

NONSMOOTH ANALYSIS ON SMOOTH MANIFOLDS

YU. S. LEDYAEV AND QIJI J. ZHU

ABSTRACT. We study infinitesimal properties of nonsmooth (nondifferentiable) functions on smooth manifolds. The eigenvalue function of a matrix on the manifold of symmetric matrices gives a natural example of such a nonsmooth function.

A subdifferential calculus for lower semicontinuous functions is developed here for studying constrained optimization problems, nonclassical problems of calculus of variations, and generalized solutions of first-order partial differential equations on manifolds. We also establish criteria for monotonicity and invariance of functions and sets with respect to solutions of differential inclusions.

1. INTRODUCTION

The subject of this paper is the study of infinitesimal properties of nondifferentiable functions on smooth manifolds. Such nondifferentiable, or nonsmooth, functions arise naturally in many problems on smooth manifolds.

Consider for example the manifold of $N \times N$ symmetric matrices. The well-known Ky Fan equality (see, for example, [40]) gives the following representation of the k -th largest eigenvalue of a matrix A :

$$\lambda_k(A) = \max_{X \in S(k+1, N)} \operatorname{tr}(X'AX) - \max_{X \in S(k, N)} \operatorname{tr}(X'AX),$$

where the Stiefel manifold

$$S(k, N) := \{X \in R^{N \times k} \mid X'X = I_k\}$$

consists of real orthogonal $N \times k$ matrices.

The importance of eigenvalue functions in applications is well known (see [18, 40]), and it is clear from the above representation that eigenvalue functions are, in general, nonsmooth. Although much research has been conducted on the spectral functions with the tools of nonsmooth analysis [9, 10, 19, 20, 50], the smooth manifold structure of the Stiefel manifold has not been systematically exploited.

The second example is the metric distance on a Riemann manifold defined by

$$\rho(m_1, m_2) := \inf\{T : c(0) = m_1, c(T) = m_2, \|\dot{c}(t)\| \leq 1\}.$$

Received by the editors June 20, 2003 and, in revised form, May 5, 2005.

2000 *Mathematics Subject Classification.* Primary 93D05, 93D20, 34D20.

Key words and phrases. Differential manifolds, nonsmooth analysis, calculus of semicontinuous functions on manifolds, differential inclusions on manifolds, monotonicity and invariance.

The first author was supported in part by NSF grant #0102496 and by the Russian Fund for Fundamental Research Grant # 02-01-00769.

The second author was supported in part by NSF grants #9704203, #0102496 and by the Faculty Research and Creative Activities Support Fund at Western Michigan University.

©2007 American Mathematical Society
 Reverts to public domain 28 years from publication

Here \inf is taken on the set of all differentiable curves connecting points m_1 and m_2 . The function $\varphi(m) := \rho(m_1, m)$ is, in general, not differentiable. In a Euclidean space the distance function can be considered as the unique continuous (generalized) solution of the following “eikonal” equation:

$$(1.1) \quad \|\partial_F \varphi(m)\| = 1, \quad \varphi(m_1) = 0,$$

where $\partial_F \varphi$ is the Fréchet subdifferential of the function φ . This point of view is quite fruitful in Euclidean space when discussing, for example, the nearest points to a set.

The third example concerns control systems on manifolds and stabilization. It is known that, if a control system on a manifold can be stabilized by continuous feedback, then this manifold is isomorphic to R^N [46]. Of course, most interesting control applications require using discontinuous feedback controls [27]. The main tool used for solving stabilization problems in R^n has been the analysis of the corresponding control Lyapunov functions, which are nonsmooth even for simple examples.

These three examples explain the need for a systematic study of tools of nonsmooth analysis on smooth manifolds. This paper is an attempt to establish a framework and some tools for studying semicontinuous functions on smooth manifolds that can be used for analyzing optimization, dynamic optimization, generalized partial differential equations, optimal control and feedback control problems on manifolds.

To begin, one needs a concept of generalized differential to replace the differential of smooth functions on manifolds. Many such concepts and related calculus for lower semicontinuous functions in R^N or Banach spaces have been the subject of intense research in the past several decades. They are important constituents of nonsmooth analysis, which in turn is a part of nonlinear analysis. Research done by Rockafellar, Pshenichnyi, Danskin and Dem’yanov was instrumental in the early studies on nondifferentiable functions and their applications in the optimization theory with emphasis on the convex and quasidifferentiable functions. In early nineteen-seventies, Clarke first introduced the concept of a generalized gradient for locally Lipschitz functions. He developed corresponding calculus and applied the generalized gradient in optimization and optimal control problems. Clarke also developed many techniques such as the scalarization of the problem, application of the perturbation and penalization techniques and variational analysis, which soon became a new paradigm in this field. This paradigm has been extremely effective in the analysis of a wide range of problems in nonsmooth nontraditional optimization and calculus of variations due to important and significant contributions made by Aubin, Borwein, Hirriart-Uruty, Ioffe, Mordukhovich, Penot, Rockafellar, Sussmann, Warga and others. The books and articles [2, 14, 25, 26, 29, 42, 45, 49, 56, 58, 75, 76] provide powerful tools and present interesting perspectives to the field of nonsmooth and variational analysis. They are the foundation of our current research, and we also recommend them for readers who are interested in the history and the state of the art of nonsmooth and variational analysis.

We shall focus on a natural generalization of the Fréchet subdifferential and its related limiting and singular subdifferentials. The related geometric concept of a normal cone to a set is defined as the subdifferential of the indicator function of the set. The Fréchet subdifferential appeared in [7]. Mordukhovich introduced the

limiting and singular subdifferentials for general lower semicontinuous functions and systematically developed their calculus [53, 54, 57, 56].

We should note here that a convex hull of the Mordukhovich limiting subdifferential coincides with the Clarke generalized gradients. Some elements of this geometric approach have appeared in Clarke's 1973 thesis [23] with its main focus on the normal cone to a set.

Our development follows more recent monographs [29, 14] and the survey [13]. These concepts of generalized subdifferentials fit very well with techniques of variational analysis and can conveniently handle semicontinuous functions. We emphasize that this is not the only possible approach. There are many other approaches and refinements in the rich literature of nonsmooth analysis that may also lead to interesting developments for the study of nonsmooth functions on smooth manifolds. For example, Warga [75, 76] defined a derivate container based on uniform approximation. It was recently refined by Sussmann [67, 68, 69]. These generalized derivative concepts are quite effective in dealing with optimal control and optimization problems with more topological methods. Treiman [72, 73] introduced the linear subdifferential that refined the limiting subdifferential. However, these are beyond the scope of the current paper.

Note that a distance metric function defined above can be considered as an optimal value function of more general optimal control of the minimizing of a functional

$$l(c(T)) + \int_0^T L_0(c(t), \dot{c}(t)) dt$$

with the differential constraints on the curve

$$\dot{c}(t) \in F(c(t))$$

where $F(x)$ is a multifunction. Formally, these dynamical constraints in the form of the differential inclusion can be taken into account if the integrand in the functional is modified by adding the indicator function of the set $F(x)$. Thus, this generalized optimal control problem is reduced to some nonclassical variational problem with a lower semicontinuous integrand which can take infinite values. This formal reduction can be very useful for the derivation of optimality conditions for such a general variational problem in the presence of an appropriate calculus for analysis of such integrands. In fact, this was one of the original motivations of Clarke's pioneering work which was in turn motivated by Rockafellar's results for completely convex nonsmooth calculus of variations problems.

Another aspect of the general optimal control problem in R^N is the fact that its optimal value function is a generalized (viscosity or minimax) solution of the Hamilton-Jacobi equation [31, 32, 66]. It is interesting that the same techniques used in nonsmooth analysis for the derivation of optimality conditions can be used for the proof of uniqueness and existence of the solution of this first-order partial differential equation. How to handle a dynamic system motivated by the above examples is another major issue discussed in this paper.

This paper contains the conceptual framework for the study of generalized derivatives for semicontinuous functions and dynamic systems on smooth manifolds. We also illustrate our methods with examples of applications. To make the paper more focused we selected applications that demonstrate the methods rather than pursue maximum generality. We hope to discuss additional applications in greater detail in subsequent papers.

The paper is arranged as follows: we collect necessary preliminaries and notation regarding smooth manifolds in Section 2; in Section 3 we introduce the concepts of sub- and superdifferentials of lower and upper semicontinuous functions. The geometric concept of a normal to an arbitrary set on a manifold is introduced and is related to the distance function to the given set. Section 4 contains the elements of the calculus for subdifferentials: sum and product rules, chain rules, subgradients of supremum functions, and a multiplier rule for an optimization problem with semicontinuous data.

We discuss several applications of the sub- and superdifferentials and their calculus in Section 5. They are subdifferential characterizations of Lipschitz properties of functions, implicit function theorems and subdifferentials for the spectral functions.

We introduce a dynamic system on smooth manifolds in the form of differential inclusions in Section 6. We then discuss related concepts of monotonicity and invariance of functions and sets with respect to solutions of differential inclusions on manifolds. We use these results to discuss elements of the theory for generalized (viscosity) solutions of Hamilton-Jacobi equations in Section 7. We also discuss applications to the distance function on Riemannian manifolds there.

2. PRELIMINARIES AND NOTATION

In this section we recall some pertinent concepts and results related to a smooth manifold. Our main references are [16, 51, 65].

Let M be an N -dimensional C^∞ complex manifold (paracompact Hausdorff space) with a C^∞ atlas $\{(U_a, \psi_a)\}_{a \in A}$. For each a , the N components (x_a^1, \dots, x_a^N) of ψ_a are called local coordinate systems on (U_a, ψ_a) . A function $g : M \rightarrow R$ is C^r at $m \in M$ if $m \in U_a$, and $g \circ \psi_a^{-1}$ is a C^r function in a neighborhood of $\psi_a(m)$. Here r is a nonnegative integer or ∞ . As usual C^0 represents the collection of continuous functions. It is well known that this definition is independent on the coordinate systems. If g is C^∞ at all $m \in M$, we say g is C^∞ on M . The collection of all $C^\infty(C^r)$ functions on M is denoted by $C^\infty(M)(C^r(M))$. A map $v : C^\infty(M) \rightarrow R$ is called a tangent vector of M at m provided that, for any $f, g \in C^\infty(M)$, (1) $v(\lambda f + \mu g) = \lambda v(f) + \mu v(g)$ for all $\lambda, \mu \in R$ and (2) $v(f \cdot g) = v(f)g(m) + f(m)v(g)$. The collection of all the tangent vectors of M at m form an (N -dimensional) vector space and is denoted by $T_m(M)$. The union $\bigcup_{m \in M}(m, T_m(M))$ forms a new space called the *tangent bundle* to M , denoted by $T(M)$. The dual space of $T_m(M)$ is called the cotangent space of M at m , denoted by $T_m^*(M)$. The *cotangent bundle* to M then is $T^*(M) := \bigcup_{m \in M}(m, T_m^*(M))$. We will use π (π^*) to denote the canonical projection on $T(M)$ ($T^*(M)$) defined by $\pi(m, T_m(M)) = m$ ($\pi^*(m, T_m^*(M)) = m$). A mapping $X : M \rightarrow T(M)$ is called a vector field provided that $\pi(X(m)) = m$. A vector field X is C^r at $m \in M$ provided so is $X(g)$ for any $g \in C^\infty$. If a vector field X is C^r for all $m \in M$ we say it is C^r on M . The collection of all C^r vector fields on M is denoted by $V^r(M)$.

In particular, if (U, ψ) is a local coordinate neighborhood with $m \in U$ and (x^1, \dots, x^N) is the corresponding local coordinate system on (U, ψ) , then $(\frac{\partial}{\partial x^n})_m$, $n = 1, \dots, N$, defined by $(\frac{\partial}{\partial x^n})_m g = \frac{\partial g \circ \psi^{-1}}{\partial x^n}(\psi(m))$ is a basis of $T_m(M)$. Let g be a C^1 function at m , the differential of g at m , $dg(m)$, is an element of $T_m^*(M)$, and is defined by

$$dg(m)(v) = v(g) \quad \forall v \in T_m(M).$$

Let M_1 and M_2 be two C^∞ manifolds. Consider a map $\phi : M_1 \rightarrow M_2$. Then for every function $g \in C^\infty(M_2)$, ϕ induces a function ϕ^*g on M_1 defined by $\phi^*g = g \circ \phi$. A map $\phi : M_1 \rightarrow M_2$ is called C^r at $m \in M_1$ (on $S \subset M_1$) provided that so is ϕ^*g for any $g \in C^\infty(M_2)$. Let $\phi : M_1 \rightarrow M_2$ be a C^1 map and let $m \in M_1$ be a fixed element. Define, for $v \in T_m(M_1)$ and $g \in C^\infty(M_2)$, $((\phi_*)_m v)(g) = v(\phi^*g)$. Then $(\phi_*)_m : T_m(M_1) \rightarrow T_{\phi(m)}(M_2)$ is a linear map. The dual map of $(\phi_*)_m$ is denoted by ϕ_m^* . It is a map from $T_{\phi(m)}^*(M_2) \rightarrow T_m^*(M_1)$ and has the property that, for any $g \in C^1(M_2)$, $\phi_m^* dg(\phi(m)) = d(\phi^*g)(m)$.

Let $v_i^* \in T_{m_i}^*(M)$, $i = 1, 2, \dots$, be a sequence of cotangent vectors of M and let $v^* \in T_m^*(M)$. We say v_i^* converges to v^* , denoted by $\lim v_i^* = v^*$, provided that $m_i \rightarrow m$ and, for any $X \in V^\infty(M)$, $\langle v_i^*, X(m_i) \rangle \rightarrow \langle v^*, X(m) \rangle$. Let (U, ψ) be a local coordinate neighborhood with $m \in U$. Since $m_i \rightarrow m$ we may assume without loss of generality that $m_i \in U$ for all i . Then $\lim v_i^* = v^*$ if and only if $\langle v_i^*, (\frac{\partial}{\partial x^n})_{m_i} \rangle \rightarrow \langle v^*, (\frac{\partial}{\partial x^n})_m \rangle$ for $n = 1, \dots, N$. Another equivalent description is $(\psi^{-1})_{\psi(m_i)}^* v_i^* \rightarrow (\psi^{-1})_{\psi(m)}^* v^*$ (in the dual of R^N).

Recall that a mapping $g : T(M) \times T(M) \rightarrow R$ is a C^∞ Riemann metric if

- (1) for each m , $g_m(v, u)$ is an inner product on $T_m(M)$;
- (2) if (U, ψ) is a local coordinate neighborhood around m with a local coordinate system (x^1, \dots, x^N) , then $g_{ij}(m) := g_m(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) \in C^\infty(M)$.

One can check that (2) is independent on local coordinate systems. The manifold M together with the Riemann metric g is called a Riemannian manifold. Since any paracompact manifold admits a positive-definite metric structure in many cases we may assume that M is a Riemannian manifold without significant loss of generality.

Let (M, g) be a Riemannian manifold. For each $m \in M$, the Riemann metric induces an isomorphism between $T_m(M)$ and $T_m^*(M)$ by

$$v^* = g_m(v, \cdot) \quad (\langle v^*, u \rangle = g_m(v, u), \quad \forall u \in T_m(M)).$$

Then we define norms on $T_m(M)$ and $T_m^*(M)$ by

$$\|v^*\|^2 = \|v\|^2 = g_m(v, v).$$

The following generalized Cauchy inequality is crucial: for any $v^* \in T_m^*(M)$ and $u \in T_m(M)$,

$$\langle v^*, u \rangle \leq \|v^*\| \|u\|.$$

Let $r : [0, 1] \rightarrow M$ be a C^1 curve. The length of r is

$$l(r) = \int_0^1 \|r'(s)\| ds.$$

Let $m_1, m_2 \in M$. Denote the collection of all C^1 curves joining m_1 and m_2 by $C(m_1, m_2)$. Then the distance between m_1 and m_2 is defined by

$$\rho(m_1, m_2) := \inf\{l(r) : r \in C(m_1, m_2)\},$$

or by (1.1). The distance between a point $m \in M$ and a set $S \subset M$ is defined by $\rho(m; S) := \inf\{\rho(m, m') : m' \in S\}$. Finally, we recall the definition of an absolutely continuous mapping $c : [a, b] \rightarrow M$. Mapping c is called absolutely continuous if a function $\varphi(c(t))$ is absolutely continuous for any smooth function $\varphi \in C^\infty(M)$. We call such a mapping a *curve* and leave it to the reader to prove that for any curve c

there exists mapping \dot{c} such that $\dot{c}(t) \in T_{c(t)}(M)$ for almost all (a.a.) $t \in [a, b]$ and, for any $\varphi \in C^\infty(M)$,

$$\frac{d(\varphi \circ c)(t)}{dt} = \dot{c}(t)(\varphi) \quad \text{a.a. } t \in [a, b].$$

3. SUB- AND SUPER-DIFFERENTIALS OF NONSMOOTH FUNCTIONS ON MANIFOLDS

Now we turn to nonsmooth functions on a manifold and their sub- and superdifferentials. We denote the extended real line by $\bar{R} := R \cup \{+\infty\}$. For an extended-valued function $f : M \rightarrow \bar{R}$ the domain of f is defined by $\text{dom}(f) := \{m \in M : f(m) < \infty\}$.

Definition 3.1. Let $f : M \rightarrow \bar{R}$ be an extended-valued lower semicontinuous function. We define the Fréchet-subdifferential of f at $m \in \text{dom}(f)$ by

$$\partial_F f(m) := \{dg(m) : g \in C^1(M) \text{ and } f - g \text{ attains a local minimum at } m\}.$$

We define the (limiting) subdifferential and singular subdifferential of f at $m \in M$ by

$$\partial f(m) := \{\lim v_i^* : v_i^* \in \partial_F f(m_i), (m_i, f(m_i)) \rightarrow (m, f(m))\}$$

and

$$\partial^\infty f(m) := \{\lim t_i v_i^* : v_i^* \in \partial_F f(m_i), (m_i, f(m_i)) \rightarrow (m, f(m)) \text{ and } t_i \rightarrow 0^+\},$$

respectively.

Remark 3.2. Note that the Fréchet-subdifferential of f depends only on the local behavior of the function f . Moreover, every local C^1 function can be extended to a C^1 function on M (see e.g. [51, Lemmas 1 and 2]). Therefore, the support function g in the definition of the Fréchet subdifferential need only be C^1 in a neighborhood of m .

The elements of a Fréchet subdifferential are called Fréchet subgradients. We define the Fréchet *superdifferential* by $\partial^F f(m) = -\partial_F(-f)(m)$, and its elements are called *supergradients*. An alternative definition of a supergradient $v^* \in \partial^F f(m)$ is $v^* = dg$ for some C^1 function g such that $f - g$ attains a local maximum at m on M .

When $M = R^N$, then a function g in the definition of a Fréchet subgradient can be chosen to be a quadratic one. In this case the definition becomes a definition of a proximal subgradient which has a natural geometric interpretation in terms of normal vectors to an epigraph of function f (it is useful to recall that in the case of a smooth function f the vector $(f'(x), -1)$ is a normal vector to its epigraph). This geometric interpretation of subgradients also explains their relationship with a concept of generalized solutions of Hamilton-Jacobi equations in view of a normal vector characterization of invariance properties of solutions of differential inclusions. For details of proximal calculus and related results we refer to the textbook [29]. The other subdifferentials defined in Definition 3.1 coincide with the usual limiting and singular subdifferentials [53, 54]. Note that $\text{co} \{\partial f(x) + \partial^\infty f(x)\}$ coincides with the original Clarke's generalized gradient.

Returning to the manifold case we note that it follows directly from the definition that $\partial_F f(m) \subset \partial f(m)$ and $0 \in \partial^\infty f(m)$. Note that $\partial_F f(m)$ may be empty.

However, if f attains a local minimum at m , then $0 \in \partial_F f(m) \subset \partial f(m)$. These are the usual properties to be expected for a subdifferential.

The geometric concept of normal cones to a closed set can be naturally established as usual by using the subdifferential for the corresponding indicator function.

Definition 3.3. Let S be a closed subset of M with $s \in S$. We define the Fréchet-normal cone of S at s by

$$N_F(s; S) := \partial_F \delta_S(s).$$

We define the (limiting) normal cone of S at s by

$$N(s; S) := \partial \delta_S(s).$$

Here δ_S is the indicator function of S defined by $\delta_S(s) = 0$ if $s \in S$ and $\delta_S(s) = \infty$ if $s \notin S$.

Remark 3.4. (a) It is easy to verify that $N_F(s; S)$ is a cone and $N(s; S) := \partial \delta_S(s) = \partial^\infty \delta_S(s)$.

(b) It follows from the definition that $\{0\} \in N_F(s; S) \subset N(s; S)$ for any $s \in S$ and $N_F(s; S) = N(s; S) = \{0\}$ for any $s \in \text{int } S$. Thus, we will be mainly interested in nonzero normal cones. They are necessarily normal cones for the boundary points of S , denoted by $\text{bdy}(S)$.

When M is a Riemannian manifold we have the following relationship.

Proposition 3.5. Let M be a Riemannian manifold and let S be a closed subset of M . Then, for any $s \in S$,

$$N_F(s; S) = \text{cone } \partial_F \rho(s; S).$$

Proof. Suppose that $\xi \in \alpha \partial_F \rho(s; S)$ for some $\alpha > 0$. Then there exists $g \in C^1(M)$ such that $dg(s) = \xi$ and $\alpha \rho(\cdot; S) - g(\cdot)$ attains a local minimum at s . Since $\delta_S(\cdot) \geq \alpha \rho(\cdot; S)$ and $\delta_S(s) = \alpha \rho(s; S) = 0$ we have that $\delta_S(\cdot) - g(\cdot)$ attains a local minimum at s . That is $\xi \in N_F(s; S)$. Since $N_F(s; S)$ is a cone we have $\text{cone } \partial_F \rho(s; S) \subset N_F(s; S)$.

To prove the inverse inclusion we consider $\xi \in N_F(s; S)$. Then there exists $g \in C^1(M)$ such that $dg(s) = \xi$ and $\delta_S(\cdot) - g(\cdot)$ attains a local minimum at s . Observing that g is also locally Lipschitz near s , we can choose a neighborhood U of s such that $g(m) - g(s') \leq \alpha \rho(m, s')$ for all $m, s' \in U$ and $-g(s) \leq -g(m)$ for all $m \in S \cap U$. Now, for any $m \in U$ close enough to s and any $\varepsilon > 0$, we can choose $s' \in S \cap U$ such that $\rho(m, s') < \rho(m; S) + \varepsilon$. Then

$$\begin{aligned} g(m) - g(s) &= g(m) - g(s') + g(s') - g(s) \\ &\leq g(m) - g(s') \leq \alpha \rho(m, s') \leq \alpha(\rho(m; S) + \varepsilon). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ we have $g(m) - g(s) \leq \alpha \rho(m; S)$. Thus, $\alpha \rho(\cdot; S) - g(\cdot)$ attains a local minimum at s and, therefore, $\xi \in \alpha \partial_F \rho(s; S)$. \square

Example 3.6 (Normal cone to a submanifold). Let S be a C^1 submanifold of M . Then, for any $s \in S$,

$$N_F(s; S) = T_s(S)^\perp := \{v^* \in T_s^*(M) : \langle v^*, v \rangle = 0, \forall v \in T_s(S)\}.$$

In fact, assume that the dimension of S is p . Then there exists a coordinate neighborhood (U, ψ) around s with a corresponding local coordinate system (x^1, \dots, x^N)

such that $s = \psi^{-1}(0)$ and $S \cap U = \psi^{-1}(x^1, \dots, x^p, 0, \dots, 0)$ for $(x^1, \dots, x^p, 0, \dots, 0) \in \psi(U)$. Then $T_s(S) = \text{span}((\frac{\partial}{\partial x^1})_s, \dots, (\frac{\partial}{\partial x^p})_s)$ and

$$T_s(S)^\perp = \text{span}(dx^{p+1}(s), \dots, dx^N(s)).$$

Let $v^* \in N_F(s; S)$. Then there exists $g \in C^1(M)$ such that $g(s) = 0$, $dg(s) = v^*$ and, for any $m \in S$, $g(m) \leq 0$. Let $v \in T_s(S)$ and $\gamma : (-r, r) \rightarrow S$ be a C^1 curve such that $\gamma(0) = s$ and $\gamma'(0) = v$. Then, in particular, $g \circ \gamma(t) \leq 0, t \in (-r, r)$ and $g \circ \gamma(0) = 0$. Thus, $d(g \circ \gamma)(0) = \langle v^*, v \rangle = 0$. That is to say $N_F(s; S) \subset T_s(S)^\perp$.

On the other hand, let $v^* \in T_s(S)^\perp = \text{span}(dx^{p+1}(s), \dots, dx^N(s))$. Then $v^* = \sum_{n=p+1}^N \alpha_n dx^n(s)$ where $\alpha_n \in \mathbb{R}, n = p+1, \dots, N$. Define $g : U \rightarrow \mathbb{R}$ by

$$g(m) = \sum_{n=p+1}^N \alpha_n x^n(m).$$

Then, for any $m \in S \cap U$, $g(m) = 0$. Thus, $v^* = \sum_{n=p+1}^N \alpha_n dx^n(s) = dg(s) \in N_F(s; S)$, which establishes $T_s(S)^\perp \subset N_F(s; S)$. \square

We turn to a discussion of the generalization of the Bouligand tangent cone from \mathbb{R}^N to a general smooth manifold. To do so we need the following simple and useful lemma.

Lemma 3.7. *Let M be a C^∞ (Riemannian) manifold, let $m \in M$ and let $v \in T_m(M)$. Then there exists a C^∞ (Lipschitz) vector field $v(\cdot)$, in a neighborhood of m such that $v(m) = v$. We call $v(\cdot)$ a local (Lipschitz) extension of v .*

Proof. Let (U, ψ) be a local coordinate system around m with local coordinates (x^1, \dots, x^N) . Then, there exist constants $\alpha_n, n = 1, \dots, N$, such that

$$v = \sum_{n=1}^N \alpha_n \left(\frac{\partial}{\partial x^n} \right)_m.$$

Then, $v(\cdot) := \sum_{n=1}^N \alpha_n \left(\frac{\partial}{\partial x^n} \right)$. satisfies our requirement. \square

Consider a subset S of M and a vector $v \in T_s(M)$. Let $v(\cdot)$ be a local extension of v and let c_v denote a local solution of the equation

$$\dot{c}(t) = v(c(t)), \quad c(0) = s.$$

Definition 3.8. The Bouligand tangent cone $T_s^B(S)$ to S at s consists of all tangent vectors v such that there exist a sequence $t_i \rightarrow 0+$ and $v_i \in T_{s_i}(M)$ such that $v_i \rightarrow v$ and $c_{v_i}(t_i) \in S$.

An easy consequence of this definition is the following proposition.

Proposition 3.9.

$$(3.1) \quad \text{co } T_s^B(S) \subset N_F^*(s; S).$$

Proof. Since the dual cone

$$N_F^*(s; S) := \{v \in T_s(M) : \langle v, n \rangle \leq 0 \ \forall n \in N_F(s; S)\}$$

is a convex set, it is enough to prove the inclusion (3.1) only for $T_s^B(S)$.

Consider the pair $\xi \in N_F(s; S)$ and $g \in C^1(M)$ such that $dg(s) = \xi$ and $\delta(\cdot; S) - g$ attains a local minimum at s . We have that if $v \in T_s^B(S)$, then there exists $t_i \rightarrow 0+$ and $v_i \rightarrow v$ such that $c_{v_i}(t_i) \in S$. This implies that

$$\int_0^{t_i} \langle dg(c_{v_i}(t)), v_i(c_{v_i}(t)) \rangle dt = \int_0^{t_i} \frac{dg(c_{v_i}(t))}{dt} dt = g(c_{v_i}(t_i)) - g(s) \leq 0.$$

Dividing the above inequality by t_i and taking limits, we obtain

$$\langle \xi, v \rangle = \langle dg(s), v \rangle \leq 0.$$

That is to say $\xi \in N_F^*(s; S)$. \square

We conclude this section with the following infinitesimal characterization of the Bouligand tangent cone in terms of the distance function to the set S .

Proposition 3.10.

$$(3.2) \quad T_s^B(S) = \{v \in T_s(M) : D\rho(s; S)(v) \leq 0\},$$

where D denotes an analogue of the lower Dini derivative along the tangent vector v :

$$Df(s)(v) := \liminf_{t \rightarrow 0+, v' \rightarrow v} \frac{f(c_{v'}(t)) - f(s)}{t}.$$

4. CALCULUS FOR SUBDIFFERENTIALS OF LOWER SEMICONTINUOUS FUNCTIONS

This section contains elements of a calculus of subdifferentials of lower semicontinuous functions on smooth manifolds: sum and product rules, chain rules, subgradients for supremum functions and necessary optimality conditions for constrained optimization problems. One can derive similar results for a superdifferential which we will not list in detail to make the exposition concise. Our proofs are accessible to readers without experience in nonsmooth analysis techniques. A useful technique in discussing properties of subdifferentials on a manifold by using known properties of Fréchet (or proximal) subdifferentials in R^N is along the following lines. First, convert the problem into one in a Euclidean space via a local coordinate system. Then apply corresponding results for subdifferentials in a Euclidean space to the problem. Finally, lift the conclusion back onto the manifold.

4.1. A chain rule. The following simple chain rule is crucial for implementing the scheme alluded to above.

Theorem 4.1. *Let M and N be smooth manifolds, let $g : N \rightarrow M$ be a C^1 mapping and let $f : M \rightarrow \bar{R}$ be a lower semicontinuous function. Suppose that $m = g(n)$. Then*

$$(4.1) \quad g_m^* \partial_F f(m) \subset \partial_F(f \circ g)(n),$$

$$(4.2) \quad g_m^* \partial f(m) \subset \partial(f \circ g)(n),$$

and

$$(4.3) \quad g_m^* \partial^\infty f(m) \subset \partial^\infty(f \circ g)(n).$$

Moreover, if g is a C^1 diffeomorphism, then both sides of (4.1), (4.2) and (4.3) are equal.

Proof. Since (4.2) and (4.3) follow directly from (4.1) by taking limits, we prove (4.1). Let $y^* \in \partial_F f(m)$. Then there exists a C^1 function h such that $dh(m) = y^*$ and $f - h$ attains a local minimum at m . It follows that $f \circ g - h \circ g$ attains a local minimum at n . Observing that $h \circ g$ is a C^1 function on N , we have

$$\partial_F(f \circ g)(n) \ni d(h \circ g)(n) = g_m^* dh(m) = g_m^* y^*.$$

Thus,

$$g_m^* \partial_F f(m) \subset \partial_F(f \circ g)(n).$$

When g is a diffeomorphism applying (4.1) to $f \circ g$ and g^{-1} yields the opposite inclusion. \square

Applying Theorem 4.1 to $g = \psi^{-1}$ for a local coordinate mapping ψ yields the following corollary.

Corollary 4.2. *Let $f : M \rightarrow \bar{R}$ be a lower semicontinuous function. Suppose that (U, ψ) is a local coordinate neighborhood and $m \in U$. Then*

$$\partial_F f(m) = \psi_m^* \partial_F(f \circ \psi^{-1})(\psi(m)),$$

$$\partial f(m) = \psi_m^* \partial(f \circ \psi^{-1})(\psi(m)),$$

and

$$\partial^\infty f(m) = \psi_m^* \partial^\infty(f \circ \psi^{-1})(\psi(m)).$$

We illustrate the usage of Corollary 4.2 by proving the density of the domain of the Fréchet subdifferential. The Banach space version of this result appeared in [8].

Theorem 4.3. *Let $f : M \rightarrow \bar{R}$ be a lower semicontinuous function and S be a closed subset of M . Then $\text{dom}(f) \subset \overline{\text{dom}(\partial_F f)} \subset \overline{\text{dom}(\partial f)}$ and $\text{bdy}(S) \subset \overline{\{s : N_F(s; S) \neq \{0\}\}} \subset \overline{\{s : N(s; S) \neq \{0\}\}}$.*

Proof. We need only to show that $\text{dom}(f) \subset \overline{\text{dom}(\partial_F f)}$. The rest follows easily. Let $m \in \text{dom}(f)$ and W be an arbitrary neighborhood of m . Without loss of generality we may assume that there is a local coordinate neighborhood (U, ψ) such that $W \subset U$. Then $f \circ \psi^{-1} : \psi(U) \rightarrow \bar{R}$ is a lower semicontinuous function and $\psi(m) \in \text{dom}(f \circ \psi^{-1})$. Since $\psi(W)$ is a neighborhood of $\psi(m)$ by the density theorem of the Fréchet-subdifferential in R^N , there exists $x \in \psi(W)$ such that $\partial_F(f \circ \psi^{-1})(x) \neq \emptyset$. Then $w = \psi^{-1}(x) \in W$ and by Corollary 4.2

$$\partial_F f(w) = \psi_m^* \partial_F(f \circ \psi^{-1})(x) \neq \emptyset.$$

\square

4.2. Fuzzy sum rules. The fuzzy sum rule is one of the equivalent forms of several fundamental principles for subdifferentials [79] which plays an important role in many applications of the subdifferentials. Its prototype appeared in [42, 43, 44]. The following is a version for the Fréchet subdifferential on a smooth manifold.

Theorem 4.4 (Fuzzy sum rule). *Let f_1, \dots, f_L be lower semicontinuous functions on a manifold M and let $\xi \in \partial_F(f_1 + \dots + f_L)(\bar{m})$. Then, for any $\varepsilon > 0$, any neighborhood V of \bar{m} and any $v \in V^\infty(M)$, there exist $m_l \in V$, $l = 1, 2, \dots, L$, and $\xi_l \in \partial_F f_l(m_l)$ such that $|f_l(m_l) - f_l(\bar{m})| < \varepsilon$ and*

$$|\langle \xi, v \rangle(\bar{m}) - \sum_{l=1}^L \langle \xi_l, v \rangle(m_l)| < \varepsilon.$$

Proof. Without loss of generality we may assume that $(f_1 + \dots + f_L)$ attains a local minimum at \bar{m} . Let (U, ψ) be a local coordinate system of M with $\bar{m} \in U$ and $\psi = (x^1, \dots, x^N)$. Without loss of generality we may assume that $U \subset V$. It is obvious that the function $f_1 \circ \psi^{-1} + \dots + f_L \circ \psi^{-1}$ attains a local minimum at $\bar{x} = \psi(\bar{m})$. Let $v \in V^\infty(M)$. Then $v(x^n) \circ \psi^{-1}, n = 1, 2, \dots, N$, are C^∞ functions on R^N . In particular, they are Lipschitz on $\psi(U)$, say with a uniform rank K . Set $\varepsilon' = \varepsilon/2(LK + \sup_{x \in \psi(U)} \|(v(x^1) \circ \psi^{-1}, \dots, v(x^N) \circ \psi^{-1})(x)\|)$.

Applying the fuzzy sum rule [13, Theorem 2.6] on R^N we have that there exists $(x_l, f_l \circ \psi^{-1}(x_l)) \in (\bar{x}, f_l \circ \psi^{-1}(\bar{x})) + \varepsilon' B_{R^{N+1}}$ satisfying

$$\|x_l^*\| \text{diam}(x_1, x_2, \dots, x_L) < \varepsilon'$$

and

$$\left\| \sum_{l=1}^L x_l^* \right\| < \varepsilon'.$$

Suppose

$$x_l^* = (a_l^1, a_l^2, \dots, a_l^N).$$

Define

$$\xi_l := \sum_{n=1}^N a_l^n dx^n.$$

Then

$$\xi_l = \psi_{x_l}^* x_l^* \in \partial_F f_l(m_l), \text{ where } m_l = \psi^{-1}(x_l) \in V.$$

We then have

$$\begin{aligned} \sum_{l=1}^L \langle \xi_l, v \rangle &= \sum_{l=1}^L \sum_{n=1}^N a_l^n \langle dx^n, v \rangle_{m_l} \\ &= \sum_{l=1}^L \sum_{n=1}^N a_l^n v(x^n)_{m_l} = \sum_{l=1}^L \sum_{n=1}^N a_l^n v(x^n) \circ \psi^{-1}(x_l) \\ &= \left\langle \sum_{l=1}^L x_l^*, (v(x^1) \circ \psi^{-1}, \dots, v(x^N) \circ \psi^{-1})(x_1) \right. \\ &\quad \left. + \sum_{l=1}^L \langle x_l^*, (v(x^1) \circ \psi^{-1}(x_l) - v(x^1) \circ \psi^{-1}(x_1), \dots, v(x^N) \right. \\ &\quad \left. \circ \psi^{-1}(x_l) - v(x^N) \circ \psi^{-1}(x_1)) \rangle \right. \end{aligned}$$

Thus,

$$\begin{aligned} &\left| \sum_{l=1}^L \langle \xi_l, v \rangle \right| \\ &\leq \left\| \sum_{l=1}^L x_l^* \right\| \cdot \sup_{x \in \psi(U)} \|(v(x^1) \circ \psi^{-1}, \dots, v(x^N) \circ \psi^{-1})(x)\| \\ &\quad + \sum_{l=1}^L \|x_l^*\| K \text{diam}(x_1, \dots, x_L) \\ &< \varepsilon. \end{aligned}$$

□

Remark 4.5. In a Banach space this fuzzy sum rule can be refined to include additional information on the ‘size’ of the subderivatives involved (see [12]). This feature is lost here due to the lack of a metric in $T(M)$. However, a similar estimate can be established on a Riemannian manifold.

4.3. A subdifferential representation of the superdifferential. Now we use the lifting method to derive a smooth manifold version of the subdifferential representation of the superdifferential due to Barron and Jenson [5].

Theorem 4.6. *Let $f : M \rightarrow R$ be a continuous function and let $\xi \in \partial^F f(m)$. Then for any $\varepsilon > 0$, any neighborhood U of m and any $v \in V^1(M)$, there exist $m_i \in U$, $\xi_i \in \partial_F f(m_i)$ and convex coefficients $\alpha_i, i = 1, \dots, k$, such that*

$$|\langle \xi, v \rangle - \sum_{i=1}^k \alpha_i \langle \xi_i, v \rangle| < \varepsilon.$$

Proof. First note that when $M = R^N$ this is a corollary of the result in [5]. To prove the general case we can assume without loss of generality that (U, ψ) is a local coordinate system. It follows from Corollary 4.2 that, for $x = \psi(m)$, there exists $x^* \in \partial^F(f \circ \psi^{-1})(x)$ with $\psi_m^* x^* = \xi$. Note that $\psi_* v \in V^1(R^N)$. Applying the R^N version of the theorem we have that there exist $x_i \in \psi(U)$, $x_i^* \in \partial_F(f \circ \psi^{-1})(x_i)$ and convex coefficients $\alpha_i, i = 1, \dots, k$, such that

$$(4.4) \quad |\langle x^*, \psi_* v \rangle - \sum_{i=1}^k \alpha_i \langle x_i^*, \psi_* v \rangle| < \varepsilon.$$

Denoting $\xi_i = \psi_{m_i}^*$ with $m_i = \psi^{-1}(x_i) \in U$ we have $\xi_i \in \partial_F f(m_i)$ and (4.4) becomes

$$|\langle \xi, v \rangle - \sum_{i=1}^k \alpha_i \langle \xi_i, v \rangle| < \varepsilon.$$

□

4.4. Mean value inequalities. The importance of mean value theorems in analysis are well known. Often what one actually uses is an inequality. In this section we discuss such mean value inequalities for extended valued lower semicontinuous functions and continuous functions on smooth manifolds. In dealing with extended valued lower semicontinuous functions, it is often important to control the value of the function at the mean value point. This is reflected in an additional estimate that does not appear in the smooth case. We discuss two different types of such estimates, and their applications will become clear in later sections. The first is a smooth manifold version of Zagrodny’s result in [77].

Theorem 4.7 (Mean value inequality I). *Let $f : M \rightarrow \bar{R}$ be a lower semicontinuous function, let $v \in V^\infty(M)$ and let c be a curve corresponding to this vector field, i.e., $\dot{c}(t) = v(c(t))$, $t \in [0, 1]$, with $f(c(0)) < \infty$. Then, for any*

$$r < f(c(1)) - f(c(0)),$$

and any open neighborhood U of $c([0, 1])$, there exists $\bar{m} \in U$, $\xi \in \partial_F f(\bar{m})$ such that, $f(\bar{m}) < \min\{f(c(0)), f(c(1)) + \max\{0, r\}\}$ and

$$r < \langle \xi, v(\bar{m}) \rangle.$$

Proof. We first consider the simple case when (U, ψ) is a local coordinate system with local coordinates (x^1, \dots, x^N) . For any $r < f(c(1)) - f(c(0))$, choose $\bar{r} \in (r, f(c(1)) - f(c(0)))$ and choose $\bar{\alpha} > 0$ such that $\alpha \in (0, \bar{\alpha})$ and

$$\frac{1}{\alpha^2} |\psi(m) - \psi(c(t))|^2 < f(c(0)) - \inf_M f + |\bar{r}|$$

for some $t \in [0, 1]$ implies that $m \in U$. Consider the function

$$\varphi_\alpha(t) := f_\alpha(c(t)) - \bar{r}t, \alpha \in (0, \bar{\alpha}),$$

where

$$f_\alpha(m) := \min_{m' \in U} \{f(m') + \frac{1}{\alpha^2} |\psi(m) - \psi(m')|^2\}.$$

Then

$$\varphi_\alpha(1) - \varphi_\alpha(0) = f_\alpha(c(1)) - f_\alpha(c(0)) - \bar{r}.$$

Since $f_\alpha \rightarrow f$ as $\alpha \rightarrow 0$, when α is sufficiently small, we have

$$\varphi_\alpha(0) < \varphi_\alpha(1).$$

Assume that φ_α attains minimum at $t_\alpha \in [0, 1]$ over $[0, 1]$. Consider the right Dini derivative

$$\begin{aligned} (\varphi_\alpha)'_+(t_\alpha) &= \liminf_{\lambda \rightarrow 0^+} \frac{\varphi_\alpha(t_\alpha + \lambda) - \varphi_\alpha(t_\alpha)}{\lambda} \\ &= \liminf_{\lambda \rightarrow 0^+} \left[\frac{f_\alpha(c(t_\alpha + \lambda)) - f_\alpha(c(t_\alpha))}{\lambda} - \bar{r} \right] \\ &\leq \liminf_{\lambda \rightarrow 0^+} \frac{1}{\lambda} \left[(f(m_\alpha) + \frac{1}{\alpha^2} |\psi(c(t_\alpha + \lambda)) - \psi(m_\alpha)|^2) \right. \\ &\quad \left. - (f(m_\alpha) + \frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2) \right] - \bar{r} \end{aligned}$$

where m_α is a point of minimum for the function

$$m \rightarrow f(m) + \frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m)|^2.$$

Observing that

$$\begin{aligned} \frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 &= \varphi_\alpha(t_\alpha) - f(m_\alpha) + \bar{r}t_\alpha \\ &\leq \varphi_\alpha(0) - f(m_\alpha) + \bar{r}t_\alpha \leq f(c(0)) - \inf_M f + |\bar{r}|, \end{aligned}$$

we can conclude that $m_\alpha \in U$. Then we have

$$\begin{aligned} (\varphi_\alpha)'_+(t_\alpha) &\leq \liminf_{\lambda \rightarrow 0^+} \frac{1}{\lambda \alpha^2} \left[|\psi(c(t_\alpha + \lambda)) - \psi(m_\alpha)|^2 - |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 \right] - \bar{r} \\ &= \frac{2}{\alpha^2} \sum_{n=1}^N \langle dx^n(c(t_\alpha)), \dot{c}(t_\alpha) \rangle (x^n(c(t_\alpha)) - x^n(m_\alpha)) - \bar{r}. \end{aligned}$$

Since $(\varphi_\alpha)'_+(t_\alpha) \geq 0$ we have

$$\frac{2}{\alpha^2} \sum_{n=1}^N \langle dx^n(c(t_\alpha)), \dot{c}(t_\alpha) \rangle (x^n(c(t_\alpha)) - x^n(m_\alpha)) \geq \bar{r}.$$

By the definition of subdifferentials we have

$$\xi_\alpha := \frac{2}{\alpha^2} \sum_{n=1}^N (x^n(c(t_\alpha)) - x^n(m_\alpha)) dx^n(c(t_\alpha)) \in \partial_F f(m_\alpha).$$

Then we have the following inequality:

$$\begin{aligned} \bar{r} &\leq \langle \xi_\alpha, v(\bar{m}_\alpha) \rangle + \frac{2}{\alpha^2} \sum_{n=1}^N (x^n(c(t_\alpha)) - x^n(m_\alpha)) \\ &\quad \times [\langle dx^n(c(t_\alpha)), v(c(t_\alpha)) \rangle - \langle dx^n(m_\alpha), v(m_\alpha) \rangle] \\ &= \langle \xi, v(m_\alpha) \rangle + \frac{2}{\alpha^2} \sum_{n=1}^N (x^n(c(t_\alpha)) - x^n(m_\alpha)) [v(x^n)(c(t_\alpha)) - v(x^n)(m_\alpha)] \\ &= \langle \xi, v(m_\alpha) \rangle + \frac{2}{\alpha^2} \sum_{n=1}^N (x^n(c(t_\alpha)) - x^n(m_\alpha)) \\ &\quad \times [(v(x^n) \circ \psi^{-1})(\psi(c(t_\alpha))) - (v(x^n) \circ \psi^{-1})(\psi(m_\alpha))]. \end{aligned}$$

Since $v(x^n) \circ \psi^{-1}$ is C^1 and, therefore, locally Lipschitz around $\varphi(c(t_\alpha))$ we can conclude that there exists a constant K such that

$$|(v(x^n) \circ \psi^{-1})(\psi(c(t_\alpha))) - (v(x^n) \circ \psi^{-1})(\psi(m_\alpha))| \leq K |\psi(c(t_\alpha)) - \psi(m_\alpha)|.$$

Thus we have

$$(4.5) \quad \bar{r} \leq \langle \xi_\alpha, v(m_\alpha) \rangle + \frac{2KN}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2.$$

Next we show that

$$(4.6) \quad \lim_{\alpha \rightarrow 0^+} \frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 = 0.$$

By the definition of φ_α it is increasing when $\alpha \rightarrow 0^+$ and so is $\varphi_\alpha(t_\alpha)$. Moreover $\varphi_\alpha(t_\alpha) \leq \varphi_\alpha(0) \leq f(c(0))$ and, therefore, $\lim_{\alpha \rightarrow 0^+} \varphi_\alpha(t_\alpha)$ exists. By the definition of t_α and m_α we have

$$\begin{aligned} \varphi_{2\alpha}(t_{2\alpha}) &\leq \varphi_{2\alpha}(t_\alpha) \leq f(m_\alpha) + \frac{1}{4\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 - \bar{r}t_\alpha \\ &= \varphi_\alpha(t_\alpha) - \frac{3}{4\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2, \end{aligned}$$

that is to say

$$\frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 \leq \frac{4}{3} (\varphi_\alpha(t_\alpha) - \varphi_{2\alpha}(t_{2\alpha})).$$

Taking limits as $\alpha \rightarrow 0^+$ yields (4.6). Now taking α small enough so that $m_\alpha \in U$ and

$$\bar{r} - \frac{2KN}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 > r$$

we have, by (4.5), that

$$r \leq \langle \xi_\alpha, v(m_\alpha) \rangle.$$

Moreover,

$$\begin{aligned} f(m_\alpha) &\leq f(m_\alpha) + \frac{1}{\alpha^2} |\psi(c(t_\alpha)) - \psi(m_\alpha)|^2 \\ &= \varphi_\alpha(t_\alpha) \leq \min\{\varphi_\alpha(0), \varphi_\alpha(1)\} \\ &\leq \min\{f(c(0)), f(c(1)) + \max\{0, r\}\}. \end{aligned}$$

It remains to take $\bar{m} = m_\alpha$ and $\xi = \xi_\alpha$.

For the general case, since there are a finite number of coordinate neighborhoods covering the compact set $c([0, 1])$ we may assume that $U = \bigcup_{i=1}^k U_i$, where $(U_i, \psi_i), i = 1, \dots, k$, are coordinate neighborhoods. Let $\{\chi_i : i = 1, \dots, k\}$ be a C^1 partition of unity U corresponding to $U_i, i = 1, \dots, k$. Then we can define, for $m \in U$,

$$f_\alpha(m) := \min_{m' \in U} \{f(m') + \frac{1}{\alpha^2} \sum_{i=1}^k \chi_i(m') |\psi_i(m) - \psi_i(m')|^2\}.$$

Applying the above argument to f_α and supposing that $c(t_\alpha) \in U_j$, for some $j \in \{1, \dots, k\}$, then everything goes through with (U, ψ) replaced by (U_j, ψ_j) . \square

When f is continuous we can get a version of the mean value inequality with a lower estimate on the function value $f(\bar{m})$ following the model in [47]. This result can also be viewed as a combination of the mean value inequality I and the subdifferential representation of the superdifferential. Our proof below is based on this observation. These mean value inequalities are also interesting in that it is not ‘local’ in the sense that the curve $c([0, 1])$ is, in general, not contained in one local coordinate system. Thus, the ‘lifting’ approach in the proof of the fuzzy sum rule does not directly apply. In the proof of Theorem 4.7 we overcome this difficulty by using a partition of unity. In the proof of the second mean value inequality we adopt a different strategy in overcoming this difficulty.

Theorem 4.8 (Mean value inequality II). *Let $f : M \rightarrow \bar{R}$ be a continuous function bounded from below, let $v \in V^\infty(M)$ and let c be a curve corresponding to this vector field, i.e., $\dot{c}(t) = v(c(t)), t \in [0, 1]$. Then, for any*

$$r < f(c(1)) - f(c(0)),$$

any $\varepsilon > 0$ and any open neighborhood U of $c([0, 1])$, there exist $\bar{m} \in U$, $\xi \in \partial_F f(\bar{m})$ such that, $f(\bar{m}) > \max\{f(c(1)), f(c(0)) - \max\{r, 0\}\} - \varepsilon$ and

$$r < \langle \xi, v(\bar{m}) \rangle.$$

Proof. Again, we first consider the simple case when (U, ψ) is a local coordinate system. Consider $-f$ and $s(t) = c(1 - t)$. Then $\dot{s}(t) = -v(s(t))$ and

$$r < (-f)(s(1)) - (-f)(s(0)).$$

Applying Theorem 4.7 we have that there exist $m \in U$, $\xi \in \partial_F (-f)(m)$ such that

$$(4.7) \quad -f(m) < \min\{-f(s(0)), -f(s(1)) + \max\{0, r\}\}$$

and

$$(4.8) \quad r < \langle \xi, -v(m) \rangle = \langle -\xi, v(m) \rangle.$$

Observe that $-\xi \in -\partial_F(-f)(m) = \partial^F f(m)$. By the subdifferential representation of the superdifferential of Theorem 4.6, there exist m_i in a small neighborhood of m with

$$(4.9) \quad |f(m_i) - f(m)| < \varepsilon,$$

$\xi_i \in \partial_F f(m_i)$, and convex coefficients $\alpha_i, i = 1, \dots, k$, such that

$$r < \sum_{i=1}^k \alpha_i \langle \xi_i, v(m_i) \rangle.$$

Then at least one i will satisfy

$$r < \langle \xi_i, v(m_i) \rangle.$$

Combining inequalities (4.7) and (4.9) we have

$$f(m_i) > \max\{f(c(1)), f(c(0)) - \max\{0, r\}\} - \varepsilon.$$

It remains to set $\bar{m} = m_i$ and $\xi = \xi_i$.

Now we consider the general case when the curve c can be covered by a finite number of open subsets $U_i \subset M, i = 0, \dots, k$, from the atlas. Then there exists a finite partition

$$0 = \tau_0 < \tau_1 < \dots < \tau_k = 1$$

such that $c([\tau_i, \tau_{i+1}]) \subset U_i$ for all i from 0 to k . Again, we assume that $U = \bigcup_{i=1}^k U_i$.

Choose $r < f(c(1)) - f(c(0))$; then there exists an index j such that

$$r(\tau_{j+1} - \tau_j) < f(c(\tau_{j+1})) - f(c(\tau_j)).$$

We can assume that j is the largest such index which implies that

$$f(c(\tau_{j+1})) \geq \max\{f(c(1)), f(c(0)) - r(1 - \tau_{j+1})\}.$$

Now we apply the mean value inequality, which we proved before, to the interval $[\tau_j, \tau_{j+1}]$ to derive the existence of $\bar{m} \in U$ and $\xi \in \partial_F f(\bar{m})$ such that

$$r(\tau_{j+1} - \tau_j) < \langle \xi, v(\bar{m}) \rangle (\tau_{j+1} - \tau_j),$$

and

$$\begin{aligned} f(\bar{m}) &> \max\{f(c(\tau_{j+1})), f(c(\tau_j)) - \max\{r, 0\}(\tau_{j+1} - \tau_j)\} - \varepsilon \\ &> \max\{f(c(1)), f(c(0)) - \max\{r, 0\}\} - \varepsilon, \end{aligned}$$

which implies the assertion of the theorem. \square

4.5. Subdifferential of sup-envelope function. We turn to discuss subdifferentials for sup-envelops of lower semicontinuous functions which is often useful in problems related to optimizations. Let $f_\gamma : M \rightarrow \bar{R}, \gamma \in \Gamma$ be a family of continuous functions and let

$$f(m) := \sup_{\gamma \in \Gamma} f_\gamma(m)$$

be the sup-envelop of $\{f_\gamma\}_{\gamma \in \Gamma}$. How to estimate the Fréchet subdifferential of f in terms of the subdifferentials of f_γ is a subject of much research (see e.g. [25, 34, 35, 37, 38, 60, 61]).

Here we prove a smooth manifold version of [47]. A recent application of this result to the proof of Helly's intersection theorem for Cartan-Hadamard manifolds can be found in [48].

We need the following set:

$$\mathcal{G}_{\delta,U}(\bar{m}) := \{(m, \gamma) \in U \times \Gamma : f_{\gamma}(m) \geq f(\bar{m}) - \delta\}$$

defined for positive δ and a neighborhood U of \bar{m} . Note that in the case of M being a Riemannian manifold we can consider a neighborhood $U = U_{\delta} := \{m \in M : \rho(m, \bar{m}) < \delta\}$. The set $\mathcal{G}_{\delta,U}(\bar{m})$ plays a role of a set of approximate γ in maximizing $\gamma \rightarrow f_{\gamma}(\bar{m})$. We consider a weak-star topology on $V^{\infty}(M)$ generated by neighborhoods

$$U_{\varepsilon, \xi, m}(v) := \{v' \in V^{\infty}(M) : |\langle \xi, v' \rangle_m - \langle \xi, v \rangle_m| < \varepsilon, \xi \in T_m^*(M)\}.$$

Note that the functional

$$v \rightarrow \sum_{k=1}^K \alpha_k \langle \xi_k, v \rangle_{m_k}$$

where $\xi_k \in T_{m_k}^*(M)$, and $\alpha_k \in R, k = 1, \dots, K$, is continuous in such a topology.

Theorem 4.9. *Let $\xi \in \partial_F f(m)$ with $f(m) = \sup_{\gamma \in \Gamma} f_{\gamma}(m)$. Then for any convex symmetric compact $W \subset V^{\infty}(M)$, any neighborhood U of m , any $\delta > 0$ and $\varepsilon > 0$ there exist convex coefficients $\{\alpha\}_{k=1}^K$, points $(m_k, \gamma_k) \in \mathcal{G}_{\delta,U}(m)$ and subgradients $\xi_k \in \partial_F f_{\gamma_k}(m_k)$, $k = 1, \dots, K$, such that*

$$|\langle \xi, v \rangle - \sum_{k=1}^K \alpha_k \langle \xi_k, v \rangle_{m_k}| < \varepsilon, \quad \forall v \in W.$$

Proof. Let $\xi \in \partial_F f(m)$. Then there exists a function $g \in C^1(M)$ such that $f - g$ attains a local minimum at m and $g'(m) = \xi$. Fix a vector field $v \in W$ and consider an integral curve

$$c'(t) = v(c(t)), c(0) = m.$$

We can assume that $c(t)$ is defined for all $t > 0$ small enough and belongs to U . We have that, for all t small enough,

$$(4.10) \quad f(c(t)) - f(c(0)) \geq g(c(t)) - g(c(0)).$$

Choose $\gamma \in \Gamma$ such that

$$(4.11) \quad f_{\gamma}(c(t)) \geq f(c(t)) - t^2.$$

Then we have

$$(4.12) \quad f_{\gamma}(c(t)) - f_{\gamma}(c(0)) \geq g(c(t)) - g(c(0)) - t^2.$$

Applying the mean value inequality of Theorem 4.8 we have that there exist $m_t \in U \subset M$ satisfying

$$(4.13) \quad f_{\gamma}(m_t) > f_{\gamma}(c(t)) - g(c(t)) - g(c(0)) - t^2$$

and subgradient $\xi_t \in \partial_F f_{\gamma}(m_t)$ such that

$$(4.14) \quad \langle \xi, v \rangle t > g(c(t)) - g(c(0)) - t^2.$$

It follows from (4.12), (4.13) and the lower semicontinuity of f that by choosing t small enough we can make $(m_t, \gamma) \in \mathcal{G}_{\delta,U}(m)$. Now we define a convex subset \mathcal{M} of the cotangent fields as follows:

$$\mathcal{M} := \left\{ \sum_{k=1}^K \alpha_k \xi_k : \xi_k \in \partial_F f_{\gamma_k}(m_k), \text{ with } (m_k, \gamma_k) \in \mathcal{G}_{\delta,U}(m), K = 1, 2, \dots \right\}.$$

For $\eta = \sum_{k=1}^K \alpha_k \xi_k$, consider the following bilinear functional:

$$\Phi(\eta, v) = \sum_{k=1}^K \alpha_k \langle \xi_k, v \rangle_{m_k}.$$

It follows from (4.14) that, for all $t > 0$ small enough,

$$\sup_{\eta \in \mathcal{M}} \Phi(\eta, v) \geq \frac{g(c(t)) - g(c(0))}{t} - t.$$

By taking limits as $t \rightarrow 0$ we obtain

$$(4.15) \quad \sup_{\eta \in \mathcal{M}} \Phi(\eta, v) \geq \langle \xi, v \rangle, \quad \forall v \in W.$$

It follows from (4.15) and the symmetricity of W that

$$(4.16) \quad \inf_{v \in W} \sup_{\eta \in \mathcal{M}} [\Phi(\eta, v) - \langle \xi, v \rangle] = 0.$$

Since the functional $\Phi(\eta, v) - \langle \xi, v \rangle$ is continuous in v in the weak-star topology and is bilinear, \mathcal{M} is convex and W is convex and compact, we can apply the nonsymmetric minimax theorem of Borwein-Zhang [15] to obtain from (4.16) that

$$\sup_{\eta \in \mathcal{M}} \inf_{v \in W} [\Phi(\eta, v) - \langle \xi, v \rangle] = 0.$$

This implies that, for any $\varepsilon > 0$, there exists $\eta = \sum_{k=1}^K \alpha_k \xi_k \in \mathcal{M}$ such that

$$\inf_{v \in W} [\Phi(\eta, v) - \langle \xi, v \rangle] > -\varepsilon,$$

which implies, for all $v \in W$, that the inequality

$$|\langle \xi, v \rangle - \sum_{k=1}^K \alpha_k \langle \xi_k, v \rangle_{m_k}| < \varepsilon$$

holds. □

4.6. Necessary conditions for constrained minimization problems. Necessary optimality conditions is one of the primary concerns in optimization problems. The relevance here is that in the general form represented below it is also closely related to the calculus of subdifferentials. There are many different versions of such necessary optimality conditions under different assumptions (see [25, 26, 49, 54, 56]). The form we present here is directly related to [11, 73].

Let $f_l : M \rightarrow \bar{R}$, $l = 0, 1, \dots, L$. Consider the following *optimization* problem on M :

$$\begin{array}{ll} \mathcal{P} & \text{minimize} \quad f_0(m) \\ & \text{subject to} \quad f_l(m) \leq 0, \quad l = 1, 2, \dots, K, \\ & \quad \quad \quad f_l(m) = 0, \quad l = K + 1, \dots, L, \\ & \quad \quad \quad m \in S. \end{array}$$

Following [11] we use the quantities $\tau_l, l = 0, 1, \dots, L$, to simplify the notation. The τ_l 's associated with the inequality constraints and the cost function are always

1, i.e., $\tau_l := 1, l = 0, 1, \dots, K$. This corresponds to nonnegative multipliers. The τ_l 's associated with the equality constraints are either 1 or -1 , corresponding to multipliers with arbitrary sign, i.e., $\tau_l \in \{-1, 1\}, l = K + 1, \dots, L$.

Theorem 4.10. *Let $f_l : M \rightarrow \bar{R}$ be lower semicontinuous for $l = 0, 1, \dots, K$ and continuous for $l = K + 1, \dots, L$ and let S be a closed subset of M . Suppose that \bar{m} is a local solution of problem \mathcal{P} . Then either:*

(A1) *there exist $v_l^\infty \in \partial^\infty(\tau_l f_l)(\bar{p})$, $l = 0, 1, \dots, L$, and $u_{L+1}^\infty \in N(\bar{m}; S)$ not all zero such that*

$$0 = \sum_{l=0}^{L+1} v_l^\infty,$$

or there exist $\mu_l \geq 0, l = 0, \dots, L$, not all zero such that

(A2)

$$0 \in \sum_{k \in \{l: \mu_l > 0\}} \mu_k \partial(\tau_k f_k)(\bar{m}) + \sum_{k \in \{l: \mu_l = 0\}} \partial^\infty(\tau_k f_k)(\bar{m}) + N(\bar{m}; S).$$

Proof. Let (U, ψ) be a local coordinate neighborhood of M with $\bar{m} \in U$. Then $\bar{x} := \psi(\bar{m})$ is a (local) solution to the following constrained optimization problem on R^N :

$$\begin{aligned} &\text{minimize} && f_0 \circ \psi^{-1}(x) \\ &\text{subject to} && f_l \circ \psi^{-1}(x) \leq 0, \quad l = 1, 2, \dots, K, \\ & && f_l \circ \psi^{-1}(x) = 0, \quad l = K + 1, \dots, L, \\ & && x \in \psi(S). \end{aligned}$$

Applying the limiting multiplier rule of [13, Theorem 7.4] and Lemma 4.2 completes the proof. \square

4.7. Calculus rules for subdifferentials. Calculus for subdifferentials are important for the effective application of subdifferentials. Many different approaches have been used in developing calculus for various subdifferentials in general Banach spaces [14, 25, 26, 29, 43, 49, 55, 56, 58, 59, 64, 75, 76]. We will establish a chain rule first and then deduce a sum rule, a product rule and a quotient rule from it. We start by establishing a general chain rule in a Euclidean space. The first general chain rule for nonsmooth functions was proved by Warga in [76]. Here we deduce a chain rule from the necessary optimality conditions in [11] using a conversion in [78]. We continue to use the τ_l notation introduced in the previous section.

Theorem 4.11 (Chain rule in R^N). *Let $f : R^L \rightarrow \bar{R}$ and $f_l : R^N \rightarrow \bar{R}, l = 1, \dots, K$, be lower semicontinuous functions and let $f_l : X \rightarrow R, l = K + 1, \dots, L$, be continuous functions. Suppose that f is nondecreasing for each of its first K variables. Then either:*

(A1) *there exist $v_n^\infty \in \partial^\infty(\tau_l f_l)(x)$, $l = 1, \dots, L$, not all zero such that*

$$0 = \sum_{l=1}^L v_l^\infty,$$

or there exists $0 \neq \mu^\infty \in \partial^\infty f(f_1(x), \dots, f_L(x))$ such that, for any $\mu_l^\infty \neq 0$, one has $0 \in \partial f_l(x)$,

or there exist $\mu = (\mu_1, \dots, \mu_L) \in \partial f(f_1(x), \dots, f_L(x))$ such that

(A2)

$$\partial f(f_1, \dots, f_L)(x) \subset \sum_{k \in \{l: \mu_l \neq 0\}} \tau_k \mu_k \partial(\tau_k f_k)(x) + \sum_{k \in \{l: \mu_l = 0\}} \partial^\infty(\tau_k f_k)(x)$$

and

$$\partial^\infty f(f_1, \dots, f_L)(x) \subset \sum_{l=1}^L \partial^\infty(\tau_l f_l)(x).$$

Proof. Let $\xi \in \partial f(f_1, \dots, f_L)(x)$ ($\xi \in \partial^\infty f(f_1, \dots, f_L)(x)$). Then there exist sequences ξ_i and x_i satisfying $(x_i, f(f_1, \dots, f_L)(x_i)) \rightarrow (x, f(f_1, \dots, f_L)(x))$, $\xi_i \in \partial_F f(f_1, \dots, f_L)(x_i)$ such that

$$\lim_{i \rightarrow \infty} \xi_i = \xi \quad (\lim_{i \rightarrow \infty} h_i \xi_i = \xi, h_i \rightarrow 0^+).$$

Let $g_i \in C^1(R^N)$ such that $dg_i(x_i) = \xi_i$ and $f(f_1, \dots, f_L) - g_i$ attains a local minimum at x_i . Then $(x_i, f_1(x_i), \dots, f_L(x_i))$ is a local solution to the following constrained minimization problem:

$$\begin{aligned} & \text{minimize} && f(y_1, \dots, y_L) - g_i(x) \\ & \text{subject to} && f_l(x) - y_l \leq 0, \quad l = 1, 2, \dots, K, \\ & && f_l(x) - y_l = 0, \quad l = K + 1, \dots, L. \end{aligned}$$

Applying the fuzzy multiplier rule of Theorem 3.1 in [11], there exist x_i^l and y_i^l , $l = 0, 1, \dots, L$, satisfying $\|g'_i(x_i^0) - \xi_i\| < 1/i$, $|y_i^l - f_l(x_i)| < 1/i$,

$$\|(x_i^l, f_l(x_i^l)) - (x_i, f_l(x_i))\| < 1/i$$

and $|f(y_i^1, \dots, y_i^L) - f(f_1, \dots, f_L)(x_i)| < 1/i$ such that

$$0 \in (\partial_F f(y_i^1, \dots, y_i^L), -g'_i(x_i^0)) + \sum_{l=1}^L \nu_i^l (-\tau_l^i e_l, \partial_F(\tau_l^i f_l)(x_i^l)) + (1/i)B_{R^{L+N}},$$

where e_l is the l th column of the $L \times L$ identity matrix and $\nu_i^l > 0$, for $l = 1, \dots, L$. Taking a subsequence if necessary we may assume that τ_l^i are independent on i . It follows that

$$(4.17) \quad \mu_i = (\mu_i^1, \dots, \mu_i^L) := (\tau_1 \nu_i^1, \dots, \tau_L \nu_i^L) \in \partial_F f(y_i^1, \dots, y_i^L) + (1/i)B_{R^L}$$

and there exists $\xi_i^l \in \partial_F(\tau_l f_l)(x_i^l)$ such that

$$(4.18) \quad \|\xi_i - \sum_{l=1}^L \nu_i^l \xi_i^l\| < 2/i.$$

Let $t_i := \sum_{l=1}^L \|\nu_i^l \xi_i^l\|$ ($t_i := \sum_{l=1}^L \|h_i \nu_i^l \xi_i^l\|$) and consider two cases:

Case 1. t_i is unbounded. Passing to a subsequence if necessary we may assume that $t_i \rightarrow +\infty$. Taking another subsequence if necessary we may assume that $\nu_i^l \xi_i^l / t_i$ ($h_i \nu_i^l \xi_i^l / t_i$) converges to $v_l^\infty \in \partial^\infty(\tau_l f_l)(x)$. Moreover $\sum_{l=1}^L \|v_l^\infty\| = 1$. Dividing (4.18) by t_i (and multiplying by h_i) and taking limits yields

$$\sum_{l=1}^L v_l^\infty = 0.$$

Case 2. t_i is bounded. Then without loss of generality we may assume that all the sequences $\nu_i^l \xi_i^l$, $(h_i \nu_i^l \xi_i^l)$, $l = 1, \dots, L$, converge. Now we check $s_i := \sum_{l=1}^L \|\mu_i^l\|$. If the sequence s_i is unbounded, then we may assume without loss of generality that $s_i \rightarrow +\infty$. Thus, $\mu_i/s_i \rightarrow \mu^\infty := (\mu_1^\infty, \dots, \mu_L^\infty) \in \partial^\infty f(f_1(x), \dots, f_L(x))$. Note that, for $\mu_i^\infty \neq 0$, $\|\mu_i^l\| \rightarrow \infty$ and, therefore, $\xi_i^l \rightarrow 0$, i.e., $0 \in \partial(\tau_l f_l)(x)$. This corresponds to the second abnormal case in (A1). If s_i is bounded, then passing to a subsequence again we may assume that $\mu_i \rightarrow \mu \in \partial f(f_1(x), \dots, f_L(x))$. Taking limits in (4.18) (multiplying h_i and taking limits in (4.18)) we arrive at alternative (A2). \square

Combining this chain rule in Euclidean spaces and Corollary 4.2 we have the following general chain rule for lower semicontinuous functions and continuous functions on a manifold.

Theorem 4.12 (Chain rule). *Let $f : R^L \rightarrow \bar{R}$ and $f_l : M \rightarrow \bar{R}$, $l = 1, \dots, K$, be lower semicontinuous functions and let $f_l : M \rightarrow R$, $l = K + 1, \dots, L$, be continuous functions. Suppose that f is nondecreasing for each of its first K variables. Then either:*

(A1) *there exist $v_l^\infty \in \partial^\infty(\tau_l f_l)(m)$, $l = 1, \dots, L$, not all zero such that*

$$0 = \sum_{l=1}^L v_l^\infty,$$

or there exists $0 \neq \mu^\infty \in \partial^\infty f(f_1(m), \dots, f_L(m))$ such that, for any $\mu_i^\infty \neq 0$, one has $0 \in \partial f_l(m)$,

or there exist $\mu = (\mu_1, \dots, \mu_L) \in \partial f(f_1(m), \dots, f_L(m))$ such that

(A2)

$$\partial f(f_1, \dots, f_L)(m) \subset \sum_{k \in \{l: \mu_l \neq 0\}} \tau_k \mu_k \partial(\tau_k f_k)(m) + \sum_{k \in \{l: \mu_l = 0\}} \partial^\infty(\tau_k f_k)(m)$$

and

$$\partial^\infty f(f_1, \dots, f_L)(x) \subset \sum_{l=1}^L \partial^\infty(\tau_l f_l)(x).$$

It is well known that as special cases of the chain rule one can deduce a sum rule, a product rule, a quotient rule and a subdifferential formula for maximum functions (see e.g. [13, 49, 55, 56, 59]).

Theorem 4.13 (Sum rule). *Let $f_1, \dots, f_L : M \rightarrow \bar{R}$ be lower semicontinuous functions. Then, for any $m \in M$, either*

$$(A1) \quad \partial\left(\sum_{l=1}^L f_l\right)(m) \subset \sum_{l=1}^L \partial f_l(m)$$

and

$$\partial^\infty\left(\sum_{l=1}^L f_l\right)(m) \subset \sum_{l=1}^L \partial^\infty f_l(m),$$

or there exist $v_l^\infty \in \partial^\infty f_l(m)$, $l = 1, \dots, L$, not all zero such that

$$(A2) \quad 0 = \sum_{l=1}^L v_l^\infty.$$

Proof. Apply the chain rule of Theorem 4.12 to $f(y_1, \dots, y_L) = \sum_{l=1}^L y_l$. Note that f is nondecreasing in any of its variables, $\tau_l = 1$ for all l . Moreover,

$$\partial f(f_1(m), \dots, f_L(m)) = \{(1, 1, \dots, 1)\}.$$

The conclusion follows from direct calculation. \square

Similarly, applying the chain rule of Theorem 4.12 to $f(y_1, \dots, y_L) = \prod_{l=1}^L y_l$ and $f(y_1, y_2) = y_1/y_2$ yields the following product rule and quotient rule.

Theorem 4.14 (Product rule). *Let $f_l : M \rightarrow \bar{R}$, $l = 1, \dots, L$, be nonnegative lower semicontinuous functions. Then either:*

(A1) *there exist $v_l^\infty \in \partial^\infty(\tau_l f_l)(m)$, $l = 1, \dots, L$, not all zero such that*

$$0 = \sum_{l=1}^L v_l^\infty,$$

or

(A2)

$$\partial(f_1 \cdots f_L)(m) \subset \sum_{l=1}^L f_1(m) \cdots f_{l-1}(m) \cdot f_{l+1}(m) \cdots f_L(m) \partial(f_l)(m)$$

and

$$\partial^\infty(f_1 \cdots f_L)(m) \subset \sum_{l=1}^L f_1(m) \cdots f_{l-1}(m) \cdot f_{l+1}(m) \cdots f_L(m) \partial^\infty(f_l)(m).$$

In the following quotient rule $\tau \in \{-1, 1\}$.

Theorem 4.15 (Quotient rule). *Let $f_1 : M \rightarrow \bar{R}$ be a lower semicontinuous function and let $f_2 : M \rightarrow R$ be a continuous function with $f_2(m) > 0$. Then either:*

(A1) *there exist $v_1^\infty \in \partial^\infty(f_1)(m)$ and $v_2^\infty \in \partial^\infty(\tau f_2)(m)$ not all zero such that*

$$0 = v_1^\infty + v_2^\infty,$$

or

(A2)

$$\partial(f_1/f_2)(m) \subset \frac{f_2(m)\partial(f_1)(m) - f_1(m)\partial(\tau f_2)(m)}{[f_2(m)]^2}$$

and

$$\partial^\infty(f_1/f_2)(m) \subset \frac{f_2(m)\partial^\infty(f_1)(m) - f_1(m)\partial^\infty(\tau f_2)(m)}{[f_2(m)]^2}.$$

When the function f in the chain rule is the maximum of its variables we have the following result.

Theorem 4.16 (Subdifferential of the maximum). *Let $f_l : M \rightarrow \bar{R}$, $l = 1, 2, \dots, L$, be lower semicontinuous functions. Then either:*

(A1) *there exist $v_n^\infty \in \partial^\infty f_l(m)$, $l = 1, \dots, L$, not all zero such that*

$$0 = \sum_{l=1}^L v_l^\infty,$$

or there exist $\mu_l \geq 0$ with $\sum_{l=1}^L \mu_l = 1$ such that

(A2)

$$\partial \max(f_1, \dots, f_L)(m) \subset \sum_{k \in \{l: \mu_l \neq 0\}} \mu_k \partial(f_k)(m) + \sum_{k \in \{l: \mu_l = 0\}} \partial^\infty(f_k)(m)$$

and

$$\partial^\infty \max(f_1, \dots, f_L)(x) \subset \sum_{l=1}^L \partial^\infty f_l(x).$$

The following is a useful corollary for the subdifferential of the positive part $f_+ := \max(f, 0)$ of a lower semicontinuous function f .

Corollary 4.17 (Subdifferential of the positive part). *Let $f : M \rightarrow \bar{R}$, $l = 1, 2, \dots, L$, be a lower semicontinuous function. Then, there exists $\lambda \in [0, 1]$ such that*

$$\partial f_+(m) \cup \partial^\infty f_+(m) \subset \lambda \partial f(m) \cup \partial^\infty f(m).$$

5. APPLICATIONS

We now turn to several applications of the sub- and superdifferential concepts and their calculus.

5.1. Lipschitz criterion. We start with a subdifferential criterion for a function on a Riemannian manifold to be Lipschitz. It is a generalization of the well known result that a C^1 function f on an open convex set U in a Banach space is Lipschitz with rank K if and only if $\|f'(x)\| \leq K$ for all $x \in U$. Forms of this criterion for nonsmooth functions can be found in [30, 63, 70]. To state this result we need a concept of convexity on a manifold. We say that an open subset U of a smooth Riemannian manifold M is convex provided that, for any two points $m_1, m_2 \in U$, there exist $\varepsilon_0 = \varepsilon_0(m_1, m_2) > 0$ such that, for any $\varepsilon \in (0, \varepsilon_0)$, there exist a C^1 regular curve $r : [0, 1] \rightarrow M$ with $\int_0^1 \|r'(t)\| dt < \rho(m_1, m_2) + \varepsilon$ and $r([0, 1]) \subset U$. We say that M is locally convex if, for any $m \in M$ and any neighborhood V of m , there exists a convex neighborhood U of m such that $U \subset V$. Roughly speaking we require that any ‘approximate’ geodesics connecting two points of a convex open set U remains in U . When M has the property that there exist geodesics between any two points, this is equivalent to requiring that for any two points in U a geodesic joining them remains in U (see [3, 21] for discussion about geodesics). When M is a Euclidean space this definition of convexity coincides with the usual convexity concept.

Definition 5.1. Let (M, g) be a Riemannian manifold with distance ρ , and U is an open subset of M . We say that $f : U \rightarrow \bar{R}$ is Lipschitz with rank K provided that, for any $m_1, m_2 \in U$,

$$|f(m_1) - f(m_2)| \leq K\rho(m_1, m_2).$$

We say that f is locally Lipschitz at $m \in M$ provided that f is Lipschitz in a neighborhood of m .

Theorem 5.2. *Let (M, g) be a Riemannian manifold with distance ρ , and U is a convex open subset of M . Let $f : U \rightarrow \bar{R}$ be a lower semicontinuous function. Then*

the following are equivalent:

- (i) f is Lipschitz with rank K ;
- (ii) for any $m \in U$ and $v^* \in \partial_F f(m)$, $\|v^*\| \leq K$;
- (iii) for any $m \in U$ and $v^* \in \partial f(m)$, $\|v^*\| \leq K$.

Proof. (i) \Rightarrow (ii): Let $v^* \in \partial_F f(m)$. Then there exists $h \in C^1(M)$ with $dh(m) = v^*$ such that $f - h$ attains a local minimum at m . Let $v \in T_m(M)$ be an arbitrary vector and let $r : (-1, 1) \rightarrow M$ be a C^1 curve with $r'(0) = v$. Then, for t close to 0,

$$h(r(t)) - h(m) \leq f(r(t)) - f(m) \leq K\rho(r(t), m) \leq K \int_0^t \|r'(s)\| ds.$$

Dividing both sides by t and taking limits as $t \rightarrow 0$, we have

$$\langle v^*, v \rangle = \langle dh(m), r'(0) \rangle \leq K\|v\|.$$

Since v is arbitrary, we have $\|v^*\| \leq K$.

(ii) \Rightarrow (iii) is obvious.

(iii) \Rightarrow (i): We prove by contradiction. Suppose f is not Lipschitz with rank K . Then there exist $m_1, m_2 \in M$ such that

$$|f(m_2) - f(m_1)| > K\rho(m_2, m_1).$$

Without loss of generality we may assume that

$$f(m_2) - f(m_1) > K\rho(m_2, m_1).$$

Then there exists a C^1 regular curve $r : [0, 1] \rightarrow M$ with $r(0) = m_1$ and $r(1) = m_2$ such that

$$f(r(1)) - f(r(0)) > K \int_0^1 \|r'(s)\| ds.$$

Set

$$\phi(t) := f \circ r(t) - K \int_0^t \|r'(s)\| ds.$$

Since $\phi(1) - \phi(0) > 0$ there exists $t \in (0, 1)$ and $\xi \in \partial_F \phi(t)$ such that $\xi > 0$ (see e.g. [13, Theorem 4.4]). Note that $\xi \in \partial_F \phi(t)$ implies that

$$\xi + K\|r'(t)\| \in \partial_F(f \circ r)(t)$$

or

$$(K + \eta)\|r'(t)\| \in \partial_F(f \circ r)(t) \subset \partial(f \circ r)(t),$$

where $\eta = \xi/\|r'(t)\| > 0$. We rewrite this inclusion as

$$(K + \eta)\|r'(t)\| \in \partial(f \circ \psi^{-1}) \circ (\psi \circ r)(t),$$

where ψ is a local coordinate system around $r(t)$.

Now apply the chain rule of Theorem 4.12 to the composition of $f \circ \psi^{-1}$ and $\psi \circ r$. Since $\psi \circ r$ is C^1 alternative (A1) is impossible and alternative (A2) gives us

$$u^* \in \partial(f \circ \psi^{-1})(\psi(r(t)))$$

such that

$$(K + \eta)\|r'(t)\| = \langle u^*, (\psi \circ r)'(t) \rangle = \langle u^*, \psi_{*r(t)} r'(t) \rangle = \langle \psi_{r(t)}^* u^*, r'(t) \rangle.$$

It follows that

$$\|\psi_{r(t)}^* u^*\| \geq K + \eta > K.$$

On the other hand, it follows from Corollary 4.2 that $\psi_{r(t)}^* u^* \in \partial f(r(t))$, which is a contradiction. \square

Applying the above theorem to a locally convex manifold we have:

Corollary 5.3. *Let (M, g) be a locally convex Riemannian manifold with distance ρ . Let $f : M \rightarrow \bar{R}$ be a lower semicontinuous function. Then the following are equivalent:*

- (i) f is locally Lipschitz at $m \in M$;
- (ii) $\partial_F f$ is bounded in a neighborhood of m ;
- (iii) ∂f is bounded in a neighborhood of m ;
- (iv) $\partial f(m)$ is bounded;
- (v) $\partial^\infty f(m) = \{0\}$.

5.2. Implicit function theorem. Next we prove an implicit function theorem for a general lower semicontinuous function on a Riemannian manifold. We also derive a metric estimate for the implicit multifunction. We need the following notation. Let M be a manifold and P a parametric set. Consider a function $f : M \times P \rightarrow \bar{R}$. We denote the positive part of f by $f_+ := \max(f, 0)$. We use $G(p)$ to denote the implicit multifunction determined by $f(m, p) \leq 0$, i.e.,

$$G(p) := \{m \in M : f(m, p) \leq 0\}.$$

In this section $\partial f(m, p)$ and $\partial_F f(m, p)$ signify the limiting subdifferential and the Fréchet subdifferential with respect to variable m . Now we can state our implicit function theorem.

Theorem 5.4. *Let (M, g) be a locally convex Riemannian manifold with distance ρ , let P be a metric space and let U be an open set of $M \times P$. Consider $f : U \times P \rightarrow \bar{R}$ that satisfies the following conditions:*

- (i) *there exists $(\bar{m}, \bar{p}) \in U$ such that*

$$f(\bar{m}, \bar{p}) \leq 0;$$

- (ii) *function $p \rightarrow f(\bar{m}, p)$ is upper semicontinuous at \bar{p} ;*
- (iii) *for any p near \bar{p} , function $m \rightarrow f(m, p)$ is lower semicontinuous;*
- (iv) *there exists $\sigma > 0$ such that for any $(m, p) \in U$ with $f(m, p) > 0$, $\xi \in \partial f(m, p)$ implies that $\|\xi\| \geq \sigma$.*

Then there exist open sets $W \subset M$ and $V \subset P$ containing \bar{m} and \bar{p} respectively such that

- (a) *for any $p \in V$, $W \cap G(p) \neq \emptyset$;*
- (b) *for any $p \in V$ and $m \in W$,*

$$d(m, G(p)) \leq \frac{f_+(m, p)}{\sigma}.$$

Proof. We need only prove (b). For if (b) holds, by condition (ii) for any $\eta > 0$ sufficiently small there exists an open set $V \subset P$ containing \bar{p} such that

$$\sigma d(\bar{m}, G(p)) \leq f_+(\bar{m}, p) < \eta, \quad \forall p \in V.$$

Take η small so that $W := \{m : \rho(m, \bar{m}) < \eta\}$ and let V satisfy $W \times V \subset U$. Then, for any $p \in V$, $W \cap G(p) \neq \emptyset$.

We prove (b) by way of contradiction. Suppose that (b) does not hold. Then, there exists a sequence $(m_n, p_n) \in U$ converging to (\bar{m}, \bar{p}) such that

$$d(m_n, G(p_n)) > \frac{1}{\sigma} f_+(m_n, p_n) = \frac{1}{\sigma} f(m_n, p_n) > 0.$$

Set $\varepsilon = f(m_n, p_n)$ and choose $\lambda \in (f_+(m_n, p_n)/\sigma, d(m_n, G(p_n)))$. Then

$$f_+(m_n, p_n) \leq \inf_{m \in M} f_+(m, p_n) + \varepsilon.$$

Note that f_+ is lower semicontinuous with respect to m . Invoking the Ekeland variational principle [39] there exists $v_n \in M$ such that

$$\rho(m_n, v_n) \leq \lambda$$

and

$$(5.1) \quad f_+(v_n, p_n) \leq f_+(m, p_n) + \frac{\varepsilon}{\lambda} \rho(m, v_n).$$

Since $\rho(m_n, v_n) \leq \lambda < d(m_n, G(p_n))$, $v_n \notin G(p_n)$. Therefore $f(v_n, p_n) > 0$. Thus, for m close enough to v_n , $f(m, p_n) > 0$. It follows from (5.1) that

$$m \rightarrow f(m, p_n) + \frac{\varepsilon}{\lambda} \rho(m, v_n)$$

attains a local minimum at $m = v_n$. By the sum rule of Theorem 4.13 we have

$$0 \in \partial f(v_n, p_n) + \frac{\varepsilon}{\lambda} \partial \rho(v_n, v_n).$$

In other words there exists $v^* \in \partial \rho(v_n, v_n)$ such that

$$u^* = -\frac{\varepsilon}{\lambda} v^* \in \partial f(v_n, p_n).$$

Note that $m \rightarrow \rho(m, v_n)$ is a Lipschitz function with rank 1. We have $\|v^*\| \leq 1$ by Theorem 5.2. Thus, we found an element $u^* \in \partial f(v_n, p_n)$ with

$$\|u^*\| \leq \frac{\varepsilon}{\lambda} < \frac{f(m_n, p_n)}{f(m_n, p_n)/\sigma} = \sigma,$$

a contradiction. □

5.3. Subdifferential of spectral functions. Let M be the space of complex $N \times N$ matrices. For any $X \in M$ we use $\lambda_1(X), \dots, \lambda_N(X)$ to denote the N (including repeated) eigenvalues of X in the lexicographical order of real and imaginary parts. We call $\lambda(X) := (\lambda_1(X), \dots, \lambda_N(X))$ the eigenvalue mapping. A spectral function is a function of the form $\phi := f \circ \lambda : M \rightarrow \bar{\mathbb{R}}$ where f is invariant under permutation of its variables. The concept of a spectral function encompasses many useful functions related to the eigenvalue mapping such as the *spectral abscissa* $\max\{\operatorname{Re} \lambda_n, n = 1, \dots, N\}$, the *spectral radius* $\max\{|\lambda_n|, n = 1, \dots, N\}$, the *determinant* \det and the *trace* tr . These functions are often intrinsic nonsmooth which makes analyzing their properties difficult. Recently in a series of papers, J. Burke, A. Lewis, and M. Overton made a breakthrough in the analysis of various subdifferentials of the spectral functions (see [19, 20, 50] and their references). In this section we illustrate how to understand their work by using the nonsmooth analysis tools on manifolds established in the previous sections.

As usual we define an inner product on M by

$$\langle X, Y \rangle := \operatorname{tr}(X^* Y) = \sum_{n,k=1}^N \overline{x_{n,k}} y_{n,k}, \quad X, Y \in M,$$

and norm $\|X\| := \sqrt{\langle X, X \rangle}$. We use $GL(N)$ and $O(N)$ to denote the set of all invertible and orthogonal matrices in M , respectively. For $U \in GL(N)$ ($U \in O(N)$) we define a mapping $u : M \rightarrow M$ by

$$u(X) = U^{-1}XU \quad (u(X) = U^*XU).$$

Then u is a diffeomorphism (in fact, a linear invertible transform). Moreover, it is easy to calculate that $u^* : T^*(M) \rightarrow T^*(M)$ is defined by $u^*(Y) = UYU^{-1}$ ($u^*(Y) = UYU^*$). Let ϕ be a spectral function on M and $Z = u(X)$ for $U \in GL(N)$ (or $U \in O(N)$). It follows directly from the smooth chain rule of Theorem 4.1 that

$$u^*\tilde{\partial}\phi(Z) = \tilde{\partial}(\phi \circ u)(X),$$

where $\tilde{\partial} = \partial_F, \partial$ or ∂^∞ . It is easy to see that $\phi \circ u = \phi$. Thus, we have the following lemma.

Lemma 5.5. *Let ϕ be a spectral function on M and let $U \in GL(N)$ ($U \in O(N)$). Then*

$$\begin{aligned} \tilde{\partial}\phi(U^{-1}XU) &= U^{-1}\tilde{\partial}\phi(X)U, \\ (\tilde{\partial}\phi(U^*XU) &= U^*\tilde{\partial}\phi(X)U), \end{aligned}$$

where $\tilde{\partial} = \partial_F, \partial$ or ∂^∞ .

Next we consider the $GL(N)$ and $O(N)$ orbit of $X \in M$ defined by $GL(N) \cdot X := \{U^{-1}XU : U \in GL(N)\}$ and $O \cdot X := \{U^*XU : U \in O(N)\}$, respectively. It is well known that $GL(N) \cdot X$ and $O(N) \cdot X$ are submanifolds of M , and their tangent and normal spaces at X are described in the lemma below.

Lemma 5.6. *Let $X \in M$. Then*

$$T_X(O(N) \cdot X) = T_X(GL(N) \cdot X) = \{XY - YX : Y \in M\}$$

and

$$N_F(O(N) \cdot X; X) = N_F(GL(N) \cdot X; X) = \{Y \in M : XY^* - Y^*X = 0\}.$$

Lemma 5.7. *Let ϕ be a spectral function on M . Then $Y \in \tilde{\partial}\phi(X)$ implies that $XY^* = Y^*X$, where $\tilde{\partial} = \partial_F, \partial$ or ∂^∞ .*

Proof. We need only to prove the case when $\tilde{\partial} = \partial_F$. The rest follows by a limiting process. Observe that by the definition of the Fréchet subdifferential we have

$$\partial_F\phi(X) \subset N_F(\phi^{-1}(-\infty, \phi(X)); X).$$

Since ϕ is a constant on $O(N) \cdot X$, we have $O(N) \cdot X \subset \phi^{-1}(-\infty, \phi(X))$. Thus,

$$\partial_F\phi(X) \subset N_F(O(N) \cdot X; X).$$

The lemma follows from the representation of $N_F(O(N) \cdot X; X)$ in Lemma 5.6. \square

Now we can deduce the key result that can help us understand the relationship between the subdifferential of a spectral function $\phi = f \circ \lambda$ and that of f . For any $X \in M$, we will use $\text{diag } X$ to denote the diagonal of X as a vector in C^N , i.e., $\text{diag } X = (x_{11}, \dots, x_{NN})$.

Theorem 5.8. *Let ϕ be a spectral function on M and let $Y \in \tilde{\partial}\phi(X)$. Then there exists a matrix $U \in O(N)$ such that $T = U^*XU$ is an upper triangle with $\text{diag } T = \lambda(X)$ and $S = U^*YU$ is a lower triangle and*

$$(5.2) \quad S \in \tilde{\partial}\phi(T),$$

where $\tilde{\partial} = \partial_F, \partial$ or ∂^∞ .

Proof. By Lemma 5.7 $XY^* = Y^*X$. The existence of U as described in the theorem is a direct consequence of the Schur Theorem. The conclusion then follows from Lemma 5.5. \square

Similar results holds when M is the space of Hermitian matrices. In this case S and T are both diagonal matrices. Suppose that $\phi = f \circ \lambda$. A. Lewis showed in [50] that if S and T are diagonal, then (5.2) is equivalent to

$$\text{diag } S \in \partial f(\lambda(X)).$$

J. Burke and M. Overton [19, 20] showed that the necessity holds for general S and T . While these relationships are highly nontrivial we can see that they can be understood as nonsmooth chain rules on the smooth manifolds of matrices.

6. DIFFERENTIAL INCLUSIONS ON MANIFOLDS: MONOTONICITY AND INVARIANCE

We now turn to discuss differential inclusions on manifolds which is a convenient generic form of representing control systems and dynamic systems with uncertainty. Consider a multifunction F defined on the manifold M whose values are compact convex sets of the corresponding tangent space, namely,

$$F(m) \subset T_m(M), \quad \forall m \in M.$$

We say F is upper semicontinuous at m provided that $m_i \rightarrow m$ and $F(m_i) \ni v_i \rightarrow v$ implies that $v \in F(m)$. In this section we always assume that F is an upper semicontinuous convex compact valued multifunction. The differential inclusion

$$(6.1) \quad \dot{c}(t) \in F(c(t)), \quad c(0) = m_0,$$

is a well known object (see [1, 33]).

Definition 6.1. An absolutely continuous function $c : [0, T] \rightarrow M$ is called a solution of the differential inclusion if the inclusion (6.1) holds for almost all (a.a.) $t \in [0, T]$.

Let (U, ψ) be a local coordinate neighborhood around m_0 . Then it is not hard to see that an absolute continuous function c is a local solution to (6.1) if and only if $\psi \circ c$ is a local solution to the differential inclusion

$$(6.2) \quad \frac{d(\psi \circ c)(t)}{dt} \in \psi_* F(\psi^{-1}((\psi \circ c)(t))), \quad \psi \circ c(0) = \psi(m_0).$$

Existence theorems of local solutions to a differential inclusion on manifolds follows directly from corresponding results for differential inclusions in Euclidean spaces. The following is an example:

Theorem 6.2. *Let F be an upper semicontinuous multifunction with compact and convex values. Then, for any $v \in F(m_0)$, the differential inclusion (6.1) has a local solution c with $c(0) = m_0$ and $c'(0) = v$.*

Proof. Let $v(m')$ be a local extension of v as described in Lemma 3.7 and let $u(m')$ be the projection of $v(m')$ on $F(m')$. Then $u(\cdot)$ is a (local) continuous selection of $F(\cdot)$ and $u(m) = v$. We need only to take c be the solution of

$$\dot{c}(t) = u(c(t)), c(0) = m_0.$$

□

We will focus on monotonicity of a differential inclusion with respect to an extended valued lower semicontinuous function. The essentially equivalent results about invariance of a differential inclusion with respect to a closed set is also discussed. These properties play important roles in many problems related to a differential inclusion. A good reference is the survey paper [28].

6.1. Strong invariance and monotonicity. For a differential inclusion the related concepts of invariance with respect to a set and monotonicity with respect to a function play important roles in the analysis of such systems. We start with the strong invariance and monotonicity.

Definition 6.3. The function $\varphi : M \rightarrow \bar{R}$ is called strong monotone decreasing with respect to solutions of (6.1) if for any initial point $m_0 \in M$ and any solution c of (6.1) we have

$$(6.3) \quad \varphi(c(t)) \leq \varphi(c(0)), \quad \forall t \in [0, \tau(c)),$$

where $[0, \tau(c))$ is the maximal interval of existence of the solution c .

Definition 6.4. $S \subset M$ is called strong invariant with respect to solutions of (6.1) if for any initial point $m_0 \in M$ and any solution c of (6.1) we have

$$(6.4) \quad \varphi(c(t)) \in S, \quad \forall t \in [0, \tau(c)),$$

where $[0, \tau(c))$ is the maximal interval of existence of the solution c .

It is easy to see that these two concepts are closely related. S is strongly invariant if its indicator function δ_S is strongly monotone decreasing and function φ is strongly monotone if all its level sets

$$\{m \in M : \varphi(m) \leq \varphi(m_0)\}$$

are strongly invariant.

Below we give an infinitesimal characterization of these properties in terms of the upper Hamiltonian

$$(6.5) \quad H(m, p) := \sup_{v \in F(m)} \langle p, v \rangle,$$

where $p \in T_m^*(M)$ and $\langle \cdot, \cdot \rangle$ is the pairing between elements of $T_m(M)$ and its dual $T_m^*(M)$.

It is well known that in the case of $M = R^N$ we need some additional assumptions to ensure an infinitesimal characterization of strong invariance, e.g. Lipschitz behavior of the multifunction F . Here we use a similar condition in terms of the upper Hamiltonian. In the reminder of this paper we always assume that M is a Riemannian manifold with a metric ρ .

Assumption (H1). For any $m \in M$ there exists a neighborhood U of m and a constant L such that, for any $m' \in U$ and any $\varphi \in C^\infty(M)$,

$$(6.6) \quad H(m', d\varphi(m')) - H(m, d\varphi(m)) \leq L\rho(m', m) \sup_{m'' \in U} \|d\varphi(m'')\|.$$

For $\varphi \in C^\infty(M)$, the strong monotonicity are obviously characterized by

$$(6.7) \quad H(m, d\varphi(m)) \leq 0, \quad \forall m \in M.$$

We show below that analogous infinitesimal conditions hold for a lower semicontinuous function φ with Fréchet subgradients replacing the derivatives of φ in (6.7).

Theorem 6.5. *Let F satisfy Assumption (H1) and let $\varphi : M \rightarrow \bar{R}$ be a lower semicontinuous function. Then the following are equivalent:*

- (i.) φ is strongly monotone decreasing with respect to the solutions of (6.1).
- (ii.) $H(m, p) \leq 0$, for any $m \in M$ and $p \in \partial_F \varphi(m)$.
- (iii.) $H(m, p) \leq 0$, for any $m \in M$ and $p \in \partial \varphi(m) \cup \partial^\infty \varphi(m)$.

Proof. (i.) implies (ii.) Suppose that φ is strongly monotone decreasing with respect to the solutions of (6.1) and $p \in \partial_F \varphi(m)$. Choose $g \in C^1(M)$ such that $\varphi - g$ attains a local minimum at m and $dg(m) = p$. For any $v \in F(m)$, by Theorem 6.2, there exists a (local) solution c of (6.1) that satisfies initial condition $c(0) = m$ and $c'(0) = v$. Then, for $t > 0$ we have $\varphi(c(t)) - g(c(t)) \geq \varphi(c(0)) - g(c(0))$ or $0 \geq g(c(t)) - g(c(0))$. Dividing by t and taking limits when $t \rightarrow 0+$ we have $\langle p, c'(0) \rangle = \langle p, v \rangle \leq 0$. Since we can take v to be any elements of $F(m)$ we have $H(m, p) \leq 0$.

(ii.) implies (iii.) Note that H is continuous and positive homogeneous in p . This follows directly from definitions of the limiting and singular subdifferentials.

(iii.) implies (i.) First, observe that it is enough to show that

$$(6.8) \quad \varphi(c(t)) \leq \varphi(c(0))$$

for $t > 0$ small enough so that $c(t)$ stays in a local coordinate neighborhood U where (U, ψ) is a local coordinate system with local coordinates (x^1, x^2, \dots, x^N) . Moreover, shrinking U if necessary we may assume that $\psi(U)$ is a convex subset of R^N . Second, by shifting a constant if necessary we may assume that $\varphi(c(0)) = 0$ without loss of generality.

For $\alpha > 0$, define

$$\varphi_\alpha(m) := \min_{m' \in U} (\varphi_+(m') + \frac{1}{2\alpha^2} |\psi(m') - \psi(m)|^2).$$

It is not difficult to see that function $\varphi_\alpha(m)$ is locally Lipschitz which implies that along any solution c of (6.1) function $\varphi_\alpha(c(t))$ is absolutely continuous. Then at a differentiable point of both $\varphi_\alpha(c(t))$ and $c(t)$ we have

$$(6.9) \quad \begin{aligned} \frac{d\varphi_\alpha(c(t))}{dt} &\leq \lim_{h \rightarrow 0+} \frac{1}{2\alpha^2} [|\psi(c(t+h)) - \psi(m')|^2 - |\psi(c(t)) - \psi(m')|^2] \\ &= \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t)) - x^n(m')) \langle dx^n(c(t)), \dot{c}(t) \rangle, \end{aligned}$$

where m' is a minimum point for the function

$$m \rightarrow \varphi_+(m) + \frac{1}{2\alpha^2} |\psi(m) - \psi(c(t))|^2.$$

This implies that

$$p' := \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t)) - x^n(m')) dx^n(m') \in \partial_F \varphi_+(m').$$

Since H is positive homogeneous in p , combining (iii.) and Corollary 4.17 we have

$$(6.10) \quad H(m', p') \leq 0.$$

Denote

$$p := \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t)) - x^n(m')) dx^n(c(t)).$$

Using (6.10) and Assumption (H1), we have the following estimate:

$$\begin{aligned} \frac{d\varphi_\alpha(c(t))}{dt} &\leq \langle p, \dot{c}(t) \rangle \leq H(c(t), p) \\ &\leq H(c(t), p) - H(m', p') \\ (6.11) \quad &= H(c(t), \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t)) - x^n(m')) dx^n(c(t))) \\ &\quad - H(m', \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t)) - x^n(m')) dx^n(m')) \\ &\leq \frac{NL}{\alpha^2} |\psi(c(t)) - \psi(m')| \rho(c(t), m') \sup_{m'' \in U} \|d\psi(m'')\|. \end{aligned}$$

We need the following simple estimate

Lemma 6.6. *For $m_1, m_2 \in U$, one has*

$$\rho(m_1, m_2) \leq \sup_{x \in \psi(U)} \|d\psi^{-1}(x)\| |\psi(m_1) - \psi(m_2)|.$$

Proof. Let $c(t) = t\psi(m_2) + (1-t)\psi(m_1)$. Since $\psi(U)$ is convex, $c[0, 1] \subset \psi(U)$. Defining $r(t) := (\psi^{-1} \circ c)(t)$ we have $r(0) = m_1$ and $r(1) = m_2$. Thus,

$$\begin{aligned} \rho(m_1, m_2) &\leq \int_0^1 \|\dot{r}(t)\| dt \\ &\leq \|\langle d\psi^{-1}(c(t)), \psi(m_2) - \psi(m_1) \rangle\| \\ &\leq \sup_{x \in \psi(U)} \|d\psi^{-1}(x)\| |\psi(m_1) - \psi(m_2)|. \end{aligned}$$

Denote

$$k := 2NL \sup_{x \in \psi(U)} \|d\psi^{-1}(x)\| \sup_{m'' \in U} \|d\psi(m'')\|.$$

It follows from (6.11) that

$$\begin{aligned} \frac{d\varphi_\alpha(c(t))}{dt} &\leq k \frac{1}{2\alpha^2} |\psi(c(t)) - \psi(m')|^2 \\ &\leq k [\frac{1}{2\alpha^2} |\psi(c(t)) - \psi(m')|^2 + \varphi_+(m')] \\ (6.12) \quad &= k\varphi_\alpha(c(t)). \end{aligned}$$

Invoking Gronwall's inequality we obtain

$$\varphi_\alpha(c(t)) \leq e^{kt} \varphi_\alpha(c(0)) = 0.$$

Since $\lim_{\alpha \rightarrow 0+} \varphi_\alpha(c(t)) = \varphi_+(c(t))$ we arrive at $\varphi_+(c(t)) \leq 0$ which implies $\varphi(c(t)) \leq 0$. \square

Applying Theorem 6.5 to the indicator function of a set we have the following characterization of the strong invariance.

Theorem 6.7. *Let F satisfy Assumption (H1) and let S be a closed subset of M . Then the following are equivalent:*

- (i.) S is strongly invariant with respect to the solutions of (6.1).
- (ii.) $H(m, p) \leq 0$, for any $m \in M$ and $p \in N_F(S, m)$.
- (iii.) $H(m, p) \leq 0$, for any $m \in M$ and $p \in N(S, m)$.

6.2. Weak invariance and monotonicity. We now discuss weak invariance and monotonicity. They require only existence of a trajectory of the differential inclusion system that satisfies the invariant or monotonicity property.

Definition 6.8. *A function $\varphi : M \rightarrow \bar{R}$ is called weak monotone decreasing with respect to solutions of (6.1) if for any initial point $m_0 \in M$ there exist a $\tau > 0$ and a solution c of (6.1) on $[0, \tau)$ such that*

$$(6.13) \quad \varphi(c(t)) \leq \varphi(c(0)), \quad \forall t \in [0, \tau).$$

Definition 6.9. *A set $S \subset M$ is called weak invariant with respect to solutions of (6.1) if for any initial point $m_0 \in M$ there exist a $\tau > 0$ and a solution c of (6.1) on $[0, \tau)$ such that*

$$(6.14) \quad c(t) \in S, \quad \forall t \in [0, \tau).$$

As stated in the definition of the strong invariance these two concepts are closely related. A set S is weakly invariant if its indicator function δ_S is weakly monotone decreasing and a function φ is weakly monotone decreasing if all of its level sets

$$\{m \in M : \varphi(m) \leq \varphi(m_0)\}$$

are weakly invariant.

We characterize the weak monotonicity and the weak invariance property in terms of the lower Hamiltonian

$$(6.15) \quad h(m, p) := \inf_{v \in F(m)} \langle p, v \rangle.$$

The following assumption is needed.

Assumption (H2). F is bounded in the sense that for any function $\varphi \in C^1(M)$,

$$H(m, d\varphi(m)) < \infty, \forall m \in M.$$

Theorem 6.10. *Let F satisfy Assumption (H2) and let $\varphi : M \rightarrow \bar{R}$ be a lower semicontinuous function. Then the following are equivalent:*

- (i.) φ is weakly monotone decreasing with respect to the solutions of (6.1).
- (ii.) $h(m, p) \leq 0$, for any $m \in M$ and $p \in \partial_F \varphi(m)$.
- (iii.) $h(m, p) \leq 0$, for any $m \in M$ and $p \in \partial \varphi(m) \cup \partial^\infty \varphi(m)$.

Proof. (i.) implies (ii.) Suppose that φ is weakly monotone decreasing with respect to the solutions of (6.1). Let $p \in \partial_F \varphi(m)$ and let $g \in C^1(M)$ such that $\varphi - g$ attains

a local minimum at m and $dg(m) = p$. Consider a solution c of (6.1) with $c(0) = m$ satisfying $\varphi(c(t)) \leq \varphi(c(0))$ for all $t \in [0, \tau]$. Then

$$\int_0^t dg(c(r)) \cdot \dot{c}(r) dr = g(c(t)) - g(c(0)) \leq 0.$$

This implies that

$$\int_0^t h(c(r), dg(c(r))) dr \leq 0.$$

Note that the function $r \rightarrow h(c(r), dg(c(r)))$ is lower semicontinuous due to the upper semicontinuity of F . Thus,

$$h(m, p) = h(m, dg(m)) \leq \liminf_{t \rightarrow 0^+} \frac{1}{t} \int_0^t h(c(r), dg(c(r))) dr \leq 0.$$

(ii.) implies (iii.) Note that h is lower semicontinuous and positive homogeneous in p . This follows directly from definitions of the limiting and singular subdifferentials.

(iii.) implies (i.) Let us fix $m_0 \in \text{dom } \varphi$; then there exists a local coordinate neighborhood U of m_0 and a local coordinate system (U, ψ) with local coordinates (x^1, \dots, x^N) . Without loss of generality we can assume that for some positive r

$$U_1 := \text{cl } \psi^{-1}(x_0 + 3rB) \subset U$$

where $x_0 = \psi(m_0)$. Note that sets U_1 and

$$U_2 := \text{cl } \psi^{-1}(x_0 + rB)$$

are compact.

Remark 6.11. Since φ is lower semicontinuous we can assume that it is bounded from below on U_1 and, moreover, shifting φ by a constant if necessary, we can assume that it is positive on U_1 .

Consider solutions of the differential inclusion

$$(6.16) \quad \dot{c}(t) \in F(c(t)), \quad c(0) = m_0.$$

We show below that under Assumption (H2) there exists a constant $\tau > 0$ such all solutions of (6.16) exist on the interval $[0, \tau]$ and stay in U .

To show it we define a multifunction

$$\tilde{F}(x) := \psi_* F(\psi^{-1}(x)).$$

It is easy to see that \tilde{F} is convex-valued and upper semicontinuous. The following lemma demonstrates that \tilde{F} is bounded.

Lemma 6.12. *Let multifunction F be upper semicontinuous and Assumption (H2) hold. Then there exists a constant C_1 such that for any $x \in x_0 + 2rB$, $p \in R^N$ and $\tilde{v} \in \tilde{F}(x)$*

$$(6.17) \quad \langle \tilde{v}, p \rangle \leq C_1 \|p\|.$$

Proof. Let $\tilde{v} \in \tilde{F}(x)$; then there exists $v \in F(\psi(x))$ such that

$$\tilde{v} = \psi_* v.$$

This implies that

$$\begin{aligned} \langle \tilde{v}, p \rangle &= \langle v, \psi^* p \rangle = \langle (\sum_{n=1}^N p_n dx^n), v \rangle = \sum_{n=1}^N p_n \langle dx^n, v \rangle \\ &\leq \sum_{n=1}^N |p_n| H(\psi(x), dx^n) \leq \|p\| \sum_{n=1}^N \max_{m \in U_1} H(m, dx^n(m)). \end{aligned}$$

But this means (6.17). \square

It follows from this lemma that if $\tau := r/C_1$, then all solutions of the differential inclusion

$$(6.18) \quad \dot{x}(t) \in \tilde{F}(x(t)), \quad x(0) = x_0$$

exists on the interval $[0, \tau]$ and satisfy

$$x(t) \in x_0 + rB, \quad \forall t \in [0, \tau].$$

Also we have that any solution $x(t)$ of the differential inclusion

$$(6.19) \quad \dot{x}(t) \in \text{co } \tilde{F}(x(t) + \varepsilon B)$$

stays in $x_0 + rB$ on the interval $[0, \tau]$ for any $\varepsilon \in [0, r]$.

Now we note that any solution $c(t)$ of (6.16) is a lifting of some solution $x(t)$ of (6.18) which implies that $c(t)$ stays in U_2 on the interval $[0, \tau]$.

For $\alpha > 0$, consider the function φ_α analogous to the one defined before

$$\varphi_\alpha(m) := \min_{m' \in U_1} (\varphi(m') + \frac{1}{2\alpha^2} |\psi(m') - \psi(m)|^2).$$

We fix an arbitrary small positive α satisfying

$$\alpha < r/(1 + \sqrt{2\varphi(m_0)}).$$

In view of the Remark 6.11 it implies that for any $m \in U_2$ minimizer m' in the definition of $\varphi_\alpha(m)$ will be an interior point of U_1 .

We construct approximate solutions of (6.1) as follows: consider a uniform partition $\pi := \{t_k\}_{k=0}^K$ of the interval $[0, \tau]$ where $t_{k+1} - t_k = \delta$, $\delta := \tau/K$. We can assume that

$$\delta < \min\{\alpha/C_1, 2\alpha^2/C_1^2\tau\}.$$

We define the approximate solution $c_\pi(t)$ of the differential inclusion (6.1) on $[0, \tau]$ recursively as a lifting of the absolutely continuous arc $x_\pi : [0, \tau] \rightarrow U_2$,

$$(6.20) \quad c_\pi(t) = \psi^{-1}(x_\pi(t)),$$

where x_π is a solution of the differential inclusion

$$(6.21) \quad \dot{x}_\pi(t) \in \tilde{F}(x_\pi(t) + \Delta(\alpha)B)$$

and function $\Delta(\alpha)$ is defined as follows:

$$(6.22) \quad \Delta(\alpha) := (\sqrt{2(\varphi(m_0) + 1)} + 1)\alpha.$$

We determine $c_\pi(t)$ as follows: assume that an approximate solution $c_\pi(t)$ has been defined on $[0, t_k]$ and satisfies

$$(6.23) \quad c_\pi(t) \in U_2, \quad \text{and } \varphi_\alpha(c_\pi(t)) \leq \varphi(m) + C_1^2 \delta t / 2\alpha^2$$

on $[0, t_k]$.

We extend c_π to the interval $[t_k, t_{k+1}]$. Let m_k denote a point such that the function

$$m' \rightarrow \varphi(m') + \frac{1}{2\alpha^2} |\psi(m') - \psi(c(t_k))|^2$$

attains a local minimum. In view of the Remark 6.11 this implies that m_k is an interior point of U_1 and

$$p_k := \sum_{n=1}^N \frac{1}{\alpha^2} (x^n(c(t_k)) - x^n(m_k)) dx^n(m_k) \in \partial_F \varphi(m_k).$$

Combining condition (iii) and Corollary 4.17 we have

$$h(m_k, p_k) \leq 0.$$

Next we find $v_k \in F(m_k)$ such that

$$(6.24) \quad \langle p_k, v_k \rangle = h(m_k, p_k) \leq 0.$$

Let $\tilde{v}_k := \psi_{*m_k} v_k$ and define

$$(6.25) \quad x_\pi(t) := x_\pi(t_k) + (t - t_k) \tilde{v}_k, \quad t \in [t_k, t_{k+1}].$$

Due to Lemma 6.12 we have that

$$\|\dot{x}_k(t)\| \leq C_1, \quad \forall t \in [0, t_{k+1}]$$

and

$$x_\pi(t) \in x_0 + rB, \quad \forall t \in [0, t_{k+1}].$$

This implies that the first relation in (6.23) holds on $[0, t_{k+1}]$.

Note that due to (6.23) and choice of δ we have

$$(6.26) \quad |x_\pi(t_k) - \psi(m_k)| = |\psi(c_\pi(t_k)) - \psi(m_k)|$$

$$(6.27) \quad \leq \sqrt{2\varphi(c_\pi(t_k))} \alpha \leq \sqrt{2(\varphi(m_0) + 1)} \alpha.$$

Now we estimate the increment of the function φ_α along this trajectory for $t \in [t_k, t_{k+1}]$. We use (6.20), (6.25), and (6.17) for this purpose

$$\begin{aligned} \varphi_\alpha(c_\pi(t)) - \varphi_\alpha(c_\pi(t_k)) &\leq \frac{1}{2\alpha^2} [|\psi(m_k) - \psi(c(t))|^2 - |\psi(m_k) - \psi(c(t_k))|^2] \\ &= \frac{1}{2\alpha^2} [2\langle \psi(c(t)) - \psi(c(t_k)), \psi(c(t_k)) - \psi(m_k) \rangle \\ &\quad + |\psi(c(t)) - \psi(c(t_k))|^2] \\ &\leq \frac{1}{2\alpha^2} [2\langle \tilde{v}_k, \psi(c(t_k)) - \psi(m_k) \rangle (t - t_k) + C_1^2 (t - t_k)^2] \\ &\leq \frac{1}{2\alpha^2} [\langle v_k, p_k \rangle (t - t_k) + C_1^2 (t - t_k) \delta]. \end{aligned}$$

Due to (6.24) we obtain that

$$(6.28) \quad \varphi_\alpha(c_\pi(t)) - \varphi_\alpha(c_\pi(t_k)) \leq C_1^2 (t - t_k) \delta / 2\alpha^2.$$

Using the second relation in (6.23) for $t = t_k$ we obtain from (6.28) that this relation is also valid for all $t \in [0, t_{k+1}]$.

To show that $x_\pi(t)$ is a solution of (6.21) on $[t_k, t_{k+1}]$ we use (6.26), (6.23) and the estimate on the choice of δ .

Thus, we proved that there exists an arc $c_\pi(t)$ satisfying (6.20), (6.25) and (6.23) on $[0, \tau]$.

Now, by choosing a sequence of partitions π_i with $\delta_i \rightarrow 0$ we can assume without loss of generality that x_{π_i} converges uniformly to some arc x which is a solution of the differential inclusion

$$\dot{x}(t) \in \text{co } \tilde{F}(x(t) + \Delta(\alpha))$$

and

$$\varphi_\alpha(c(t)) \leq \varphi(m_0),$$

where $c(t) = \psi^{-1}(x(t))$. Then by choosing a sequence of α_i and arcs x_i such that on $[0, \tau]$, $c_i(t) = \psi^{-1}(x_i(t))$ satisfies

$$\varphi_{\alpha_i}(x_i(t)) \leq \varphi(m_0).$$

Again without loss of generality we can assume that the sequence x_i converges uniformly to some arc x which is a solution of the differential inclusion (6.18). The corresponding lifting $c(t) = \psi^{-1}(x(t))$ is a solution of (6.1) satisfying (6.13). The theorem is proved. \square

For a closed set S , letting $\varphi = \delta_S$ we deduce characterizations of weak invariance as a corollary.

Corollary 6.13. *Let F satisfy Assumption (H2) and let S be a closed subset of M . Then the following are equivalent:*

- (i.) S is weakly invariant with respect to the solutions of (6.1).
- (ii.) $h(m, p) \leq 0$, for any $m \in M$ and $p \in N_F(m; S)$.
- (iii.) $h(m, p) \leq 0$, for any $m \in M$ and $p \in N(m; S)$.

7. OPTIMAL CONTROL PROBLEMS AND HAMILTON-JACOBI EQUATIONS ON MANIFOLDS

In this section we establish the existence and uniqueness of generalized lower semicontinuous solutions of Hamilton-Jacobi equations on smooth manifolds for the initial value problem

$$(7.1) \quad V_t(t, x) + h(x, V_x(t, x)) = 0, \quad V(\theta, x) = l(x),$$

and for the boundary value problem

$$(7.2) \quad h(x, V_x(x)) = -1, \quad V|_S = 0.$$

Here, h is the lower hamiltonian corresponding to a multifunction F , θ is a fixed time moment and S is a closed subset of M .

We show that infinitesimal criteria for invariance and monotonicity from the previous section can be used for comparing lower semicontinuous semisolutions of Hamilton-Jacobi equations (7.1) and (7.2) and to prove uniqueness of generalized solutions. Also these criteria are used to demonstrate that optimal value functions for corresponding optimal control problems coincide with these generalized solutions, which ensures their existence. We should mention that this approach and nonsmooth analysis tools developed in this paper can be used for studying generalized nonsmooth solutions of more general first-order partial differential equations on manifolds to provide analogues of the viscosity solution theory [4] and minimax solutions theory [66]. Our approach is based on methods from [29] which, in turn, is based on Subbotin's idea to use the invariance in the theory of generalized solutions of first-order partial differential equations.

As an application of these results we demonstrate that for any closed subset S of M the distance function $d_S(m)$ is a unique bilateral solution of the following eikonal equation:

$$(7.3) \quad \|\zeta\| = -1, \quad \forall \zeta \in \partial_F V(m), \quad V|_S = 0.$$

In particular, when $S = \{m_1\}$ this result provides an infinitesimal characterization of the Riemann metric $m \rightarrow \rho(m_1, m)$.

7.1. Initial value problem for a Hamilton-Jacobi equation. We consider a basic fixed time optimal control problem of minimizing the functional

$$l(c(\theta))$$

on the set of trajectories of a differential inclusion

$$(7.4) \quad \dot{c}(t) \in F(c(t)),$$

where the function $l : M \rightarrow R$ is lower semicontinuous and θ is a fixed moment of time.

We define an *optimal value function* V for this optimal control problem as follows:

$$(7.5) \quad V(t, m) = \inf_{c(\cdot)} l(c(\theta))$$

where infimum is taken over the set of all solutions c of (7.4) such that $c(t) = m$. Under sufficiently general assumptions V is lower semicontinuous and we demonstrate that V is a unique generalized solution of the initial value problem (7.1) for a Hamilton-Jacobi equation with the hamiltonian h defined in (6.15).

The concept of a generalized solution that we use here is based on the concept of a *bilateral* lower semicontinuous solution introduced by Barron and Jensen [5, 6] in the case of $M = R^N$. Let us we define super- and subsolutions of a Hamilton-Jacobi equation (7.1). The important distinction between them lies in the fact that subsolutions are also required to satisfy some differential inequality for $t = \theta$. We will often use the product space $R \times M$ (where M is a smooth manifold) in the sequel. For $(t_0, m_0) \in R \times M$ local coordinate neighborhoods are introduced naturally as follows: for any local ordinate system (U, φ) for m_0 and any $\delta > 0$ we define $((t_0 - \delta, t_0 + \delta) \times U, (t, \varphi))$ to be a local coordinate system for (t_0, m_0) .

Definition 7.1. A lower semicontinuous function $u : (-\infty, \theta] \times M \rightarrow R$ is called a supersolution of the initial value problem (7.1) if it satisfies the initial condition $u(\theta, m) = l(m)$ and the following inequality for all $(t, m) \in (-\infty, \theta) \times M$:

$$(7.6) \quad \langle u_t, \frac{\partial}{\partial t} \rangle + h(m, u_m) \leq 0, \quad \forall (u_t, u_m) \in \partial_F u(t, m).$$

Definition 7.2. A lower semicontinuous function $v : (-\infty, \theta] \times M \rightarrow R$ is called a subsolution of the initial value problem (7.1) if it satisfies the initial condition $v(\theta, m) = l(m)$ and the following inequality for all $(t, m) \in (-\infty, \theta] \times M$:

$$(7.7) \quad \langle v_t, \frac{\partial}{\partial t} \rangle + h(m, v_m) \geq 0, \quad \forall (v_t, v_m) \in \partial_F v(t, m).$$

Remark 7.3. We assume that super- and subsolutions take value $+\infty$ for all $t > \theta$.

The following definition of a bilateral lower semicontinuous solution implies that a subgradient of a solution should satisfy the Hamilton-Jacobi equation at points with $t < \theta$ in a remarkable resemblance of classical smooth solutions.

Definition 7.4. A lower semicontinuous function $u : (-\infty, \theta] \times M \rightarrow R$ is called a bilateral solution of the initial value problem (7.1) if it satisfies the initial condition $u(\theta, m) = l(m)$ and is a super- and subsolution simultaneously.

We prove an existence and uniqueness theorem for the bilateral lower semicontinuous solution for the initial value problem (7.1) under the following assumptions which combine compactness properties of a set of trajectories of differential inclusion and their prolongation.

Assumptions (H3). For any sequence of trajectories $c_i(\cdot)$ defined on $[0, t_i]$ with $c_i(0) \rightarrow m$ there exists a trajectory $c(\cdot)$ of (7.4) with $c(0) = m$ defined on $[0, \limsup t_i]$ and a subsequence $c_{i_j}(\cdot)$ of $c_i(\cdot)$ converging uniformly to $c(\cdot)$ on $[0, \tau]$ for any $\tau < \limsup t_i$.

We note that these assumptions can be replaced by some growth assumption on F and Assumptions (H1) as it is often done in the case of $M = R^N$. Moreover, Assumption (H3) also implies the existence of optimal trajectory for any initial point (t, m) if $V(t, m) < +\infty$. Finally, it is not hard to establish that the optimal value function is lower semicontinuous under these assumptions.

Theorem 7.5. Under Assumptions (H1), (H2) and (H3), for any lower semicontinuous function $l : M \rightarrow R$ bounded from below, the optimal value function V in (7.5) is the unique bilateral solution of the initial value problem for the Hamilton-Jacobi equation (7.1).

Proof. We start with a comparison result for a super- and a subsolution u and v , namely, we always have

$$(7.8) \quad v \leq u.$$

Obviously, this comparison result implies the uniqueness of the bilateral solution which is simultaneously a super- and a subsolution. \square

Let us consider on the manifold $\tilde{M} := R \times M$ the multifunction \tilde{F} consisting at any $(t, m) \in \tilde{M}$ of all pairs $(\frac{\partial}{\partial t}, \mathbf{v})$ where \mathbf{v} belongs to $F(m)$. It is clear that \tilde{F} satisfies Assumptions (H1), (H2) and (H3). Then the following lemma follows immediately from Theorems 6.10 and 6.5.

Lemma 7.6. The lower semicontinuous functions $u : (-\infty, \theta] \rightarrow R$ and $v : (-\infty, \theta] \rightarrow R$ are respectively super- and subsolutions if and only if they satisfy the initial condition $u(\theta, m) = l(m)$, $v(\theta, m) = l(m)$, and u is weakly decreasing and v is strongly predecreasing with respect to solutions $(t, c(t))$ of the differential inclusion with \tilde{F} on $(-\infty, \theta] \times M$.

Proof. It follows from Theorem 6.10 that u is weakly monotone decreasing with respect to solutions of differential inclusion \tilde{F} if and only if for the hamiltonian $h_{\tilde{F}}$ the following inequality holds for any $(t, m) \in (-\infty, \theta) \times M$:

$$h_{\tilde{F}}(t, m, \zeta) \leq 0, \quad \forall \zeta \in \partial_F u(t, m).$$

Computing this hamiltonian, we have that this condition coincides with the differential inequality in Definition 7.1.

Analogously, we observe that the strong monotone predecrease property (decreasing in reversed time) of the function v is equivalent to the strong monotone

decrease property in reverse time which means that v is strongly monotone decreasing with respect to the differential inclusion with the right-hand side $-\tilde{F}$. Thus, by Theorem 6.5, strong monotone predecrease property is equivalent to the following differential inequality which holds for any $(t, m) \in (-\infty, \theta] \times M$:

$$H_{(-\tilde{F})}(t, m, \zeta) \leq 0, \quad \forall \zeta \in \partial_F v(t, m).$$

But this inequality is exactly the differential inequality from Definition 7.2 of a subsolution. The lemma is proven. \square

Now we use this lemma to establish the following comparison result for super- and subsolutions and the optimal value function V which implies (7.8).

Lemma 7.7. *Let u and v be a supersolution and a subsolution, respectively; then, for any $(t, m) \in (-\infty, \theta] \times M$,*

$$(7.9) \quad v(t, m) \leq V(t, m) \leq u(t, m).$$

Proof. Consider an arbitrary point $(t, m) \in (-\infty, \theta] \times M$ and assume that $V(t, m) < +\infty$. It follows from the weak monotone decrease property of u that for any point there exists a trajectory of the differential inclusion \tilde{F} starting from (t, m) such that

$$u(\theta, c(\theta)) \leq u(t, m).$$

Due to the initial condition for u this means that

$$l(c(\theta)) \leq u(t, m)$$

which implies the inequality $V(t, m) \leq u(t, m)$. It is clear that in the case $V(t, m) = +\infty$ the same monotone decrease argument will imply that $u(t, m) = +\infty$.

Now we use the strong monotone predecrease property of v to derive that for any solution $c(\cdot)$ of the differential inclusion (7.4) with $c(t) = m$

$$v(t, m) \leq v(\theta, c(\theta)).$$

Due to the initial condition for v the last inequality implies

$$v(t, m) \leq l(c(\theta))$$

for any solution c starting from m at t . Of course, it means that $v(t, m) \leq V(t, m)$ which finishes the proof of this lemma.

As we have mentioned this comparison result (7.9) implies the uniqueness of the bilateral lower semicontinuous solution.

Now we demonstrate that the optimal value function V (7.5) is such a bilateral solution of the initial value problem (7.1) for a Hamilton-Jacobi equation. Namely, we show that V is weakly monotone decreasing and strongly monotone predecreasing with respect to solutions of the differential inclusion \tilde{F} and use Lemma 7.6 to derive that V is a super- and subsolution.

Indeed, for a given point (t, m) consider the optimal trajectory c of F such that $c(t) = m$. This implies that $V(t, c(t)) = l(c(\theta)) = V(t', c(t'))$ for any $t' \in [t, \theta]$. Thus, V has weak decrease property.

Now we consider an arbitrary solution which is a prolongation of this optimal trajectory c on the interval $(-\infty, t]$. For any $t' < t$ sufficiently close to t we have that $V(c(t')) \leq l(c(\theta)) = V(t, c(t))$ which implies that V is strongly monotone predecreasing.

Thus, V is a supersolution and subsolution simultaneously which implies that it is the unique bilateral solution. \square

7.2. Boundary value problem for a Hamilton-Jacobi equation. In this subsection, generalized lower semicontinuous solutions of the “eikonal” equation (7.2) are related with optimal value function for the following optimal time control problem of minimization of the functional:

$$(7.10) \quad \theta_S(c(\cdot)) := \min\{T > 0 : c(T) \in S\}$$

on a set of trajectories of a differential inclusion (7.4) where S is a closed subset of M . This functional takes the value $+\infty$ if a trajectory c doesn't intersect with the set S .

The optimal value, or *minimal time function*, V for this optimal control problem is defined as follows:

$$(7.11) \quad V(m) = \inf_{c(\cdot)} \theta(c(\cdot))$$

where the infimum is taken over the set of all solutions $c(\cdot)$ of (7.4) such that $c(0) = m$. Under sufficiently general assumptions function V is lower semicontinuous, and we demonstrate that V is a unique generalized solution of the boundary value problem (7.2) for the Hamilton-Jacobi equation with the hamiltonian h defined in (6.15).

The following concepts of sub- and supersolutions for the boundary value problem (7.2) will be used to define a bilateral lower semicontinuous solution. We show that these semisolutions provide lower and upper estimates for the optimal time function V .

Definition 7.8. A lower semicontinuous function $u : M \rightarrow R$ is called a supersolution of the boundary value problem (7.2) if it satisfies the boundary condition $u(m) = 0$ on S and the following inequality for all $m \in S^c := M \setminus S$:

$$(7.12) \quad h(m, u_m) \leq -1, \quad \forall u_m \in \partial_F u(m).$$

Definition 7.9. A lower semicontinuous function $v : M \rightarrow (-\infty, +\infty]$ is called a subsolution of the boundary value problem (7.2) if it satisfies the boundary condition $u(m) = 0$ and the following inequality for all $m \in \overline{S^c}$:

$$(7.13) \quad h(m, u_m) \geq -1, \quad \forall u_m \in \partial_F u(m).$$

The following definition of a bilateral lower semicontinuous solution implies that a subgradient of a solution should satisfy the Hamilton-Jacobi equation at points within S^c in a remarkable resemblance to the classical smooth solution of the eikonal equation (7.2).

Definition 7.10. A lower semicontinuous function $u : M \rightarrow (-\infty, +\infty]$ is called a bilateral solution of the boundary value problem (7.2) if it satisfies the boundary condition $u(m) = 0$ on S and is a sub- and a supersolution simultaneously.

The following assumption will provide existence and uniqueness of such bilateral solutions. It combines some variant of compactness assumption and escape time properties of trajectories to imply lower semicontinuity of minimal time function V .

For any trajectory $c(\cdot)$ and open subset U of M , define

$$\text{Esc}(c(\cdot), U) := \sup\{t > 0 : c(\cdot) \text{ exists on } [0, t] \text{ and lies in } U\}.$$

Assumptions (H4). (a) For any sequence of trajectories $c_i(\cdot)$ defined on $[0, t_i]$ and such that $c_i(0) \rightarrow m$ there exists a trajectory $c(\cdot)$ of (7.4) with $c(0) = m$ and a subsequence $c_{i_j}(\cdot)$ converging uniformly to $c(\cdot)$ on $[0, \theta']$ for any $\theta' < \limsup t_i$.

(b) For any trajectory $c(\cdot)$ such that $\text{Esc}(c(\cdot), M) < +\infty$,

$$\text{Esc}(c(\cdot), S^c) < \text{Esc}(c(\cdot), M).$$

It is easy to verify that under this assumption that the optimal value function V is lower semicontinuous. The next theorem establishes that this optimal value function is a unique bilateral lower semicontinuous solution for the boundary value problem (7.2) in the class of lower semicontinuous functions bounded from below.

Theorem 7.11. Under Assumptions (H1), (H2) and (H4), the minimal time function V in (7.11) is the unique bilateral solution bounded from below of the initial value problem for Hamilton-Jacobi equation (7.1).

Proof. As in the proof of 7.5 we need only to show that, for a supersolution u and a subsolution v , we always have

$$(7.14) \quad v \leq u.$$

Let us consider on the manifold $\tilde{M} := R \times M$ the multifunction \tilde{F} consisting at any $(t, m) \in \tilde{M}$ of all pairs $(\frac{\partial}{\partial t}, \mathbf{v})$ where \mathbf{v} belongs to $F(m)$. It is clear that \tilde{F} satisfies Assumptions (H1) and (H2). Then the following lemma follows immediately from Theorems 6.10 and 6.5.

Lemma 7.12. Lower semicontinuous functions $u : M \rightarrow R$ and $v : M \rightarrow R$ are super- and subsolutions, respectively, if and only if they satisfy the boundary condition $u(m) = 0$, $v(m) = 0$ for $m \in S$, and $u(m) + t$ is weakly decreasing and $v(m) + t$ is strongly predecreasing with respect to solutions $(t, c(t))$ of the differential inclusion with \tilde{F} on S^c and \bar{S}^c , respectively.

Proof. It follows from Theorem 6.10 that u is weakly monotone decreasing with respect to solutions of differential inclusion \tilde{F} if and only if for the hamiltonian $h_{\tilde{F}}$ the following inequality holds for any $(t, m) \in (-\infty, \theta) \times M$:

$$h_{\tilde{F}}(t, m, \zeta) \leq 0 \quad \forall (\zeta_t, \zeta_m) \in \partial_F(u(m) + t),$$

where $\langle \zeta_t, \frac{\partial}{\partial t} \rangle = 1$. Computing this hamiltonian, we have that this condition coincides with the differential inequality in Definition 7.8.

Analogously, we observe that the strong monotone predecrease property of the function v is equivalent to the strong monotone decrease property in reverse time which means that v is strong monotone decreasing with respect to the differential inclusion with the right-hand side $-\tilde{F}$. Thus, the strong monotone predecrease property is equivalent to the following differential inequality which holds for any $m \in \bar{S}^c$:

$$H_{(-\tilde{F})}(t, m, \zeta) \leq 0 \quad \forall \zeta \in \partial_F v(t, m).$$

But this inequality is exactly the differential inequality from Definition 7.2 of a subsolution. The lemma is proven. \square

Now we use this lemma to establish the following comparison result for super- and subsolutions and the optimal value function V .

Lemma 7.13. *Let u and v be respectively a bounded from below supersolution and a subsolution; then for any $m \in M$*

$$(7.15) \quad v(m) \leq V(m) \leq u(m).$$

Proof. It follows from the weak monotone decrease property of $u(m) + t$ that for any initial point $(0, m) \in R \times M$ there exists a trajectory of the differential inclusion \tilde{F} starting from $(0, m)$ such that

$$u(c(t)) + t \leq u(m)$$

while $c(t)$ stays in S^c . Since u is bounded below the previous inequality implies in view of Assumption (H4) that there exists a moment T such that $c(T) \in S$. By the lower semicontinuity of u we have that $0 \leq u(m) - T$ which means that $T \leq u(m)$ and $V(m) \leq u(m)$.

Now we fix $m \in S^c$ such that $V(m) < +\infty$ and consider an arbitrary trajectory $c(\cdot)$ such that $T := \theta(c(\cdot)) < +\infty$. This means that $c(T) \in S$, and we use the strong monotone predecrease property of $v(m) + t$ to derive that

$$v(m) \leq v(c(T)) + T.$$

Due to the boundary condition for v we have that $v(c(T)) = 0$ and since $c(\cdot)$ is an arbitrary solution the last inequality implies

$$v(m) \leq V(m)$$

which finishes the proof of this lemma. \square

As we have mentioned this comparison result (7.15) implies the uniqueness of bilateral lower semicontinuous solution.

Now we demonstrate that the optimal value function V (7.11) is such a bilateral solution of the boundary value problem (7.2) for the Hamilton-Jacobi equation. Namely, we show that $V + t$ is weakly monotone decreasing and strongly monotone predecreasing with respect to solutions of the differential inclusion \tilde{F} and use Lemma 7.12 to derive that V is both a super- and subsolution.

Indeed, for a given point m consider the optimal solution c of F such that $c(0) = m$ and $T := V(m)$. This implies that $V(c(t)) + t = V(m)$ for any $t \in [0, T]$ which means that $V(c(t)) + t$ is weak monotone decreasing.

Now we consider an arbitrary solution which is a local prolongation of this optimal trajectory c on the interval $t < 0$. For any $t < 0$, we have that $V(c(t)) + t \leq V(m)$, which implies that V is strong monotone predecreasing.

Thus, due to Lemma 7.12, V is a supersolution and a subsolution simultaneously which implies that it is the unique bilateral solution. This completes the proof of Theorem 7.11. \square

As an example of an application of this theorem we consider the following optimal time problem for the closed set $S \subset M$ of minimization of the functional $\theta_S(c(\cdot))$ (7.10) on the set of trajectories

$$(7.16) \quad \dot{c}(t) \in B(c(t))$$

where $B(m) := \{v \in T_m(M) : \|v\| \leq 1\}$.

It is obvious that the minimal time function V for such a control problem coincides with the distance function d_S . Thus, we obtain from Theorem 7.11 the following corollary.

Corollary 7.14. *Under Assumptions (H4) on trajectories of (7.16) there exists a unique bilateral solution of the boundary value problem for the eikonal equation (7.3) which coincides with the distance function d_S to the set S .*

When $S = \{m_1\}$ this result gives the infinitesimal characterization of the Riemann metric $m \rightarrow \rho(m, m_1)$.

Remark 7.15. The dynamical programming approach to free time optimal control problems in terms of generalized solutions to a Hamilton-Jacobi equation has been investigated by I. Chrysoschoos and R. B. Vinter in [22], where one can also find an interesting concrete example.

ACKNOWLEDGEMENT

The idea of developing a nonsmooth analysis on manifolds came from discussions with E. Sontag. We thank J. M. Borwein, F. H. Clarke and E. Sontag for stimulating conversations, and we thank J. S. Treiman for helpful discussions.

REFERENCES

1. J. P. Aubin and A. Cellina, *Differential Inclusions*. Springer-Verlag, Berlin, 1984. MR0755330 (85j:49010)
2. J. P. Aubin and H. Frankowska, *Set-Valued Analysis*. Birkhäuser, Boston, 1990. MR1048347 (91d:49001)
3. W. Ballmann, M. Gromov and V. Schroeder, *Manifolds of Nonpositive Curvature*, Progress in Mathematics, 61, Birkhäuser Boston Inc. Boston MA, 1985. MR0823981 (87h:53050)
4. M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, With appendix by M. Falcone and P. Soravia, Systems and Control: Foundations and Applications, Birkhäuser Boston, Inc. Boston, MA, 1997. MR1484411 (99e:49001)
5. E. N. Barron and R. Jensen, *Semicontinuous viscosity solutions for Hamilton-Jacobi equations with convex Hamiltonians*, Commun. In Partial Differential Equations, **15** (1990), 1713-1742. MR1080619 (91h:35069)
6. E. N. Barron and R. Jensen, *Optimal control and semicontinuous viscosity solutions*, Proc. Amer. Math. Soc. **113** (1991), 396-402. MR1076572 (91m:49027)
7. M. S. Bazaraa, J. J. Goode and M. Z. Nashed, On the cones of tangents with applications to mathematical programming, J. Optim. Theo. Appli., **13** (1974), 389-426. MR0366398 (51:2645)
8. J. M. Borwein and D. Preiss, *A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions*, Trans. Amer. Math. Soc., **303** (1987), 517-527. MR0902782 (88k:49013)
9. J. M. Borwein, J. Read, A. S. Lewis and Q. J. Zhu, *Convex spectral functions of compact operators*, J. Nonlinear Convex Anal. **1** (2000), 17-35. MR1751726 (2003a:49045)
10. J. M. Borwein, A. S. Lewis and Q. J. Zhu, *Convex spectral functions of compact operators, Part 2: Lower semicontinuity and rearrangement invariance*, Proceedings of the Optimization and related topics (Ballarat/Melbourne, 1999), Applied Optim. **47**, Kluwer Academic Publ. 2001, pp. 179-196. MR1893622 (2004c:49075)
11. J. M. Borwein, J. S. Treiman and Q. J. Zhu, *Necessary conditions for constrained optimization problems with semicontinuous and continuous data*, Trans. Amer. Math. Soc. **350** (1998), 2409-2429. MR1433112 (98h:90108)
12. J. M. Borwein and Q. J. Zhu, *Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity*, SIAM J. Control and Optimization, **34** (1996), 1568-1591. MR1404847 (97g:49037)
13. J. M. Borwein and Q. J. Zhu, *A survey of subdifferential calculus with applications*, Nonlinear Analysis, TMA **38** (1999), 687-773. MR1710152 (2000j:49024)
14. J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, CMS Springer-Verlag Books, Springer-Verlag, New York, 2005. MR2144010

15. J. M. Borwein and D. Zhuang, *On Fan's minimax theorem*, Mathematical Programming, **34** (1996), 232-234. MR0838482 (87i:90216)
16. F. Brickell and R. S. Clark, *Differentiable manifolds: an introduction*, V.N. Reinhold Co., London, New York, 1970.
17. R. W. Brockett, *Nonlinear control theory and differential geometry*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983), PWN, Warsaw, 1984, pp. 1357-1368. MR0804784 (86k:93068)
18. R. W. Brockett, *Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems*, Linear Algebra Appl. **146** (1991), 79-91. MR1083465 (92j:90043)
19. J. V. Burke and M. L. Overton, *On the subdifferentiability of functions of a matrix spectrum. II. Subdifferential formulas*, Nonsmooth optimization: methods and applications (Erice, 1991), Gordon and Breach, Montreux, 1992, pp. 19-29. MR1263488 (95d:49025)
20. J. V. Burke and M. L. Overton, *Differential properties of the spectral abscissa and the spectral radius for analytic matrix-valued mappings*, Nonlinear Anal. **23** (1994), 467-488. MR1294349 (95f:49017)
21. H. Busemann, *The Geometry of Geodesics*, Academic Press Inc., New York, N.Y., 1955. MR0075623 (17:779a)
22. I. Chrysoschoos and R. B. Vinter, *Optimal control problems on manifolds: a dynamical programming approach*, J. Math. Anal. Appl., **287** (2003) 118-140. MR2010261 (2004j:49045)
23. F. H. Clarke, *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations*, Ph. D. thesis, Univ. of Washington, 1973.
24. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., **205** (1975), 247-262. MR0367131 (51:3373)
25. F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983, Russian edition Nauka, Moscow, 1988; Reprinted as Vol. 5 of the series Classics in Applied Mathematics, SIAM, Philadelphia, 1990. MR1058436 (91e:49001)
26. F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional conference series in applied mathematics, Vol. 57 SIAM, Philadelphia, 1989. MR1085948 (91j:49001)
27. F. H. Clarke, Yu. S. Ledyaeu, E. D. Sontag, and A. I. Subbotin, *Asymptotic controllability implies feedback stabilization*, I.E.E.E. Trans. Aut. Control, **42** (1999), 1394-1407. MR1472857 (98g:93003)
28. F. H. Clarke, Yu. S. Ledyaeu, R. J. Stern and P. R. Wolenski, *Qualitative properties of trajectories of control systems: A survey*, J. Dynamical and Contr. Sys. **1** (1995), 1-48. MR1319056 (95k:49002)
29. F. H. Clarke, Yu. S. Ledyaeu, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Graduate Texts in Mathematics Vol. 178, Springer-Verlag, New York, 1998. MR1488695 (99a:49001)
30. F. H. Clarke, R. J. Stern and P. R. Wolenski, *Subgradient criteria for monotonicity, the Lipschitz condition, and convexity*, Can. J. Math., **45** (1993), 1167-1183. MR1247540 (94j:49018)
31. M. G. Crandall, L. C. Evans and P. L. Lions *Some properties of viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc., **282** (1984), 487-502. MR0732102 (86a:35031)
32. M. G. Crandall and P. L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. **277** (1983), 1-42. MR0690039 (85g:35029)
33. K. Deimling, *Multivalued Differential Equations*. de Gruyter, Berlin, 1992. MR1189795 (94b:34026)
34. J. M. Danskin, *The theory of max-min, with applications*, SIAM J. Appl. Math., **14** (1966), 641-664. MR0210456 (35:1349)
35. J. M. Danskin, *The theory of max-min and its application to weapons allocation problems*, Econometrics and Operations Research, Vol. V. Springer-Verlag New York, Inc., New York, 1967. MR0228260 (37:3843)
36. V. F. Dem'yanov, *On the directional differentiation of the maximin function*, Dokl. Akad. Nauk SSSR, **179** (1968), 1032-1033. MR0227832 (37:3416)
37. V. F. Dem'yanov, *Minimax: directional differentiability*. Izdat. Leningrad. Univ., Leningrad, 1974. MR0445825 (56:4159)

38. V. F. Dem'yanov and V. N. Malozemov, *Introduction to minimax*, Translated from the Russian by D. Louvish. Reprint of the 1974 edition. Dover Publications, Inc., New York, 1990. MR1088479 (91k:49001)
39. I. Ekeland, *On the variational principle*, J. Math. Anal. Appl., **47** (1974), 324-353. MR0346619 (49:11344)
40. U. Helmke and J. B. Moore, *Optimization and dynamical systems*, With a foreword by R. Brockett, Springer-Verlag London Ltd., London, 1994. MR1299725 (95j:49001)
41. R. Hermann, *Differential Geometry and the Calculus of Variations*, Vol. 49 in Mathematics in Science and Engineering, ed. R. Bellman, Academic Press, New York, London, 1968. MR0233313 (38:1635)
42. A. D. Ioffe, *On subdifferentiability spaces*, Ann. N.Y. Acad. Sci., **410** (1983), 107-119. MR0775520 (86g:90124)
43. A. D. Ioffe, *Subdifferentiability spaces and nonsmooth analysis*, Bull. Amer. Math. Soc. **10** (1984), 87-90. MR0722857 (85c:58014)
44. A. D. Ioffe, *Proximal analysis and approximate subdifferentials*, J. London Math. Soc. **41** (1990), 175-192. MR1063554 (91i:46045)
45. A. D. Ioffe, *Variational analysis in local and global nonsmooth analysis*, In F. H. Clarke and R. J. Stern, editors, Nonlinear analysis, differential equations and control, Vol. 528 of NATO Sci. Ser. C Math. Phys Sci., pages 447-502. Kluwer Acad. Publ. Dordrecht, 1999. MR1695012 (2000e:49022)
46. A. Isidori, *Nonlinear Control Systems, Third Edition*, Springer-Verlag, London, 1989. MR1410988 (97g:93003)
47. Yu. S. Ledyae and J.S.Treiman, *Sub- and supergradients of envelopes, closures and limits*, J.Functional Analysis, submitted.
48. Yu.S.Ledyae, J.S.Treiman and Q.J.Zhu, *Helly's intersection theorem for manifolds of non-positive curvature*, J.Convex Analysis, submitted.
49. P. D. Loewen, *Optimal Control via Nonsmooth Analysis*, CRM Lecture Notes Series, Amer. Math. Soc., Summer School on Control, CRM, Université de Montréal, 1992, Amer. Math. Soc., Providence, 1993. MR1232864 (94h:49003)
50. A. Lewis, *Nonsmooth analysis of eigenvalues*, Mathematical Programming, **84** (1999), 1-24. MR1687292 (2000d:49029)
51. Y. Matsushima, *Differential Manifolds*, Marcel Dekker, New York, 1972. MR0346831 (49:11553)
52. P. Michel and J. P. Penot, *Calcul sous-différentiel pour des fonctions Lipschitziennes et non-Lipschitziennes*, C. R. Acad. Sci. Paris, Ser. I Math. **298** (1985), 269-272. MR0745320 (85i:49027)
53. B. S. Mordukhovich, *Maximum principle in problems of time optimal control with nonsmooth constraints*, J. Appl. Math. Mech. **40** (1976), 960-969. MR0487669 (58:7284)
54. B. S. Mordukhovich, *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems*, Soviet Math. Dokl., **22** (1980), 526-530. MR0592682 (82b:90104)
55. B. S. Mordukhovich, *Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings*, Dokl. Akad. Nauk. BSSR, **28** (1984), 976-979. MR0771737 (86c:49018)
56. B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988. (Russian; English transl. to appear in Wiley-Interscience.) MR0945143 (89m:49001)
57. B. S. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, J. Math. Anal. Appl. **183** (1994), 250-288. MR1273445 (95i:49029)
58. B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I and II*, Springer-Verlag, Berlin, 2006. MR2191744; MR2191745
59. B. S. Mordukhovich and Y. Shao, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc., **348** (1996), 1235-1280. MR1333396 (96h:49036)
60. B. N. Pshenichnyi, *Convex programming in a normalized space*, Cybernetics, **1** (1966), 46-57. MR0234732 (38:3048)
61. B. N. Pshenichnyi, *Necessary conditions for an extremum*. Marcel Dekker Inc., New York, 1971. Russian original edition: Nauka, Moscow, 1969. MR0276845 (43:2585)
62. R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, N.J., 1970. MR0274683 (43:445)

63. R. T. Rockafellar, *Clarke's tangent cones and boundaries of closed sets in \mathbb{R}^n* , Nonlinear Analysis: TMA, **3** (1979), 145-154. MR0520481 (80d:49032)
64. R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 1997. MR1491362 (98m:49001)
65. S. Sternberg, *Lectures on differential geometry*, Prentice-Hall, Englewood Cliffs, N.J., 1964. MR0193578 (33:1797)
66. A. I. Subbotin, *Generalized Solutions of First-Order PDEs: The Dynamical Optimization Perspective*, Birkhäuser, Boston, 1995. MR1320507 (96b:49002)
67. H. J. Sussmann, *A strong version of the maximum principle under weak hypotheses*, Proceedings of the 33rd IEEE conference on decision and control, Lake Buena Vista, FL, December 1994.
68. H. J. Sussmann, *A strong maximum principle for systems of differential inclusions*, Proceedings of the 35th IEEE conference on decision and control, Kobe, Japan, December 1996.
69. H. J. Sussmann, *Transversality conditions and a strong maximum principle for systems of differential inclusions*, Proceedings of the 37th IEEE conference on decision and control, Tampa, FL, December 1998.
70. J. S. Treiman, *Generalized gradients, Lipschitz behavior and directional derivatives*, Can. J. Math., **37** (1985), 1074-1084. MR0828835 (87h:90263)
71. J. S. Treiman, *Finite dimensional optimality conditions: B-gradients*, J. Optim. Theory Appl. **62** (1989), 139-150. MR1006610 (90f:49011)
72. J. S. Treiman, *The linear nonconvex generalized gradients and Lagrange multipliers*, SIAM J. Optim. **5** (1995), 670-680. MR1344675 (96i:49032)
73. J. S. Treiman, *Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints*, SIAM J. Control Optim. **37** (1999), 1313-1329. MR1710060 (2000j:49026)
74. R. B. Vinter, *Optimal control*, Birkhäuser Boston Inc., Boston, MA, 2000. MR1756410 (2001c:49001)
75. J. Warga, *Derivate containers, inverse functions, and controllability*, in Calculus of Variations and Control Theory, D. L. Russell, Ed., Academic Press, New York, 1976. MR0427561 (55:592)
76. J. Warga, *Fat homeomorphisms and unbounded derivate containers*, J. Math. Anal. Appl. **81** (1981), 545-560. MR0622836 (83f:58007)
77. D. Zagrodny, *Approximate mean value theorem for upper subderivatives*, Nonlinear Anal. TMA, **12** (1988), 1413-1428. MR0972409 (89k:58034)
78. Q. J. Zhu, *Subderivatives and their applications*, Proceedings of International Conference on Dynamical Systems and Differential Equations, Edited by W. Chen and S. Hu, Springfield, MO, June, 1996, pp. 379-394. MR1722485 (2000j:49027)
79. Q. J. Zhu, *The equivalence of several basic theorems for subdifferentials*, Set-Valued Analysis, **6** (1998), 171-185. MR1646482 (2000c:49035)

DEPARTMENT OF MATHEMATICS, WESTERN MICHIGAN UNIVERSITY, KALAMAZOO, MICHIGAN 49008 – AND – STEKLOV INSTITUTE OF MATHEMATICS, MOSCOW 117966, RUSSIA
E-mail address: ledyaev@wmich.edu

DEPARTMENT OF MATHEMATICS, WESTERN MICHIGAN UNIVERSITY, KALAMAZOO, MICHIGAN 49008
E-mail address: zhu@wmich.edu