made to extend the theory to generalized convexity. Such extensions were necessary to understand optimization models in a wide spectrum of applications.

The book *Pseudolinear Functions and Optimization* is to my knowledge the first book that is dedicated to a specific class of generalized convex functions that are called pseudolinear functions. This is an in-depth study of the mathematics of pseudolinear functions and their applications. Most of the recent results on pseudolinear functions are covered in this book.

The writing is pleasant and rigorous and the presentation of material is very clear. The book will definitely will be useful for the optimization community and will have a lasting effect.

*Panos M. Pardalos*
*Distinguished Professor*
*Paul and Heidi Brown Preeminent Professor in Industrial and Systems Engineering*
*University of Florida*
*Gainesville, USA*
Preface

In 1967 Kortanek and Evans [149] studied the properties of the class of functions, which are both pseudoconvex and pseudoconcave. This class of functions were later termed as pseudolinear functions. In 1984 Chew and Choo [47] derived first and second order characterizations for pseudolinear functions. The linear and quadratic fractional functions are particular cases of pseudolinear functions. Several authors have studied pseudolinear functions and their characterizations, see Cambini and Carosi [34], Schaible and Ibaraki [246], Rapcsak [233], Komlosi [147], Kaul et al. [139], Lu and Zhu [171], Dinh et al. [67], Zhao and Tang [298], Ansari and Rezaei [4] and Mishra et al. [200].

Chapter 1 is introductory and contains basic definitions and concepts needed in the book.

Chapter 2, presents basic properties and characterization results on pseudolinear functions. Further, it includes semilocal pseudolinear functions, Dini differentiable pseudolinear functions, locally Lipschitz pseudolinear functions, h-pseudolinear functions, directionally differentiable pseudolinear functions, weakly pseudolinear functions and their characterizations.

Chapter 3, presents characterizations of solution sets of pseudolinear optimization problems, linear fractional optimization problems, directionally differentiable pseudolinear optimization problems, h-pseudolinear optimization problems and locally Lipschitz optimization problems.

Chapter 4, presents characterizations of solution sets in terms of Lagrange multipliers for pseudolinear optimization problems and its other generalizations given in Chapter 3.

Chapter 5, considers multiobjective pseudolinear optimization problems and multiobjective fractional pseudolinear optimization problems and presents optimality conditions and duality results for these two problems.

Chapter 6, extends the results of Chapter 5 to locally Lipschitz functions using the Clarke subdifferentials.

Chapter 7, considers static minmax pseudolinear optimization problems and static minmax fractional pseudolinear optimization problems and presents optimality conditions and duality results for these two problems.

Chapter 8, extends the results of Chapter 7 to locally Lipschitz functions using the Clarke subdifferentials.

Chapter 9, presents optimality and duality results for h-pseudolinear optimization problems.
Chapter 10, presents optimality and duality results for semi-infinite pseudolinear optimization problems.

Chapter 11, presents relationships between vector variational inequalities and vector optimization problems involving pseudolinear functions. Moreover, relationships between vector variational inequalities and vector optimization problems involving locally Lipschitz pseudolinear functions using the Clarke subdifferentials are also presented.

Chapter 12, presents an extension of pseudolinear functions are used to establish results on variational inequality problems.

Chapter 13, presents results on \(\eta\)-pseudolinear functions and characterizations of solution sets of \(\eta\)-pseudolinear optimization problems.

Chapter 14, presents pseudolinear functions on Riemannian manifolds and characterizations of solution sets of pseudolinear optimization problems on Riemannian manifolds. Moreover, \(\eta\)-pseudolinear functions and characterizations of solution sets of \(\eta\)-pseudolinear optimization problems on differentiable manifolds are also presented.

Chapter 15, presents results on pseudolinearity of quadratic fractional functions.

Chapter 16, extends the class of pseudolinear functions and \(\eta\)-pseudolinear functions to pseudolinear and \(\eta\)-pseudolinear fuzzy mappings and characterizations of solution sets of pseudolinear fuzzy optimization problems and \(\eta\)-pseudolinear fuzzy optimization problems.

Finally, in Chapter 17, some applications of pseudolinear optimization problems to hospital management and economics are given.

The authors are thankful to Prof. Nicolas Hadjisavvas for his help and discussion in Chapter 4. The authors are indebted to Prof. Juan Enrich Martinez-Legaz, Prof. Pierre Marechal, Prof. Dinh The Luc, Prof. Le Thi Hoai An, Prof. Sy-Ming Gun, Prof. King Keung Lai, and Prof. S.K. Neogy for their help, support, and encouragement in the course of writing this book. The authors are also thankful to Ms. Aastha Sharma from CRC Press for her patience and effort in handling the book.

\textit{Shashi Kant Mishra}
\textit{Balendu Bhooshan Upadhyay}
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Epigraph of a Function $f$</td>
<td>9</td>
</tr>
<tr>
<td>1.2</td>
<td>Convex and Nonconvex Sets</td>
<td>24</td>
</tr>
<tr>
<td>1.3</td>
<td>Convex Hull of a Set $X$</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>Linearization of a Convex Function $f$</td>
<td>39</td>
</tr>
<tr>
<td>1.5</td>
<td>Local and Global Minima for CMP</td>
<td>75</td>
</tr>
<tr>
<td>4.1</td>
<td>Objective Function Plot for (COP)</td>
<td>160</td>
</tr>
</tbody>
</table>
# List of Tables

2.1 Geometrical Meaning of the Functional Class . . . . . . . . 132

17.1 Iterations in Case of Degeneracy . . . . . . . . . . . . . . 450

17.2 Iterations in Phase I and Phase II of the Algorithm . . . . . 451
Symbol Description

:= equal to by definition
φ empty set
∀ for all
∞ infinity
⟨, ⟩ Euclidean inner product
∥ ⋅ ∥ Euclidean norm
∃ there exists
0 zero element in the vector space \( \mathbb{R}^n \)
\( 2^X \) family of all subsets of a set \( X \)
\( A^T \) transpose of a matrix \( A \)
lin(\( A \)) linear hull of a set \( A \)
aff(\( A \)) affine hull of a set \( A \)
arg \( \min \) \( f \) set of all minima of a function \( f \)
\( B_r(x) \) open ball with center at \( x \) and radius \( r \)
\( B_r[x] \) closed ball with center at \( x \) and radius \( r \)
\( B \) open unit ball
\( [x, y] \) closed line segment joining \( x \) and \( y \)
\( ]x, y[ \) open line segment joining \( x \) and \( y \)
bd(\( A \)) boundary of a set \( A \)
cl(\( A \)) closure of a set \( A \)
co(\( A \)) convex hull of a set \( A \)
con(\( A \)) conic hull of a set \( A \)
d(\( x, y \)) distance between \( x \) and \( y \)
d(\( f \)) domain of a function \( f \)
dom(\( f \)) effective domain of map \( f \)
epi(\( f \)) epigraph of a function \( f \)
\( \nabla f(x) \) gradient of a function \( f \) at \( x \)
\( \nabla^2 f(x) \) Hessian matrix of a function \( f \) at \( x \)
f(\( x; d \)) directional derivative of \( f \) at \( x \) in the direction \( d \)
f(\( x; d \)) right sided directional derivative of \( f \) at \( x \) in the direction \( d \)
\( D^+ f(x; d) \) Dini upper directional derivative of \( f \) at \( x \) in the direction \( d \)
\( D^- f(x; d) \) Dini lower directional derivative of \( f \) at \( x \) in the direction \( d \)
\( D^{DH} f(x; d) \) Dini-Hadamard upper directional derivative of \( f \) at \( x \) in the direction \( d \)
\( D_{DH} f(x; d) \) Dini-Hadamard lower directional derivative of \( f \) at \( x \) in the direction \( d \)
\( \partial f(x) \) subdifferential of a function \( f \) at \( x \)
d(\( C \)) distance function of a set \( C \)
d(\( f, C \)) hypograph of a function \( f \)
\( \delta(f,C) \) indicator function of a set \( C \)
\( \{ u \} _{α} \) \( α \) - cut set of fuzzy set \( u \)
\( \supp u \) support of fuzzy set \( u \)
\( \{ u \} _{0} \) family of fuzzy numbers
\( X^⊥ \) orthogonal complement of \( X \)
\( H^+ \) upper closed half-space
\( H^{++} \) upper open half-space
\( H^- \) lower closed half-space
\( H^{--} \) lower open half-space
\( \text{int}(A) \) interior of a set \( A \)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Set</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ri($A$)</td>
<td>relative interior of a set $A$</td>
<td>$\mathbb{R}_+$</td>
<td>set of all nonnegative real numbers</td>
</tr>
<tr>
<td>$J(f)(x)$</td>
<td>Jacobian matrix of a function at $x$</td>
<td>$\sigma_C(.)$</td>
<td>support function of a set $C$</td>
</tr>
<tr>
<td>$X^*$</td>
<td>dual cone of the cone $X$</td>
<td>$\mathbb{R}_{++}$</td>
<td>set of all positive real numbers</td>
</tr>
<tr>
<td>$\Lambda(f, \alpha)$</td>
<td>lower level set of a function $f$ at level $\alpha$</td>
<td>$\mathbb{R}^n$</td>
<td>$n$-dimensional Euclidean space</td>
</tr>
<tr>
<td>$\mathbb{N}$</td>
<td>set of all natural numbers</td>
<td>$\mathbb{R}^n_+$</td>
<td>$(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0 \ \forall \ i$</td>
</tr>
<tr>
<td>$N_C(x)$</td>
<td>normal cone to a set $C$ at $x$</td>
<td>$\mathbb{R}^n_++$</td>
<td>$(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i &gt; 0 \ \forall \ i$</td>
</tr>
<tr>
<td>$0^+X$</td>
<td>recession cone of a set $X$</td>
<td>$\mathbb{R}$</td>
<td>upper level set of a function $f$ at level $\alpha$</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>set of all real numbers</td>
<td>$\Omega(f, \alpha)$</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Basic Concepts in Convex Analysis

1.1 Introduction

Optimization is everywhere, as nothing at all takes place in the universe, in which some rule of maximum or minimum does not appear. It is the human nature to seek for the best among the available alternatives. An optimization problem is characterized by its specific objective function that is to be maximized or minimized, depending upon the problem and, in the case of a constrained problem, a given set of constraints. Possible objective functions include expressions representing profits, costs, market share, portfolio risk, etc. Possible constraints include those that represent limited budgets or resources, nonnegativity constraints on the variables, conservation equations, etc.

The concept of convexity is of great importance in the study of optimization problems. It extends the validity of a local solution of a minimization problem to global one and the first order necessary optimality conditions become sufficient for a point to be a global minimizer. We mention the earlier work of Jensen [112], Fenchel [80, 81] and Rockafellar [238]. However, in several real-world applications, the notion of convexity does no longer suffice. In many cases, the nonconvex functions provide more accurate representation of reality. Nonconvex functions preserve one or more properties of convex functions and give rise to models which are more adaptable to the real-world situations, than convex models. This led to the introduction of several generalizations of the classical notion of convexity.

In 1949, the Italian mathematician Bruno de Finetti [62] introduced one of the fundamental generalized convex functions, known as quasiconvex function having wider applications in economics, management sciences and engineering. Mangasarian [176] introduced the notion of pseudoconvex and pseudoconcave functions as generalizations of convex and concave functions, respectively. In 1969, Kortanek and Evans [149] studied the properties of a class of functions, which are both pseudoconvex and pseudoconcave, later termed as pseudolinear functions. In case of stationary points the behavior of the class of pseudolinear functions is as good as linear functions. Several other generalizations of these functions have been introduced to find weakest conditions in order to establish sufficient optimality conditions and duality results for optimization problems.

The main aim of this chapter is to explore the properties of convex sets
and convex functions. We give some basic definitions and preliminary results from algebra, geometry and topology, that will be used throughout the book to develop some important results. Taking into consideration that, a function \( f \) is concave if and only if \( -f \) is convex, the proofs of the results are provided just for convex functions.

### 1.2 Basic Definitions and Preliminaries

Throughout the book, suppose that \( \mathbb{R}^n \) be the \( n \)-dimensional Euclidean space. Let \( \mathbb{R}^n_+ \) and \( \mathbb{R}^n_{++} \) denote the nonnegative and positive orthants of \( \mathbb{R}^n \), respectively. Let \( \mathbb{R} \) denote the real number system and \( \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm \infty\} \) be an extended real line. The number 0 will denote either the real number zero or the zero vector in \( \mathbb{R}^n \), all components of which are zero.

If \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n \), then, the following convention for equalities and inequalities will be adopted:

\[
\begin{align*}
  x &= y \iff x_i = y_i, \forall i = 1, \ldots, n; \\
  x &< y \iff x_i < y_i, \forall i = 1, \ldots, n; \\
  x &\leq y \iff x_i \leq y_i, \forall i = 1, \ldots, n; \\
  x &\leq y \iff x_i \leq y_i, \forall i = 1, \ldots, n, \text{ but } x \neq y.
\end{align*}
\]

The following rules are natural extensions of the rules of arithmetic:

- For every \( x \in \mathbb{R} \),
  \[
  x + \infty = \infty \quad \text{and} \quad x - \infty = -\infty;
  \]
  \[
  x \times \infty = \infty, \quad \text{if } x > 0; \quad x \times \infty = -\infty, \quad \text{if } x < 0.
  \]

- For every extended real number \( x \),
  \[
  x \times 0 = 0.
  \]

The expression \( \infty - \infty \) is meaningless.

Let \( \langle \cdot, \cdot \rangle \) denote the Euclidean inner product, that is, for \( x, y \in \mathbb{R}^n \)

\[
\langle x, y \rangle = x^T y = x_1y_1 + \ldots + x_ny_n.
\]

Norm of any point \( x \in \mathbb{R}^n \) is given by

\[
\|x\| = \sqrt{\langle x, x \rangle}.
\]

The symbol \( \|\cdot\| \) will denote the Euclidean norm, unless otherwise specified.

The following proposition illustrates that any two norms on \( \mathbb{R}^n \) are equivalent:
Proposition 1.1 If \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) are any two norms on \( \mathbb{R}^n \), then there exist constants \( c_1 \geq c_2 > 0 \), such that

\[
c_1 \| x \|_1 \leq \| x \|_2 \leq c_2 \| x \|_1, \forall x \in \mathbb{R}^n.
\]

The Euclidean distance between two points \( x, y \in \mathbb{R}^n \), is given by

\[
d(x, y) := \| x - y \| = \langle x - y, x - y \rangle^{\frac{1}{2}}.
\]

The closed and open balls with center at any point \( \bar{x} \) and radius \( \varepsilon \), denoted by \( B_{\varepsilon}[\bar{x}] \) and \( B_{\varepsilon}(\bar{x}) \), respectively are defined as follows:

\[
B_{\varepsilon}[\bar{x}] := \{ x \in \mathbb{R}^n : \| x - \bar{x} \| \leq \varepsilon \}
\]

and

\[
B_{\varepsilon}(\bar{x}) := \{ x \in \mathbb{R}^n : \| x - \bar{x} \| < \varepsilon \}.
\]

Moreover, the open unit ball is the set

\[
B := \{ x \in \mathbb{R}^n : \| x \| < 1 \}.
\]

Definition 1.1 (Lower and upper bound) Let \( X \subseteq \mathbb{R} \) be a nonempty set. Then a number \( \alpha \in \mathbb{R} \), is called a lower bound of \( X \), if

\[
x \geq \alpha, \forall x \in X.
\]

A number \( \beta \in \mathbb{R} \), is called an upper bound of \( X \), if

\[
x \leq \beta, \forall x \in X.
\]

A set \( X \subseteq \mathbb{R} \) is said to be bounded if lower and upper bound of the set exists.

Definition 1.2 (Supremum and infimum of set of real numbers) Let \( X \subseteq \mathbb{R}^n \) be a nonempty set. Then a number \( M \) is called the least upper bound or the supremum of the set \( X \), if

(i) \( M \) is the upper bound of the set, i.e.,

\[
x \leq M, \forall x \in X.
\]

(ii) No number less than \( M \) can be an upper bound of \( X \), i.e., for every \( \varepsilon > 0 \) however small, there exists a number \( y \in X \) such that

\[
y > M - \varepsilon;
\]

A number \( m \) is called the greatest lower bound or the infimum of set \( X \), if

(i) \( m \) is the lower bound of the set, i.e.,

\[
x \geq m, \forall x \in X.
\]
(ii) No number greater than \( m \) can be a lower bound of \( X \), i.e., for every \( \epsilon > 0 \), however small, there exists a number \( z \in X \), such that

\[ z < m + \epsilon. \]

By the completeness axiom it is known that every nonempty set \( X \subseteq \mathbb{R} \), which has a lower (upper) bound has a greatest (least) lower (upper) bound in \( \mathbb{R} \).

**Definition 1.3 (\( \varepsilon \)-Neighborhood of a point)** Let \( \bar{x} \in \mathbb{R}^n \) and \( \varepsilon > 0 \) be given. Then, a subset \( N \) of \( \mathbb{R}^n \) is said to be an \( \varepsilon \)-neighborhood of \( \bar{x} \), if there exists an open ball \( B_\varepsilon(\bar{x}) \), with center at \( \bar{x} \) and radius \( \varepsilon \), such that

\[ \bar{x} \in B_\varepsilon(\bar{x}) \subseteq N. \]

**Definition 1.4 (Interior of a set)** A point \( \bar{x} \in X \subseteq \mathbb{R}^n \) is said to be an interior point of \( X \), if there exists a neighborhood \( N \) of \( \bar{x} \) contained in \( X \), i.e.,

\[ \bar{x} \in N \subseteq X. \]

The set of all interior points of \( X \) is called the interior of \( X \) and is denoted by \( \text{int}(X) \). For example, interior of \( \mathbb{R}^n_+ \) is the set

\[ \mathbb{R}^n_+ := \{ x \in \mathbb{R}^n : x_i > 0, \forall i = 1, \ldots, n \}. \]

**Definition 1.5 (Open set)** Any set \( X \subseteq \mathbb{R}^n \) is said to be open, if every point of \( X \) is an interior point and conversely. In other words, \( X = \text{int}(X) \).

For example, every open ball \( B_\varepsilon(x) \) is an open set. It is easy to see that \( \text{int}(X) \) is the largest open set contained in \( X \).

**Definition 1.6 (Closure of a set)** A point \( \bar{x} \in \mathbb{R}^n \) is said to be a point of closure of the set \( X \subseteq \mathbb{R}^n \), denoted by \( \text{cl}(X) \), if for each \( \varepsilon > 0 \),

\[ B_\varepsilon(\bar{x}) \cap X \neq \emptyset. \]

In other words, \( \bar{x} \in \text{cl}(X) \), if and only if every ball around \( \bar{x} \) contains at least one point of \( X \).

**Definition 1.7 (Closed set)** A set \( X \subseteq \mathbb{R}^n \) is said to be a closed set if every point of closure of \( X \) is in \( X \). In other words, \( X = \text{cl}(X) \). For example, every closed ball \( B_\varepsilon[\bar{x}] \) is a closed set.

**Definition 1.8 (Relatively open and closed)** Let \( X \) and \( Y \) be two nonempty subsets of \( \mathbb{R}^n \), such that \( X \subseteq Y \subseteq \mathbb{R}^n \). Then, \( X \) is said to be relatively open with respect to \( Y \), if

\[ X = Y \cap \Omega, \]
where $\Omega$ is some open set in $\mathbb{R}^n$.

Furthermore, the set $X$ is said to be relatively closed with respect to $Y$, if

$$X = Y \cap \Lambda,$$

where $\Lambda$ is some closed set in $\mathbb{R}^n$.

**Definition 1.9 (Boundary of a set)** Let $X \subseteq \mathbb{R}^n$ be a nonempty set. Any point $\bar{x} \in \mathbb{R}^n$ is said to be a point of boundary of the set $X$, if for each $\varepsilon > 0$, open ball $B_\varepsilon(\bar{x})$ contains points of $X$ as well as points not belonging to $X$.

The set of all boundary points of $X$ is called the boundary of $X$ and it is denoted by $\text{bd}(X)$.

**Definition 1.10 (Angle between two vectors)** Let $x$ and $y$ be two nonzero vectors in $\mathbb{R}^n$. Then the angle $\theta$ between $x$ and $y$, is given by

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\|\|y\|}, \quad 0 \leq \theta \leq \pi.$$

The nonzero vector $x$ and $y$ are said to

(i) be orthogonal if $\langle x, y \rangle = 0$, that is, if $\theta = \frac{\pi}{2}$;

(ii) form an acute (strictly acute) angle if $\langle x, y \rangle \geq 0$, $0 \leq \theta \leq \frac{\pi}{2} \left( \langle x, y \rangle > 0, \left(0 < \theta < \frac{\pi}{2}\right) \right)$;

(iii) form an obtuse (strictly obtuse) angle if $\langle x, y \rangle \leq 0$, $\frac{\pi}{2} \leq \theta \leq \pi \left( \langle x, y \rangle < 0, \left(\frac{\pi}{2} < \theta < \pi\right) \right)$.

Given a subspace $X$ of $\mathbb{R}^n$, the orthogonal complement of $X$, denoted by $X^\perp$, is defined as

$$X^\perp := \{ x \in \mathbb{R}^n : \langle x, y \rangle = 0, \ \forall \ y \in \mathbb{R}^n \}.$$

It is easy to see that $X^\perp$ is another subspace of $\mathbb{R}^n$ and $\dim X + \dim X^\perp = n$. Furthermore, any vector $x$ can be uniquely expressed as the sum of a vector from $X$ and a vector from $X^\perp$.

**Proposition 1.2 (Pythagorean theorem)** For any two orthogonal vectors $x$ and $y$, in $\mathbb{R}^n$, we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

**Proposition 1.3 (Schwarz inequality)** Let any two vectors $x, y \in \mathbb{R}^n$. Then,

$$|\langle x, y \rangle| \leq \|x\|\|y\|.$$

The above inequality holds as equality, if and only if $x = \lambda y$, for some $\lambda \in \mathbb{R}$. 

1.3 Matrices

A matrix is a rectangular array of numbers called elements. Any matrix $A$ with $m$ rows and $n$ columns is called a $m \times n$ matrix and is written as:

$$A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}.$$ 

The $ij$th element of $A$ is denoted by $a_{ij}$. The transpose of $A$, denoted by $A^T$ is defined as $A^T = [a_{ij}]^T = [a_{ji}]$. A matrix $A$ is said to be nonvacuous, if it contains at least one element.

The $i$th row of the matrix $A$ is denoted by $A_i$ and is given by $A_i = (a_{i1}, \ldots, a_{in}), i = 1, \ldots, m$.

The $j$th column of the matrix $A$ is denoted by $A_j$ and is given by

$$A_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}, j = 1, \ldots, n.$$ 

The rank of a matrix is equal to the number of linearly independent columns of $A$, which is also equal to the maximum number of linearly independent rows of $A$. The matrix $A$ and transpose of $A$, that is $A^T$ have the same rank. The matrix $A$ is said to have full rank, if

$$\text{rank}(A) = \min\{m, n\}.$$ 

In other words, for a matrix $A$ having full rank either all the rows are linearly independent or all the columns of $A$ are linearly independent.

Let $A$ be a square matrix of order $n \times n$. Then $A$ is said to be symmetric if $A^T = A$. The square matrix $A$ is said to be diagonal, if $a_{ij} = 0$, for $i \neq j$. The diagonal matrix $A$ is said to be identity, if $a_{ij} = 1$, for $i = j$, i.e. diagonal elements are equal to 1.

**Definition 1.11** Let $A$ be a square matrix of order $n \times n$. Then $A$ is said to be

(i) Nonsingular, if

$$\text{rank} (A) = n.$$ 

(ii) Positive semidefinite matrix, if

$$\langle x, Ax \rangle \geq 0, \forall x \in \mathbb{R}^n.$$
(iii) Negative semidefinite, if
\[ \langle x, Ax \rangle \leq 0, \forall x \in \mathbb{R}^n. \]

**Definition 1.12** Let \( A \) be a square matrix of order \( n \times n \). Then, \( A \) is said to be

(i) Positive definite matrix, if
\[ \langle x, Ax \rangle > 0, \forall x \in \mathbb{R}^n, x \neq 0. \]

(ii) Negative definite matrix, if
\[ \langle x, Ax \rangle < 0, \forall x \in \mathbb{R}^n, x \neq 0. \]

**Remark 1.1** It is clear by the definitions, that every positive (negative) definite matrix is positive (negative) semidefinite matrix. Furthermore, the negative of a positive definite (positive semidefinite) matrix is negative definite (negative semidefinite) and vice versa. It is easy to see that each positive (negative) definite matrix is nonsingular.

Next, we state the following result from Bertsekas et al. [23].

**Proposition 1.4** Let \( A \) and \( B \) be two square matrices of order \( n \times n \). Then,

(i) The matrix \( A \) is symmetric and positive definite if and only if it is invertible and its inverse is symmetric and positive definite.

(ii) If \( A \) and \( B \) are symmetric positive semidefinite matrices, then, their sum \( A + B \) is positive semidefinite matrix. In addition, if \( A \) or \( B \) is positive definite, then \( A + B \) is positive definite matrix.

(iii) If \( A \) is symmetric positive semidefinite matrix and \( X \) be any \( m \times n \) matrix, then, the matrix \( XAX^T \) is positive semidefinite matrix. If \( A \) is positive definite matrix and \( X \) be invertible, then, \( XAX^T \) is positive definite matrix.

(iv) If \( A \) is symmetric positive definite matrix, then there exists positive real numbers \( \alpha \) and \( \beta \), such that
\[ \alpha \| x \|^2 \leq \langle x, Ax \rangle \leq \beta \| x \|^2. \]

(v) If \( A \) is symmetric positive definite matrix, then there exists unique symmetric positive definite matrix that yields \( A \), when multiplied with itself. This matrix is called square root of \( A \) and is denoted by \( A^{1/2} \) and its inverse is denoted by \( A^{-1/2} \).
Definition 1.13 (Mapping) Let $X$ and $Y$ be two sets. Then a correspondence $f: X \to Y$ which associates to each $x \in X$, a subset of $Y$ is called a mapping. For each $x \in X$ the set $f(x)$ is called image of $x$. The subset of points of $X$, for which the image $f(x)$ is nonempty is called the domain of $f$, i.e., the set

$$d(f) = \{ x \in X : f(x) \neq \emptyset \}$$

is called domain of $f$. The union of image of points of $d(f)$ is called range of $f$, denoted by $f(d(f))$, i.e.,

$$f(d(f)) = \bigcup_{x \in d(f)} f(x).$$

Definition 1.14 (Function) Let $X$ and $Y$ be two sets. A single valued mapping $f: X \to Y$ is called a function. In other words, for each $x \in X$, the image set $f(x)$ consists of a single element of $Y$. The domain of $f$ is $X$ and the range of $f$ is

$$f(x) = \bigcup_{x \in X} f(x).$$

Definition 1.15 (Affine and linear functions) A function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be affine, if it has the form

$$f(x) = \langle a, x \rangle + b,$$

where $a, x \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Similarly, a function $f: \mathbb{R}^n \to \mathbb{R}^m$ is called affine, if it has the form

$$f(x) = \langle A, x \rangle + b,$$

where $A$ is any $m \times n$ matrix, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. If $b = 0$, then $f$ is said to be a linear function or a linear transformation.

In optimization problems, we often encounter the objective functions, which can take values on an extended real line, referred to as an extended real-valued function.

Definition 1.16 (Effective domain) Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $f: X \to \mathbb{R}$ be an extended real-valued function. The effective domain of $f$, denoted by $\text{dom}(f)$, is defined by

$$\text{dom}(f) := \{ x \in X : f(x) < \infty \}.$$

Definition 1.17 Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $f: X \to \mathbb{R}$ be an extended real-valued function.

(i) The graph of $f$ is defined by

$$\text{graph}(f) := \{ (x, \alpha) : x \in \mathbb{R}^n, \alpha \in \mathbb{R} : f(x) = \alpha \}.$$
(ii) The epigraph of the function $f$, denoted by $\text{epi}(f)$, is a subset of $\mathbb{R}^{n+1}$ and is defined as follows

$$\text{epi}(f) := \{(x, \alpha) : x \in X, \alpha \in \mathbb{R} : f(x) \leq \alpha\}.$$ 

(iii) The hypograph of the function $f$, denoted by $\text{hyp}(f)$, is a subset of $\mathbb{R}^{n+1}$ and is defined as follows

$$\text{hyp}(f) := \{(x, \alpha) : x \in X, \alpha \in \mathbb{R} : f(x) \geq \alpha\}.$$ 

For the epigraph of a function $f$, see Figure 1.1.

![Figure 1.1: Epigraph of a Function $f$](image)

**Definition 1.18** Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $f : X \rightarrow \mathbb{R}$ be an extended real-valued function.

(i) Lower level set at the level $\alpha$, is defined as

$$\Lambda(f, \alpha) := \{x \in \mathbb{R}^n : f(x) \leq \alpha\}.$$ 

(ii) Upper level set at the level $\alpha$, is defined as

$$\Omega(f, \alpha) := \{x \in \mathbb{R}^n : f(x) \geq \alpha\}.$$ 

It can be seen that

$$\text{dom}(f) := \{x \in X : \exists \ y \in \mathbb{R} \text{ such that } (x, y) \in \text{epi}(f)\}.$$ 

In other words, $\text{dom}(f)$ is the projection of epigraph of $f$ on $\mathbb{R}^n$. 
Definition 1.19 (Proper function) Let \( f : X \to \mathbb{R} \) be an extended real-valued function. The function \( f \) is said to be proper, if \( f(x) < \infty \), for at least one \( x \in X \) and \( f(x) > -\infty \), for all \( x \in X \). In other words, a function \( f \) is proper if \( \text{dom}(f) \) is nonempty and \( f \) is finite valued on \( \text{dom}(f) \).

Definition 1.20 (Sequence) A sequence on a set \( X \subseteq \mathbb{R}^n \) is a function \( f \) from the set \( \mathbb{N} \) of natural number to the set \( X \). If \( f(n) = x_n \in X, \) for \( n \in \mathbb{N} \), then the sequence, \( f \) is denoted by the symbol \( \{x_n\} \) or by \( x_1, x_2, \ldots \).

Definition 1.21 (Bounded sequence) Let \( \{x_n\} \) be any sequence in \( \mathbb{R}^n \). Then, \( \{x_n\} \) is said to be bounded above, if there exists \( \bar{x} \in \mathbb{R}^n \), such that
\[
x_n \leq \bar{x}, \forall n \in \mathbb{N}.
\]
The sequence \( \{x_n\} \) is said to be bounded below, if there exists \( \hat{x} \in \mathbb{R}^n \), such that
\[
x_n \geq \hat{x}, \forall n \in \mathbb{N}.
\]
A sequence \( \{x_n\} \) is said to be bounded, if it is bounded above as well as bounded below.

Definition 1.22 (Monotonic sequence) Let \( \{x_n\} \) be any sequence in \( \mathbb{R}^n \). Then, \( \{x_n\} \) is said to be monotonically increasing (respectively, decreasing), if
\[
x_n \leq x_{n+1}, \forall n \quad \text{(respectively, } x_n \geq x_{n+1}, \forall n)\]
The sequence \( \{x_n\} \) is said to be a monotonic sequence, if it is either monotonically increasing or monotonically decreasing.

Definition 1.23 (Limit point) Let \( \{x_n\} \) be any sequence in \( \mathbb{R}^n \). Then, a number \( \bar{x} \in \mathbb{R}^n \) is said to be the limit point (accumulation point) of the sequence \( \{x_n\} \), if for given \( \varepsilon > 0 \), one has
\[
\|x_n - \bar{x}\| < \varepsilon,
\]
for infinitely many values of \( n \).

Definition 1.24 (Limit) Let \( \{x_n\} \) be any sequence in \( \mathbb{R}^n \). Then a number \( \bar{x} \in \mathbb{R}^n \), is said to be limit of the sequence, if for each \( \varepsilon > 0 \), there exists a positive integer \( m \) (depending on \( \varepsilon \)), such that
\[
\|x_n - \bar{x}\| < \varepsilon, \forall n \geq m.
\]
Moreover, the sequence \( \{x_n\} \) is said to converge to \( \bar{x} \), and is denoted by \( x_n \to \bar{x} \), whenever \( n \to \infty \) or \( \lim_{n \to \infty} x_n = \bar{x} \).
Remark 1.2 It is obvious, that a limit of a sequence is also limit point of the sequence, but not conversely. For example, the sequence \( \{1, -1, 1, -1, \ldots\} \) has limit points 1 and \(-1\) but having no limit.

Definition 1.25 (Subsequence) Let \( \{x_n\} \) be any sequence. If there exists a sequence of positive integers \( n_1, n_2, \ldots \), such that \( n_1 \leq n_2 \leq \ldots \), then, the sequence \( \{x_{n_k}\} \) is called a subsequence of the sequence \( \{x_n\} \).

It is known that if \( \bar{x} \) is a limit of a sequence \( \{x_n\} \), then, it is limit of every subsequence of \( \{x_n\} \).

Now, we state the following classical result on bounded sequences.

Proposition 1.5 (Bolzano-Weierstrass theorem) Every bounded sequence in \( \mathbb{R}^n \), has a convergent subsequence.

Now, we have the following important results.

Proposition 1.6 Every bounded and monotonically nondecreasing (nonincreasing) sequence in \( \mathbb{R}^n \), has a limit.

Proposition 1.7 (Cauchy convergence criteria) A sequence \( \{x_n\} \) in \( \mathbb{R}^n \) converges to a limit, if and only if for each \( \varepsilon > 0 \), there exists a positive integer \( p \), such that

\[
||x_m - x_n|| < \varepsilon, \forall m, n \geq p.
\]

The following notions of limit infimum and limit supremum of sequences in \( \mathbb{R} \) are important to understand the concept of continuity of real-valued functions. Let \( \{x_n\} \) be any sequence in \( \mathbb{R} \). Define

\[
y_k := \inf \{x_n : n \geq k\} \quad \text{and} \quad z_k := \sup \{x_n : n \geq k\}.
\]

By the definitions, it is clear that \( \{y_k\} \) and \( \{z_k\} \) are nondecreasing and nonincreasing sequences, respectively. If the sequence \( \{x_n\} \) is a bounded below or above, then by Proposition 1.6, the sequences \( \{y_k\} \) and \( \{z_k\} \) respectively, have a limit. The limit of \( \{y_k\} \) is called lower limit or limit infimum of \( \{x_n\} \) and is denoted by \( \lim_{n \to \infty} \inf \{x_n\} \), while, the limit of \( \{z_k\} \) is called the upper limit or limit supremum of \( \{x_n\} \) and is denoted by \( \lim_{n \to \infty} \sup \{x_n\} \).

If a sequence \( \{x_n\} \) is unbounded below, then we write \( \lim_{n \to \infty} \inf \{x_n\} = - \infty \), while if the sequence \( \{x_n\} \) is unbounded above, then we write \( \lim_{n \to \infty} \sup \{x_n\} = \infty \).

Definition 1.26 (Compact set) A set \( X \subseteq \mathbb{R}^n \) is said to be a compact set, if it satisfies, any one of the following equivalent conditions:

(i) \( X \) is closed and bounded.
(ii) **(Bolzano-Weierstrass property)** Every sequence of points in $X$, has a limit point in $X$.

(iii) **(Finite intersection property)** For any family $\{X_i, i \in I\}$ of sets closed relative to $X$, if  
$$ \bigcap_{i \in I} X_i = \emptyset \Rightarrow X_{i_1} \cap \ldots \cap X_{i_k} = \emptyset,$$  
for some $i_1, \ldots, i_k \in I$.

(iv) **(Heine-Borel property)** Every open cover of $X$ admits a finite subcover. In other words for every family $\{X_i, i \in I\}$ of open sets such that $\bigcup_{i \in I} X_i \subseteq X$, there exists a finite subfamily $\{X_{i_1}, \ldots, X_{i_k}\}$, such that  
$$ X_{i_1} \cup \ldots \cup X_{i_k} \subseteq X.$$

(v) Every sequence of points in $X$, has a subsequence, that converges to a point of $X$.

Now, we state the following proposition, that will be used frequently.

**Proposition 1.8** Let $\{x_n\}$ and $\{y_n\}$ be any two sequences in $\mathbb{R}$. Then,

(i) We have  
$$ \sup\{x_n : n \geq 0\} \geq \limsup_{n \to \infty} x_n \geq \liminf_{n \to \infty} x_n \geq \{x_n : n \geq 0\}. $$

(ii) The sequence $\{x_n\}$ converges if and only if  
$$ \infty > \limsup_{n \to \infty} x_n = \liminf_{n \to \infty} x_n > \infty.$$  
The common value of $\limsup_{n \to \infty} x_n$ and $\liminf_{n \to \infty} x_n$ is the limit of the sequence $\{x_n\}$.

(iii) If $x_n \leq y_n$, for all $n$, then  
$$ \limsup_{n \to \infty} x_n \leq \limsup_{n \to \infty} y_n $$  
and  
$$ \liminf_{n \to \infty} x_n \leq \liminf_{n \to \infty} y_n.$$ 

(iv) We have  
$$ \limsup_{n \to \infty} (x_n + y_n) \leq \limsup_{n \to \infty} x_n + \limsup_{n \to \infty} y_n $$  
and  
$$ \liminf_{n \to \infty} (x_n + y_n) \geq \liminf_{n \to \infty} x_n + \liminf_{n \to \infty} y_n.$$
Definition 1.27 (Limit infimum and limit supremum of extended real-valued functions) Let \( f : \mathbb{R}^n \to \mathbb{R} \) be an extended real-valued function, then we define
\[
\liminf_{y \to x} = \sup_{\varepsilon > 0} \inf_{y \in B_\varepsilon(x)} f(y)
\]
and
\[
\limsup_{y \to x} = \inf_{\varepsilon > 0} \sup_{y \in B_\varepsilon(x)} f(y).
\]
It is clear that
\[
\liminf_{y \to x} (-f(y)) = -\limsup_{y \to x} f(y)
\]
and
\[
\limsup_{y \to x} (-f(y)) = -\liminf_{y \to x} f(y).
\]
Moreover, \( \lim f(y) \) is said to exist, if
\[
\liminf_{y \to x} f(y) = \limsup_{y \to x} f(y).
\]

The following results from Rockafellar and Wets [239], present characterizations for limit infimum and limit supremum, of an extended real-valued function.

Proposition 1.9 Let \( f : \mathbb{R}^n \to \mathbb{R} \) be an extended real-valued function, then
\[
\liminf_{y \to x} = \min\{\alpha \in \mathbb{R} : \exists x_n \to \bar{x} \text{ with } f(x_n) \to \alpha\}
\]
and
\[
\limsup_{y \to x} = \max\{\beta \in \mathbb{R} : \exists x_n \to \bar{x} \text{ with } f(x_n) \to \beta\}.
\]

Next, we state the following results for infimum and supremum operations from Rockafellar and Wets [239], that will be used in sequel.

Proposition 1.10 Let \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R} \) be extended real-valued functions and let \( \Omega, \Omega_1 \) and \( \Omega_2 \) be subsets of \( \mathbb{R}^n \). Then,
\[ (i) \text{ For } \alpha \geq 0, \text{ we have } \]
\[
\inf_{x \in \Omega} (\alpha f(x)) = \alpha \inf_{x \in \Omega} (f(x))
\]
and
\[
\sup_{x \in \Omega} (\alpha f(x)) = \alpha \sup_{x \in \Omega} (f(x)).
\]
(ii) We have
\[
\sup_{x \in \Omega} f(x) + \sup_{x \in \Omega} g(x) \geq \sup_{x \in \Omega} (f(x) + g(x)) \geq \inf_{x \in \Omega} (f(x) + g(x)) \\
\geq \inf_{x \in \Omega} f(x) + \inf_{x \in \Omega} g(x).
\]

(iii) If \( \Omega_1 \subseteq \Omega_2 \), then,
\[
\sup_{x_1 \in \Omega_1} f(x_1) \leq \sup_{x_2 \in \Omega_2} f(x_2)
\]
and
\[
\inf_{x_1 \in \Omega_1} f(x_1) \geq \inf_{x_2 \in \Omega_2} f(x_2).
\]

**Definition 1.28** Let \( X \subseteq \mathbb{R}^n \) be a nonempty set. Then, a function \( f : X \to \mathbb{R} \) is said to be

(i) **Positively homogeneous**, if for all \( x \in X \) and all \( r \geq 0 \), we have
\[
f(rx) = rf(x).
\]
It is clear that positive homogeneity is equivalent to the epigraph of \( f \) being a cone on \( \mathbb{R}^{n+1} \).

(ii) **Subadditive**, if for all \( x, y \in X \), we have
\[
f(x + y) \leq f(x) + f(y).
\]

(iii) **Sublinear**, if it is a positively homogeneous and subadditive function.

(iv) **Subodd**, if for all \( x \in \mathbb{R}^n / \{0\} \), we have
\[
f(x) \geq -f(-x)
\]
or equivalently, if \( f(x) + f(-x) \geq 0 \). The function \( f \) is said to be odd, if \( f \) and \( -f \) are both subodd.

**Definition 1.29 (Continuous function)** Let \( X \subseteq \mathbb{R}^n \) be a nonempty set and let \( f : X \to \mathbb{R} \) be a real-valued function. Then, \( f \) is said to be continuous at \( \bar{x} \in X \), if either of the following two conditions are satisfied

(i) For given \( \varepsilon > 0 \) there exists \( \delta > 0 \), such that for each \( x \in X \), one has
\[
|f(x) - f(\bar{x})| < \varepsilon, \text{ whenever } \|x - \bar{x}\| < \delta.
\]

(ii) For each sequence \( \{x_n\} \in X \) converging to \( \bar{x} \), one has
\[
\lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n) = f(\bar{x}).
\]
The function $f$ is said to be continuous on $X$, if it is continuous at each point $\bar{x} \in X$.

**Proposition 1.11** The function $f : X \to \mathbb{R}$ is said to be continuous on $X$, if the following equivalent conditions are satisfied:

(i) The sets $\{ x : x \in X, f(x) \geq \alpha \}$ and $\{ x : x \in X, f(x) \geq \alpha \}$ are closed relative to $X$, for each real number $\alpha$.

(ii) The sets $\{ x : x \in X, f(x) < \beta \}$ and $\{ x : x \in X, f(x) > \beta \}$ are open relative to $X$, for each real number $\beta$.

(iii) The epigraph of $\text{epi}(f) = \{ (x, \alpha) : x \in X, \alpha \in \mathbb{R}, f(x) \leq \alpha \}$ and the hypograph of $\text{hyp}(f) = \{ (x, \alpha) : x \in X, \alpha \in \mathbb{R}, f(x) \geq \alpha \}$ are closed relative to $X \times \mathbb{R}$.

Now, we present the following theorem from Bazaraa et al. [17].

**Theorem 1.1 (Weierstrass theorem)** Let $X \subseteq \mathbb{R}^n$ be a nonempty compact set and let $f : X \to \mathbb{R}$ be continuous on $X$, then, the problem

$$\min \{ f(x) : x \in X \}$$

attains its minimum.

Proof Since, $X$ is both nonempty closed and bounded and $f$ is continuous on $X$, therefore, $f$ is bounded below on $X$. Moreover, there exists a greatest lower bound $\alpha = \inf \{ f(x) : x \in X \}$. Let $0 < \varepsilon < 1$ and for each, $k = 1, 2, \ldots$, consider the set

$$X_k = \{ x \in X : \alpha \leq f(x) \leq \alpha + \varepsilon^k \}.$$  

By the definition of an infimum, the set $X_k \neq \phi$, for each $k$. Therefore, by selecting a point $x_k \in X_k$, for each $k = 1, 2, \ldots$, we may construct a sequence $\{ x_k \} \subseteq X$. Since, $X$ is bounded, there exists a convergent subsequence $\{ x_k \} \rightarrow \bar{x}$ indexed by the set $K$. Since, $X$ is closed, $\bar{x} \in X$.  

Since,

$$\alpha \leq f(x) \leq \alpha + \varepsilon^k, \forall k.$$  

By continuity of $f$, we get

$$\alpha = \lim_{k \to \infty, k \in K} f(x_k) = f(\bar{x}).$$  

Thus, we have established that there exists a solution $\bar{x}$, such that

$$f(\bar{x}) = \alpha = \inf \{ f(x) : x \in X \}.$$  

Hence, $\bar{x}$ is a minimizing solution. This completes the proof.
Definition 1.30 (Lower semicontinuous function) Let $X \subseteq \mathbb{R}^n$ be a nonempty set. Then a function $f : X \to \mathbb{R}$ is said to be lower semicontinuous at $\bar{x} \in X$, if either of the following two conditions is satisfied:

(i) For each $\varepsilon > 0$, there exists $\delta > 0$, such that,

$$-\varepsilon < f(x) - f(\bar{x}), \text{ whenever } |x - \bar{x}| < \delta.$$

(ii) For each sequence $\{x_n\}$ in $X$, converging to $\bar{x}$,

$$\liminf_{n \to \infty} f(x_n) \geq f(\lim_{n \to \infty} x_n) = f(\bar{x}),$$

where $\liminf f(x_n)$ denotes the infimum of the limit point of the sequence of real numbers $f(x^1), f(x^2), \ldots$.

Example 1.1 For example, the function $f : X = \mathbb{R} \to \mathbb{R}$, given by

$$f(x) = \begin{cases} x, & \text{if } x \neq 1, \\ \frac{1}{3}, & \text{if } x = 1, \end{cases}$$

is lower semicontinuous on $\mathbb{R}$.

The following theorem relates the closedness of the epigraph and lower level sets with the lower semicontinuity of a function $f$. A simple proof can be found in Bertsekas et al. [23].

Theorem 1.2 The function $f : X \to \mathbb{R}$ is lower semicontinuous on $X$, if the following equivalent conditions are satisfied:

(i) The set $\{x : x \in X, f(x) \leq \alpha\}$ is closed relative to $X$, for each real number $\alpha$.

(ii) The set $\{x : x \in X, f(x) > \beta\}$ is open relative to $X$, for each real number $\beta$.

(iii) The epigraph of $f$, i.e.,

$$\text{epi}(f) = \{(x, \alpha) : x \in X, \alpha \in \mathbb{R}, f(x) \leq \alpha\}$$

is closed relative to $X \times \mathbb{R}$.

Definition 1.31 (Upper semicontinuous function) Let $X \subseteq \mathbb{R}^n$ be a nonempty set. Then a function $f : X \to \mathbb{R}$ is said to be upper semicontinuous at $\bar{x} \in X$, if either of the following two conditions are satisfied:

(i) For each $\varepsilon > 0$, there exists $\delta > 0$, such that for each $x \in X$,

$$f(x) - f(\bar{x}) < \varepsilon, \text{ whenever } |x - \bar{x}| < \delta.$$
(ii) For each sequence \( \{x_n\} \) in \( X \), converging to \( \bar{x} \),
\[
\limsup_{n \to \infty} f(x_n) \leq f(\lim_{n \to \infty} x_n) = f(\bar{x}),
\]
where \( \limsup_{n \to \infty} f(x_n) \) denotes the supremum of the limit point of the sequence of real numbers \( f(x^1), f(x^2), \ldots \).

Example 1.2 For example, the function \( f : X = \mathbb{R} \to \mathbb{R} \), given by
\[
f(x) = \begin{cases} 
x^2, & \text{if } x \neq 0, \\
\frac{1}{2}, & \text{if } x = 0
\end{cases}
\]
is upper semicontinuous on \( \mathbb{R} \).

The following theorem relates the closedness of the hypograph and upper level sets with the upper semicontinuity of a function \( f \).

Theorem 1.3 The function \( f : X \to \mathbb{R} \) is lower semicontinuous on \( X \), if the following equivalent conditions are satisfied:

(i) The set \( \{x : x \in X, f(x) \geq \alpha\} \) is closed relative to \( X \), for each real number \( \alpha \).

(ii) The set \( \{x : x \in X, f(x) < \beta\} \) is open relative to \( X \), for each real number \( \beta \).

(iii) The hypograph of \( f \), i.e.,
\[
\text{hyp}(f) = \{(x, \alpha) : x \in X, \alpha \in \mathbb{R}, f(x) \geq \alpha\}
\]
is a closed relative to \( X \times \mathbb{R} \).

1.4 Derivatives and Hessians

In optimization theory, derivatives and Hessians play a very important role. We recall the following definitions from Mangasarian [176].

Definition 1.32 (Partial derivative and gradient) Let \( X \subseteq \mathbb{R}^n \) be an open set and let \( \bar{x} \in X \). Then the partial derivative of a function \( f : X \to \mathbb{R} \) at a point \( \bar{x} \), with respect to \( x_i, i = 1, \ldots, n \), denoted by \( \frac{\partial f(\bar{x})}{\partial x_i} \), is given by
\[
\frac{\partial f(\bar{x})}{\partial x_i} :=
\]
Suppose \( f(\bar{x}) \) is a real-valued function. Suppose that \( f(\bar{x}) \) is differentiable at \( \bar{x} \) and a function \( \alpha : \mathbb{R}^{n} \rightarrow \mathbb{R} \), such that

\[
    f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \|x - \bar{x}\| \alpha(\bar{x}, x - \bar{x}), \forall x \in X,
\]

where \( \lim_{x \to \bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0 \). The function \( f \) is said to be differentiable on \( X \), if it is differentiable at each \( \bar{x} \in X \).

**Definition 1.34 (Jacobian of a vector function)** Let \( X \subseteq \mathbb{R}^{m} \) be an open set and let \( \bar{x} \in X \). Suppose \( f : X \rightarrow \mathbb{R}^{m} \) be a vector valued function such that the partial derivative \( \frac{\partial f(\bar{x})}{\partial x} \) exists at \( \bar{x} \), for each \( i = 1, \ldots, m \) and \( j = 1, \ldots, n \). Then the Jacobian (matrix) of \( f \) at \( \bar{x} \) denoted by \( J(f)(\bar{x}) \) is a \( m \times n \) matrix defined by

\[
    J(f)(\bar{x}) = \begin{bmatrix}
        \frac{\partial f_1(\bar{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\bar{x})}{\partial x_n} \\
        \vdots & \ddots & \vdots \\
        \frac{\partial f_m(\bar{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\bar{x})}{\partial x_n}
    \end{bmatrix}.
\]

**Proposition 1.12 (Chain rule theorem)** Let \( X \subseteq \mathbb{R}^{n} \) be an open set and let \( \bar{x} \in X \). Suppose \( f : X \rightarrow \mathbb{R}^{m} \) be a vector valued function and let \( g : \mathbb{R}^{n} \rightarrow \mathbb{R} \) be a real-valued function. Suppose that \( f \) is differentiable at \( \bar{x} \) and \( g \) is differentiable at \( y = f(\bar{x}) \). Then the composite function

\[
    \varphi(x) := g(f(x))
\]

is also differentiable at \( \bar{x} \), and

\[
    \nabla \varphi(\bar{x}) = \nabla g(y) \nabla f(\bar{x}).
\]

**Definition 1.35 (Twice differentiable function and Hessian)** Let \( X \subseteq \mathbb{R}^{n} \) be an open set and let \( \bar{x} \in X \). Suppose \( f : X \rightarrow \mathbb{R}^{n} \) be a vector valued function. Then \( f \) is said to be twice differentiable at \( \bar{x} \), if there exists a vector \( \nabla f(\bar{x}) \), \( n \times n \) symmetric matrix \( \nabla^2 f(\bar{x}) \) and a function \( \alpha : X \rightarrow \mathbb{R} \), such that

\[
    f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \langle x - \bar{x}, \nabla^2 f(\bar{x})(x - \bar{x}) \rangle + \|x - \bar{x}\|^2 \alpha(\bar{x}, x - \bar{x}), \forall x \in \mathbb{R}^{n},
\]
where \( \lim_{x \to \bar{x}} \alpha(x, x - \bar{x}) = 0 \). The \( n \times n \) matrix \( \nabla^2 f(\bar{x}) \), called the Hessian matrix of \( f \) at \( \bar{x} \), is given by

\[
\nabla^2 f(\bar{x}) := \begin{bmatrix}
\frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \ldots & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f(\bar{x})}{\partial x_n \partial x_1} & \ldots & \frac{\partial^2 f(\bar{x})}{\partial x_n^2}
\end{bmatrix}.
\]

**Remark 1.3** Let \( X \subseteq \mathbb{R}^n \times \mathbb{R}^m \) be an open set and let \( f : X \to \mathbb{R} \) be a real-valued function, differentiable at \((\bar{x}, \bar{y}) \in X\). Then, we define

\[
\nabla_x f(\bar{x}, \bar{y}) := \left[ \frac{\partial f(\bar{x}, \bar{y})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x}, \bar{y})}{\partial x_n} \right]
\]

and

\[
\nabla_{\bar{x}} f(\bar{x}, \bar{y}) := \left[ \frac{\partial f(\bar{x}, \bar{y})}{\partial \bar{x}_1}, \ldots, \frac{\partial f(\bar{x}, \bar{y})}{\partial \bar{x}_m} \right].
\]

If \( f : X \to \mathbb{R}^m \) be a vector valued function, differentiable at \((\bar{x}, \bar{y}) \in X\), then, we define

\[
\nabla_x f(\bar{x}, \bar{y}) = \begin{bmatrix}
\frac{\partial f_1(\bar{x}, \bar{y})}{\partial x_1} & \ldots & \frac{\partial f_1(\bar{x}, \bar{y})}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m(\bar{x}, \bar{y})}{\partial x_1} & \ldots & \frac{\partial f_m(\bar{x}, \bar{y})}{\partial x_n}
\end{bmatrix}
\]

and

\[
\nabla_y f(\bar{x}, \bar{y}) = \begin{bmatrix}
\frac{\partial f_1(\bar{x}, \bar{y})}{\partial y_1} & \ldots & \frac{\partial f_1(\bar{x}, \bar{y})}{\partial y_m} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m(\bar{x}, \bar{y})}{\partial y_1} & \ldots & \frac{\partial f_m(\bar{x}, \bar{y})}{\partial y_m}
\end{bmatrix}.
\]

Now, we state one of the most fundamental theorems from classical theory of mathematical analysis.

**Proposition 1.13 (Mean value theorem)** Let \( X \subseteq \mathbb{R}^n \) be an open convex set and let \( f : X \to \mathbb{R} \) be a differentiable function on \( X \). Then, for all \( x, \bar{x} \in X \),

\[
f(x) = f(\bar{x}) + \langle \nabla f(x + \delta(x - \bar{x})), x - \bar{x} \rangle,
\]

for some real number \( \delta, 0 < \delta < 1 \).

**Proposition 1.14 (Taylor’s theorem (second order))** Let \( X \subseteq \mathbb{R}^n \) be an open convex set and let \( f : X \to \mathbb{R} \) be a twice continuously differentiable function on \( X \). Then, for all \( x, \bar{x} \in X \),

\[
f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{\langle (x - \bar{x}), \nabla^2 f(x + \delta(x - \bar{x}))(x - \bar{x}) \rangle}{2},
\]

for some real number \( \delta, 0 < \delta < 1 \).
Proposition 1.15 (Implicit function theorem) Let \( X \subseteq \mathbb{R}^n \times \mathbb{R}^m \) be an open set and let \( f : X \to \mathbb{R}^m \) be a vector valued function. Suppose that \( f \) has a continuous partial derivative at \((\bar{x}, \bar{y}) \in X\), such that \( f(\bar{x}, \bar{y}) = 0 \) and \( \nabla_y f(\bar{x}, \bar{y}) \) are nonsingular. Then, there exists a ball \( B_\varepsilon(\bar{x}, \bar{y}) \) with the radius \( \varepsilon > 0 \) in \( \mathbb{R}^{n+m} \), an open set \( \Omega \subseteq \mathbb{R}^n \) containing \( \bar{x} \), and an \( m \)-dimensional vector function \( \varphi \) with continuous first partial derivative on \( \Omega \), such that \( \nabla_y f(x, y) \) is nonsingular for all \((x, y) \in B_\varepsilon(\bar{x}, \bar{y}), \bar{y} = \varphi(\bar{x}) \) and
\[
 f(x, \varphi(x)) = 0, \forall x \in \Omega.
\]

1.5 Convex Sets and Properties

The notion of convexity of the sets is core of optimization theory. We start this section with the definition of a subspace of \( \mathbb{R}^n \).

**Definition 1.36 (Subspace)** A set \( X \subseteq \mathbb{R}^n \) is said to be a subspace if for each \( x, y \in X \) and \( p, q \in \mathbb{R} \), one has
\[
 px + qy \in X.
\]

Each subspace of \( \mathbb{R}^n \) contains the origin.

**Definition 1.37 (Affine set)** Let \( x \) and \( y \) be any two points in \( \mathbb{R}^n \). Then, the set of all points of the form
\[
 z = (1 - \lambda)x + \lambda y = x + \lambda(y - x), \lambda \in \mathbb{R}
\]
is called the line through \( x \) and \( y \). A subset \( X \) of \( \mathbb{R}^n \) is referred to as an affine set (or affine manifold) if it contains every line through any two points of it. Affine sets, which contain the origin are subspaces of \( \mathbb{R}^n \). It is easy to prove that if the set \( X \) is affine, then there exist \( x \in X \) and a subspace \( X_0 \) of \( X \), such that \( X = x + X_0 \). The subspace \( X_0 \) is said to be parallel to the affine set \( X \) and is uniquely determined for a given nonempty affine set \( X \).

The dimension of a nonempty affine set is defined as the dimension of the subspace parallel to it. The affine sets of dimensions 0, 1 and 2 are points, lines and planes, respectively. Every affine subset of \( \mathbb{R}^n \) can be characterized as the solution sets \( F \) (say) of the system of simultaneous linear equations in \( n \) variables, that is,
\[
 F := \{ x \in \mathbb{R}^n : Ax = b \},
\]
where \( A \) is an \( m \times n \) real matrix and a vector \( b \in \mathbb{R}^m \).
Definition 1.38 (Hyperplanes) Any \((n-1)\)-dimensional affine set in \(\mathbb{R}\) is called a hyperplane or a plane, for short. Any hyperplane is a set of the form
\[H = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = \beta \},\]
where \(\alpha \in \mathbb{R}^n \setminus \{0\}\) and \(\beta \in \mathbb{R}\).

The vector \(\alpha\) is called a normal to the hyperplane \(H\). Every other normal is either a positive or negative scalar multiple of \(\alpha\). Angle between two hyperplanes is the angle between their normal vectors. Every affine subset is the intersection of finite collection of hyperplanes. If \(\bar{x} \in H\), then the hyperplane \(H\) can be expressed as
\[H = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = \langle \alpha, \bar{x} \rangle \} = \bar{x} + \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = 0 \}.
\]
Therefore, \(H\) is an affine set parallel to \(\{x \in \mathbb{R}^n : \langle \alpha, x \rangle = 0 \}\). It is easy to see that the intersection of arbitrary collection of affine sets is again affine. This led to the concept of affine hull.

Definition 1.39 (Half-spaces) Let \(\alpha\) be a nonzero vector in \(\mathbb{R}^n\) and \(\beta\) be a scalar. Then, the sets
\[K^+ := \{x \in \mathbb{R}^n : \langle \alpha, x \rangle \geq \beta \} \text{ and } K^- := \{x \in \mathbb{R}^n : \langle \alpha, x \rangle \leq \beta \}\]
are called upper and lower closed half-spaces. The sets
\[H^+ := \{x \in \mathbb{R}^n : \langle \alpha, x \rangle > \beta \} \text{ and } H^- := \{x \in \mathbb{R}^n : \langle \alpha, x \rangle < \beta \}\]
are called upper and lower open half-spaces.

It is clear that all four sets are nonempty and convex sets. We note that, if we replace \(\alpha\) and \(\beta\), respectively by \(\lambda \alpha\) and \(\lambda \beta\), for some \(\lambda \neq 0\), we will get the same quartet of half-spaces. Thus half-spaces depend only on the hyperplane
\[H = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = \beta \}.
\]

Definition 1.40 (Affine hull) Given any \(X \subseteq \mathbb{R}^n\), the intersection of collections of affine sets containing \(X\), is called affine hull of \(X\), denoted by \(\text{aff}(X)\). It can be proved that \(\text{aff}(X)\) is the unique smallest affine set containing \(X\).

Definition 1.41 (Supporting hyperplane) Let \(X \subseteq \mathbb{R}^n\) be a nonempty set and \(\bar{x} \in \text{bd}(X)\). A hyperplane
\[H = \{x \in \mathbb{R}^n : \langle \alpha, x - \bar{x} \rangle = 0, \alpha \neq 0, \alpha \in \mathbb{R}^n \}\]
is called a supporting hyperplane to \(X\) at \(\bar{x}\), if either
\[X \subseteq H^+, \text{ that is, } \langle \alpha, x - \bar{x} \rangle \geq 0, \forall x \in X\]
Pseudolinear Functions and Optimization

or

\[ X \subseteq H^{-}, \text{ i.e., } \langle \alpha, x - \bar{x} \rangle \leq 0, \forall x \in X. \]

The hyperplane \( H \) is called a proper supporting hyperplane, if in addition to the above properties \( X \nsubseteq H \) is also satisfied.

The following proposition provides a characterization for the affine hull of a set.

**Proposition 1.16** The affine hull of a set \( X \), that is, \( \text{aff}(X) \) is a set consisting of all points of the form

\[ z : z = \lambda_1 x_1 + \ldots + \lambda_n x_n, \lambda_i \in \mathbb{R}, \sum_{i=1}^{k} \lambda_i = 1, \]

where \( k \) is a natural number.

Proof Let \( C \) be the set of collection of all the points of the above form. Let \( x, y \in C \), then, we have

\[ x = \sum_{i=1}^{k} \lambda_i x_i \]

and

\[ y = \sum_{j=1}^{k} \mu_j y_j, \]

where \( x_i, y_j \in X, \lambda_i, \mu_j \geq 0, \forall i, j = 1, \ldots, k \) and

\[ \sum_{i=1}^{k} \lambda_i = 1 = \sum_{j=1}^{k} \mu_j. \]

Now, for any \( \alpha \in ]0,1[ \), we get

\[ \alpha x + (1 - \alpha)y = \sum_{i=1}^{k} \alpha \lambda_i x_i + \sum_{j=1}^{k} (1 - \alpha) \mu_j y_j \]

with \( \sum_{i=1}^{k} \alpha \lambda_i + \sum_{j=1}^{k} (1 - \alpha) \mu_j = \alpha + (1 - \alpha) = 1. \) Thus, \( \alpha x + (1 - \alpha)y \in C. \) Hence, \( C \) is an affine set. Therefore, \( \text{aff}(X) \subseteq C. \)

On the other hand, it can be seen that if \( x = \sum_{i=1}^{k} \lambda_i x_i, \lambda_i \geq 0, \forall i = 1, \ldots, k \) with \( \sum_{i=1}^{k} \lambda_i = 1 \) then \( x \in \text{aff}(X). \) For \( k = 2, \) the result follows by the definition of affine set and for \( k = 3, \) the result follows by induction. Hence, \( C \subseteq \text{aff}(X). \) Hence, we get \( C = \text{aff}(X). \)

**Definition 1.42 (Affine independence)** Any \( k \) points \( x_1, \ldots, x_k \) in \( \mathbb{R}^n \) are said to be affinely independent, if \( \{x_1, \ldots, x_k\} \) has dimension \( k - 1, \) that is the vectors \( x_2 - x_1, \ldots, x_k - x_{k-1} \) are linearly independent.
Definition 1.43 (Line segment) Let \( x \) and \( y \) be any two points in \( \mathbb{R}^n \). We define the following line segments joining any two points of it:

(i) closed line segment \([x, y]\) := \{z : z = (1 - \lambda)x + \lambda y, \lambda \in [0, 1]\};

(ii) open closed line segment \([x, y]\) := \{z : z = (1 - \lambda)x + \lambda y, \lambda \in ]0, 1]\};

(iii) closed open line segment \([x, y]\) := \{z : z = (1 - \lambda)x + \lambda y, \lambda \in [0, 1]\};

(iv) open line segment \([x, y]\) := \{z : z = (1 - \lambda)x + \lambda y, \lambda \in ]0, 1]\}.

Definition 1.44 (Convex set) A set \( X \subseteq \mathbb{R}^n \) is said to be convex, if it contains closed line segment joining any two points of it. In other words, \( X \) is a convex set, if \( \forall x, y \in X \Rightarrow \lambda x + (1 - \lambda)y \in X, \forall \lambda \in [0, 1] \).

It is obvious that affine sets are convex sets. However, the converse is not true. For example, solid ellipsoid and cube in \( \mathbb{R}^3 \) are convex sets, but not affine. See Figure 1.2.

Next, we prove the following important results for convex sets. The proof follows along the lines of Mangasarian [176].

Theorem 1.4 Let \( \{X_i, i \in I\} \) be a family (finite or infinite) of convex sets in \( \mathbb{R}^n \), then, their intersection \( \bigcap_{i \in I} X_i \) is a convex set. Moreover, if \( C, D \) are convex sets in \( \mathbb{R}^n \), then,

\[
C + D := \{x + y : x \in C, y \in D\}
\]

and

\[
\alpha C := \{\alpha x : x \in C\}, \alpha \in \mathbb{R},
\]

are convex sets.

Proof Let \( a, b \in \bigcap_{i \in I} X_i \). Then \( a, b \in X_i \), for each \( i \in I \). Now, for any \( \lambda \in [0, 1] \), \( \lambda a + (1 - \lambda)b \in X_i \), for each \( i \in I \). Therefore, \( \lambda a + (1 - \lambda)b \in \bigcap_{i \in I} X_i \), for any \( \lambda \in [0, 1] \). Hence, \( \bigcap_{i \in I} X_i \) is a convex set.

If \( C, D \) are convex sets and \( a, b \in C + D \), then \( a = x_1 + y_1 \) and \( b = x_2 + y_2 \), where \( x_1, x_2 \in C \) and \( y_1, y_2 \in D \). Now, for each \( \lambda \in [0, 1] \), we have

\[
\lambda a + (1 - \lambda)b = \lambda(x_1 + y_1) + (1 - \lambda)(x_2 + y_2)
= (\lambda x_1 + (1 - \lambda)x_2) + (\lambda y_1 + (1 - \lambda)y_2) \in C + D.
\]

Hence, \( C + D \) is a convex set. Similarly, we can prove that \( \alpha C, \alpha \in \mathbb{R} \) is a convex set.

The following proposition from Mangasarian [176] will be used to prove the strict separation theorem.
Proposition 1.17  Let $X$ and $Y$ be two sets in $\mathbb{R}^n$ such that $X$ is compact and $Y$ is closed, then the sum $Z = X + Y$ is compact.

Proof  Let $\bar{z} \in \text{cl}(Z)$. Then there exists a sequence $\{z_n\}$ in $Z$, which converges to $\bar{z}$. Then there exists a sequence $\{x_n\} \in X$ and $\{y_n\} \in Y$, such that $z_n = x_n + y_n$, for $n = 1, 2, \ldots$. Since, $X$ is compact, there exists a subsequence $\{x_{n_i}\}$, which converges to $\bar{x} \in X$. Then, we have

$$\lim_{i \to \infty} z_{n_i} = \bar{z}$$
$$\lim_{i \to \infty} x_{n_i} = \bar{x} \in X.$$  

Since, $Y$ is closed, we have

$$\lim_{i \to \infty} y_{n_i} = \bar{z} - \bar{x} \in Y.$$ 

Therefore, $\bar{z} = \bar{x} + (\bar{z} - \bar{x}) \in Z$. Hence, $Z$ is closed.

The following definition and characterization are from Mangasarian [176].

Definition 1.45  (Convex hull)  Let $X \subseteq \mathbb{R}^n$ be a given set. Then, the intersection of family of all convex sets containing $X$ is the smallest convex set containing $X$. This set is called the convex hull of $X$ and denoted by $\text{co}(X)$. It can be shown that the convex hull of the set $X$ can be expressed as

$$\text{co} \ X = \left\{ z : z = \lambda_1 x_1 + \ldots + \lambda_n x_n, x_i \in X, \lambda_i \in \mathbb{R}_+, \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$ 

For example, see Figure 1.3.
Definition 1.46 (Convex combination) A point $x \in \mathbb{R}^n$, such that

$$x = \sum_{i=1}^{k} \lambda_i a_i \text{ with } a_i \in X \subseteq \mathbb{R}^n, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1$$

is called a convex combination of $a_1, \ldots, a_n \in \mathbb{R}^n$.

The following theorem provides a characterization for convex sets. The proof can be found in Mangasarian [176], Rockafellar [238], and Tuy [273].

**Theorem 1.5** A subset $X \subseteq \mathbb{R}^n$ is convex if and only if convex combinations of points of $X$ are contained in $X$. Equivalently, $X$ is convex if and only if

$$x_1, x_2, \ldots, x_k \in X, \lambda_1, \lambda_2, \ldots, \lambda_k \geq 0, \lambda_1 + \lambda_2 + \ldots + \lambda_k = 1$$

$$\Rightarrow \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k \in X. \quad (1.1)$$

From the above discussion, it is clear that if $X$ is a convex set, then any point of $X$ can be expressed as a convex combination of $m$ points of $X$, where $m \geq 1$ is arbitrary. However, for an arbitrary set $X$, if any point is a convex combination of $m$ points of $X$, then $m \leq n + 1$. This is one of the most applicable results in optimization theory known as Caratheodory’s theorem. There are several approaches to prove Caratheodory’s theorem. See Mangasarian [176], Rockafellar [238], and Tuy [273].

**Theorem 1.6** (Caratheodory’s theorem) Let $X \subseteq \mathbb{R}^n$ be a nonempty set. Let $x$ be a convex combination of points of $X$. Then $x$ can be expressed as a convex combination of $n + 1$ or fewer points of $X$. 
The following result is from Bertsekas et al. [23] and Rockafellar [238].

**Proposition 1.18** Let $X \subseteq \mathbb{R}^n$ be a compact set, then $\text{co}(X)$ is a compact set.

**Proof** To show that $\text{co}(X)$ is a compact set, it is sufficient to show that every sequence in $\text{co}(X)$ has a convergent subsequence, whose limit is in $\text{co}(X)$. Consider a sequence $\{z_r\} \in \text{co}(X)$. By Theorem 1.6, for all $r$ there exist sequences $\{\lambda^r_i\}, \lambda^r_i \geq 0, i = 1, 2, \ldots, n + 1, \sum_{i=1}^{n+1} \lambda^r_i = 1$ and $\{x^r_i\}, x^r_i \in X, r = 1, 2, \ldots, n + 1$, such that

$$z_r = \sum_{i=1}^{n+1} \lambda^r_i x^r_i.$$

Since, $X$ is a compact set, the sequence $\{x^r_i\}$ is bounded. Furthermore, since $\lambda^r_i \geq 0, i = 1, 2, \ldots, n + 1$ and $\sum_{i=1}^{n+1} \lambda^r_i = 1$, the sequence, $\{\lambda^r_i\}$ is also bounded. Therefore, the sequence

$$\{\lambda^r_1, \ldots, \lambda^r_{n+1}, x^r_1, \ldots, x^r_{n+1}\}$$

is bounded. From the Bolzano-Weierstrass theorem, the bounded sequence

$$\{\lambda^1_1, \ldots, \lambda^{n+1}_1, x^1_1, \ldots, x^{n+1}_1\}$$

has a limit point say $\{\lambda_1, \ldots, \lambda_{n+1}, x_1, \ldots, x_{n+1}\}$, which must satisfy

$$\lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1 \text{ and } x_i \in X, \forall i = 1, \ldots, n + 1.$$

Let us define $z = \sum_{i=1}^{n+1} \lambda_i x_i$. Thus, $z$ is a limit point of the sequence $\{z_r\}$ and the sequence $z_r \to z \in \text{co}(X)$. Hence, $\text{co}(X)$ is a compact set.

**Remark 1.4** However, Proposition 1.18 does not hold true, if we assume that the set $X$ is a closed set, rather than a compact set. The following example from Bertsekas et al. [23] illustrates the fact. Consider the closed set $X$, given by

$$X = \{(0, 1)\} \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 \geq 0, x_1 > 0, x_2 > 0 \}.$$

Then, the convex hull of $X$, is given by

$$X = \{(0, 1)\} \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0 \},$$

which is not a closed set.

Now, we give the definition of a radially continuous function.
Definition 1.47 (Radially continuous function) Let $X$ be a nonempty convex subset of $\mathbb{R}^n$. A function $f : X \to \mathbb{R}$ is said to be radially upper(lower) semicontinuous on $X$ (also, known as upper (lower) hemicontinuous on $X$), if for every pair of distinct points $x, y \in X$, the function $f$ is upper(lower) semicontinuous on the line segment $[x, y]$.

Moreover, the function $f$ is said to be radially continuous on $X$ (also, known as hemicontinuous on $X$), if it is both radially upper semicontinuous and radially lower semicontinuous on $X$.

The following example from Ansari et al. [2] shows that a function may be radially continuous at a point, even not being continuous at that point.

Example 1.3 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(x, y) := \begin{cases} \frac{2x^2y}{x^2+y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

We note that the function $f$ is continuous at every point of $\mathbb{R}^2$, except at $(0, 0)$. If we approach $x$ along the path $y = mx^2$, then, we have

$$\lim_{(x,y) \to (0,0)} f(x, y) = \frac{2m}{1+m^2}$$

which is different for different values of $m$. However, $f$ is radially continuous, because if we approach $(0,0)$ along the line $y = mx$, then,

$$\lim_{(x,y) \to (0,0)} f(x, y) = \frac{2mx}{x^2 + m^2} = 0 = f(0, 0).$$

Definition 1.48 (Cone) A subset $X \subseteq \mathbb{R}^n$ is called a cone, if it is closed under positive scalar multiplication, i.e., if

$$x \in X, \lambda > 0 \Rightarrow \lambda x \in X.$$ 

The origin $0$ itself may or may not be included in the set. However, the origin always lies in the closure of a nonempty cone. A set $\alpha + X$, which is a translation of a cone $X$ by $\alpha \in \mathbb{R}^n$, is called a cone with apex $\alpha$.

A cone $X$ is said to be pointed, if it contains a line. In other words, $X$ is said to be pointed, if

$$X \cap (-X) = \{0\}.$$ 

A subset $X \subseteq \mathbb{R}^n$ is a convex cone, if $X$ is a convex set and cone, that is,

$$\forall x, y \in X \text{ and } \lambda, \mu \geq 0, \lambda x + \mu y \in X.$$ 

Subspaces of $\mathbb{R}^n$, closed and open half-spaces corresponding to a hyperplane through the origin are examples of convex cones. The nonnegative orthant of $\mathbb{R}^n$, that is,

$$\{x = (x_1, \ldots, x_n) : x_i \geq 0, i = 1, \ldots, n\}$$
is an example of a closed, convex and pointed cone. Like convex sets, intersection of an arbitrary collection of convex cones is a convex cone.

**Theorem 1.7** A subset \( X \subseteq \mathbb{R}^n \) is a convex cone, if and only if it is closed under addition and positive scalar multiplication.

Proof Let \( X \) be a cone and let \( x, y \in X \). If \( X \) is convex, the vector \( z = \frac{1}{2}(x + y) \in K \), hence, the vector \( x + y = 2z \in X \). By definition of a cone \( X \) is closed under positive scalar multiplication. To prove the converse, we assume that cone \( X \) is closed under addition and positive scalar multiplication. Then for \( x, y \in X \) and \( 0 < \lambda < 1, \lambda x \in X, (1 - \lambda)y \in X \) and therefore, \( \lambda x + (1 - \lambda)y \in X \). Hence, \( X \) is convex.

The following corollaries are a direct consequence of the above theorem:

**Corollary 1.1** A subset \( X \subseteq \mathbb{R}^n \) is a convex cone if and only if it contains all the positive linear combinations of its elements.

**Corollary 1.2** Let \( X \subseteq \mathbb{R}^n \) be an arbitrary set and let \( S \) be the set of all convex combinations of the elements of \( X \). Then \( S \) is the smallest convex cone containing \( X \). Furthermore, if \( X \) is a convex set, then the set

\[
S := \{\lambda x : x \in X, \lambda > 0\}
\]

is the smallest convex cone, which includes \( X \).

**Definition 1.49 (Conic combination and polyhedral cone)** Let \( S = \{x_1, x_2, \ldots, x_n\} \) be a given set of vectors. Then, a vector

\[
x = \lambda_1 x_1 + \ldots + \lambda_n x_n, \lambda_i \geq 0, i = 1, \ldots, n
\]

is called the conic combination of the vectors. The set of all conic combination of the elements of \( S \) is called the polyhedral cone generated by \( x_1, \ldots, x_n \).

**Definition 1.50 (Conic hull)** The intersection of all the convex cones containing the given set \( X \) is called the conic hull of \( X \) and is denoted by \( \text{con}(X) \). In fact, \( \text{con}(X) \) is the smallest convex set containing the set \( X \).

**Definition 1.51 (Polar or dual cone)** Let \( X \subseteq \mathbb{R}^n \) be a nonempty set. Then the cone

\[
X^* := \{y : \langle y, x \rangle \leq 0, \forall x \in X\}
\]

is called the polar cone to \( X \). It is clear that polar cone \( X^* \) is the intersection of closed half-spaces and therefore, it is closed and convex, independent of the set closedness or convexity of the set \( X \). It is easy to see that if \( X \) is a subspace, then \( X^* \) is equal to the orthogonal complement of \( X \), that is, \( X^* = X^\perp \).

The proof of the following proposition can be found in Bertsekas et al. [23].
Proposition 1.19 Let $X \subseteq \mathbb{R}^n$ be a nonempty set. Then, we have

(i) $X^* = (\text{cl}(X))^* = (\text{co}(X))^* = (\text{con}(X))^*$;

(ii) (Polar cone theorem) In addition, if $X$ is a cone, then

$(X^*)^* = \text{cl}(\text{co}(X))$.

In particular, if $X$ is closed and convex, we have $(X^*)^* = X$.

Definition 1.52 (Normal cone) Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set. A vector $\xi \in \mathbb{R}^n$ is said to be normal to $X$ at $x \in X$, if

$\langle \xi, y - x \rangle \leq 0, \forall y \in X$.

The set of all such vectors is called the normal cone to $X$ at $x$ and is denoted by $N_X(x)$.

Next, we state an important result about the interior of a convex set. The proof can be found in Bertsekas et al. [23].

Proposition 1.20 A point $x \in X \subseteq \mathbb{R}^n$ is an interior point of $X$, if for every $y \in \mathbb{R}^n$, there exists $\alpha > 0$, such that

$x + \alpha(y - x) \in X$.

We know that the interior of a nonempty convex set may be empty. For instance, the interior of a line or a triangle in $\mathbb{R}^3$ is an empty set. The interior of these convex sets is nonempty relative to the affine hull of the sets. In fact, a convex set has a nonempty interior relative to the smallest affine set containing it. This led to the introduction of the concept of relative interiors.

1.6 Relative Interiors

Definition 1.53 (Relative interior) The relative interior of a convex set $X$, denoted by $\text{ri}(X)$ is defined as the interior of $X$, relative to the affine hull of $X$. In other words, $x \in \text{ri}(X)$ if there exists an open ball $\mathbb{B}_\varepsilon(x)$, centered at $x$, such that

$\mathbb{B}_\varepsilon(x) \cap \text{aff}(X) \subseteq X$.

More precisely,

$\text{ri}(X) := \{x \in \text{aff}(X) : \text{there exists } \varepsilon > 0, \text{ such that } B_\varepsilon(x) \cap \text{aff}(X) \subseteq X\}$. 
A vector in the closure of $X$, which is not a relative interior point of $X$ is called relative boundary of $X$. We know that if $X \subseteq \mathbb{R}^n$ is a convex set, then $\text{aff}(X) = \mathbb{R}^n$. Therefore, in this case interior and relative interior of the set $X$ coincide, that is, $\text{int}(X) = \text{ri}(X)$.

For nonempty convex sets $X$ and $Y$, we know that $X \subseteq Y \Rightarrow \text{int}(X) \subseteq \text{int}(Y)$ and $\text{cl}(X) \subseteq \text{cl}(Y)$.

However, unlike the concept of closure and interior of nonempty convex sets, we have $X \subseteq Y \not\Rightarrow \text{ri}(X) \subseteq \text{ri}(Y)$, which is a drawback of the concept of relative interiors. The following example from Dhara and Dutta [63] justifies the statement.

**Example 1.4** Let $X = \{(0,0)\}$ and $Y = \{(0,y) \in \mathbb{R}^2 : y \geq 0\}$. It is clear that $X \subseteq Y$. However, $\text{ri}(X) = \{(0,0)\}$ and $\text{ri}(Y) = \{(0,y) \in \mathbb{R}^2 : y > 0\}$. Thus, $\text{ri}(X)$ and $\text{ri}(Y)$ are nonempty and disjoint.

The following proposition gives some basic facts about relative interiors and closures of a nonempty convex set. The proof can be found in Bertsekas et al. [23].

**Proposition 1.21** Let $X \subseteq \mathbb{R}^n$ be a nonempty convex set. Then the following hold:

(i) *(Nonemptiness)* $\text{ri}(X)$ and $\text{cl}(X)$ are nonempty convex set with same affine hull as $X$.

(ii) *(Line segment principle)* Let $x \in \text{ri}(X)$ and $y \in \text{cl}(X)$. Then,

$$x + \lambda(y - x) \in \text{ri}(X), \forall \lambda \in [0,1].$$

(iii) *(Prolongation principle)* $x \in \text{ri}(X)$ if and only if every line segment in $X$ with $x$ has one end point that can be prolonged beyond $x$ without leaving $X$. In other words, for all $y \in X$, there exists $\lambda > 1$, such that

$$x + (1 - \lambda)(y - x) \in X.$$

The following result from Tuy [273] illustrates a very important property of relative interiors.

**Proposition 1.22** Let $X \subseteq \mathbb{R}^n$ be a nonempty convex set. Then $\text{ri}(X)$ is nonempty.

Proof Let $S = \text{aff}(X)$ and $\text{dim} S = k$. Therefore, there exists $k + 1$ affinely independent elements $x_0, x_1, \ldots, x_k$ of $X$. We show that the point

$$a = \frac{1}{k+1} \sum_{i=1}^{k} x_i \in \text{ri}(X).$$
Let \( x \in S \), then
\[
x = \sum_{i=1}^{k+1} \lambda_i x_i \quad \text{with} \quad \sum_{i=1}^{k+1} \lambda_i = 1.
\]
Now, we consider
\[
a + \alpha(x - a) = \frac{1}{k+1} \sum_{i=1}^{k+1} x_j \alpha + \alpha \left( \sum_{i=1}^{k+1} \lambda_i x_i - \frac{1}{k+1} \sum_{i=1}^{k+1} x_i \right)
\]
\[
= \sum_{i=1}^{k+1} \left( (1 - \alpha) \frac{1}{k+1} + \alpha \lambda_i \right) x_i = \sum_{i=1}^{k+1} \mu_i x_i
\]
where \( \mu_i = (1 - \alpha) \frac{1}{k+1} + \alpha \lambda_i \) and \( \sum_{i=1}^{k+1} \mu_i = 1 \). For \( \alpha > 0 \) sufficiently small, we have
\[
\mu_i > 0, \quad i = 1, \ldots, k+1.
\]
Hence, \( a + \alpha(x - a) \in X \). Therefore, by Proposition 1.21, \( a \) is an interior point of \( X \) relative to \( S = \text{aff}(X) \). Therefore, \( a \in \text{ri}(X) \).

**Corollary 1.3** Let \( X \subseteq \mathbb{R}^n \) be a nonempty convex set and let \( C \) be the nonempty set generated by \( \{(1, x) : x \in C\} \). Then \( \text{ri}(C) \) consists of the pairs \( (\lambda, x) \) such that \( \lambda > 0 \) and \( x \in \lambda \text{ri}(C) \).

The following corollaries from Rockafellar [238] will be needed in the sequel.

**Corollary 1.4** Let \( C \subseteq \mathbb{R}^n \) be a convex set and let \( M \) be an affine set, which contains a point of \( \text{ri}(C) \). Then,
\[
\text{ri}(M \cap C) = M \cap \text{ri}(C) \quad \text{and} \quad \text{cl}(M \cap C) = M \cap \text{cl}(C).
\]

**Corollary 1.5** Let \( C_1 \) be a convex set and let \( C_2 \) be the convex set contained in the \( \text{cl}(C_1) \) but not contained in the relative boundary of \( C_1 \). Then
\[
\text{ri}(C_2) \subseteq \text{ri}(C_1).
\]

**Definition 1.54 (Recession cone)** Let \( X \subseteq \mathbb{R}^n \) be a nonempty set. Then \( X \) is said to recede in the direction \( y, y \neq 0 \) if and only if
\[
x + \lambda y \in X, \forall \lambda \geq 0 \quad \text{and} \quad x \in C.
\]
The set of all vectors \( y \in \mathbb{R}^n \), satisfying the condition, including \( y = 0 \), denoted by \( 0^+ X \), is called the recession cone of \( X \). Directions in which \( X \), recedes is referred to as direction of recessions of \( X \).

Now, we present the following example of recession cone from Rockafellar [238].
Example 1.5 We consider the following convex set in $\mathbb{R}^n$, given by

$$X = \{(x_1, x_2) | x_1^2 + x_2^2 \leq 1\}.$$ 

Then the recession cone of $X$, is given by $0^+ X := \{(x_1, x_2) | x_1 = x_2 = 0\} \bigcup \{(0,0)\}$.

An unbounded closed convex set contains at least one point at infinity, that is, it recedes in at least one direction.

The following theorem is from Rockafellar [238].

Theorem 1.8 Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $0^+ X$ will be closed and it consists of all possible limits of sequences of the form $\lambda_1 x_1, \lambda_2 x_2, \ldots$, where $x_i \in X$ and $\lambda_i \downarrow 0$. In fact, for the convex cone $C$ in $X \subseteq \mathbb{R}^{n+1}$ generated by $\{(1,x) : x \in X\}$, one has

$$\text{cl}(C) = C \bigcup \{(0,x) : x \in 0^+ X\}.$$ 

Proof By Corollary 1.3, the hyperplane $M = \{(1,x) : x \in \mathbb{R}^n\}$ must intersect with $\text{ri}(C)$. Therefore, by the closure rule in Corollary 1.4, we have

$$M \cap \text{cl}(C) = \text{cl}(M \cap C) = M \cap C = \{(1,x) : x \in X\}.$$ 

Therefore, the cone

$$\bar{C} := \{(\lambda,x) : \lambda > 0, x \in \lambda x\} \bigcup \{(0,x) : x \in 0^+ X\}$$

must contain $\text{cl}(C)$, because of the maximality property. Again since, the cone $\bar{C}$ is contained in the half-space

$$H := \{(\lambda,x) : \lambda \geq 0\}$$

and meets $\text{int}(H)$, therefore, by Corollary 1.5, $\text{ri}(\bar{C})$ must be entirely contained in $\text{int}(H)$. Hence, $\text{ri}(\bar{C})$ must be contained in $C$. Now, we have

$$\text{cl}(C) \subseteq \bar{C} \subseteq \text{cl}(\text{ri}(\bar{C})) \subseteq \text{cl}(C),$$

which implies that $\text{cl}(C) = \bar{C}$. Again, the set $\{(0,x) : x \in 0^+ X\}$ is the intersection of $\text{cl}(C)$ with $\{(0,x) : x \in \mathbb{R}^n\}$, therefore, it is closed set and consists of the limits of the sequences of the form $\lambda_1(1,x_1), \lambda_2(1,x_2), \ldots$, where $x_i \in X$ and $\lambda_i \downarrow 0$.

The following theorem provides a characterization of boundedness of a closed convex set in terms of its recession cone. The proof follows along the lines of Rockafellar [238].

Theorem 1.9 Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $X$ is bounded if and only if its recession cone contains $0^+ X$ consists of the zero vector alone.
Proof If \( X \) is bounded, it certainly contains no half-line, so that \( 0^+ X = \{0\} \). Suppose conversely that \( X \) is unbounded. Then it contains a sequence of nonzero vectors \( x_1, x_2, \ldots \) whose Euclidean norms \( |x_i| \) increase without bound. The vectors \( \lambda_i x_i \), where \( \lambda_i = \frac{1}{|x_i|} \), all belong to the unit sphere \( S = \{x : |x| = 1\} \). Since, \( X \) is closed and bounded, there exists some subsequence of \( \lambda_1 x_1, \lambda_2 x_2, \ldots \), will converge to a certain \( y \in X \). By Theorem 1.8, this is a nonzero vector of \( 0^+ X \).

**Definition 1.55 (Faces and vertex/extreme points)** Let \( X \subseteq \mathbb{R}^n \) be a convex set. Then a face \( \bar{X} \) of the set \( X \) is a convex subset of \( X \), such that every closed line segment in \( X \) with relative interior points in \( \bar{X} \) has both end points in \( \bar{X} \). The empty set and the set \( X \) are faces of \( X \). A face of zero dimension is called a vertex or extreme point. In other words, any point \( x \in X \) is called a vertex (or extreme point) if there exist no two distinct points \( x_1, x_2 \), such that \( x \in [x_1, x_2] \). In other words, if \( x \) is a vertex, then

\[
x \in [x_1, x_2] \Rightarrow x = x_1 \text{ or } x = x_2.
\]

**Definition 1.56 (Exposed faces and exposed points)** Let \( X \subseteq \mathbb{R}^n \) be a convex set. Then the exposed faces of the set \( X \) are the set of the form \( X \cap H \), where \( H \) is a nontrivial supporting hyperplane to \( X \). An exposed point of \( X \) is an exposed face, which is a point. In other words, exposed points of a convex set \( X \) is a point, through which there is a supporting hyperplane, which contains no other points of \( X \).

The following proposition from Rockafellar [238] relates extreme points and the exposed points of a convex set.

**Proposition 1.23 (Straszevicz’s theorem)** Let \( X \subseteq \mathbb{R}^n \) be a convex set. Then, the set of exposed points of \( X \) is a dense subset of the set of extreme points. Thus every exposed point is the limit of some sequence of exposed points.

**Remark 1.5** A convex set \( X \subseteq \mathbb{R}^n \) may have no vertices (for instance, the open ball \( B_\varepsilon(\bar{x}) \) and the hyperplane \( H = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = \beta\} \)), finite number of vertices, (for instance, the set \( \{x \in \mathbb{R}^n : x \geq (e, x) = 1\} \), where \( e = (1, \ldots, 1) \) is \( n \)-vectors of 1s with \( e_i = 1 \) and \( e_j = 0 \), for \( i \neq j \)) or an infinite number of vertices, (for instance, the closed ball \( B_\varepsilon[\bar{x}] \) has an infinite number of vertices given by \( \{x : x \in \mathbb{R}^n, ||x - \bar{x}|| = \varepsilon\} \).

**Definition 1.57 (Polyhedron and polytope)** A set \( X \subseteq \mathbb{R}^n \) which is the intersection of a finite number of closed half-spaces in \( \mathbb{R}^n \) is called a polyhedron. A polyhedron is a convex hull of finite many points. A polyhedron, which is bounded is called a polytope. Hence, a polytope is a closed and bounded convex set. Polyhedral convex sets are more well behaved than ordinary convex sets due to lack of curvature. An extreme point (0-dimensional face) of a polyhedron is also called a vertex and 1-dimensional face is called an edge.
Definition 1.58 (Simplex) Let \( x_0, x_1, \ldots, x_k \) be \((k+1)\) distinct point in \( \mathbb{R}^n \) with \( k \leq n \). If \( \{x_0, x_1, \ldots, x_k\} \) are affinely independent, then its convex hull of the set, that is,

\[
\text{co } X := \left\{ z : z = \lambda_0 x_0 + \lambda_1 x_1, \ldots, \lambda_n x_n, x_i \in X, \lambda_i \in \mathbb{R}_+, \sum_{i=0}^{k} \lambda_i = 1 \right\}.
\]

is called \( k \)-simplex with vertices \( x_0, x_1, \ldots, x_k \). In other words, \( k \)-simplex in \( \mathbb{R}^n \) is a convex polyhedron having \( k+1 \) vertices. For instance, a \( 0 \)-simplex is a point, \( 1 \)-simplex is a line, \( 2 \)-simplex is a triangle and \( 3 \)-simplex is a tetrahedron.

Now, we state the following result from Rockafellar [238].

**Corollary 1.6 (Minkowski theorem)** Every closed and bounded convex set in \( \mathbb{R}^n \) is the convex hull of its extreme points.

### 1.7 Convex Functions and Properties

In this section, we will study convex functions and their properties. We start with the definition of a convex function.

**Definition 1.59 (Convex function)** Let \( X \subseteq \mathbb{R}^n \) be a nonempty convex set. Then a function \( f : X \to \mathbb{R} \) is said to be convex function on \( X \), if for all \( x, y \in X \), we have

\[
f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y), \forall \lambda \in [0, 1]. \quad (1.2)
\]

A convex function is proper if and only if its epigraph is nonempty and does not contain a vertical line. A function \( f : X \to \mathbb{R} \) is said to be concave if \(-f\) is convex on \( X \). In other words, the function \( f \) is said to be concave, if

\[
f((1 - \lambda)x + \lambda y) \geq (1 - \lambda)f(x) + \lambda f(y), \forall \lambda \in [0, 1]. \quad (1.3)
\]

An affine function i.e., the function of the form

\[
f(x) = \langle c, x \rangle + \alpha, c, x \in \mathbb{R}^n
\]

and \( \alpha \in \mathbb{R} \). The function \( f \) is said to be strictly convex (strictly concave), if the above inequality (1.3) (respectively, (1.4)) is strict for \( x \neq y \) and \( \lambda \in ]0, 1[ \). Every strictly convex (strictly concave) function is convex (concave), but the converse is not true. For example, the affine functions are convex (concave) but not strictly convex (strictly concave).

The following proposition presents the characterization of convex functions in terms of the epigraph of \( f \).
Proposition 1.24 Let \( X \subseteq \mathbb{R}^n \) be a convex set. The function \( f : X \to \mathbb{R} \) is convex if and only if \( \text{epi}(f) \) is a convex subset of \( \mathbb{R}^{n+1} \).

Proof Suppose that \( f : X \to \mathbb{R} \) is a convex function. Let \((x, \alpha)\) and \((y, \beta)\) be any two points of \( \text{epi}(f) \). By the convexity of \( f \), we have
\[
 f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y).
\]
From which, it follows that
\[
 (1 - \lambda)(x, \alpha) + \lambda (y, \beta) \in \text{epi}(f), \forall \lambda \in [0, 1].
\]
Hence, \( \text{epi}(f) \) is a convex subset of \( \mathbb{R}^{n+1} \).

Suppose conversely that \( \text{epi}(f) \) is a convex subset of \( \mathbb{R}^{n+1} \). To prove that \( f \) is a convex function on \( X \), let \( x, y \in X \). For \( \lambda \in [0, 1] \), we have
\[
 (1 - \lambda)(x, f(x)) + \lambda (y, f(y)) \in \text{epi}(f),
\]
which implies that,
\[
 f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y), \forall \lambda \in [0, 1].
\]
Hence, \( f \) is a convex function on \( \mathbb{R}^n \). This completes the proof.

An important property of convex and concave functions is that they are continuous on the interior of their domains. Though one may find various approaches to prove the result in Bertsekas et al. [23], Mangasarian [176], Rockafellar [238], and Tuy [273], we present a simple proof from Bazaraa et al. [17].

Proposition 1.25 Let \( X \subseteq \mathbb{R}^n \) be a nonempty set and let \( f : X \to \mathbb{R}^n \) be convex. Then \( f \) is continuous on the interior of \( X \).

Proof Let \( \bar{x} \in \text{int}(X) \). We have to prove that \( f \) is continuous at \( \bar{x} \). Let \( \varepsilon > 0 \) be given. Now, we have to show that for given \( \varepsilon > 0 \), there exists \( \delta > 0 \) (depending on \( \varepsilon \)), such that
\[
 |f(x) - f(\bar{x})| \leq \varepsilon, \text{ whenever } \|x - \bar{x}\| \leq \delta.
\]
First, we show that \( f \) is upper semicontinuous at \( \bar{x} \), i.e.,
\[
 f(x) - f(\bar{x}) \leq \varepsilon, \text{ whenever } \|x - \bar{x}\| \leq \delta.
\]
Since, \( \bar{x} \in \text{int}(X) \), therefore, there exists a \( \delta' > 0 \) such that
\[
 |x - \bar{x}| \leq \delta \Rightarrow x \in X.
\]
Now, we define \( \alpha \) as follows
\[
 \alpha = \max_{1 \leq i \leq n} \left\{ \max_{1 \leq i \leq n} \{ f(\bar{x} + \delta e_i) - f(\bar{x}), f(\bar{x} - \delta e_i) - f(\bar{x}) \} \right\},
\]
(1.4)
where $e_i$ is a vector of zeros except having 1 at the $i$th position. We note that $0 \leq \alpha \leq \infty$. Setting
\[
\delta = \min \left( \frac{\delta'}{n}, \frac{\alpha}{n\alpha} \right).
\]
Choose an $x$ with $\|x - \bar{x}\| \leq \delta$. Let
\[
z_i = \begin{cases} 
\delta' e_i, & \text{if } x_i - \bar{x}_i \geq 0, \\
-\delta' e_i, & \text{otherwise.}
\end{cases}
\]
Then $x - \bar{x} = \sum_{i=1}^{n} \lambda_i z_i$, where $\lambda_i \geq 0$, $i = 1, 2, ..., n$. Moreover,
\[
\|x - \bar{x}\| = \delta' \left( \sum_{i=1}^{n} \lambda_i^2 \right)^{1/2}
\]
(1.6)
Since, $\|x - \bar{x}\| \leq \delta$, from (1.7), it follows that $\lambda_i \leq 1/n$, for $i = 1, 2, ..., n$.
Since, $0 \leq n\lambda_i \leq 1$, by the continuity of $f$, we have
\[
f(x) = f \left( \bar{x} + \sum_{i=1}^{n} \lambda_i z_i \right) = f \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \bar{x} + \sum_{i=1}^{n} n\lambda_i z_i \right) \right]
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{n} f \left[ (1 - n\lambda_i) \bar{x} + n\lambda_i (\bar{x} + z_i) \right]
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{n} \left[ (1 - n\lambda_i) f(\bar{x}) + n\lambda_i f((\bar{x} + z_i)) \right].
\]
Therefore,
\[
f(x) - f(\bar{x}) \leq \sum_{i=1}^{n} \lambda_i [f(\bar{x} + z_i) - f(\bar{x})].
\]
By (1.5), we get
\[
f(\bar{x} + z_i) - f(\bar{x}) \leq \alpha, \forall i.
\]
Since, $\lambda_i \geq 0$, it follows that
\[
f(x) - f(\bar{x}) \leq \alpha \sum_{i=1}^{n} \lambda_i.
\]
(1.7)
From (1.7) and (1.8), it follows that $\lambda_i \leq \frac{\alpha}{n\alpha}$. Therefore, by (1.8), we get
\[
f(x) - f(\bar{x}) \leq \varepsilon,
\]
whenever
\[
\|x - \bar{x}\| \leq \delta.
\]
Thus, $f$ is upper semicontinuous at $\bar{x}$. 

---

*Pseudolinear Functions and Optimization*
Now, we have to prove that $f$ is lower semicontinuous at $\bar{x}$, that is, we have to show that $f(\bar{x}) - f(x) \leq \varepsilon$. Let $y = 2\bar{x} - x$, then we note that $\|y - \bar{x}\| \leq \delta$. Therefore, as above

$$f(y) - f(\bar{x}) \leq \varepsilon.$$  \hspace{1cm} (1.8)

Now, we have $\bar{x} = \frac{1}{2}y + \frac{1}{2}x$, therefore, by the continuity of $f$, we have

$$f(\bar{x}) \leq \frac{1}{2}f(y) + \frac{1}{2}f(x).$$  \hspace{1cm} (1.9)

Combining, (1.9) and (1.10), it follows that $f(\bar{x}) - f(x) \leq \varepsilon$. Hence, the proof is complete.

**Remark 1.6** However, it should be noted that a convex function may not be continuous everywhere on their domain. Consider the following example, from Mangasarian [176]. Let $f : X \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} x^2, & \text{if } x > -1, \\ 2, & \text{if } x = -1, \end{cases}$$

where, $X := \{x : x \in \mathbb{R}, x > -1\}$. Then, it is clear that $f$ is a convex function on $X$, but not continuous on $X$.

The following theorems illustrate that once or twice differentiable convex functions have their specific characterizations. We present a simple proof from Mangasarian [176].

**Theorem 1.10** Let $X \subseteq \mathbb{R}^n$ be an open convex set. Then $f : X \to \mathbb{R}$ is a convex function over $X$, if and only if

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle, \forall x, y \in X.$$ 

Proof (Necessary) Assume that $f$ is convex over $X$ and let $x, y \in \mathbb{R}^n$ be such that $x \neq y$. For $\lambda \in [0, 1]$, we consider the function

$$g(\lambda) = \frac{f(x + \lambda(y - x)) - f(x)}{\lambda}.$$

Now, we will show that $g(\lambda)$ is monotonically increasing with $\lambda$. We consider any $\lambda_1, \lambda_2$ with $0 < \lambda_1 < \lambda_2$ and let

$$\bar{\lambda} = \frac{\lambda_1}{\lambda_2}, \quad \bar{y} = x + \lambda(y - x).$$  \hspace{1cm} (1.10)

Now, by the convexity of $f$, we have

$$f(x + \bar{\lambda}(\bar{y} - x)) \leq \bar{\lambda}f(\bar{y}) + (1 - \bar{\lambda})f(x),$$
which can be rewritten as
\[ \frac{f(x + \bar{\lambda}(\bar{y} - x)) - f(x)}{\bar{\lambda}} \leq f(\bar{y}) - f(x). \] (1.11)

Using (1.11) into (1.12), it may be shown that
\[ \frac{f(x + \lambda_1(y - x)) - f(x)}{\lambda_1} \leq \frac{f(x + \lambda_2(y - x)) - f(x)}{\lambda_2}. \]

Hence, for \(0 < \lambda_1 < \lambda_2 < 1\), we get
\[ g(\lambda_1) \leq g(\lambda_2). \]

This will imply that
\[ \langle \nabla f(x), y - x \rangle = \lim_{\lambda \to 0} g(\lambda) \leq g(1) = f(y) - f(x). \]

(Sufficiency) Let \(x, y \in X\) and let \(0 \leq \lambda \leq 1\). Since, \(X\) is convex, it follows that \((1 - \lambda)x + \lambda y \in X\). Now, we have
\[ f(x) - f((1 - \lambda)x + \lambda y) \geq \lambda \langle \nabla f((1 - \lambda)x + \lambda y), y - x \rangle, \]
\[ f(y) - f((1 - \lambda)x + \lambda y) \geq -\lambda \langle \nabla f((1 - \lambda)x + \lambda y), y - x \rangle. \]

Multiplying the first inequality by \((1 - \lambda)\) and the second by \(\lambda\) and adding them, we get
\[ (1 - \lambda)f(x) + \lambda f(y) \geq f((1 - \lambda)x + \lambda y), \forall \lambda \in [0,1]. \]

Hence, the proof is complete.

**Remark 1.7** (i) In a similar way, by using the definition of concave functions, we can prove that a function \(f : X \to \mathbb{R}\) is concave over a convex set \(X\), if and only if
\[ f(y) - f(x) \leq \langle \nabla f(x), y - x \rangle, \forall x, y \in X. \]

Again taking \(x \neq y\) and \(\lambda \in [0, 1]\), we can prove that a function \(f : X \to \mathbb{R}\) is strictly convex (strictly concave) over a convex set \(X\), if and only if
\[ f(y) - f(x) > (\leq) \langle \nabla f(x), y - x \rangle. \]

(ii) Geometrically, the above theorem illustrates that, for a convex function, the linearization \(f(x) + \langle \nabla f(x), y - x \rangle\) never overestimates \(f(x)\) for any \(x \in X\). Similarly, for a concave function, the linearization \(f(x) + \langle \nabla f(x), y - x \rangle\) never underestimates \(f(x)\) for any \(x \in X\) (for example, see Figure 1.4).

(iii) Furthermore, it is clear from the above theorem that if \(f : X = \mathbb{R}^n \to \mathbb{R}\) is a convex function and \(\nabla f(x) = 0\), then \(x\) minimizes \(f\) over \(\mathbb{R}^n\). This is classical sufficient optimality condition for unconstrained optimization known as Fermat’s rule in one dimension.
**Definition 1.60 (Distance function)** Let $C \subseteq \mathbb{R}^n$ be any nonempty set. Then the distance function of $C$, denoted by $d_C(\cdot)$, is defined by

$$d_C(x) = \inf \{ \|x - c\| : c \in C\}.$$

It is clear that if $C$ is a closed set, then $x \in C$ if and only if $d_C(x) = 0$. Distance function $d_C(\cdot)$ satisfies the following global Lipschitz property, i.e.,

$$|d_C(x) - d_C(y)| \leq \|x - y\|, \forall x, y \in \mathbb{R}^n.$$

If the set $C$ is a convex set, then $d_C(\cdot)$ is a convex function.

**Definition 1.61 (Indicator function)** Let $C \subseteq \mathbb{R}^n$ be any set. Then indicator function of $C$ denoted by $\delta(\cdot|C)$, is defined as

$$\delta(x|C) = \begin{cases} 0, & \text{if } x \in C, \\ \infty, & \text{if } x \notin C. \end{cases}$$

It is clear that the set $C$ is a convex set if and only if $\delta(\cdot,C)$, is a convex function. It can be seen that the epigraph of an indicator function is a “half-cylinder” with the cross section as $C$.

**Definition 1.62 (Support function)** The support function of a convex set $C \subseteq \mathbb{R}^n$ is a function $\sigma_C(\cdot) : X \to \mathbb{R} \cup \{\infty\}$, defined by

$$\sigma_C(x) = \sup \{ \langle x, y \rangle : y \in X \}.$$
For instance, the support function of the convex set $C = [-1, 1]$ is the absolute value function $f(x) = |x|$.

Next, we state some important properties of the support function from Clarke [49].

**Theorem 1.11** Let $C, D$ be nonempty closed convex subset of $\mathbb{R}^n$. Then,

(i) $C \subseteq D \iff \sigma_C(x) \leq \sigma_D(x), \forall x \in \mathbb{R}^n$.

(ii) $C$ is compact if and only if $\sigma_C(\cdot) : (\mathbb{R})^n \rightarrow \mathbb{R}$ is finite valued on $\mathbb{R}^n$.

(iii) A given function $\sigma : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is positively homogeneous, subadditive, lower semicontinuous function and not identically $+\infty$, if and only if there is a nonempty closed convex subset $C$ of $\mathbb{R}^n$ such that $\sigma = \sigma_C$. Any such $C$ is unique.

**Definition 1.63 (Monotone function)** Let $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a vector valued function, then $f$ is said to be monotone (strictly monotone) function on $X$, if

$$\langle f(y) - f(x), y - x \rangle \geq (>)0, \forall x, y \in \mathbb{R}.$$

**Theorem 1.12** Let $X \subseteq \mathbb{R}^n$ be an open convex set. Then $f : X \rightarrow \mathbb{R}$ is a convex (concave) function over $X$, if and only if

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0(\leq 0), \forall x, y \in X.$$

Proof We prove the theorem for the convex case. The proof for the concave case follows similarly.

(Necessary) Let $f$ be a convex function and let $x, y \in X$. Then by Theorem 1.10, we have

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$$

and

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle.$$

Adding these two inequalities, we get

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

(Sufficiency) Let $x, y \in X$. Since, $X$ is convex, then,

$$(1 - \lambda)x + \lambda y \in X, \forall \lambda \in [0, 1].$$

Now, by mean value theorem, there exists $\bar{\lambda}, 0 < \bar{\lambda} < 1$, such that

$$f(y) - f(x) = \bar{\lambda} \langle \nabla f(x + \bar{\lambda}(y - x)), y - x \rangle. \quad (1.12)$$

By our assumption, we have

$$\langle \nabla f(x + \bar{\lambda}y - x) - \nabla f(x), y - x \rangle \geq 0,$$
which is equivalent to
\[ \langle \nabla f(x + \lambda(y - x)), y - x \rangle - \langle \nabla f(x), y - x \rangle \geq 0. \] (1.13)
Hence, by (1.13) and (1.14), we get
\[ f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle, \forall x, y \in X. \]
Hence, \( f \) is a convex function over \( X \).

**Remark 1.8** It is clear from the above theorem, that a differentiable function \( f : X \to \mathbb{R} \) is convex on open convex set \( X \) if and only if \( \nabla f \) is monotone on \( X \). Similarly, we can show that differential function \( f : X \to \mathbb{R} \) is strictly convex on open convex set \( X \) if and only if \( \nabla f \) is strictly monotone on \( X \).

The following theorem presents a characterization for twice continuously differential convex functions.

**Theorem 1.13** Let \( X \subseteq \mathbb{R}^n \) be an open convex set and let \( f \) be a twice differentiable function on \( X \). Then \( f \) is a convex (concave) function over \( X \) if and only if for all \( x \in X \), the Hessian matrix \( \nabla^2 f(x) \) is positive (negative) semidefinite matrix on \( X \).

Proof (Necessary) By Proposition 1.14, for all \( x, y \in X \), we have
\[ f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle y - x, \nabla^2 f(x + \lambda(y - x))(y - x) \rangle, \]
for some \( \lambda \in [0,1] \). Since, \( \nabla^2 f(x) \) is positive semidefinite, we must have
\[ f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in X. \]
Hence, \( f \) is convex over \( X \). (Sufficiency) Suppose \( f \) is convex and \( \bar{x} \in X \). We have to prove that \( \nabla^2 f(x) \) is a positive semidefinite matrix at \( \bar{x} \in X \), that is, for each \( y \in \mathbb{R}^n \), we must have
\[ \langle y, \nabla^2 f(\bar{x})y \rangle \geq 0. \]
Since, \( X \) is open, there exists a \( \bar{\lambda} > 0 \), such that \( \bar{x} + \lambda y \in X \), for \( 0 < \lambda < \bar{\lambda} \). Taking into consideration, the convexity and twice differentiability of \( f \), for \( 0 < \lambda < \bar{\lambda} \), we have
\[ f(\bar{x} + \lambda y) \geq f(\bar{x}) + \lambda \langle y, \nabla f(\bar{x})y \rangle, \]
\[ f(\bar{x} + \lambda y) \geq f(\bar{x}) + \lambda \langle \nabla f(\bar{x}), y \rangle + \frac{1}{2} \lambda^2 \langle y, \nabla^2 f(\bar{x})y \rangle + \lambda^2 \alpha(\bar{x}, \lambda y)(\|y\|)^2. \]
Hence, we get
\[ \frac{1}{2} \lambda^2 \langle y, \nabla^2 f(\bar{x})y \rangle + \lambda^2 \alpha(\bar{x}, \lambda y)(\|y\|)^2 \geq 0, \text{ for } 0 < \lambda < \bar{\lambda}. \]
Taking the limit as \( \lambda \to 0 \), we get
\[
\langle y, \nabla^2 f(\bar{x}) y \rangle \geq 0, \forall y \in \mathbb{R}^n.
\]
This completes the proof for the convex case.

Taking into consideration the concavity of \( f \) and proceeding on the same line, we can prove the theorem for concave case.

**Remark 1.9**  (i) The above theorem is useful in checking the convexity and concavity of the twice differentiable functions. Moreover a quadratic function
\[
f(x) = \frac{1}{2} \langle x, Qx \rangle + \langle x, a \rangle + b
\]
where \( x, a \in \mathbb{R}^n, b \in \mathbb{R} \) and \( Q \) is a \( n \times n \) symmetric matrix, is convex on \( \mathbb{R}^n \) if and only if \( Q \) is a positive definite matrix.

(ii) In a similar way, we can show that if \( f : X \to \mathbb{R} \) be a twice differentiable function over an open convex set \( X \subseteq \mathbb{R}^n \) and if the Hessian \( \nabla^2 f(x) \) is positive definite for all \( x \in X \), then \( f \) is a strictly convex function over \( X \). Hence, for each \( x \in X \), the Hessian matrix \( \nabla^2 f(x) \) is positive (negative) semidefinite matrix but not necessarily positive definite on \( X \). For example the function \( f : \mathbb{R} \to \mathbb{R} \) given by \( f(x) = x^4 \) is strictly convex but \( \nabla^2 f(x) = 12x^2 \) is not positive definite, since \( \nabla^2 f(0) = 0 \).

**Proposition 1.26** Let \( X \subseteq \mathbb{R}^n \) be a convex set. If the function \( f : X \to \mathbb{R} \) is convex on \( X \), then the lower level set \( \Lambda(f, \alpha) := \{ x \in X : f(x) \leq \alpha \} \) is a convex subset of \( X \), for arbitrary \( \alpha \in \mathbb{R} \).

Proof Suppose \( f \) be convex on \( X \) and for arbitrary \( \alpha \in \mathbb{R} \), let \( x, y \in \Lambda(f, \alpha) \). Then by the convexity of \( f \), we have
\[
f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y), \forall \lambda \in [0, 1].
\]
\[
\leq (1 - \lambda)\alpha + \lambda \alpha = \alpha.
\]
Thus, for all \( \lambda \in [0, 1] \), it follows that \( (1 - \lambda)x + \lambda y \in \Lambda(f, \alpha) \). Hence, \( \Lambda(f, \alpha) \) is a convex subset of \( X \), for each \( \alpha \in \mathbb{R} \).

**Remark 1.10** The converse of the above proposition is not true in general. For example, consider the function \( f : \mathbb{R} \to \mathbb{R} \) given by \( f(x) = x^3 \). Clearly, \( f \) is not a convex function. However, the lower level set
\[
\Lambda(f, \alpha) = \{ x : x \in \mathbb{R}, x^3 \leq \alpha \} = \{ x : x \in \mathbb{R}, x \leq \alpha^{\frac{1}{3}} \}
\]
is a convex set. This led to the generalization of the notion of convexity to quasiconvexity.
1.8 Generalized Convex Functions

The Italian mathematician Bruno de Finetti [62] introduced one of the fundamental generalized convex functions characterized by convex lower level sets, now known as quasiconvex functions. From an analytical point of view, we have the following definition:

**Definition 1.64 (Mangasarian [176])** Let $f : X \rightarrow \mathbb{R}$ be a real-valued function, defined on a convex set $X$. The function $f$ is said to be *quasiconvex* on $X$, if

$$f(\lambda x + (1 - \lambda) y) \leq \max\{f(x), f(y)\}, \forall x, y \in X \text{ and } \lambda \in [0, 1].$$

The function $f$ is said to be *quasiconcave* on $X$, if $-f$ is quasiconvex on $X$. A function $f$ is said to be *quasilinear* on $X$, if it is both quasiconvex and quasiconcave on $X$.

**Example 1.6** The function $f(x) = x^3$ is a quasilinear function on $\mathbb{R}$.

If the above inequality is strict, we have the following notion of strictly quasiconvex functions.

**Definition 1.65 (Strictly quasiconvex function)** Let $f : X \rightarrow \mathbb{R}$ be a real-valued function defined on a convex set $X$. Then, the function $f$ is said to be strictly quasiconvex function on $X$, if for $x \neq y$, $\lambda \in ]0, 1[,$ we have

$$f(\lambda x + (1 - \lambda) y) < \max\{f(x), f(y)\}, \forall x, y \in X.$$

It is clear from the definitions that every strictly quasiconvex function is quasiconvex, but not conversely. However, the converse is not true as illustrated by the following example, from Cambini and Martein [36].

**Example 1.7** The function $f(x) = \begin{cases} |x|, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0 \end{cases}$ is quasiconvex but not strictly quasiconvex.

The following theorem shows that the quasiconvex functions, can be completely characterized in terms of lower level sets. The proof is along the lines of Mangasarian [176].

**Theorem 1.14** Let $f : X \rightarrow \mathbb{R}$ be a real-valued function defined on a convex set $X$. The function $f$ is quasiconvex on $X$, if and only if the lower level set

$$\Lambda(f, \alpha) = \{x \in X : f(x) \leq \alpha\}$$

is convex for each $\alpha \in \mathbb{R}$.
Pseudolinear Functions and Optimization

Proof Suppose that $\Lambda(f,\alpha)$ is convex for each $\alpha \in \mathbb{R}$. To show that $f$ is quasiconvex on $X$, let $x, y \in X$ be such that $f(x) \geq f(y)$ and let $0 \leq \lambda \leq 1$. Setting, $f(x) = \alpha$, by convexity of $\Lambda(f,\alpha)$, we get

$$f(\lambda x + (1-\lambda)y) \leq \alpha = f(x),$$

Hence, $f$ is quasiconvex on $X$.

Suppose conversely that $f$ is quasiconvex on $X$. Let $\alpha$ be any real number and let $x, y \in \Lambda_\alpha$. Without loss of generality, assume that $f(y) \leq f(x)$. Since, $x, y \in \Lambda_\alpha$, we get

$$f(y) \leq f(x) \leq \alpha.$$

Since, $f$ is quasiconvex and $X$ is convex, for $0 \leq \lambda \leq 1$, we have

$$f(\lambda x + (1-\lambda)y) \leq \alpha.$$

Hence, $\lambda x + (1-\lambda)y \in \Lambda(f,\alpha)$ and $\Lambda(f,\alpha)$ is convex.

Taking into account the properties of quasiconvex and quasiconcave functions, the following theorem from Cambini and Martein [36] characterizes the properties of quasilinear functions.

**Corollary 1.7** Let $f : X \subseteq \mathbb{R}^n$ be a function defined on the convex set $X$. Then $f$ is quasilinear on $X$ if and only if one of the following conditions holds:

1. For $x, y \in X$, and $\lambda \in [0, 1]$, one has
   $$\min \{f(x), f(y)\} \leq f((1-\lambda)x + \lambda y) \leq \max \{f(x), f(y)\}$$

2. The lower level and upper level sets of $f$ are convex.

3. Any restriction of $f$ on a line segment is a nonincreasing or a nondecreasing function.

The following theorem states the relationship between convexity, strict convexity, quasiconvexity and strict quasiconvexity. The proof may be found in Cambini and Martein [36] and Mangasarian [176].

**Theorem 1.15** Let $X \subseteq \mathbb{R}^n$ be a convex set and $f : X \to \mathbb{R}$ any function. Then

1. If $f$ is convex on $S$, then $f$ is quasiconvex on $S$.
2. If $f$ is strictly convex on $S$, then $f$ is strictly quasiconvex on $S$.
3. If $f$ is strictly quasiconvex on $S$, then $f$ is quasiconvex on $S$.

It is obvious that the class of quasiconvex functions contains properly the classes of convex as well as strictly quasiconvex functions. However, there is no inclusion relation between the classes of convex and strictly quasiconvex functions.

The following theorem characterizes differentiable quasiconvex functions. The proof may be found in Cambini and Martein [36] and Mangasarian [176].
Theorem 1.16 Let \( X \subseteq \mathbb{R}^n \) be an open convex set and let \( f : X \to \mathbb{R} \) be a differentiable function on \( X \). Then, \( f \) is quasiconvex on \( X \), if and only if the following implication holds:

\[
f(y) \leq f(x) \implies \langle \nabla f(x), y - x \rangle \leq 0, \forall x, y \in X.
\]

However, if \( x \) is not a critical point, then on both sides of (1.15), we have a strict inequality.

Theorem 1.17 Let \( X \subseteq \mathbb{R}^n \) be an open convex set and let \( f : X \to \mathbb{R} \) be a differentiable function on \( X \). Then, \( f \) is quasiconvex on \( X \), if and only if the following implication holds:

\[
f(y) < f(x), \nabla f(x) \neq 0 \implies \langle \nabla f(x), y - x \rangle < 0, \forall x, y \in X.
\]

Proof For the proof, we refer to Cambini and Martein [36].

The following theorem from Cambini and Martein [36] gives a characterization for quasilinear functions. The characterization is based on the behavior of the differentiable quasilinear functions on the same level sets.

Theorem 1.18 Let \( f : X \to \mathbb{R} \) be a differentiable function on an open convex set \( X \). Then \( f \) is quasilinear on \( X \) if and only if the following condition holds:

\[
f(x) = f(y) \implies \langle \nabla f(x), y - x \rangle = 0, \forall x, y \in X.
\]

Mangasarian [176] introduced the concepts of pseudoconvex and pseudoconcave functions as generalizations of convex and concave functions, respectively. Analytically, pseudoconvex and pseudoconcave functions are defined as follows:

Definition 1.66 Let \( f : X \to \mathbb{R} \) be a differentiable function on open convex set \( X \). The function \( f \) is said to be pseudoconvex at \( x \in X \), if for all \( y \in X \), one has

\[
\langle \nabla f(x), y - x \rangle \geq 0 \implies f(y) \geq f(x);
\]

or equivalently,

\[
f(y) < f(x) \implies \langle \nabla f(x), y - x \rangle < 0.
\]

The function \( f \) is called pseudoconvex on \( X \), if the above property is satisfied for all \( x \in X \). The function \( f \) is called pseudoconcave on \( X \), if \(-f\) is pseudoconvex on \( X \). The function \( f \) is called pseudolinear on \( X \), if \( f \) is both pseudoconvex and pseudoconcave on \( X \).

In the subsequent chapters, we study in detail about pseudolinear functions and their properties.

Definition 1.67 Let \( f : X \to \mathbb{R} \) be a differentiable function on an open convex set \( X \). The function \( f \) is said to be strictly pseudoconvex at \( x \in X \), if for all \( y \in X, x \neq y \), one has

\[
f(y) \leq f(x) \implies \langle \nabla f(x), y - x \rangle < 0.
\]
The function $f$ is called strictly pseudoconvex on $X$, if the above property is satisfied for all $x \in X$.

It is obvious that every strictly pseudoconvex function is pseudoconvex, but the converse is not true, in general. For example, every constant function is pseudoconvex, but not strictly pseudoconvex.

The following theorem states the relationship between convex, pseudoconvex, and quasiconvex functions. For the proof, we refer to Mangasarian [176].

**Theorem 1.19** Let $f : X \to \mathbb{R}$ be a differentiable function on the open convex set $X$.

(i) If $f$ is convex on $X$, then $f$ is pseudoconvex on $X$.

(ii) If $f$ is strictly pseudoconvex on $X$, then $f$ is strictly quasiconvex function on $X$, and therefore quasiconvex function on $X$.

However, the converse is not necessarily true. For example the function $f(x) = x + x^3$ is pseudoconvex, but not convex on $\mathbb{R}$. However, there is no relation between the classes of pseudoconvex and strictly quasiconvex functions. For example the function $f(x) = x^3$ is strictly quasiconvex on $\mathbb{R}$, but not pseudoconvex on $\mathbb{R}$. Moreover, the constant function is pseudoconvex but not strictly quasiconvex on $\mathbb{R}$.

**Definition 1.68 (Karamardian [131])** Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $F : X \to \mathbb{R}^n$ be any map. Then $F$ is said to be pseudomonotone, if for all $x, y \in X$, we have

$$\langle F(x), y - x \rangle \geq 0 \Rightarrow \langle F(y), y - x \rangle \geq 0.$$

Now, we need the following lemma from Cambini and Martein [36].

**Lemma 1.1** Let $X \subseteq \mathbb{R}^n$ be a convex set and $f : X \to \mathbb{R}$ be a differentiable function on $X$. Suppose that $\nabla f$ is pseudomonotone on $X$ and $x, y \in X$ such that

$$\langle \nabla f(x), y - z \rangle \geq 0.$$

Then, the restriction of $f$ on $[x, y]$ is nondecreasing.

The following theorem from Cambini and Martein [36] characterizes pseudoconvex functions in terms of the pseudomonotonicity of the gradient map.

**Theorem 1.20** Let $X \subseteq \mathbb{R}^n$ be a convex set and $f : X \to \mathbb{R}$ be a differentiable function on $X$. Then $f$ is pseudoconvex on $X$, if and only if $\nabla f$ is pseudomonotone on $X$.

Proof Assume that $f$ is pseudoconvex on $X$ and let $x, y \in X$. Suppose that

$$\langle \nabla f(x), y - x \rangle \geq 0.$$
Since, $f$ is pseudoconvex on $X$, it is also quasiconvex on $X$, therefore, for all $x, y \in X$, we get

$$\langle \nabla f(y), x - y \rangle \leq 0,$$

which is equivalent to

$$\langle \nabla f(y), y - x \rangle \geq 0.$$

Hence, $\nabla f$ is pseudomonotone on $X$. Conversely, suppose that $\nabla f$ is pseudomonotone on $X$. Suppose contrary that there exists $x, y \in X$, such that

$$f(x) < f(y) \text{ and } \langle \nabla f(y), y - x \rangle \geq 0.$$

Then, by Lemma 1.1, $f$ is nondecreasing on the restriction $[x, y]$. Therefore, we must have

$$f(x) \geq f(y),$$

which is a contradiction. This completes the proof.

The following theorem gives a very important characterization of pseudoconvex and strictly pseudoconvex functions.

**Theorem 1.21** Let $f : X \to \mathbb{R}$ be a differentiable function on the open convex set $X$. Then $f$ is pseudoconvex (strictly pseudoconvex) if and only if for every $\bar{x} \in X$ and $u \in \mathbb{R}^n$, such that

$$\langle u, \nabla f(\bar{x}) \rangle = 0,$$

the function $\varphi(t) = f(\bar{x} + tu)$ attains a local minimum (strict local minimum) at $t = 0$. 

### 1.9 Separations and Alternative Theorems

It is obvious that the hyperplane $H$ divides the space $\mathbb{R}^n$, into two half-spaces either closed or open. This led to the development of the idea of separation of spaces. We start the section with the concept of a separating plane.

**Definition 1.69** *(Separating plane)* The plane $\{ x : x \in \mathbb{R}^n, \langle a, x \rangle = \alpha, \alpha \neq 0 \}$ is said to separate two nonempty sets $X$ and $Y$ in $\mathbb{R}^n$, if

$$\langle a, x \rangle \leq \alpha \leq \langle a, y \rangle, \forall x \in X \text{ and } \forall y \in Y.$$

The separation is said to be strict if

$$\langle a, x \rangle < \alpha < \langle a, y \rangle, \forall x \in X \text{ and } \forall y \in Y.$$

The sets $X$ and $Y$ are said to be separable, if such a plan exists.
It can be shown that two disjoint sets may not be separable and also two separable sets need not be disjoint. However, under the convexity assumptions on the set, we have certain theorems known as separation theorems, which imply that the disjoint convex sets are separable. The separation theorems play a fundamental role in establishing several optimality conditions. Following are some basic and useful separation theorems. There are several approaches to prove these results. See Mangasarian [176], Rockafellar [238], and Tuy [273].

First, we state the following lemma from Mangasarian [176].

**Lemma 1.2 (Separation of a point and a plane)** Let $X$ be a nonempty convex sets in $\mathbb{R}^n$ not containing the origin. Then, there exists a plane

$$\{x : x \in \mathbb{R}^n, \langle a, x \rangle = 0, a \neq 0\}$$

separating $X$ and $0$. That is,

$$x \in X \Rightarrow \langle a, x \rangle \geq 0.$$

Using Lemma 1.2, we have the following separation theorem known as the fundamental separation theorem from Mangasarian [176].

**Theorem 1.22 (Separation of two planes)** Let $X$ and $Y$ be two nonempty disjoint convex sets in $\mathbb{R}^n$. Then, there exists a plane

$$\{x : x \in \mathbb{R}^n, \langle a, x \rangle = \gamma, a \neq 0\},$$

which separates them, that is,

$$x \in X \Rightarrow \langle a, x \rangle \leq \gamma$$

and

$$x \in Y \Rightarrow \langle a, x \rangle \geq \gamma.$$

**Proof** We consider the set

$$X - Y = \{z : z = x - y, x \in X, y \in Y\}.$$  

By Theorem 1.1, the set $X - Y$ is convex. Since, $X \cap Y = \emptyset$, hence, $0 \notin X - Y$. By Lemma 1.2, there exists a plane $\{z : z \in \mathbb{R}^n, \langle a, z \rangle = 0, a \neq 0\}$, such that

$$z \in X - Y \Rightarrow \langle a, z \rangle \geq 0$$

or

$$x \in X, y \in Y \Rightarrow \langle a, x - y \rangle \geq 0.$$

Hence,

$$\alpha = \inf_{x \in X} \langle a, x \rangle \geq \sup_{y \in Y} \langle a, y \rangle = \beta.$$
Define 
\[ \gamma = \frac{\alpha + \beta}{2}. \]
Then, 
\[ x \in X \Rightarrow \langle a, x \rangle \leq \gamma \]
and 
\[ x \in Y \Rightarrow \langle a, x \rangle \geq \gamma. \]
This completes the proof.

The following corollary and lemma can be proved with the help of Theorem 1.22.

**Corollary 1.8** Let \( X \subseteq \mathbb{R}^n \) be a nonempty convex set. If \( 0 \notin \text{cl}(X) \), then there exists a hyperplane 
\[ \{ x : x \in \mathbb{R}^n, \langle a, x \rangle = \alpha, a \neq 0, a \in \mathbb{R}^n \}, \]
which strictly separates \( X \) and 0 conversely. In other words, 
\[ 0 \notin \text{cl}(X) \iff \exists a \neq 0, a > 0 \text{ such that } x \in X \Rightarrow \langle a, x \rangle > \alpha. \]

The following lemma is very important to prove a strict separation theorem. The proof may be found in Bazaraa et al. [17] and Mangasarian [176].

**Lemma 1.3** Let \( X \subseteq \mathbb{R}^n \) be a nonempty, closed and convex set. If \( 0 \notin X \), then, there exists a nonzero vector \( a \) and a scalar \( \alpha \), such that 
\[ \langle a, y \rangle > \alpha \text{ and } \langle a, x \rangle \leq \alpha, \text{ for each } x \in X. \]

Using Corollary 1.6 and Lemma 1.3, we have the following strict separation theorem. We present a simple proof from Mangasarian [176].

**Theorem 1.23 (Strict separation theorem)** Let \( X \) and \( Y \) be two nonempty disjoint convex sets in \( \mathbb{R}^n \) with \( X \) compact and \( Y \) closed. Then, there exists a hyperplane 
\[ \{ x : x \in \mathbb{R}^n, \langle a, x \rangle = \alpha, a \neq 0, a \in \mathbb{R}^n \}, \]
which strictly separates \( X \) and \( Y \) and conversely. In other words, 
\[ X \cap Y = \emptyset \iff \exists a \neq 0, a > 0 \text{ such that } \langle a, x \rangle < \gamma, \forall x \in X \text{ and } \langle a, y \rangle > \gamma, \forall y \in Y. \]
Proof (Necessary) We consider the set
\[ X - Y := \{ z : z = x - y, x \in X, y \in Y \}. \]

By Theorem 1.1, the set \( X - Y \) is convex. Since, \( X \cap Y = \emptyset \), hence, \( 0 \notin X - Y \).

By Lemma 1.1, there exists a plane \( \{ x : x \in \mathbb{R}^n, \langle a, x \rangle = \mu \} \), \( a \neq 0, \mu > 0 \), such that
\[ z \in X - Y \Rightarrow \langle a, z \rangle > \mu > 0 \]
or
\[ x \in X, y \in Y \Rightarrow \langle a, x - y \rangle > \mu > 0. \]

Hence,
\[ \alpha = \inf_{x \in X} \langle a, x \rangle \geq \sup_{y \in Y} \langle a, y \rangle + \mu > \sup_{y \in Y} \langle a, y \rangle = \beta. \]

Define
\[ \gamma := \frac{\alpha + \beta}{2}. \]

Then, we have
\[ x \in X \Rightarrow \langle a, x \rangle < \gamma \]
and
\[ x \in Y \Rightarrow \langle a, x \rangle > \gamma. \]

(Sufficiency) Suppose to the contrary that \( z \in X \cap Y \), then \( \langle a, x \rangle < \gamma < \langle a, x \rangle \), which is a contradiction. This completes the proof.

The alternative theorems play a very important role in establishing the necessary optimality conditions for linear as well as nonlinear programming problems. Now, we state the two most applicable alternative theorems, that is, Farkas’s and Gordan’s alternative theorems. The proof can be found in Bazaraa et al. [17] and Mangasarian [176].

**Theorem 1.24 (Farkas’s theorem)** For each \( p \times n \) matrix \( A \) and each fixed vector \( b \in \mathbb{R}^n \), either

(I) \( Ax \leq 0, \ b^T x > 0 \) has a solution \( x \in \mathbb{R}^n \);

or

(II) \( A^T y = b \) has a solution \( y \in \mathbb{R}^p \);

but never both.

Proof Assume that the system (II) has a solution. Then there exists \( y \geq 0 \), such that \( A^T y = b \). Let there exist some \( x \in \mathbb{R}^n \), such that \( Ax \leq 0 \). Then, we have
\[ b^T x = y^T A x \leq 0. \]

Hence, system (I) has no solution.

Now, we assume that the system (II) has no solution. Consider the set
\[ X := \{ x \in \mathbb{R}^n : x = A^T y, y \geq 0 \}. \]
It is clear that $X$ is a closed convex set and that $b \in X$. Therefore, by Lemma 1.3, there exists vector $p \in \mathbb{R}^n$ and scalar $\alpha$, such that $p^T b > \alpha$ and $p^T x \leq \alpha$, for all $x \in X$. Since, $0 \in X$, $\alpha \geq 0$, hence, $p^T b > 0$. Moreover,

$$\alpha \geq p^T x = p^T A^T y = y^T Ap, \forall y \geq 0.$$  

Since, $y \geq 0$, can be made arbitrary large, the last inequality implies that $Ap \leq 0$. Thus, we have constructed a vector $p \in \mathbb{R}^n$, such that $Ap \leq 0$ and $b^T p > 0$. Hence, the system (I) has a solution. This completes the proof.

**Remark 1.11** To understand geometrically, we may rewrite the system (I) and (II) as follows:

(I*) $A_j x \leq 0, j = 1, 2, ..., p, bx > 0$;

(II*) $\sum_{j=1}^{p} A^T_j y_j = \sum_{j=1}^{p} A_j y_j = b, y_j \geq 0$, where $A^T_j$ denotes the $j$th column vector of $A^T$ and $A_j$ the $j$th row of $A$. Then, it is clear that system (I*) requires that there exists a vector $x \in \mathbb{R}^n$, that makes an obtuse angle ($\geq \pi/2$) with the vectors $A_1$ to $A_p$ and a strictly acute angle ($< \pi/2$) with the vector $b$. The system (II*) requires that the vector $b$ be a nonnegative linear combination of the vectors $A_1$ to $A_p$.

**Theorem 1.25 (Gordan’s alternative theorem)** Let $A$ be a $p \times n$ matrix, then either

(i) $Ax < 0$ has a solution $x \in \mathbb{R}^n$;

or

(ii) $A^T y = b, y \geq 0$ for some nonzero $y \in \mathbb{R}^p$, but never both.

**Theorem 1.26 (Motzkin’s alternative theorem)** Let $A, B, C$ be given $p_1 \times n, p_2 \times n, p_3 \times n$ matrices, with $A$ being nonvacuous. Then, either the system of inequalities (i) $Ax > 0, Bx \geq 0, Cx = 0$ has a solution $x \in \mathbb{R}^n$; or (ii) $A^T y_1 + B^T y_2 + CT y_3 = 0, y_1 \geq 0, y_3 \geq 0$ has a solution $y_1, y_2, y_3$ but never both.

The following theorem by Slater [256] presents a fairly general theorem of the alternative. The proof can be found in Mangasarian [176].

**Theorem 1.27 (Slater’s alternative theorem)** Let $A, B, C$ and $D$ be given $p_1 \times n, p_2 \times n, p_3 \times n$ and $p_4 \times n$ matrices with $A$ and $B$, being nonvacuous. Then, either

(i) $Ax > 0, Bx \geq 0, Cx \geq 0$ has a solution $x \in \mathbb{R}^n$;

or

(ii) $A^T y_1 + B^T y_2 + CT y_3 + D^T y_4 = 0$, with $y_1 \geq 0, y_2 \geq 0, y_3 \geq 0$ or $y_1 \geq 0, y_2 > 0, y_3 \geq 0$ has a solution $y_1, y_2, y_3, y_4$ but never both.

The following theorem from Mangasarian [176] provides the alternative theorem for the most general systems:
**Theorem 1.28** Let $A$, $B$, $C$, and $D$ be given $p_1 \times n, p_2 \times n, p_3 \times n$, and $p_4 \times n$ matrices with $A$ and $B$ being nonvacuous. Then, either

(i) $Ax \geq 0, Bx \geq 0, Cx \geq 0, Dx = 0$ or $Ax \geq 0, Bx > 0, Cx \geq 0, Dx = 0$ has a solution $x \in \mathbb{R}^n$ or

(ii) $A^T y_1 + B^T y_2 + C^T y_3 + D^T y_4 = 0$, with $y_1 \geq 0, y_2 \geq 0, y_3 \geq 0$ or $y_1 \geq 0, y_2 > 0, y_3 \geq 0$ has a solution $y_1, y_2, y_3, y_4$ but never both.

Now, we state the following alternative result for the system of convex and linear vector valued functions. For the proof, we refer to Mangasarian [176].

**Lemma 1.4** Let $X \subseteq \mathbb{R}^n$ be a convex set and $f_1 : X \rightarrow \mathbb{R}^{m_1}$, $f_2 : X \rightarrow \mathbb{R}^{m_2}$, and $f_3 : X \rightarrow \mathbb{R}^{m_3}$ be convex vector valued functions and $f_4 : X \rightarrow \mathbb{R}^{m_4}$ be a linear vector valued function. If

\[
\begin{cases}
    f_1(x) < 0, f_2(x) \leq 0, f_3(x) \leq 0, \\
    f_4(x) = 0,
\end{cases}
\]

has no solution $x \in X$, then, there exist $p_1 \in \mathbb{R}^{m_1}, p_2 \in \mathbb{R}^{m_2}, p_3 \in \mathbb{R}^{m_3}, p_4 \in \mathbb{R}^{m_4}$, such that

\[
\begin{cases}
    p_1, p_2, p_3 \geq 0, (p_1, p_2, p_3, p_4) \neq 0, \\
    p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x) + p_4 f_4(x) \geq 0, \forall x \in X
\end{cases}
\]

\]

### 1.10 Subdifferential Calculus

The nonsmooth phenomena occur naturally and frequently in optimization theory. For example, the well-known absolute value function $f(x) = |x|$ is an example of convex function, which is differentiable everywhere except at $x = 0$. Moreover, $x = 0$ is the global minimizer of $f$ on $\mathbb{R}$. We know that for a function to be differentiable, both the left and right handed derivative must exist and be equal. But this is not the case with nondifferentiable functions. Therefore, the characterization of the convex function in terms of derivative (gradient) is not applicable for nondifferentiable convex functions. This led to the generalization of the notion of gradients to subgradients. However, for nondifferentiable convex functions, the one sided derivatives exist universally.

In this section, we will study subdifferentials for convex functions, Dini, Dini-Hadamard derivatives, and the Clarke subdifferentials for locally Lipschitz functions.

#### 1.10.1 Convex Subdifferentials

We start this section, with an important notion of the Lipschitzian property, that plays a very important role in the theory of nonsmooth analysis:
Definition 1.70 (Locally Lipschitz function) A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be locally Lipschitz (of rank $M$) near $z \in X$, if there exist a positive constant $M$ and a neighborhood $N$ of $z$, such that for any $x, y \in N$, one has

$$|f(y) - f(x)| \leq M \|y - x\|.$$ 

The function $f$ is said to be a locally Lipschitz on $X$ if the above condition is satisfied for all $z \in X$.

The following theorem presents that if a convex function is bounded above in a neighborhood of some point, then it is bounded as well as a locally Lipschitz in that neighborhood. The proof is along the lines of Roberts and Verberg [237].

Theorem 1.29 Let $X \subseteq \mathbb{R}^n$ be an open convex set. Let $f : X \to \mathbb{R}$ be a convex function on $X$ and bounded above on a neighborhood of some point of $X$. Then $f$ is bounded as well as Lipschitz near every $x \in X$.

Proof Without loss of generality, we assume that $f$ is bounded above by $M$ on the set $B_{\varepsilon}(0) \subseteq X$. Let $x \in X$ be arbitrary. First, we prove that $f$ is bounded on a neighborhood of $x$. Choose $\mu > 0$ such that $y = \mu x \in X$. If $\lambda = \frac{1}{\mu}$, then the set

$$\Omega := \{\nu : \nu = (1 - \lambda) \bar{x} + \lambda y, \bar{x} \in B_{\varepsilon}(0)\}$$

is a neighborhood of $x = \lambda y$ with radius $(1 - \lambda)\varepsilon$. Now, for all $\nu \in \Omega$, by the convexity of $f$, we have

$$f(\nu) \leq (1 - \lambda) f(\bar{x}) + \lambda f(y) \leq K + \lambda f(y).$$

Therefore, $f$ is bounded above on a neighborhood of $x$. If $z$ is any point in the $B_{(1-\lambda)\varepsilon}(x)$, there exist another such point $\bar{z}$ such that $x = (z + \bar{z})/2$. Therefore,

$$f(x) \leq \frac{1}{2} f(z) + \frac{1}{2} f(\bar{z}),$$

which implies that

$$f(z) \geq 2f(x) - f(\bar{z}) \geq 2f(x) - M - \lambda f(y).$$

Thus $f$ is bounded below near $x$, hence, we have established that $f$ is bounded near $x$.

Now, to prove that $f$ is Lipschitz near $x$. Suppose that $|f| \leq N$ on the set $B_{2\delta}(x)$, where $\delta > 0$. Let $x_1$ and $x_2$ be any two distinct points in $B_{2\delta}(x)$. Setting

$$x_3 = x_2 + \frac{\delta}{\alpha} (x_2 - x_1),$$

where $\alpha = \|x_2 - x_1\|$. It is clear that $x_3 \in B_{2\delta}(x)$. Solving for $x_2$, we get

$$x_2 = \frac{\delta}{\alpha + \delta} x_1 + \frac{\alpha}{\alpha + \delta} x_3.$$
Therefore, by convexity of \( f \), we get
\[
f(x_2) \leq \frac{\delta}{\alpha + \delta} f(x_1) + \frac{\alpha}{\alpha + \delta} f(x_3).
\]
Then,
\[
f(x_2) - f(x_1) \leq \frac{\alpha}{\alpha + \delta} [f(x_3) - f(x_1)] \leq \frac{\alpha}{\delta} |f(x_3) - f(x_1)|.
\]
Since, \( |f| \leq N \) and \( \alpha = \|x_2 - x_1\| \), it results that
\[
f(x_2) - f(x_1) \leq \frac{2N}{\delta} |x_2 - x_1|.
\]
Since, the role of \( x_1 \) and \( x_3 \) can be changed, we conclude that \( f \) is Lipschitz near \( x \).

**Definition 1.71 (Directional derivatives)** Let \( f : \mathbb{R}^n \to \bar{\mathbb{R}} \) be a function and \( x \in \mathbb{R}^n \) be a point, where \( f \) be finite. Then,

(i) The right sided directional derivative of \( f \) at \( x \) in direction of the vector \( d \in \mathbb{R}^n \), is defined as the limit
\[
f'_+ (x; d) = \lim_{\lambda \to 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda},
\]
if it exists, provided that \( +\infty \) and \( -\infty \) are allowed as limits.

(ii) The left sided directional derivative of \( f \) at \( x \) in direction of the vector \( d \in \mathbb{R}^n \), is defined as the limit
\[
f'_- (x; d) = \lim_{\lambda \to 0^-} \frac{f(x + \lambda d) - f(x)}{\lambda},
\]
if it exists, provided that \( +\infty \) and \( -\infty \) are allowed as limits.

It is easy to see that
\[
f'_+ (x; -d) = -f'_- (x; d) , \forall d \in \mathbb{R}^n.
\]
The two sided directional derivative \( f'(x; d) \) is said to exist, if \( f'_+ (x; d) \) and \( f'_- (x; -d) \) both exist and are equal, i.e.,
\[
f'_+ (x; d) = f'_- (x; -d) = f'(x; d).
\]
We know that if \( f \) is differentiable at \( x \), then the two sided directional derivative exists and is finite. Moreover,
\[
f'(x; d) = \langle \nabla f (x) , d \rangle , \forall d \in \mathbb{R}^n,
\]
where \( \nabla f (x) \) is the gradient of \( f \) at \( x \).

The following theorem from Rockafellar [238] establishes that for a convex function, the one sided directional derivatives always exist.
Theorem 1.30 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and $x \in \mathbb{R}^n$ be a point, where $f$ be finite. Then, for each $d \in \mathbb{R}^n$, the difference quotient
\[
\frac{f(x + \lambda d) - f(x)}{\lambda}
\]
is a nondecreasing function of $\lambda > 0$, so that $f'_+(x; d)$ and $f'_-(x; d)$ both exist and
\[
f'_+(x; d) = \inf_{\lambda > 0} \frac{f(x + \lambda d) - f(x)}{\lambda},
\]
\[
f'_-(x; d) = \sup_{\lambda < 0} \frac{f(x + \lambda d) - f(x)}{\lambda}.
\]
Moreover, $f'(x; d)$ is a positively homogeneous convex function of $d$, with $f'(x; 0) = 0$ and
\[
f'_+(x; d) \geq -f'_-(x; -d), \forall d \in \mathbb{R}^n.
\]

To deal with nondifferentiable convex functions, the usual notion of gradient has been replaced by the notion of subgradients.

Definition 1.72 (Subgradient) Let $f : X \subseteq \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $x \in \mathbb{R}^n$ be a point, where $f$ be finite. Then, a vector $\xi \in \mathbb{R}^n$ is said to be subgradient of $f$ at $x$, if
\[
f(y) \geq f(x) + \langle \xi, y - x \rangle, \forall y \in \mathbb{R}^n.
\]

We note that the affine function $f(x) + \langle \xi, y - x \rangle$ is a supporting hyperplane to the epigraph of $f$ at $(x, f(x))$ with slope $\xi$. Moreover, at a point, where the function is not differentiable, there can be an infinite number of such supporting hyperplanes. The collection of slopes of all these supporting hyperplanes of the function $f$ at a point $x$ is called subdifferential at $x$ and is denoted by $\partial f(x)$, that is,
\[
\partial f(x) = \{ \xi \in \mathbb{R}^n : f(y) - f(x) \geq \langle \xi, y - x \rangle, \forall y \in \mathbb{R}^n \}.
\]
The subdifferential $\partial f(x)$ of a convex function $f$ at a point $x$ is the image of a set valued map $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, known as subdifferential mapping. For more details about $\partial f$ as a set valued map or multifunction, we refer to John [122] and Rockafellar [238].

We know that if $C$ is a convex set, then the indicator function $\delta_C : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex function. Then the subdifferential of $\delta_C$ at $x \in C$, is given by
\[
\partial \delta_C(x) = \{ \xi \in \mathbb{R}^n : \delta_C(x) - \delta_C(y) \geq \langle \xi, y - x \rangle, \forall y \in \mathbb{R}^n \}
\]
\[
= \{ \xi \in \mathbb{R}^n : 0 \geq \langle \xi, y - x \rangle, \forall x \in \mathbb{R}^n \},
\]
which is nothing but the normal cone to the set $C$ at $x$, i.e., $N_C(x)$. Hence,
\[
\partial \delta_C(x) = N_C(x).
\]
The following proposition summarizes some basic properties of the convex subdifferential. The proof may be found in Bazaraa et al. [17], Bertsekas et al. [23], and Rockafellar [238].

**Proposition 1.27** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a convex function. Then

1. \( \partial f (x) \) is nonempty, convex and compact set for all \( x \in \mathbb{R}^n \).
2. \( \partial f (x) = \{ \xi \in \mathbb{R}^n : f'_+(x;d) \geq \langle \xi, d \rangle, \forall d \in \mathbb{R}^n \} \).
3. \( f'_+(x;\cdot) \) is the support function of the set \( \partial f (x) \) at \( x \in \mathbb{R}^n \).
4. If \( f \) is differentiable at \( x \) with gradient \( \nabla f (x) \), then unique subgradient is the gradient, i.e., \( \partial f (x) = \{ \nabla f (x) \} \).

Now, we present the mean value theorem for convex subdifferentiable functions. The proof may be found in Hiriart-Urruty and Lemarechal [105].

**Theorem 1.31 (Mean value theorem)** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a convex function. Then, for \( x, y \in \mathbb{R}^n \), there exists \( z \in [x, y] \) and \( \xi \in \partial f (z) \), such that

\[
 f (y) - f (x) = \langle \xi, y - x \rangle .
\]

In other words,

\[
 f (y) - f (x) \in \langle \partial f (z), y - x \rangle .
\]

The following theorem from Rockafellar [238] states that for a proper convex function \( f \), the multifunction \( \partial f \) is a monotone map.

**Theorem 1.32** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a proper convex function. Then the subdifferential is a monotone map, i.e.,

\[
 \langle \xi_1 - \xi_2, x - y \rangle \geq 0, \forall \xi_1 \in \partial f (x), \xi_2 \in \partial f (y) .
\]

Moreover, \( \partial f \) is a maximal monotone map, that is, its graph is not properly contained in the graph of any other monotone map.

Now, we state without proof, the following theorem on subdifferential calculus. The proof may be found in Bertsekas et al. [23] and Rockafellar [238].

**Theorem 1.33** Let \( X \subseteq \mathbb{R}^n \) be a nonempty convex set and \( f, f_i : X \to \mathbb{R}, i = 1, 2, ..., m \) be convex functions. Then,

(i) For \( \lambda > 0 \) and \( x \in X \), \( \partial (\lambda f) (x) = \lambda \partial f (x) \).

(ii) \( \partial (f_1 + f_2 + ... + f_m) (x) = \partial f_1 (x) + \partial f_2 + ... + \partial f_m (x) \).
1.10.2 Dini and Dini-Hadamard Derivatives

For convex functions, the one sided directional derivatives exist universally. However, this is not the case with the nonconvex functions. For nonconvex functions, the limit in the definitions of directional derivatives may not exist. This led to the generalizations of the notion of directional derivatives to the Dini and Dini-Hadamard directional derivatives.

**Definition 1.73** Let $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$ be an open convex set. Let $x \in \mathbb{R}^n, \nu \in \mathbb{R}^n$ with $\nu^T \nu = 1$. Dini derivatives of $f$ in the direction $\nu$ are defined as follows:

$$D^+ f(x, \nu) := \lim_{t_n \to \infty} \sup_{\{t_n\}} \left\{ \frac{f(x + t_n \nu) - f(x)}{t_n} : 0 < t_n \leq \frac{1}{n} \right\};$$

$$D_+ f(x, \nu) := \lim_{t_n \to \infty} \inf_{\{t_n\}} \left\{ \frac{f(x + t_n \nu) - f(x)}{t_n} : 0 < t_n \leq \frac{1}{n} \right\};$$

$$D^- f(x, \nu) := \lim_{t_n \to \infty} \sup_{\{t_n\}} \left\{ \frac{f(x - t_n \nu) - f(x)}{-t_n} : 0 < t_n \leq \frac{1}{n} \right\};$$

$$D_- f(x, \nu) := \lim_{t_n \to \infty} \inf_{\{t_n\}} \left\{ \frac{f(x - t_n \nu) - f(x)}{-t_n} : 0 < t_n \leq \frac{1}{n} \right\},$$

where $D^+ f(x; \nu)$ is the upper right derivative, $D_+ f(x; \nu)$ is the lower right derivative, $D^- f(x; \nu)$ is the upper left derivative and $D_- f(x; \nu)$ is the lower left derivative evaluated at $x$ in the direction $\nu$. Limits can be infinite in the above definition. By using the definitions, we may prove that

1. Dini derivatives always exist (finite or infinite) for any function $f$ and satisfy

$$D^+ f(x; \nu) \geq D_+ f(x; \nu), D^- f(x; \nu) \geq D_- f(x; \nu).$$

Since, $D^- f(x; \nu) = -D_+ f(x; -\nu), D_- f(x; \nu) = -D^+ f(x; -\nu)$, therefore, it is quite obvious to deal with the directional Dini derivatives $D^+ f(x; \nu)$ and $D_+ f(x; \nu)$.

2. If $D^+ f(x; \nu) = D_+ f(x; \nu)$, then, the common value, written as $f'_+(x; \nu)$ is clearly the right derivative of $f$ at $x$ in the direction $\nu$, given by

$$f'_+(x; \nu) := \lim_{t \to 0^+} \frac{f(x + tv) - f(x)}{t}.$$

Moreover, if $f'_+(x; \nu)$ exists and is finite, then the function $f$ is called differentiable (or Dini differentiable) at $x$ in the direction $\nu$. The function $f$ is called Dini directionally differentiable at the point $x$, if $f$ is Dini differentiable at $x$, for every $\nu \in \mathbb{R}^n$. If $f : \mathbb{R} \to \mathbb{R}$ be any function, then $D^+ f(x; 1), D^- f(x; 1), D_+ f(x; 1)$, and $D_- f(x; 1)$ are denoted by $D^+ f(x), D^- f(x), D_+ f(x),$ and $D_- f(x)$, respectively. Indeed, we have

$$D^+ f(x) := \limsup_{t \to x^+} \frac{f(t) - f(x)}{t - x}$$

and

$$D^+ f(x) := \liminf_{t \to x^+} \frac{f(t) - f(x)}{t - x}.$$
A continuous function may not have even a one sided directional derivative at a point but it may have the Dini directional derivative at that point. The following example from Ansari et al. [2] illustrates the fact:

**Example 1.8** Let us consider the function \( f : \mathbb{R} \to \mathbb{R} \), defined by

\[
f(x) := \begin{cases} 
|x| \cos(\frac{1}{x}), & \text{if } x \neq 0, \\
0, & \text{if } x = 0.
\end{cases}
\]

Since, \(|\cos(\frac{1}{x})| \leq 1\) for all \( x \neq 0 \), we have

\[
\lim_{x \to 0} f(x) = 0 = f(0).
\]

Therefore, \( f \) is continuous at \( x = 0 \). It is clear that \( f \) is also continuous at all the other points of \( \mathbb{R} \), so \( f \) is continuous function. However, inspection of the difference quotient reveals that

\[
\limsup_{x \to 0^+} \frac{f(x) - f(0)}{x - 0} = 1, \quad \liminf_{x \to 0^+} \frac{f(x) - f(0)}{x - 0} = 0,
\]

hence, \( f'_+(0) \) does not exist. Similarly, \( f'_-(0) \) does not exist.

However, we have

\[
D^+ f(0) = 1, \quad D^- f(0) = 0, \quad D'_+ f(0) = 1, \quad D'_- f(0) = 0,
\]

Moreover, we can see that \( f'(x) \) exists everywhere, except at \( x = 0 \), and all the four Dini derivatives have the same value.

It is known from Giorgi and Komlosi [88] that for a function \( f : I \to \mathbb{R} \), defined on real interval \( I \), if the Dini derivatives \( D^+ f(0), D^- f(0), D'_+ f(0), D'_- f(0) \), are finite at each point of \( I \), then \( f \) is differentiable, almost everywhere on \( I \). However, if all of the Dini derivatives are not finite, then continuity is in general not ensured. The following example illustrates the fact.

**Example 1.9** (Giorgi and Komlosi [88]) Consider the function \( f : \mathbb{R} \to \mathbb{R} \), defined by

\[
f(x) := \begin{cases} 
0, & \text{if } x > 0 \text{ and } x \text{ is irrational}, \\
1, & \text{if } x < 0 \text{ or } x \text{ is rational}.
\end{cases}
\]

Then, for \( x = 0 \), we have

\[
D^+ f(x) = 0, \quad D^- f(x) = 0, \quad D'_+ f(x) = -\infty, \quad D'_- f(x) = 0.
\]

Moreover, it is easy to see that the function \( f \) is not continuous at \( x = 0 \).
The important feature of the Dini derivatives is that they always exist and admit useful calculus rules. The following theorem summarizes some elementary properties and calculus rules for the Dini upper (Dini lower) directional derivatives. The proof follows directly from the definitions of the Dini upper (Dini lower) directional derivatives. For further study, we refer to McShane [181].

**Theorem 1.34** Let \( f, g : \mathbb{R}^n \to \mathbb{R} \) be real-valued functions. The following assertion holds:

1. **(Homogeneity)** \( D^+ f(x; \nu) \) is positively homogeneous in \( \nu \), that is,
   \[
   D^+ f(x; r\nu) = rD^+ f(x; \nu), \quad \forall r > 0.
   \]

2. **(Scalar multiplication)** For \( r > 0 \), \( D^+ (rf)(x; \nu) = rD^+ f(x; \nu) \), and for \( r < 0 \), \( D^+ (rf)(x; \nu) = rD^- f(x; \nu) \).

3. **(Sum rule)**
   \[
   D^+(f + g)(x; \nu) \leq D^+(x; \nu) + D^+ g(x; \nu),
   \]
   provided that the sum on the right hand side exists.

4. **(Product rule)**
   \[
   D^+(fg)(x; \nu) \leq D^+[g(x)f](x; \nu) + D^+[f(x)g](x; \nu),
   \]
   provided that the sum on the right hand side exists, the functions \( f \) and \( g \) are continuous at \( x \) and that one of the following conditions is satisfied: \( f(x) \neq 0; g(x) \neq 0; D^+ f(x; \nu) \) is finite; and \( D^+ g(x; \nu) \) is finite.

5. **(Quotient rule)**
   \[
   \left( D^+ \left( \frac{f}{g} \right) \right)(x; \nu) \leq \frac{D^+[g(x)f](x; \nu) + D^+[-f(x)g](x; \nu)}{[g(x)]^2},
   \]
   provided that the right hand side exists and the function \( g \) is continuous at \( x \).

In addition, if the functions \( f \) and \( g \) are directionally differentiable at \( x \), then, the inequalities in the last three assertions become equalities.

Similarly, the properties and calculus rules for the Dini lower directional derivatives can be obtained.

Next, we present the following theorem, which shows that the Dini upper and Dini lower directional derivatives can be used conveniently for characterizing an extremum of a function.

**Theorem 1.35** For a function \( f : \mathbb{R}^n \to \mathbb{R} \), the following assertions hold:
Pseudolinear Functions and Optimization

(a) If \( f(x) \leq f(x + tv) \) (respectively, \( f(x) \geq f(x + tv) \)) for all \( t > 0 \) are sufficiently small, then \( D_+ f(x; \nu) \geq 0 \) (respectively, \( D_+ f(x; \nu) \leq 0 \)). In particular, if \( f \) is directionally differentiable at \( x \) and \( f(x) \leq f(x + tv) \) (respectively, \( f(x) \geq f(x + tv) \)), for all \( t > 0 \) are sufficiently small, then, \( f'(x; \nu) \geq 0 \) (respectively, \( f'(x; \nu) \leq 0 \)).

(b) If \( D^+ f(x + tv; \nu) \geq 0 \) for all \( x, \nu \in \mathbb{R}^n \) and \( t \in [0,1[ \) and if the function \( t \mapsto f(x + tv) \) is continuous on \([0,1[\), then, \( f(x) \leq f(x + v) \).

Proof (a) From the definitions of the Dini lower and Dini upper directional derivatives, the proof follows.

(b) On the contrary, suppose that \( f(x) > f(x + v) \) for some \( x, \nu \in \mathbb{R}^n \). We consider the function \( h : [0,1] \to \mathbb{R} \), defined by

\[
    h(t) = f(x + tv) - f(x) + t(f(x) - f(x + v)).
\]

Clearly, \( h \) is continuous on \([0,1]\) and \( h(0) = h(1) = 0 \). Hence, there exists \( \bar{t} \in [0,1[ \), such that \( h \) has a maximum value at \( \bar{t} \). Set \( y := x + \bar{t}v \). Then,

\[
    h(\bar{t}) \geq h(\bar{t} + t), \forall t \in [0,1 - \bar{t}].
\]

Hence, we have

\[
    f(y + tv) - f(y) \leq t(f(x + v) - f(x)),
\]

for all \( t > 0 \) are sufficiently small.

Dividing the above inequality by \( t \) and taking the limit superior as \( t \to 0^+ \), it follows that

\[
    D^+ f(y; \nu) = \lim_{t \to 0^+} \frac{f(y + tv) - f(y)}{t} \leq f(x + v) - f(x) < 0,
\]

which is a contradiction to our assumption. This completes the proof.

We know that the mean value theorem plays a key role in the classical theory of the differential calculus. The following mean value theorem for upper semicontinuous functions of one variable from Diewert [65] provided an extension for nondifferentiable functions.

**Theorem 1.36 (Diewert's mean value theorem)** If \( f : [a,b] \to \mathbb{R} \) is an upper semicontinuous function, then, there exists \( c \in [a,b[ \), such that

\[
    D_+ f(c) \leq D^+ f(c) \leq \frac{f(b) - f(a)}{b - a}.
\]

Proof Let \( \alpha = \frac{f(b) - f(a)}{b - a} \). Let us define a function \( h : [a,b] \to \mathbb{R} \) by

\[
    h(t) = f(t) - \alpha t.
\]
Since $f$ is upper semicontinuous on $[a, b]$, it follows that $h$ is also upper semicontinuous on the compact set $[a, b]$. By Berge’s maximum theorem [22], there exists a point $c \in [a, b]$ such that $h$ attains its maximum at $c$. Then, we get

$$h(c) \geq h(t), \forall t \in [a, b].$$

That is,

$$f(t) - f(c) \leq \alpha (t - c), \forall t \in [a, b].$$

If $c \in [a, b]$, then from the above inequality, we get

$$\frac{f(t) - f(c)}{t - c} \leq \alpha, \forall t \in ]c, b[.$$

From the above inequality, it follows that

$$D^+ f(c) \leq D^+ f(c) \leq \alpha.$$  

This completes the proof.

The following mean value theorem for lower semicontinuous functions can be proved by using the same argument as in Theorem 1.36.

**Corollary 1.9** Let $f : [a, b] \rightarrow \mathbb{R}$ be a lower semicontinuous function. Then, there exists $c \in [a, b]$, such that

$$D^+ f(c) \geq D^+ f(c) \geq \frac{f(b) - f(a)}{b - a}.$$  

The following mean value theorem for continuous functions may be obtained by using Theorem 1.36 and Corollary 1.8.

**Corollary 1.10** If $f : [a, b] \rightarrow \mathbb{R}$ is a continuous function, then there exist $c, d \in [a, b]$ such that

$$D^+ f(c)(b - a) \leq f(b) - f(a) \leq D^+ f(d)(b - a).$$

Now, we state the following important result. The proof is based on Theorem 1.36.

**Corollary 1.11** Let $f : [a, b] \rightarrow \mathbb{R}$ be an upper semicontinuous function. If for all $c \in [a, b[, D^+ f(c) \geq 0$ (respectively, $D^+ f(c) \leq 0$), then $f$ is a nondecreasing (respectively, nonincreasing) function on $[a, b]$.

Proof Assume that $D^+ f(c) \geq 0$, for all $c \in [a, b[$. On the contrary, suppose that $f$ is strictly decreasing on $[a, b]$. Then there exist $t_1, t_2 \in [a, b]$ such that $t_1 \leq t_2$ and $f(t_1) > f(t_2)$. By Theorem 1.36 there exists $t \in [t_1, t_2]$ such that

$$D^+ f(t) \leq \frac{f(t_2) - f(t_1)}{t_2 - t_1},$$

which implies that $D^+ f(t) < 0$, which is a contradiction to our assumption. By using Corollary 1.8, we can show that $f$ is nonincreasing on $[a, b]$ if $D^+ f(c) \leq 0$ for all $c \in [a, b]$.

We now give Diewert’s mean value theorem for functions defined on a convex subset of $\mathbb{R}^n$. The proof is along the lines of Ansari et al. [2].
Theorem 1.37 (Diewert’s mean value theorem) Let $X$ be a nonempty convex subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$ be a real-valued function. Then, for every pair of distinct points $x$ and $y$ in $X$, the following assertions hold.

(i) If $f$ is radially upper semicontinuous on $X$, then there exists $z \in [x, y]$, such that
\[ D^+ f(z; y - x) \leq f(y) - f(x). \]

(ii) If $f$ is radially lower semicontinuous on $X$, then there exists $w \in [x, y]$, such that
\[ D_+ f(w; y - x) \geq f(y) - f(x). \]

(iii) If $f$ is radially continuous on $X$, then there exists $\nu \in [x, y]$ such that
\[ D^+ f(\nu; y - x) \leq f(y) - f(x) \leq D_+ f(\nu; y - x). \]

Moreover, if the Dini upper directional derivative $f^D(\nu; y - x)$ is continuous in $\nu$ on the line segment $[x, y]$, then there exists a point $w \in [x, y]$, such that
\[ D_+ f(w; y - x) = f(y) - f(x). \]

Proof Let $x$ and $y$ be two distinct points in $X$. Define a function $h : [0, 1] \to \mathbb{R}$ by
\[ h(t) = f(x + t(y - x)). \]

(a) If $f$ is radially upper semicontinuous on $X$, then $h$ is an upper semicontinuous function on $[0, 1]$. By Theorem 1.36, there exists $\bar{t} \in [0, 1]$ such that
\[ D^+ h(\bar{t}) \leq h(1) - h(0). \]

If we set $w = x + \bar{t}(y - x)$, then we have
\[ D^+ h(\bar{t}) = D^+ f(w; y - x), \]
which lead to the results.

(b) By using Corollary 1.8, the proof follows.

(c) By using Corollary 1.9, the proof follows. Moreover, if $D^+ f(\nu; y - x)$ is continuous in $\nu$ on the line segment $[x, y]$, then by the intermediate value theorem, there exists a point $w \in [x, y]$ such that
\[ D^+ f(w; y - x) = f(y) - f(x). \]

A generalization of the Dini (upper and lower) directional derivative is the Dini-Hadamard (upper and lower) directional derivative.

Definition 1.74 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $x \in \mathbb{R}^n$ be a point where $f$ is finite.
Basic Concepts in Convex Analysis

(a) The Dini-Hadamard upper directional derivative of $f$ at $x$ in the direction $\nu u \in \mathbb{R}^n$ is defined by

$$f^{DH}(x; \nu) := \limsup_{u \to \nu, t \to 0^+} \frac{f(x + tu) - f(x)}{t}.$$ 

(b) The Dini-Hadamard lower directional derivative of $f$ at $x$ in the direction $\nu \in \mathbb{R}^n$ is defined by

$$f^{DH}(x; \nu) := \liminf_{u \to \nu, t \to 0^+} \frac{f(x + tu) - f(x)}{t}.$$ 

If $f^{DH}(x; \nu) = f^{DH}(x; \nu)$, then we denote it by $f^{DH*}(x; \nu)$, that is,

$$f^{DH*}(x; \nu) := \lim_{u \to \nu, t \to 0^+} \frac{f(x + tu) - f(x)}{t}.$$ 

From the definitions of the Dini upper(lower) directional and the Dini-Hadamard upper (lower) directional derivative, we can easily obtain the following relations:

$$(-f)^{DH}(x; \nu) = -f^{DH}(x; \nu), \quad (-f)^{DH}(x; \nu) = -f^{DH}(x; \nu),$$

$$f^{DH}(x; \nu) \leq f^D(x; \nu) \leq f^{DH}(x; \nu) \leq f^{DH}(x; \nu).$$

The following example illustrates that the Dini (upper and lower) directional derivative and the Dini-Hadamard (upper and lower) directional derivative are different.

**Example 1.10** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by

$$f(x_1, x_2) = \begin{cases} 0, & \text{if } x_2 = 0, \\ x_1 + x_2, & \text{if } x_2 \neq 0. \end{cases}$$

Let $x = (0, 0)$ and $\nu = e_1 = (1, 0)$. Then, by an easy calculation, we get

$$f^{DH}(x; d) = 1 \text{ and } f^D(x; d) = 0.$$ 

The following result shows that the Dini upper (lower) directional derivative and the Dini-Hadamard upper (lower) directional derivative at a point coincide, if $f$ is locally Lipschitz around that point.

**Theorem 1.38** Let $f$ be locally Lipschitz around a point $x \in \mathbb{R}^n$. Then, for every $d \in \mathbb{R}^n$,

$$f^{DH}(x; d) = f^D(x; d) \text{ and } f^{DH}(x; d) = f^D(x; d).$$
Pseudolinear Functions and Optimization

Proof Let \( N(x) \) be a neighborhood of \( x \) and \( f \) be Lipschitz continuous on \( N(x) \) with Lipschitz constant \( k \). Let \( d \in \mathbb{R}^n \) be arbitrary. Then there exist \( \delta > 0 \) and \( \tau > 0 \) such that \( \tau \) for all \( v \in \mathbb{R}^n \) and \( t \in \mathbb{R} \) satisfying the conditions \( \| v - d \| < \delta \) and \( 0 < t < \tau \), we have \( x + td, x + tv \in N(x) \). Consequently,

\[
|f(x + tv) - f(x + td)| \leq kt\|v - d\|.
\]

Therefore,

\[
\limsup_{v \to d, t \to 0+} \frac{f(x + td) - f(x + tv)}{t} = 0.
\]

By applying the properties of lim sup, we obtain

\[
f^{DH}(x; d) = \limsup_{v \to d, t \to 0+} \frac{f(x + td) - f(x)}{t} \leq \limsup_{v \to d, t \to 0+} \frac{f(x + td) - f(x)}{t} + \limsup_{v \to d, t \to 0+} \frac{f(x + tv) - f(x + td)}{t} = f^D(x; d).
\]

Since \( f^D(x; d) \leq f^{DH}(x; d) \), we have the equality.

Similarly, we can prove that \( f^D(x; d) = f^{DH}(x; d) \).

Next, we state the following result about upper (lower) semicontinuity of the Dini-Hadamard derivatives.

**Theorem 1.39** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function. Then \( f_{DH}(x; d) \) (respectively, \( f^{DH}(x; d) \)) is a lower (respectively, upper) semicontinuous function of \( d \).

For further details about the Dini and Dini-Hadamard derivatives, we refer to Giorgi and Komlosi [88, 89, 90] and Schirotzek [245].

### 1.10.3 Clarke Subdifferentials

The concept of subdifferentials for general nonconvex, locally Lipschitz functions was introduced by Clarke [49]. To deal with the problems in which the smoothness of the data is not necessarily postulated, Clarke [49] has developed the following basic concepts:

**Definition 1.75** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a locally Lipschitz function near \( x \). The Clarke generalized directional derivative of \( f \) at \( x \) in the direction of a vector \( \nu \in \mathbb{R}^n \), denoted by \( f^0(x; \nu) \), is defined as

\[
f^0(x; \nu) := \limsup_{y \to x \atop t \downarrow 0} \frac{f(y + t\nu) - f(y)}{t}.
\]

**Definition 1.76** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a locally Lipschitz function on \( X \). The Clarke generalized subdifferential of \( f \) at \( x \), denoted by \( \partial^c f(x) \), is defined as

\[
\partial^c f(x) := \{ \xi \in \mathbb{R}^n : f^0(x; \nu) \geq \langle \xi, \nu \rangle, \forall \nu \in \mathbb{R}^n \}.
\]
In fact, a locally Lipschitz function $f$ is not differentiable everywhere. Rademacher’s theorem (see Evans and Gariepy [76]) states that a locally Lipschitz function is differentiable almost everywhere (a.e.) in the sense of Lebesgue measure, i.e., a set of points where $f$ is not differentiable, forms a set of measure zero.

**Theorem 1.40** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz function near $x$. Let $S \subseteq \mathbb{R}^n$ be a set with Lebesgue measure zero. Then, at any $x \in \mathbb{R}^n$, the Clarke generalized subdifferential is given by

$$\partial^c f (x) = \text{co} \{ \xi \in \mathbb{R}^n : \xi = \lim_{i \to \infty} \nabla f (x_i) , x_i \to x , x_i \notin \Omega_f \cup S \},$$

where $\Omega_f$ denotes the set of points, where $f$ is not differentiable.

The following proposition from Clarke [49], summarizes some basic properties for the Clarke generalized gradient.

**Proposition 1.28** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz function near $x$, with Lipschitz constant $M$. Then

1. The function $f^0 (x; \cdot)$ is a finite, positively homogeneous function of $\nu$ and satisfies
   $$|f^0 (x; \nu)| \leq M \|\nu\|.$$

2. $f^0 (x; \nu)$ is Lipschitz of rank $M$ as a function of $\nu$ and upper semicontinuous as a function of $(x; \nu)$.

3. $f^0 (x; -\nu) = (-f)^0 (x; \nu)$.

The following proposition states some important properties of the Clarke generalized subdifferential.

**Proposition 1.29** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz function near $x$ with Lipschitz constant $M$. Then,

1. For each $x$, $\partial^c f (x)$ is nonempty, convex and compact subset of $\mathbb{R}^n$. Moreover, $$\|\xi\| \leq M , \ \forall \xi \in \partial^c f (x).$$

2. For every $\nu \in \mathbb{R}^n$, $$f^0 (x; \nu) = \max \{ \langle \xi , \nu \rangle : \xi \in \partial^c f (x) \}.$$

That is, $f^0 (x; \cdot)$ is the support function of the set $\partial^c f (x)$.

The following lemma states that the Clarke subdifferential for locally Lipschitz functions are closed sets.
Lemma 1.5 Let \( f : X \rightarrow \mathbb{R} \) be a locally Lipschitz function at \( x \in K \). If \( \{x_n\} \) and \( \{\zeta_n\} \) are two sequences in \( \mathbb{R}^n \) such that, \( \zeta_n \in \partial c f(x_n) \), for all \( n \) and if \( x_n \rightarrow x \) and \( \zeta \) is a cluster point of \( \{\zeta_n\} \), then \( \zeta \in \partial c f(x) \).

The following theorem states that for convex functions, the Clarke and convex subdifferentials coincide. The proof follows along the lines of Clarke [49].

Theorem 1.41 Let \( X \subseteq \mathbb{R}^n \) be an open convex set and \( f : X \rightarrow \mathbb{R} \) is convex and locally Lipschitz with rank \( M \) on \( X \). Then \( \partial c f(x) \) coincides with \( \partial f(x) \), for each \( x \in X \). Moreover, \( f^0(x;\nu) \) coincides with \( f'(x;\nu) \), for each \( x \in X \) and \( \nu \in \mathbb{R}^n \).

Proof From Theorem 1.30, it is clear that if \( f \) is a convex function, then \( f' + (x;\nu) \) exists for each \( \nu \in \mathbb{R}^n \) and \( f' + (x;) \) is the support function of \( \partial f \) at \( x \). Therefore, we need to prove only that

\[
 f'_{+}(x;\nu) = f^0(x;\nu), \quad \forall \nu \in \mathbb{R}^n.
\]

Now, the Clarke generalized gradient \( f^0(x;\nu) \), may be written as

\[
 f^0(x;\nu) = \lim_{\varepsilon \downarrow 0} \sup_{\|\bar{x} - x\| \leq \varepsilon \delta} \sup_{0 < \lambda < \varepsilon} \frac{f(\bar{x} + \lambda \nu) - f(\bar{x})}{\lambda},
\]

where \( \delta \) is any positive real number. By Theorem 1.30, for a convex function, the difference quotient

\[
 \frac{f(\bar{x} + \lambda \nu) - f(\bar{x})}{\lambda}
\]

is a nondecreasing function of \( \lambda \). Hence,

\[
 f^0(x;\nu) = \lim_{\varepsilon \downarrow 0} \sup_{\|\bar{x} - x\| \leq \varepsilon \delta} \frac{f(\bar{x} + \varepsilon \nu) - f(\bar{x})}{\varepsilon},
\]

Now, by the locally Lipschitz property of \( f \), for any \( \bar{x} \in B_x(\varepsilon) \), we have

\[
 \left| \frac{f(\bar{x} + \varepsilon \nu) - f(\bar{x})}{\varepsilon} - \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \right| \leq 2\delta M.
\]

Therefore,

\[
 f^0(x;\nu) \leq \lim_{\varepsilon \downarrow 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} + 2M\delta = f'_{+}(x;\nu) + 2M\delta.
\]

Since, \( \delta \) is arbitrary, we get \( f^0(x;\nu) \leq f'_{+}(x;\nu) \). Moreover, \( f'_{+}(x;\nu) \leq f^0(x;\mu) \) is always true. Hence, the equality follows.

The following theorem provides a characterization for convexity in terms of the Clarke subdifferential.
Theorem 1.42 Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz on open convex set $X$. Then $f$ is a convex function on $X$ if and only if $\partial^c f$ is a monotone map on $X$.

Next, we present some basic calculus rules, which will be used frequently. Before that, we give the definition of a strictly differentiable function.

Definition 1.77 (Strictly differentiable function) Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ be any function. Then, strict directional derivative of $f$ at the point $x$ in the direction $\nu \in \mathbb{R}^n$, is given by

$$f^s(x; \nu) = \lim_{\bar{x} \to x \lambda \downarrow 0} \frac{f(\bar{x} + \lambda \nu) - f(\bar{x})}{\lambda},$$

provided the limit exists. The function $f$ is said to admit a strict derivative at $x \in X$, denoted by $D^s f(x)$, if for each $\nu \in \mathbb{R}^n$ the following holds

$$f^s(x; \nu) = \langle D^s f(x), \nu \rangle.$$

It is known that if $f$ is a continuously differentiable function at $x$, then it is strictly differentiable at $x$.

The following proposition characterizes the relationship between a locally Lipschitz property and strict differentiability.

Proposition 1.30 Let $X \subseteq \mathbb{R}^n$ be an open set and $f : X \to \mathbb{R}$ be any map. Let $\xi$ be any vector in $\mathbb{R}^n$. Then the following statements are equivalent:

1. $f$ is strictly differentiable at $x$ and $D^s f(x) = \xi$.

2. $f$ is locally Lipschitz around $x$ and $f^s(x; \nu) = \langle \xi, \nu \rangle, \forall \nu \in \mathbb{R}^n$.

Proposition 1.31 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function near $x$. Then, for any scalar $s$, one has

$$\partial^c (s \sum_{i=1}^{m} \lambda_i f_i) (x) \subseteq \lambda \partial^c f(x).$$

Moreover, the equality holds, if all but at most one of the function $f_i$ are strictly differentiable at $x$.

Proposition 1.32 Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$ be a finite family of locally Lipschitz functions near $x$. Then,

$$\partial^c (f_1 + f_2 + ... + f_m) (x) \subseteq \partial^c f_1(x) + \partial^c f_2(x) + ... + \partial^c f_m(x).$$

Moreover, the equality holds if all but at most one of the function $f_i$ are strictly differentiable at $x$.

Proposition 1.33 Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$ be a finite family of locally Lipschitz function near $x$. Then, for any scalar $\lambda_i, i = 1, ..., m$, one has

$$\partial^c \left( \sum_{i=1}^{m} \lambda_i f_i \right)(x) \subseteq \lambda \sum_{i=1}^{m} \partial^c f_i(x).$$

Moreover, the equality holds, if all but at most one of the function $f_i$ are strictly differentiable at $x$. 

The following notion of regular functions helps to sharpen the subdifferential calculus rules by turning inclusions to equality.

**Definition 1.78** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a locally Lipschitz function near \( x \). Then \( f \) is said to be regular at \( x \), if

1. For all \( \nu \in \mathbb{R}^n \), the one sided directional derivative \( f'_+ (x; \nu) \) exists.
2. For all \( \nu \in \mathbb{R}^n \), \( f'_+ (x; \nu) = f^0 (x; \nu) \).

Under the regularity assumption, Propositions 1.32 and 1.33 take the following form.

**Proposition 1.34** Let \( f_i : \mathbb{R}^n \to \mathbb{R} \), \( i = 1, \ldots, m \) be a finite family of locally Lipschitz function near \( x \). Suppose, each \( f_i \), \( i = 1, \ldots, m \) be regular at \( x \). Then,

\[
\partial^c (f_1 + f_2 + \ldots + f_m) (x) = \partial^c f_1 (x) + \partial^c f_2 (x) + \ldots + \partial^c f_m (x).
\]

In addition, if each \( \lambda_i \) is nonnegative, then,

\[
\partial^c \left( \sum_{i=1}^{m} \lambda_i f_i \right) (x) = \lambda_i \partial^c f_i (x).
\]

**Proposition 1.35** Let \( f, g : \mathbb{R}^n \to \mathbb{R} \) be a locally Lipschitz function near \( x \) and suppose that \( g(x) \neq 0 \). Then \( \frac{f}{g} \) is locally Lipschitz near \( x \) and one has

\[
\partial^c \left( \frac{f}{g} \right) (x) \subseteq \frac{g(x)\partial^c f(x) - f(x)\partial^c g(x)}{(g(x))^2}.
\]

If in addition \( f(x) \geq 0, g(x) > 0 \) and if \( f \) and \( -g \) are regular at \( x \), then the equality holds and \( \frac{f}{g} \) is regular at \( x \).

Next, we state the following important chain rule for the Clarke subdifferential.

**Proposition 1.36 (Chain rule)** Let \( h : \mathbb{R}^n \to \mathbb{R}^m \) and \( g : \mathbb{R}^m \to \mathbb{R} \) be the given functions, such that each \( h \) be strictly differentiable at \( x \) and \( g \) is Lipschitz near \( h(x) \). Then \( f = g \circ h \) is Lipschitz near \( x \) and one has

\[
\partial^c f (x) \subseteq \partial^c g (h (x)) \circ D_h (x).
\]

The equality holds, if \( g(oh - g) \) is regular at \( h(x) \).

Next, we state the following well-known mean value theorem. The proof may be found in Clarke [49] and Schirotzek [245].
Theorem 1.43 (Lebourg mean value theorem) Let \( x \) and \( y \) be points in \( X \) and suppose that \( f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be locally Lipschitz on an open set containing the line segment \([x, y]\). Then, there exists a point \( u \in [x, y[ \), such that
\[
f(y) - f(x) \in \langle \partial^c f(u), y - x \rangle,
\]
where \([x, y]\) denotes the line segment joining \( x \) and \( y \) excluding the end points.

The following proposition describes the role of the distance function in exact penalization. The proof may be found in Clarke [49].

Definition 1.79 Let \( C \subseteq \mathbb{R}^n \) be a nonempty set. The Clarke tangent cone of \( C \) denoted by \( T_C(x) \) is defined as
\[
T_C(x) := \{ y \in \mathbb{R}^n : d_C^0(x, y) = 0 \}
\]
where \( d_C(.,.) \) denotes the distance function related to \( C \). It is easy to show that \( T_C(x) \) is a closed convex cone containing 0.

The normal cone to the set \( C \) at \( x \) is defined as a cone polar to the tangent cone \( T_C(x) \), as follows.

Definition 1.80 (Clarke normal cone) Let \( C \subseteq \mathbb{R}^n \) be a nonempty set. The Clarke normal cone of \( C \) at \( x \in C \), denoted by \( N_C(x) \), is defined as
\[
N_C(x) := \{ \xi \in \mathbb{R}^n : \langle \xi, \nu \rangle \leq 0, \forall \nu \in T_C(x) \}.
\]
It follows that, \( N_C(x) \) is the closed convex cone generated by \( \partial^c d_C(x) \). Moreover, if \( C \) is a convex set, then \( N_C(x) \) coincides with the normal cone in the sense of convex analysis.

The following proposition provides the characterization of a normal cone in terms of a generalized gradient.

Proposition 1.37 Let \( C \subseteq \mathbb{R}^n \) be a nonempty set and let \( x \in C \), then,
\[
N_C(x) = \text{cl} \left\{ \bigcup_{\lambda \geq 0} \lambda \partial^c d_C(x) \right\}.
\]

Proposition 1.38 Let \( f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be locally Lipschitz of rank \( M \) on \( X \) and let \( \bar{x} \) belong to a set \( C \subseteq X \). Suppose that \( f \) attains a minimum at \( \bar{x} \). Then for any \( M' \geq M \), the function \( g(y) = f(y) + M d_C(y) \) attains a minimum over \( X \) at \( \bar{x} \). If \( M > M' \) and \( C \) is closed, then any other point minimizing \( g \) over \( X \) lies within \( C \).
1.11 Optimality Criteria

In the simplest case, an optimization problem consists of maximizing or minimizing a real-valued single objective function, by systematically choosing input values from within an allowed set, known as the feasible set and computing the value of the function. Depending upon the feasible set, the optimization problems have been classified as unconstrained and constrained optimization problems.

1.11.1 Unconstrained Minimization Problem

Consider an optimization problem of the form

\[(UMP) \quad \min f(x)\]

subject to \(x \in \mathbb{R}^n\),

where \(f : \mathbb{R}^n \to \mathbb{R}\) is a given function. The problem (UMP) is referred to as an unconstrained minimization problem. Unconstrained minimization problems arise seldom in practical applications. However, optimality conditions for constrained minimization problems become logical extension of the optimality conditions for unconstrained minimization problems.

Definition 1.81 (Local and global minimum) A vector \(\bar{x} \in \mathbb{R}^n\) is said to be a local minimum for (UMP), if there exists an \(\varepsilon > 0\), such that

\[f(\bar{x}) \leq f(x), \forall x \in B_\varepsilon(\bar{x}) \cap \mathbb{R}^n.\]

It is said to be global minimum, if

\[f(\bar{x}) \leq f(x), \forall x \in \mathbb{R}^n.\]

It is clear that a global minimum is also a local minimum.

Theorem 1.44 Suppose that \(f : \mathbb{R}^n \to \mathbb{R}\) is differentiable at \(\bar{x}\) and there exists a vector \(d \in \mathbb{R}^n\) such that \(\langle \nabla f(\bar{x}), d \rangle < 0\). Then there exists some \(\lambda > 0\), such that

\[f(\bar{x} + \lambda d) < f(\bar{x}), \forall \lambda \in ]0, \lambda[.\]

Proof Since, \(f\) is differentiable at \(\bar{x}\), we have

\[f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \langle \nabla f(\bar{x}), d \rangle + \lambda \|d\| \varphi(\bar{x}, \lambda d),\]

(1.15)

where \(\varphi(\bar{x}, \lambda d) \to 0\) as \(\lambda \to 0\). Now, the equation (1.15) can be rewritten as

\[\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \langle \nabla f(\bar{x}), d \rangle + \|d\| \varphi(\bar{x}, \lambda d).\]

(1.16)
By our assumption, \( \langle \nabla f (\bar{x}), d \rangle < 0 \) and \( \varphi (\bar{x}, \lambda d) \to 0 \) as \( \lambda \to 0 \). Therefore, there exists some \( \bar{\lambda} > 0 \), such that
\[
\langle \nabla f (\bar{x}), d \rangle + \|d\| \varphi (\bar{x}, \lambda d) < 0, \quad \forall \lambda \in ]0, \bar{\lambda}].
\]
Hence, by (1.17), the required result follows.

Now, we state the following well-known optimality conditions. The proof may be found in Luenberger [173].

**Proposition 1.39** Suppose that \( \bar{x} \) is a local minimum for (UMP) and \( f \) is differentiable at \( \bar{x} \). Then,
\[
\nabla f (\bar{x}) = 0.
\]
Proof Suppose to the contrary that \( \nabla f (\bar{x}) \neq 0 \). Then setting \( d = -\nabla f (\bar{x}) \), we get
\[
\langle \nabla f (\bar{x}), d \rangle = -\|\nabla f (\bar{x})\|^2 < 0.
\]
Therefore, by Theorem 1.38, there exists \( \bar{\lambda} > 0 \), such that
\[
f (\bar{x} + \lambda d) < f (\bar{x}), \quad \forall \lambda \in ]0, \bar{\lambda}[:
\]
a contradiction to our assumption, that \( \bar{x} \) is a local minimum for (UMP). Hence, \( \nabla f (\bar{x}) = 0 \).

**Proposition 1.40** Suppose that \( \bar{x} \) is a local minimum for (UMP) and \( f \) is a twice differentiable function at \( \bar{x} \). Then, \( \nabla^2 f (\bar{x}) \) is positive semidefinite.

Proof Since, \( f \) is a twice differentiable function at \( \bar{x} \), therefore, for arbitrary \( d \in \mathbb{R}^n \), we have
\[
f (\bar{x} + \lambda d) = f (\bar{x}) + \lambda \langle \nabla f (\bar{x}), d \rangle + \frac{1}{2} \lambda^2 \langle d, \nabla^2 f (\bar{x}) d \rangle + \lambda^2 \|d\|^2 \varphi (\bar{x}, \lambda d),
\]
where \( \varphi (\bar{x}, \lambda d) \to 0 \) as \( \lambda \to 0 \). Since, \( \bar{x} \) is a local minimum, by Proposition 1.39, we get
\[
\nabla f (\bar{x}) = 0.
\]
Using (1.19), the equation (1.18) can be rewritten as
\[
\frac{f (\bar{x} + \lambda d) - f (\bar{x})}{\lambda^2} = \frac{1}{2} \langle d, \nabla^2 f (\bar{x}) d \rangle + \|d\|^2 \varphi (\bar{x}, \lambda d).
\]
Since, \( \bar{x} \) is a local minimum for (UMP), \( f (\bar{x} + \lambda d) \geq f (\bar{x}) \), for \( \lambda \) sufficiently small. Therefore, from (1.20), it follows that
\[
\frac{1}{2} \langle d, \nabla^2 f (\bar{x}) d \rangle + \|d\|^2 \varphi (\bar{x}, \lambda d) \geq 0,
\]
for \( \lambda \) sufficiently small.
Pseudolinear Functions and Optimization

By taking the limit as $\lambda \to 0$, it follows that
\[
\langle d, \nabla^2 f (\bar{x}) d \rangle \geq 0.
\]
Since, $d$ is arbitrary, hence, $\nabla^2 f (\bar{x})$ is positive semidefinite.

The following results provide an optimality condition for the unconstrained optimization problem (UMP) in terms of the Dini-Hadamard derivative.

**Theorem 1.45** If $x \in \mathbb{R}$ is a local minimum of the function (UMP), then, we have
\[
f_{DH}(x; d) \geq 0, \forall d \in \mathbb{R}^n.
\]
Proof Since $x \in \mathbb{R}^n$ is a local minimum point of the function $f$, then for each $d \in \mathbb{R}^n$, each $v$ in the neighborhood of $d$ and each $t > 0$ sufficiently small, we have
\[
\frac{f(x + tv) - f(x)}{t} \geq 0.
\]
Thus,
\[
f_{DH}(x; d) = \lim \inf_{v \to d, t \to 0^+} \frac{f(x + tv) - f(x)}{t} \geq 0.
\]
This completes the proof.

In terms of the Dini derivatives, we have the following optimality conditions for the unconstrained optimization problem (UMP).

**Corollary 1.12** If $x \in \mathbb{R}^n$ is a local minimum of the function $f : \mathbb{R}^n \to \mathbb{R}$, then,
\[
D^+ f(x; d) \geq 0, \forall d \in \mathbb{R}^n.
\]

Next, we state the following necessary and sufficient optimality conditions, which describe one of the most important properties possessed by convex as well as pseudoconvex functions, that is, every stationary point is a global minimum.

**Theorem 1.46** Suppose that $f$ is a differentiable convex function at $\bar{x}$. Then $\bar{x}$ is a global minimum for (UMP) if and only if
\[
\nabla f (\bar{x}) = 0.
\]
Proof If $\bar{x}$ is a global minimum for (UMP), then it is local minimum for (UMP). Therefore, by Proposition 1.39, we get
\[
\nabla f (\bar{x}) = 0.
\]
Now, suppose conversely that $\nabla f (\bar{x}) = 0$. Therefore, for each $x \in \mathbb{R}^n$, we have
\[
\langle \nabla f (\bar{x}), x - \bar{x} \rangle = 0.
\]
By the convexity of $f$ at $\bar{x}$, we get
\[
f(x) - f(\bar{x}) \geq \langle \nabla f (\bar{x}), x - \bar{x} \rangle, \forall x \in \mathbb{R}^n.
\]
From (1.21) and (1.22), we get
\[ f(x) \geq f(\bar{x}), \forall x \in \mathbb{R}^n. \]
Hence, \( \bar{x} \) is a global minimum for (UMP).

**Remark 1.12** If \( f \) is a differentiable pseudoconvex function at \( \bar{x} \), then, from (1.21) and the pseudoconvexity of \( f \) at \( \bar{x} \), we get
\[ f(x) \geq f(\bar{x}), \forall x \in \mathbb{R}^n. \]

Next, we present the following theorem, which states that for a strictly quasiconvex function, every local minimizer is a global one.

**Theorem 1.47** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a differentiable function. Let \( \bar{x} \in \mathbb{R}^n \) be a local minimizer of (UMP). If \( f \) is a strictly quasiconvex function at \( \bar{x} \), then \( \bar{x} \) is a global minimum of (UMP).

**Proof** If \( \bar{x} \) is a local minimum, then there exists an open ball \( B_\varepsilon(\bar{x}) \), such that
\[ f(\bar{x}) \leq f(x), \forall x \in B_\varepsilon(\bar{x}) \cap \mathbb{R}^n. \]
On the contrary, suppose that there exist \( \hat{x} \in \mathbb{R}, \hat{x} \notin B_\varepsilon(\bar{x}) \), such that
\[ f(\hat{x}) < f(\bar{x}). \]
By the strict quasiconvexity of \( f \), we get
\[ f(\lambda \hat{x} + (1 - \lambda) \bar{x}) < f(\bar{x}), \text{ for } \lambda \in ]0,1[. \]  \hfill (1.22)
For \( \lambda < \delta / \|\hat{x} - \bar{x}\| \), we have that
\[ (1 - \lambda) \bar{x} + \lambda \hat{x} \in B_\varepsilon(\bar{x}) \cap \mathbb{R}^n. \]
Then, we have
\[ f(\bar{x}) \leq f((1 - \lambda) \bar{x} + \lambda \hat{x}), \text{ for } 0 < \lambda < \delta / \|\hat{x} - \bar{x}\|. \]
which is a contradiction to (1.23). This completes the proof.

The following theorem states that the set of global minimizers of a quasi-convex function is a convex set.

**Theorem 1.48** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a quasiconvex function on the convex set \( X \). Let \( \bar{X} \) be the set of all global minimizers of \( f \). Then \( \bar{X} \) is a convex set.

**Proof** If \( \bar{X} = \phi \), then the result is trivially true. Let \( \alpha \) be the minimum value of \( f \) on \( \mathbb{R}^n \). By the definition of global minimizers, we have
\[ \bar{X} = \{ x \in X : f(x) = \alpha \} = \{ x \in X : f(x) \leq \alpha \}. \]
The convexity of \( \bar{X} \) follows by Theorem 1.14.

The following theorem gives the optimality condition for the unconstrained minimization problem (UMP), if the objective function \( f \) is convex.
Theorem 1.49 Let the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function in the unconstrained minimization problem (UMP). Then $\bar{x} \in \mathbb{R}^n$ is a global minimizer of (UMP) if and only if

$0 \in \partial f (\bar{x})$.

Proof Suppose that $\bar{x} \in \mathbb{R}^n$ is a global minimizer of (UMP). Therefore, by the definition of minimizer, we have

$f (x) - f (\bar{x}) \geq 0, \forall x \in \mathbb{R}^n$.

By the definition of subdifferential, we get

$0 \in \partial f (\bar{x})$.

Again, using the definition of subdifferentials, the converse can be proved.

1.11.2 Constrained Minimization Problems

Consider the optimization problem

$$(CMP) \quad \min f (x)$$

subject to $x \in X \subseteq \mathbb{R}^n$,

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a given function. The problem (CMP) is referred to as a constrained minimization problem.

Definition 1.82 (Local and global minimum) A vector $\bar{x} \in X$ is said to be a local minimum for (CMP), if there exists an $\varepsilon > 0$, such that

$f (\bar{x}) \leq f (x), \forall x \in B_\varepsilon (\bar{x}) \cap X$.

It is said to be a global minimum, if

$f (\bar{x}) \leq f (x), \forall x \in X$.

From Figure 1.5, it is clear that a global minimum is also a local minimum for (CMP), but not conversely.

When $X$ is a convex set, the following optimality conditions for the problem (CMP) holds. The proof can be found in Mangasarian [176], Luenberger [173], and Bazarra et al. [17].

Proposition 1.41 Suppose that $X$ is a convex set and for some $\varepsilon > 0$ and $\bar{x} \in X$, the function $f$ is a continuously differentiable function over $B_\varepsilon (\bar{x})$. Then, the following statement holds:

1. If $\bar{x}$ is a local minimum for (CMP), then

$$\langle \nabla f (\bar{x}), x - \bar{x} \rangle \geq 0, \forall x \in X.$$ (1.23)
2. In addition if \( f \) is a convex function over \( X \) and (1.24) holds, then \( \bar{x} \) is a global minimum for (CMP).

The constrained minimization problem (CMP) can be converted to an unconstrained minimization (UMP) by using indicator functions. For (CMP), associated unconstrained minimization problem may be formulated as:

\[
\min \tilde{f}(x)
\]

subject to \( x \in \mathbb{R}^n \),

where \( \tilde{f} : \mathbb{R}^n \to \mathbb{R} \) is a function given by \( \tilde{f}(x) = f(x) + \delta_X(x) \), that is,

\[
\tilde{f}(x) = \begin{cases} 
  f(x), & \text{if } x \in X, \\
  \infty, & \text{if otherwise.}
\end{cases}
\]

Using indicator functions, the optimality condition for constrained optimization problem (CMP), may be given as follows.

**Theorem 1.50** Let \( X \) be a convex set and the function \( f : X \subseteq \mathbb{R}^n \to \mathbb{R} \) be a convex function in the constrained minimization problem (CMP). Then \( \bar{x} \in \mathbb{R}^n \) is a minimizer of (CMP) if and only if

\[
0 \in \partial (f + \delta_X)(\bar{x}).
\]

Next, we state the following theorem, which gives the necessary optimality condition for (CMP), when the objective function is locally Lipschitz.

**Theorem 1.51** Suppose that \( f : \mathbb{R}^n \to \mathbb{R} \) be locally Lipschitz of rank \( M \) near \( \bar{x} \) and let \( \bar{x} \) belong to a set \( C \subseteq \mathbb{R}^n \). Suppose that \( f \) attains a minimum over \( C \) at \( \bar{x} \). Then

\[
0 \in \partial^c f(\bar{x}) + N_C(\bar{x}).
\]
Proof Let \( N \) be a neighborhood of \( \bar{x} \) upon which \( f \) is Lipschitz of rank \( M \). Since \( C \) and \( C \cap N \) have the same normal cone at \( \bar{x} \), we may assume that \( C \subseteq X \). By Proposition 1.38, \( \bar{x} \) is a local minimizer of the function \( f(x) + M d_C(x) \). Therefore,

\[
0 \in \partial (f + M d_C)(\bar{x}) \subseteq \partial f(\bar{x}) + M \partial d_C(\bar{x}).
\]

By Proposition 1.37, we get

\[
0 \in \partial f(\bar{x}) + N_C(\bar{x}).
\]

### 1.11.3 Scalar Optimization Problems and Optimality Criteria

The optimization problems, where the constrained set \( X \) is described explicitly by inequality constraints, arise frequently in optimization theory. Analytically, a single objective (scalar) optimization problem with explicit inequality constraints may be formulated as follows:

\[
(SOP) \quad \min f(x)
\]

subject to \( g_j(x) \leq 0, \quad j = 1, 2, \ldots, m, \)

where \( f : X \to \mathbb{R}, g_j : X \to \mathbb{R}, \quad j = 1, 2, \ldots, m \) are differentiable functions on an open set \( X \subseteq \mathbb{R}^n \). Let \( S = \{x : x \in X, g(x) \leq 0\} \) denote the set of all feasible solutions of the problem \((SOP)\). Suppose that \( J(\bar{x}) = \{j \in \{1, 2, \ldots, m\} : g_j(\bar{x}) = 0\} \) denotes the set of active constraint index set at \( \bar{x} \).

Optimality criteria form the foundation of mathematical programming both theoretically and computationally. The conditions that must be satisfied at the optimum point are called necessary optimality conditions. If any point does not satisfy the necessary optimality condition, it cannot be optimum. However, not every point that satisfies the necessary condition is optimal. The well-known necessary optimality conditions for a mathematical programming problem are Fritz John [85] and Kuhn-Tucker [151] type necessary optimality conditions. The latter condition is referred to as Karush-Kuhn-Tucker optimality conditions to give credit to Karush [135], who had derived these conditions in his Master’s thesis in 1939 (see Boyd and Vandenberghe [29]).

Now, we state the following Fritz John necessary optimality conditions. For the proof we refer to Mangasarian [176].

**Theorem 1.52 (Fritz John necessary optimality conditions)** Let \( \bar{x} \) be a local minimum of the \((SOP)\). Then, there exist multipliers \( \lambda_0 \in \mathbb{R} \) and \( \mu \in \mathbb{R}^m \), such that

\[
\lambda_0 \nabla f(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0,
\]

(1.24)
\[ \mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, ..., m, \quad (1.25) \]
\[ \lambda_0 \geq 0, \mu \geq 0 \text{ and } (\lambda, \mu) \neq 0. \quad (1.26) \]

The conditions (1.25)–(1.26) are known as Fritz John condition and the condition (1.27) is known as complementary slackness condition.

We note that in case \( \lambda_0 = 0 \), the objective function \( f \) disappears and we have a degenerate case. To avoid degeneracy, we apply some regularity conditions on the constraints. These regularity conditions, which ensure that \( \lambda_0 \neq 0 \), are referred to as constraint qualifications.

Next, we state Slater’s constraint qualification, which is frequently used in nonlinear optimization.

**Definition 1.83 (Slater’s constraint qualification)** The problem (SOP) is said to satisfy the Slater constraint qualification if each \( g_j \) is convex (or pseudoconvex) and there exists a feasible point \( \bar{x} \in S \), such that
\[ g_j(\bar{x}) < 0, \quad j = 1, 2, ..., m. \]

Besides Slater’s constraint qualification, there are several other constraint qualifications for (SOP), such as the Mangasarian Fromovitz constraint qualification, Karlin’s constraint qualification, linear independent constraint qualification, strict constraint qualification, and others. For further details about different constraint qualifications for (SOP) and relations between them, we refer to Mangasarian [176].

If \( \lambda_0 \neq 0 \), we can take \( \lambda_0 = 1 \). The Fritz John condition then reduces to the famous Karush-Kuhn-Tucker optimality conditions stated as follows:

**Theorem 1.53 (Karush-Kuhn-Tucker necessary optimality conditions)** Let \( \bar{x} \) be a local minimum of the (SOP) and a suitable constraint qualification is satisfied at \( \bar{x} \). Then, there exist multiplier \( \mu \in \mathbb{R}^m \), such that
\[ \nabla f(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0, \]
\[ \mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, ..., m. \]
\[ \mu_j \geq 0, \quad j = 1, 2, ..., m. \]

Of course, one would like to have the same criterion be both necessary as well as sufficient. The validity of the Karush-Kuhn-Tucker necessary optimality condition does not guarantee the optimality of \( \bar{x} \). However, these conditions become sufficient under somewhat ideal conditions which are rarely satisfied in practice. The above necessary optimality conditions become sufficient if \( f \) and \( g \) have some kind of convexity. In the presence of convexity it is very convenient to find an optimal solution for the (SOP).

Now, we state the Karush-Kuhn-Tucker sufficient optimality conditions for (SOP).
Theorem 1.54 (Karush-Kuhn-Tucker sufficient optimality condition) Let \( \bar{x} \) be a feasible solution for the (SOP) and let the functions \( f : X \rightarrow \mathbb{R} \) and \( g_j, j = 1, 2, ..., m \) be convex and continuously differentiable at \( \bar{x} \). If there exist multipliers \( 0 \leq \mu \in \mathbb{R}^m \), such that

\[
\nabla f(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0,
\]

\[
\mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, ..., m.
\]

then, \( \bar{x} \) is a global minimum for the (SOP).

1.11.4 Multiobjective Optimization and Pareto Optimality

In practice, we usually encounter with the optimization problems, having several conflicting objectives rather than a single objective, known as multiobjective optimization problems. Multiobjective optimization provides a flexible modeling framework that allows for simultaneous optimization of more than one objective over a feasible set. These problems occur in various areas of modern research such as in analyzing design trade-offs, selecting optimal product or process designs, or any other application, where we need an optimal solution with trade-offs between two or more conflicting objectives.

Because these objectives conflict naturally, a trade-off exists. The set of trade-off designs that cannot be improved upon according to one criterion without hurting another criterion is known as the Pareto set.

Analytically, a nonlinear multiobjective optimization problem (MOP) may be formulated as follows:

\[ (MOP) \quad \min f(x) = (f_1(x), f_2(x), ..., f_k(x)) \]

subject to \( g_j(x) \leq 0, \quad j = 1, 2, ..., m \),

where \( f : X \rightarrow \mathbb{R}^k, g_j : X \rightarrow \mathbb{R}, \quad j = 1, 2, ..., m \) are differentiable functions on an open set \( X \). Let \( S = \{ x \in X : g(x) \leq 0 \} \) denote the set of feasible solutions of the problem (MOP). Suppose that \( J(\bar{x}) = \{ j \in \{ 1, 2, ..., m \} : g_j(\bar{x}) = 0 \} \) denotes the set of active constraint index set at \( \bar{x} \).

Remark 1.13 For \( k = 1 \), the problem (MOP) reduces to the scalar optimization problem (SOP).

Edgeworth [71] and Pareto [219] have given the definition of the standard optimality concept via the usage of utility functions. Then it has been extended to the classical notion of Pareto efficiency/optimality defined via an ordering cone. The allocation of resources is Pareto optimal, often called the Pareto efficient, if it is not possible to change the allocation of resources in such a way as to make some people better off without making others worse off.
Definition 1.84 (Pareto optimal) A feasible point \( \bar{x} \), is said to be Pareto optimal (Pareto efficient) solution for (MOP), if there does not exist another feasible point \( x \), such that
\[
f(x) \leq f(\bar{x})
\]
or equivalently,
\[
f_i(x) \leq f_i(\bar{x}), \forall i = 1, 2, ..., k, i \neq r,
\]
\[
f_r(x) < f_r(\bar{x}), \text{ for some } r.
\]
All Pareto optimal points lie on the boundary of the feasible criterion space (see Chen et al. [44]). Often, algorithms provide solutions that may not be Pareto optimal, but may satisfy other criteria, making them significant for practical applications. For instance, weakly Pareto optimality is defined as follows:

Definition 1.85 A feasible point \( \bar{x} \), is said to be weakly Pareto optimal (weakly Pareto efficient) solution for (MOP), if there does not exist another feasible point \( x \), such that
\[
f(x) < f(\bar{x}),
\]
or equivalently,
\[
f_i(x) < f_i(\bar{x}), \forall i = 1, 2, ..., k.
\]
A point is weakly Pareto optimal if there is no other point that improves all of the objective functions, simultaneously. In contrast, a point is Pareto optimal if there is no other point that improves at least one objective function without detriment to another function. Pareto optimal points are weakly Pareto optimal, but weakly Pareto optimal points are not Pareto optimal. All Pareto optimal points may be categorized as being either proper or improper.

Kuhn and Tucker [151] noticed that some of the Pareto optimal solutions had some undesirable properties. To avoid such properties they divided the class of Pareto optimal solutions into properly and improperly Pareto optimal solutions. Proper Pareto optimal solutions are those Pareto optimal solutions, that do not allow the unbounded trade-offs between the objectives. Although there are several definitions of proper Pareto optimal solutions, we present the easiest one by Geoffrion [86].

Definition 1.86 (Proper Pareto optimality) A feasible solution \( \bar{x} \) is said to be proper Pareto optimal (properly efficient) solution of (MOP), if it is efficient and if there exists a scalar \( M > 0 \), such that for each \( i \),
\[
\frac{f_i(x) - f_i(\bar{x})}{f_r(x) - f_r(\bar{x})} \leq M,
\]
for some \( r \), such that
\[
f_r(x) > f_r(\bar{x}), \text{ whenever } x \in S \text{ with } f_i(x) < f_i(\bar{x}).
\]
If a Pareto optimal point is not proper, it is improper.

The above quotient is referred to as a trade-off, and it represents the increment in objective function \( r \), resulting from a decrement in objective function \( i \). The definition requires that the trade-off between each function and at least one other function be bounded in order for a point to be proper Pareto optimal. In other words, a solution is properly Pareto optimal if there is at least one pair of objectives, for which a finite decrement in one objective is possible only at the expense of some reasonable increment in the other objective.

### 1.11.5 Necessary and Sufficient Conditions for Pareto Optimality

Now, we state the following Fritz John type necessary optimality conditions. The proof may be found in Da Cunha and Polak [59].

**Theorem 1.55 (Fritz John type necessary optimality conditions)** Let \( f : X \to \mathbb{R}^k \) and \( g : X \to \mathbb{R}^m \) be continuously differentiable functions at \( \bar{x} \in X \). Then a necessary condition for \( \bar{x} \) to be an efficient solution for (MOP) is that there exist multipliers \( \lambda \in \mathbb{R}^k \) and \( \mu \in \mathbb{R}^m \), such that

\[
\sum_{i=1}^{k} \lambda_i \nabla f_i(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0,
\]

\[
\mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, ..., m,
\]

\[
\lambda \geq 0, \quad \mu \geq 0, \quad (\lambda, \mu) \neq 0.
\]

We know that for scalar optimization problem (SOP), the multiplier \( \lambda \) of the objective function in the Karush-Kuhn-Tucker type optimality conditions is assumed to be positive rather than being nonnegative. Likewise, in the case of multiobjective optimization problems (MOP), we further need some constraint qualification, so that all the multipliers of the objective functions are not equal to zero.

Next, we present the Kuhn-Tucker constraint qualification.

**Definition 1.87 (Kuhn-Tucker constraint qualification)** Suppose that the constraint functions \( g_j, j = 1, 2, ..., m \) of the problem (MOP) be continuously differentiable at \( \bar{x} \in S \). The problem (MOP) is said to satisfy the Kuhn-Tucker constraint qualification at \( \bar{x} \), if for any \( d \in \mathbb{R}^n \) such that

\[
\langle \nabla g_j(\bar{x}), d \rangle \leq 0, \quad \forall j \in J(\bar{x}),
\]

there exists a vector function \( \alpha : [0, 1] \to \mathbb{R}^n \), which is continuously differentiable at 0 and some real number \( \gamma > 0 \), such that

\[
\alpha(0) = \bar{x}, \quad g(\alpha(t)) \leq 0, \quad \forall t \in [0, 1] \quad \text{and} \quad \alpha'(t) = \gamma d.
\]
Next, we present the Karush-Kuhn-Tucker necessary optimality conditions:

**Theorem 1.56 (Karush-Kuhn-Tucker necessary condition for Pareto optimality)** Let \( f : X \to \mathbb{R}^k \) and \( g : X \to \mathbb{R}^m \) be continuously differentiable functions at \( \bar{x} \in X \). Suppose that the Kuhn-Tucker constraint qualification is satisfied at \( \bar{x} \). Then, a necessary condition for \( \bar{x} \) to be an efficient solution for (MOP) is that there exist multipliers \( \lambda \in \mathbb{R}^k \) and \( \mu \in \mathbb{R}^m \), such that

\[
\sum_{i=1}^{k} \lambda_i \nabla f_i (\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j (\bar{x}) = 0, \tag{1.27}
\]

\[
\mu_j g_j (\bar{x}) = 0, \quad j = 1, 2, \ldots, m, \tag{1.28}
\]

\[
\lambda \geq 0, \quad \lambda \neq 0, \quad \mu \geq 0.
\]

Proof Suppose that \( \bar{x} \) is an efficient solution for (MOP). Consider the following system

\[
\begin{align*}
\langle \nabla f_i (\bar{x}), d \rangle &< 0, \quad \forall i = 1, 2, \ldots, k \\
\langle \nabla g_j (\bar{x}), d \rangle &\leq 0, \quad \forall j \in J (\bar{x}).
\end{align*}
\tag{1.29}
\]

First, we prove that the system (1.30) has no solution \( d \in \mathbb{R}^n \). On the contrary suppose that \( d \in \mathbb{R}^n \) solves the above system. Then from the Kuhn-Tucker constraint qualification, we know that there exists a function \( \alpha : [0, 1] \to \mathbb{R}^n \) which is continuously differentiable at 0 and some real number \( \gamma > 0 \) such that

\[
\alpha (0) = \bar{x}, \quad g (\alpha (t)) \leq 0, \forall t \in [0, 1] \text{ and } \alpha' (t) = \gamma d.
\]

Since, the functions \( f_i, i = 1, 2, \ldots, k \) are continuously differentiable, we can approximate \( f_i (\alpha (t)) \) linearly as

\[
\begin{align*}
f_i (\alpha (t)) &= f_i (\bar{x}) + \langle \nabla f_i (\bar{x}), \alpha (t) - \bar{x} \rangle + \| \alpha (t) - \bar{x} \| \varphi (\alpha (t), \bar{x}) \\
&= f_i (\bar{x}) + \langle \nabla f_i (\bar{x}), \alpha (t) - \alpha (0) \rangle + \| \alpha (t) - \alpha (0) \| \varphi (\alpha (t), \alpha (0)) \\
&= f_i (\bar{x}) + t \left( \frac{\alpha (0 + t) - \alpha (0)}{t} \right) \| \alpha (t) - \alpha (0) \| \varphi (\alpha (t), \alpha (0)),
\end{align*}
\]

where, \( \varphi (\alpha (t), \alpha (0)) \to 0 \) as \( \| \alpha (t) - \alpha (0) \| \to 0 \). As \( t \to 0 \), \( \| \alpha (t) - \alpha (0) \| \to 0 \) and

\[
\frac{\alpha (0 + t) - \alpha (0)}{t} \to \alpha' (0) = \gamma d.
\]

Using the assumption \( \langle \nabla f_i (\bar{x}), d \rangle < 0 \), for all \( i = 1, 2, \ldots, k \) and \( t \geq 0 \), we have \( f_i (\alpha (t)) < f_i (\bar{x}) \), for all \( i = 1, 2, \ldots, k \) and for sufficiently small \( t \). This contradicts the Pareto optimality of \( \bar{x} \). Hence, the system (1.30) has no solution \( d \in \mathbb{R}^n \). Therefore, invoking the Motzkin’s theorem (Theorem 1.26), there exist multipliers \( \lambda_i \geq 0, \quad i = 1, 2, \ldots, k, \quad \lambda \neq 0 \) and \( \mu_j \geq 0, \quad j \in J (\bar{x}) \) such that

\[
\sum_{i=1}^{k} \lambda_i \nabla f_i (\bar{x}) + \sum_{j \in J (\bar{x})} \mu_j \nabla g_j (\bar{x}) = 0,
\]
Setting $\mu_j = 0, \forall j \in \{1, 2, ..., m\} \setminus J(\bar{x})$, the equality (1.28) of the theorem follows.

If $g_j(\bar{x}) < 0$, for some $j = 1, 2, ..., m$, then by the above setting $\mu_j = 0$, the equality (1.29) holds.

When objective and constraint functions are convex, we have the following Karush-Kuhn-Tucker type sufficient optimality conditions (see, Miettinen [184]).

**Theorem 1.57 (Karush-Kuhn-Tucker sufficient conditions for Pareto optimality)** Let $\bar{x}$ be a feasible solution for the (MOP) and let $f_i : X \to \mathbb{R}, i = 1, 2, ..., k$ and $g_j : X \to \mathbb{R}, j = 1, 2, ..., m$ be continuously differentiable and convex function at $\bar{x}$. If there exist multipliers $0 < \lambda \in \mathbb{R}^k$ and $0 \leq \mu \in \mathbb{R}^m$, such that

$$\sum_{i=1}^{k} \lambda_i \nabla f_i (\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j (\bar{x}) = 0,$$

(1.30)

$$\mu_j g_j (\bar{x}) = 0, \ j = 1, 2, ..., m,$$

(1.31)

then, $\bar{x}$ is an efficient solution for the (MOP).

Proof Suppose there exists vectors $\lambda$ and $\mu$ such that the conditions of the theorem are satisfied. Define a function $F : \mathbb{R}^n \to \mathbb{R}$ as $F(x) = \sum_{i=1}^{k} \lambda_i f_i (x)$, where $x \in S$. Since, each $f_i, i = 1, 2, ..., k$ is convex at $\bar{x}$, and $\lambda > 0$, therefore, the function $F$ is also convex at $\bar{x}$.

Now, from (1.31) and (1.32), we get

$$F(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j (\bar{x}) = 0,$$

$$\mu_j g_j (\bar{x}) = 0, \ j = 1, 2, ..., m.$$

Therefore, by Theorem 1.47, it follows that $F$ attains its minimum at $\bar{x}$. Therefore, we get

$$F(\bar{x}) \leq F(x), \forall x \in S.$$

In other words,

$$\sum_{i=1}^{k} \lambda_i f_i (\bar{x}) \leq \sum_{i=1}^{k} \lambda_i f_i (x), \forall x \in S.$$  

(1.32)

On the contrary, suppose that $\bar{x}$ is not Pareto optimal. Then, there exists some point $x \in S$, such that

$$f_i(x) \leq f_i(\bar{x}), \forall i = 1, 2, ..., k, i \neq r,$$

$$f_r(x) < f_r(\bar{x}), \text{ for some } r.$$
Since, each $\lambda_i > 0$, $i = 1, 2, ..., k$, it results that
\[
\sum_{i=1}^{k} \lambda_i f_i(x) \leq \sum_{i=1}^{k} \lambda_i f_i(\bar{x}),
\]
which is a contradiction to (1.33). Hence $\bar{x}$ is an efficient solution for the (MOP).

In optimization theory, the notion of convexity is just a convenient sufficient condition. In fact, most of the time it is not necessary and it is a rather rigid assumption, often not satisfied in real-world applications. In many cases, nonconvex functions provide a more accurate representation of reality. For instance, their presence may ensure that usual first order necessary optimality conditions are also sufficient or that a local minimum is also a global one. This led to the introduction of several generalizations of the classical notion of convexity. See Mishra and Giorgi [189], Mishra et al. [202, 203], and the references cited therein.

The following theorem shows that the necessary Karush-Kuhn-Tucker optimality conditions become sufficient under suitable generalized convexity assumptions.

**Theorem 1.58 (Miettinen [184])** Let $\bar{x}$ be a feasible point for the (MOP). Suppose that $f_i$, $i = 1, 2, ..., p$ be pseudoconvex at $\bar{x}$ and $g_j$, $j = 1, 2, ..., m$ be quasiconvex at $\bar{x}$. If there exist $\lambda_i \in \mathbb{R}$, $i = 1, 2, ..., p$ and $\mu_j \in \mathbb{R}$, $j = 1, 2, ..., m$, such that
\[
\sum_{i=1}^{p} \lambda_i \nabla f_i (\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j (\bar{x}) = 0,
\]
\[
\mu_j g_j (\bar{x}) = 0, \quad j = 1, 2, ..., m,
\]
\[
\lambda \geq 0, \quad \sum_{i=1}^{p} \lambda_i = 1, \quad \mu \geq 0.
\]
Then, $\bar{x}$ is an optimal solution of (MOP).

### 1.11.6 Nondifferentiable Optimality Conditions

Here, we present necessary and sufficient conditions for Pareto optimality, if the objective and constraint functions of the problem (MOP) are not necessarily differentiable. We assume that the functions are locally Lipschitz and use the Clarke subdifferential and their properties to provide optimality conditions.

First, we state the following Fritz John necessary conditions for Pareto optimality. The proof may be found in Miettinen [184].
Theorem 1.59 Let \( \bar{x} \) be a feasible solution for the (MOP). Assume that the functions \( f_i : X \to \mathbb{R}, i = 1, 2, \ldots, k \) and \( g_j : X \to \mathbb{R}, j = 1, 2, \ldots, m \) be locally Lipschitz at \( \bar{x} \). Then \( \bar{x} \) is a Pareto optimal solution for (MOP) if there exists multipliers \( \lambda \in \mathbb{R}^p \) and \( \mu \in \mathbb{R}^m \), such that

\[
0 \in \sum_{i=1}^{k} \lambda_i \partial f_i(\bar{x}) + \sum_{j=1}^{m} \mu_j \partial g_j(\bar{x}) ,
\]

\[
\mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, \ldots, m,
\]

\[
\lambda \geq 0, \quad \mu \geq 0, \quad (\lambda, \mu) \neq 0.
\]

Remark 1.14 For \( k = 1 \), the above Fritz John optimality condition for (MOP) reduces to that for (SOP). See for example, Clarke [49] or Kiwiel [143].

To move from the Fritz John to the Karush-Kuhn-Tucker optimality condition, we need constraint qualifications. The constraint qualifications for non-differentiable case are different from those for differentiable case. Frequently used constraint qualifications for nondifferentiable (MOP)s are Calmness, Mangasarian Fromovitz, and Cottle constraint qualifications. For details about these and other constraint qualifications, we refer to Dolezal [69], Ishizuka and Shimijhu [109].

Here, we present the Cottle constraint qualification.

Definition 1.88 (Cottle constraint qualification) Suppose that the functions \( f_i : X \to \mathbb{R}, i = 1, 2, \ldots, k \) and \( g_j : X \to \mathbb{R}, j = 1, 2, \ldots, m \) are locally Lipschitz at \( \bar{x} \). Then, (MOP) is said to satisfy the Cottle constraint qualification at \( \bar{x} \), if either

\[
g_j(\bar{x}) < 0, \quad j = 1, 2, \ldots, m
\]

or

\[
0 \in \operatorname{co} \{ \partial g_j(\bar{x}) : g_j(\bar{x}) = 0 \}.
\]

Under Cottle constraint qualification and convexity assumptions on the functions, the following Karush-Kuhn-Tucker necessary and sufficient optimality conditions hold for nondifferentiable (MOP):

Theorem 1.60 (Miettinen [184]) Let \( \bar{x} \) be a feasible solution for the (MOP) and the Cottle constraint qualification be satisfied at \( \bar{x} \). Let \( f_i : X \to \mathbb{R}, i = 1, 2, \ldots, k \) and \( g_j : X \to \mathbb{R}, j = 1, 2, \ldots, m \) be locally Lipschitz and convex function at \( \bar{x} \). A necessary and sufficient optimality condition for \( \bar{x} \) to be an efficient solution is that there exist multipliers \( \lambda \in \mathbb{R}^p \) and \( \mu \in \mathbb{R}^m \), such that

\[
0 \in \sum_{i=1}^{k} \lambda_i \partial f_i(\bar{x}) + \sum_{j=1}^{m} \mu_j \partial g_j(\bar{x}) ,
\]

\[
\mu_j g_j(\bar{x}) = 0, \quad j = 1, 2, \ldots, m,
\]

\[
\lambda \geq 0, \quad \sum_{i=1}^{k} \lambda_i = 1, \quad \mu \geq 0.
\]
1.12 Duality

One of the most interesting, useful and fundamental aspects of linear as well as nonlinear programming is duality theory. It provides a theoretical foundation for many optimization algorithms. Duality can be used to directly solve nonlinear programming problems as well as to derive lower bounds in other high-level search techniques of the solution quality. In constrained optimization, it is often used in a number of constraint decomposition schemes such as separable programming and in space decomposition algorithms such as branch and bound.

Wolfe [285] used the Karush-Kuhn-Tucker conditions to formulate a dual program for a nonlinear optimization problem in the spirit of duality in linear programming, that is, with the aim of defining a problem whose objective value gives lower bound on the optimal value of the original or primal problem and whose optimal solution yields an optimal solution for the primal problem under certain regularity conditions. See Mishra et al. [203]. For (MOP), the Wolfe dual model (WD) may be formulated as follows:

\[
(WD) \quad \max f(u) + \sum_{j=1}^{m} \mu_j g_j(u)
\]

subject to

\[
\sum_{i=1}^{p} \lambda_i \nabla f_i(u) + \sum_{j=1}^{m} \mu_j \nabla g_j(u) = 0,
\]

\[
(\lambda, \mu) \geq 0, u \in X \subseteq \mathbb{R}^n.
\]

Mangasarian [176] has pointed out that whereas some results such as sufficiency and converse duality hold for (P) and Wolfe dual (WD), if \( f \) is only pseudoconvex and \( g \) quasiconvex. However, by means of a counterexample, Mangasarian [176] also showed that weak and strong duality theorems do not hold for these functions for Wolfe dual (WD). We present an example from Mond [206]:

\[
(P1) \quad \min x + x^3
\]

subject to \( x \geq 1 \).

It is easy to show that \( x = 1 \) is the optimal solution and the optimal value is 2.

The Wolfe dual for the problem (P1) may be formulated as

\[
(WD1) \quad \max u + u^3 + \mu (1 - u)
\]

subject to \( 3u^2 + 1 - \mu \geq 0, \mu \geq 0 \).

It is obvious that the objective function of (WD1) tends to \(+\infty\), when \( u \to -\infty \).
One of the reasons that in Wolfe duality the convexity cannot be weakened to pseudoconvexity is that unlike for convex functions, the sum of two pseudoconvex functions is not pseudoconvex. However, Wolfe duality holds if the Lagrangian $f(u) + \mu^T g(u), \mu \geq 0$ is pseudoconvex.

In order to weaken the convexity assumption, Mond and Weir [207] proposed a new type of dual based on the Wolfe dual (WD). Mond-Weir type dual (MWD) to the above nonlinear multiobjective optimization problem (MOP) is given by:

\[
\text{(MWD) max } f(u) \\
\text{subject to } \sum_{i=1}^{p} \lambda_i \nabla f_i(u) + \sum_{j=1}^{m} \mu_j \nabla g_j(u) = 0, \\
\mu_j g_j(u) \geq 0, \ j = 1, ..., m, \\
(\lambda, \mu) \geq 0, u \in X \subseteq \mathbb{R}^n.
\]

In a nonsmooth setting, using the Clarke subdifferential, the Mond-Weir dual may be formulated as follows:

\[
\text{(MWD) max } f(u) \\
\text{subject to } \sum_{i=1}^{p} \lambda_i \partial^c f_i(u) + \sum_{j=1}^{m} \mu_j \partial^c g_j(u) = 0, \\
\mu_j g_j(u) \geq 0, \ j = 1, ..., m, \\
(\lambda, \mu) \geq 0, u \in X \subseteq \mathbb{R}^n.
\]

For further details about different types of duals and duality theory, we refer to Craven [52], Craven and Glover [55], Eguido and Mond [73], Preda [223], Mishra et al. [202, 203], and the references therein.
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


[230] Pripoae, C.L., Pripoae, G.T., Preda, V.: $\eta$-Pseudolinearity on differentiable manifolds (2013)


Bibliography


