Springer Optimization and Its Applications 64
Nonconvex Optimization and Its Applications

Vladimir Shikhman

Topological Aspects of Nonsmooth Optimization

Springer

Springer Optimization and Its Applications 64

Series Editor:
Panos M. Pardalos

Subseries:
Nonconvex Optimization and Its Applications

For further volumes:
http://www.springer.com/series/7393

Vladimir Shikhman

Topological Aspects of Nonsmooth Optimization

Springer

Vladimir Shikhman
RWTH Aachen University
Mathematics, Section C
Templergraben 55
52056 Aachen, Germany
shikhman@mathc.rwth-aachen.de

ISSN 1931-6828
ISBN 978-1-4614-1896-2 e-ISBN 978-1-4614-1897-9
DOI 10.1007/978-1-4614-1897-9
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011940294

Mathematics Subject Classification (2010): 90C30, 90C31, 90C33, 90C34, 90C26, 57R45, 58K05, 49J52

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

for my grandmother Ina Kapilevich

Preface

The main goal of our study is an attempt to understand and classify nonsmooth structures arising within the optimization setting,

$$P(f,F): \quad \min f(x) \text{ s.t. } x \in M[F],$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a smooth real-valued objective function, $F : \mathbb{R}^n \longrightarrow \mathbb{R}^l$ is a smooth vector-valued function, and $M[F] \subset \mathbb{R}^n$ is a feasible set defined by F in some structured way. The nonsmoothness is given by the structure that fits the smooth function F to define the feasible set $M[F]$. The following optimization problems with particular types of nonsmoothness are considered (Chapters 2–5):

- mathematical programming problems with complementarity constraints,
- general semi-infinite programming problems,
- mathematical programming problems with vanishing constraints,
- bilevel optimization.

The basis of our study is the topological approach introduced in detail in Chapter 1. It encompasses the following questions:

(a) Under which conditions on F is $M[F]$ a Lipschitz manifold of an appropriate dimension?
(b) Under which conditions on F is $M[F]$ stable (i.e., $M[F]$ remains invariant up to a homomorphism w.r.t. smooth perturbations of F)?
(c) How does the homotopy type of lower-level set

$$M[f,F]^a := \{x \in M[F] \,|\, f(x) \le a\}$$

change (as $a \in \mathbb{R}$ varies)?

Questions (a) and (b) deal with topological invariants of $M[F]$ and, more precisely, its structure. They lead to suitable constraint qualifications. Topological changes of $M[f,F]^a$ give rise to defining stationary points and developing critical point theory for $P(f,F)$ in the sense of Morse. In so doing, we get new topologically relevant optimization notions in terms of derivatives of f and F. It is worth pointing out that the same topological questions provide different (analytical) optimization concepts when applied to the particular problems above. The difference between these analytically described optimization concepts is a key point in understanding and comparing different kinds of nonsmoothness.

In Chapter 6, we discuss the impact of the topological approach on nonsmooth analysis. Topologically regular points of a min-type nonsmooth mapping $F : \mathbb{R}^n \longrightarrow \mathbb{R}^l$ are introduced. The crucial property is that for a topologically regular value $y \in \mathbb{R}^l$ of F the nonempty set $F^{-1}(y)$ is an $(n-l)$-dimensional Lipschitz manifold. Corresponding nonsmooth versions of Sard's Theorem are given.

We point out that the topological approach in the optimization context was introduced by H. Th. Jongen in the early 1980s ([61], [62]). The introduction of topological issues turned out to be extremely fruitful for establishing an adequate optimization theory in the smooth setting ([63]). The present book sheds light on nonsmooth optimization from the topological point of view, continuing to exploit the ideas of H. Th. Jongen.

I would like to thank my teacher H. Th. Jongen for sharing with me his insights on optimization and steering my studies toward its topological nature. This book originated mainly from a collaboration with him. I also thank my other coauthors, D. Dorsch, F. Guerra-Vázquez, Jan-J. Rückmann, S. Steffensen, and O. Stein, for fruitful collaborations. I am very grateful to H. Günzel, A. Ioffe, D. Klatte, B. Kummer, B. Mordukhovich, Yu. Nesterov, and D. Pallaschke for many interesting and helpful discussions.

Aachen, April 2011 *Vladimir Shikhman*

Contents

Notation

Our notation is standard. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$ with the norm $\|\cdot\|$, its nonnegative orthant by $\mathbb{H}^n$, and its nonpositive orthant by $\mathbb{R}^n_-$. $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x > 0\}$. For $\varepsilon > 0$ and $\bar{x} \in \mathbb{R}^n$, the set $B_\varepsilon(\bar{x})$ (or $B(\bar{x}, \varepsilon)$) stands for the open Euclidean ball in $\mathbb{R}^n$ with radius ε and center $\bar{x}$. A closed ball with radius $\varepsilon > 0$ and center $\bar{x} \in \mathbb{R}^n$ is denoted by $\bar{B}(\bar{x}, \varepsilon)$.

Given an arbitrary set $K \subset \mathbb{R}^n$, $\overline{K}$, $\operatorname{int}(K)$, and ∂K denote the topological closure, interior, and boundary of K, respectively. $\operatorname{span}(K)$, $\operatorname{conv}(K)$ (or $\operatorname{co}(K)$), and $\operatorname{cone}(K)$ denote the set of all linear, convex, and nonnegative combinations of elements of K, respectively. CK denotes the complement of $K \subset \mathbb{R}^n$. By $\operatorname{span}\{a_1, \ldots, a_t\}$ we denote the vector space over $\mathbb{R}$ generated by the finite number of vectors $a_1, \ldots, a_t \in \mathbb{R}^n$, and $\dim\{\operatorname{span}\{a_1, \ldots, a_t\}\}$ stands for its dimension. The polar of K is defined by $K^\circ := \{v \in \mathbb{R}^n \,|\, v^T w \leq 0 \text{ for all } w \in K\}$. The distance from $x \in \mathbb{R}^n$ to $K \subset \mathbb{R}^n$ is denoted by $\operatorname{dist}(x, K) = \inf\limits_{y \in K} \|x - y\|$ with the convention $\operatorname{dist}(x, \emptyset) = \infty$.

$T : \mathbb{R}^n \rightrightarrows \mathbb{R}^k$ denotes a multivalued map defined on $\mathbb{R}^n$ with $T(x) \subset \mathbb{R}^k$, $x \in \mathbb{R}^n$. The graph of T is $\operatorname{gph} T = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^k \,|\, y \in T(x)\}$, and the inverse of T is $T^{-1} : \mathbb{R}^k \rightrightarrows \mathbb{R}^n$, given by $T^{-1}(y) = \{x \in \mathbb{R}^n \,|\, y \in T(x)\}$.

Given a differentiable function $F : \mathbb{R}^n \longrightarrow \mathbb{R}^k$, DF denotes its $k \times n$ Jacobian matrix. Given a differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, Df denotes its gradient as a row vector, and $D^T f$ (or ∇f) stands for the transposed vector. Given a twice continuously differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $D^2 f$ stands for its Hessian. $C^l(\mathbb{R}^n, \mathbb{R}^k)$ denotes the space of l-times continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}^k$. $C^\infty(\mathbb{R}^n, \mathbb{R}^k)$ denotes the space of smooth functions from $\mathbb{R}^n$ to $\mathbb{R}^k$. $C^l(\mathbb{R}^n)$ stands for $C^l(\mathbb{R}^n, \mathbb{R})$, and $C^\infty(\mathbb{R}^n)$ stands for $C^\infty(\mathbb{R}^n, \mathbb{R})$.

Chapter 1
Introduction

We state mathematical programming problems with complementarity constraints, general semi-infinite programming problems, mathematical programming problems with vanishing constraints and bilevel optimization. The topological approach for studying problems above is introduced. It encompasses the study of topological properties of corresponding feasible sets, as well as the critical point theory in the sense of Morse. Finally, we describe the application of the topological approach for standard nonlinear programming problems.

1.1 Nonsmooth optimization framework

We consider the nonsmooth optimization framework

$$P(f,F): \quad \min f(x) \text{ s.t. } x \in M[F], \tag{1.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a real-valued objective function, $F : \mathbb{R}^n \longrightarrow \mathbb{R}^l$ is a vector-valued function, and $M[F] \subset \mathbb{R}^n$ is a feasible set defined by F in some structured way.

Within this general framework, the nonsmoothness might be caused by

(a) the objective function f,
(b) the defining function F, or
(c) the structure according to which F defines $M[F]$.

Here, we assume functions f, F to be sufficiently smooth, and we restrict our study to the nonsmoothness given by (c). Thus, we focus rather on the underlying nonsmooth structures that fit the smooth function F to define the feasible set $M[F]$. We give some examples of particular optimization problems of type (1.1) to illustrate possible nonsmooth structures.

Example 1 (MPCC). The mathematical programming problem with complementarity constraints (MPCC) is defined as

V. Shikhman, *Topological Aspects of Nonsmooth Optimization*, Springer Optimization and Its Applications 64, DOI 10.1007/978-1-4614-1897-9_1,

$$\text{MPCC:} \quad \min f(x) \text{ s.t. } x \in M[h,g,F_1,F_2]$$

with

$$\begin{aligned} M[h,g,F_1,F_2] := \{x \in \mathbb{R}^n \mid\ & F_{1,m}(x) \geq 0, F_{2,m}(x) \geq 0, \\ & F_{1,m}(x)F_{2,m}(x) = 0,\ m = 1,\dots,k, \\ & h_i(x) = 0,\ i \in I,\ g_j(x) \geq 0,\ j \in J\}, \end{aligned}$$

where f, h_i, $i \in I$, g_j, $j \in J$, $F_{1,i}$, $F_{2,i}$, $i = 1,\dots,k$ are real-valued and smooth functions, $|I| \leq n$, $|J| < \infty$.

Here, the nonsmoothness comes into play due to the complementarity constraints:

$$F_{1,m}(x) \geq 0, F_{2,m}(x) \geq 0,\ F_{1,m}(x)F_{2,m}(x) = 0,\ m = 1,\dots,k.$$

Indeed, the basic complementarity relation

$$u \geq 0,\ v \geq 0,\ u \cdot v = 0,$$

defines the boundary of the nonnegative orthant in $\mathbb{R}^2$.

Example 2 (GSIP). Generalized semi-infinite programming problems (GSIPs) have the form

$$\text{GSIP:} \quad \text{minimize } f(x) \text{ s.t. } x \in M$$

with

$$M := \{x \in \mathbb{R}^n \mid g_0(x,y) \geq 0 \text{ for all } y \in Y(x)\}$$

and

$$Y(x) := \{y \in \mathbb{R}^m \mid g_k(x,y) \leq 0,\ k = 1,\dots,s\}.$$

All defining functions $f, g_k, k = 0,\dots,s$, are assumed to be real-valued and smooth on their respective domains.

Note that testing feasibility for x means that $\inf\limits_{y \in Y(x)} g_0(x,y) \geq 0$. The appearance of the optimal value function $\inf\limits_{y \in Y(x)} g_0(x,y)$ causes nonsmoothness.

Example 3 (MPVC). We consider the mathematical programming problem with vanishing constraints (MPVC)

$$\text{MPVC:} \quad \min f(x) \text{ s.t. } x \in M[h,g,H,G]$$

with

$$\begin{aligned} M[h,g,H,G] := \{x \in \mathbb{R}^n \mid\ & H_m(x) \geq 0,\ H_m(x)G_m(x) \leq 0,\ m = 1,\dots,k, \\ & h_i(x) = 0,\ i \in I,\ g_j(x) \geq 0,\ j \in J\}, \end{aligned}$$

where f, h_i, $i \in I$, g_j, $j \in J$, H_m, G_m, $m = 1,\dots,k$ are real-valued and smooth functions, $|I| \leq n$, $|J| < \infty$.

Here, the difficulty is due to the vanishing constraints:

$$H_m(x) \geq 0, H_m(x)G_m(x) \leq 0, m = 1, \ldots, k.$$

Note that for those x with $H_m(x) = 0$ the sign of $G_m(x)$ is not restricted.

Example 4 (Bilevel optimization). We consider bilevel optimization from the optimistic point of view

$$U: \quad \min_{(x,y)} f(x,y) \quad \text{s.t.} \quad y \in \text{Argmin}\, L(x),$$

where

$$L(x): \quad \min_{y} g(x,y) \quad \text{s.t.} \quad h_j(x,y) \geq 0, \ j \in J.$$

Above we have $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and the real-valued mappings $f, g, h_j, j \in J$ are smooth, $|J| < \infty$. Argmin $L(x)$ denotes the solution set of the optimization problem $L(x)$.

Here, the nonsmoothness comes from the fact that we deal with a parametric nonlinear programming problem $L(x)$ at the lower level. Moreover, to ensure feasibility for (x,y) at the upper level U, the problem $L(x)$ should be solved up to global optimality.

1.2 Topological approach

The main goal of our study is an attempt to understand and classify nonsmooth structures arising in (1.1) within the optimization setting. The basis of such a comparison is the topological approach. It encompasses two objects of study:

the feasible set $M[F]$

and

the lower-level sets $M[f,F]^a := \{x \in M[F] \,|\, f(x) \leq a\}, a \in \mathbb{R}$.

These objects are considered along the levels of study due to topology, optimization and stability issues as outlined in the following scheme (see Figure 1).

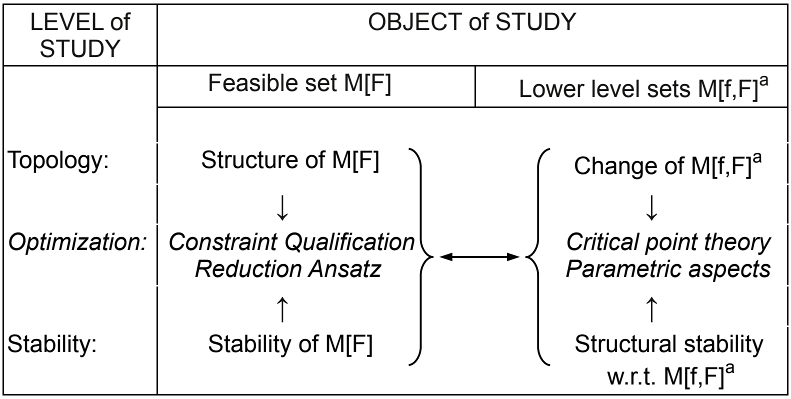

Figure 1 Topological approach

On the topology and stability levels we deal with topological invariants of $M[F]$ and $M[f,F]^a$, $a \in \mathbb{R}$. The questions mainly arise from here. They lead to establishment of an adequate theory on the optimization level. It is worth pointing out that the same topological questions provide different (analytical) optimization concepts when applied to particular problems (e.g., MPCC, GSIP, MPVC, and bilevel optimization). The difference between these analytically described optimization concepts is a key point in understanding and comparing different kinds of nonsmoothness. In what follows, we introduce the notions from the scheme in detail.

For the **structure of** $M[F]$, it is crucial to study under which conditions on F the feasible set is a **topological** or **Lipschitz manifold** (with boundary) of an appropriate dimension.

Definition 1 (Topological and Lipschitz manifolds [103]). A subset $\mathscr{M} \subseteq \mathbb{R}^n$ is called a topological (resp. Lipschitz) manifold (with boundary) of dimension $m \geq 0$ if for each $\bar{x} \in \mathscr{M}$ there exist open neighborhoods $U \subseteq \mathbb{R}^n$ of $\bar{x}$ and $V \subseteq \mathbb{R}^n$ of 0 and a homeomorphism $H : U \to V$ (resp. with H, H^{-1} being Lipschitz continuous) such that

(i) $H(\bar{x}) = 0$

and

(ii) either in the first case

$$H(\mathscr{M} \cap U) = (\mathbb{R}^m \times \{0_{n-m}\}) \cap V$$

or in the second case

$$H(\mathscr{M} \cap U) = (\mathbb{H} \times \mathbb{R}^{m-1} \times \{0_{n-m}\}) \cap V$$

occur.

If for all $x \in \mathscr{M}$ the first case in (ii) holds, then $\mathscr{M}$ is called a topological (resp. Lipschitz) manifold of dimension m. In the second case, $\bar{x}$ is said to be a boundary point of $\mathscr{M}$ (see Figure 2).

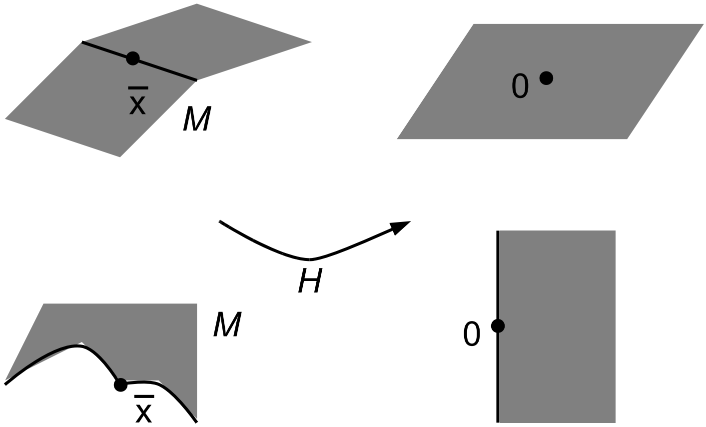

Figure 2 First and second cases for Lipschitz manifold

We shall use the **tools of nonsmooth and variational analysis** to tackle the question of $M[F]$ being a Lipschitz manifold. In particular, the application of nonsmooth versions of the **implicit function theorem** (see Section B.1) plays a major role.

Another issue for the structure of $M[F]$ is the **(topological) stability** of the feasible set under smooth perturbations of F (see Figure 3).

Definition 2 (Topological stability). The feasible set $M[F]$ from (1.1) is called (topologically) stable if there exists a C^1-neighborhood U of F in $C^1(\mathbb{R}^n, \mathbb{R}^l)$ (w.r.t. the strong or Whitney topology; see [42, 63], and Sections 1.3 and A.2 of the present volume) such that, for every $\widetilde{F} \in U$, the corresponding feasible set $M[\widetilde{F}]$ is homeomorphic with $M[F]$.

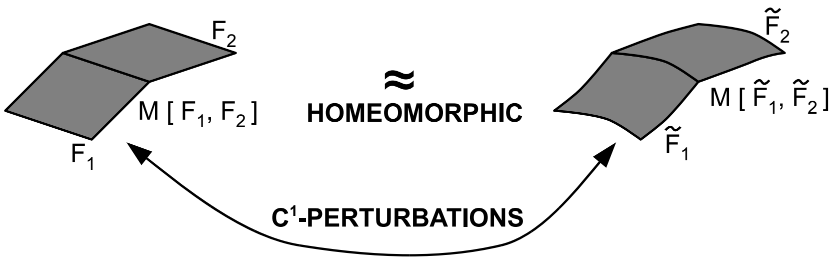

Figure 3 Topological stability

The stability of the feasible set is tightly connected with its Lipschitz manifold property. Addressing both of them will immediately lead us to suitable **constraint qualifications** for $M[F]$.

Actually, the list of topological invariants for $M[F]$ that is worth studying usually depends on particular problem realization. For example, having in mind GSIPs and bilevel optimization, an important issue for the description of the feasible set $M[F]$ becomes the so-called **reduction ansatz**. It deals with possibly infinite index sets that can be equivalently reduced to their finite subsets, at least at stationary points. Moreover, the feasible set in GSIPs need not be closed in general. This fact leads to the topological study of its closure instead. Next, the MPVC feasible set is not a Lipschitz manifold but a set glued together from manifold pieces of different dimensions along their strata.

Regarding the **behavior of the lower-level sets** $M[f,F]^a$, we study changes of their topological properties as $a \in \mathbb{R}$ varies. The smooth (un-) constrained case refers to the classical Morse theory and is well-known (see [63, 93]). We illustrate it in Figure 4.

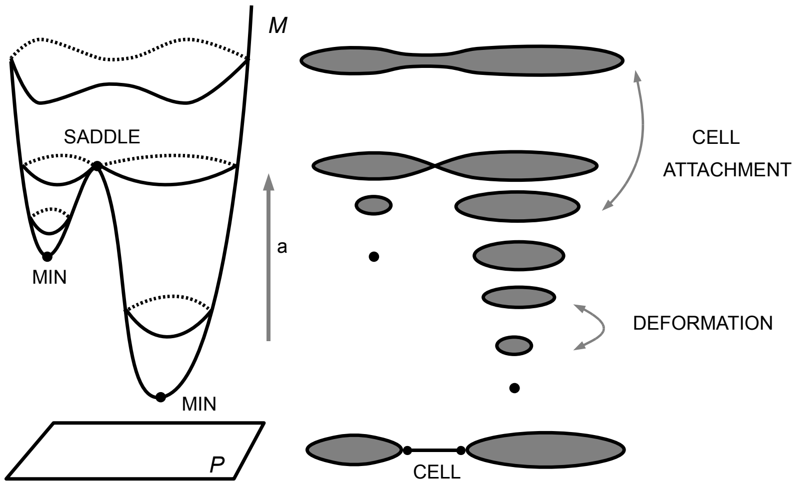

Figure 4 Deformation and cell attachment

Here, f is the height function from the plane P to the smooth manifold $M \subset \mathbb{R}^3$. Clearly, f has two local minima and one saddle point. We see that the topological changes of $M^a := \left\{x \in \mathbb{R}^2 \,|\, f(x) \le a\right\}$, $a \in \mathbb{R}$ happen only when passing these three critical values. More precisely, new components of M^a are created passing local minima and, in addition, two components are attached together passing the saddle point. Note that the dimension of the cell attached corresponds to the number of negative eigenvalues of the Hessian of f.

Coming to the nonsmooth case, an adequate stationarity concept of **(topologically) stationary points** will be introduced. The analytical description of this concept depends certainly on a particular realization of (1.1). The definition of stationary points will be given in **dual terms** using Lagrange multipliers. Additionally, it will be shown that local minimizers are stationary points under some suitable constraint qualifications.

Within this context, two basic theorems from **Morse theory** (see [63, 93] and Section A.1) are crucial.

Theorem 1 (Deformation theorem). *If for $a < b$ the (compact) set $M[f,F]_a^b := \{x \in M \,|\, a \leq f(x) \leq b\}$ does not contain stationary points, then the set $M[f,F]^a$ is a strong deformation retract of $M[f,F]^b$.*

As a consequence, the homotopy types of the lower-level sets $M[f,F]^a$ and $M[f,F]^b$ are equal. This means that the connectedness structure of the lower-level sets does not change when passing from level a to level b. In particular, the number of connected (path) components remains invariant.

For the second result, a notion of a nondegenerate stationary point, along with its index, will be introduced. Note that a nondegenerate stationary point is a local minimizer if and only if its index vanishes.

Theorem 2 (Cell-attachment theorem). *If $M[f,F]_a^b$ contains exactly one nondegenerate stationary point, then $M[f,F]^b$ is homotopy-equivalent to $M[f,F]^a$ with a q-cell attached. Here, the dimension q is the so-called index of the nondegenerate stationary point.*

The latter two theorems on homotopy equivalence show that **Morse relations**, such as Morse inequalities (see [63]), are valid. Roughly speaking, Morse relations relate the existence of stationary points of various indices with the topology of the feasible set. In fact, the cell attachment of a k-dimensional cell can either generate a hole or cancel it (see Figure 5).

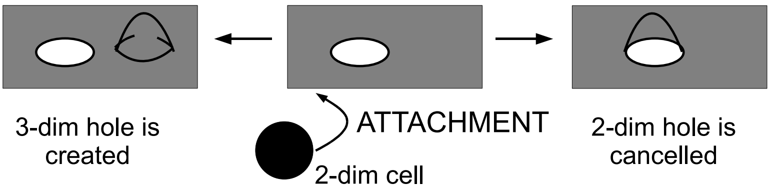

Figure 5 Generation or cancellation of holes

A **global interpretation** of the deformation and cell-attachment theorems is the following. Suppose that the feasible set is compact and connected and that all stationary points are nondegenerate with pairwise different functional values. Then, passing a level corresponding to a local minimizer, a connected component of the lower-level set is created. Different components can only be connected by attaching 1-cells. This shows the existence of at least $(k-1)$ stationary points with index equal to 1, where k is the number of local minimizers; see also [26, 63]. This issue is closely related to the global aspects of optimization theory, in particular to the existence of $0-1-0$ and $0-n-0$ graphs. The latter connect local minimizers with stationary points having index equal to 1 and the former with local maximizers [63]. Finally, we refer the reader to [2, 6, 92] for the results with Morse theory for piecewise smooth functions.

The **structural stability w.r.t. lower-level sets** is defined via special equivalence relation on $P(f,F)$ as follows.