Deutsche Forschungsgemeinschaft
Priority Programme 1962

Intrinsic Formulation of KKT Conditions and Constraint Qualifications on Smooth Manifolds

Ronny Bergmann, Roland Herzog

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-052

received on April 17, 2018

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

INTRINSIC FORMULATION OF KKT CONDITIONS AND CONSTRAINT QUALIFICATIONS ON SMOOTH MANIFOLDS*

RONNY BERGMANN† AND ROLAND HERZOG†

Abstract. Karush-Kuhn-Tucker (KKT) conditions for equality and inequality constrained optimization problems on smooth manifolds are formulated. Under the Guignard constraint qualification, local minimizers are shown to admit Lagrange multipliers. The linear independence, Mangasarian–Fromovitz, and Abadie constraint qualifications are also formulated, and the chain "LICQ implies MFCQ implies ACQ implies GCQ" is proved. Moreover, classical connections between these constraint qualifications and the set of Lagrange multipliers are established, which parallel the results in Euclidean space. The constrained Riemannian center of mass on the sphere serves as an illustrating numerical example.

Key words. nonlinear optimization, smooth manifolds, KKT conditions, constraint qualifications

AMS subject classifications. 90C30, 90C46, 49Q99, 65K05

1. Introduction.

We consider constrained, nonlinear optimization problems

$$\left\{\begin{aligned} \text{Minimize} \quad & f(\boldsymbol{p}), \quad \boldsymbol{p} \in \mathcal{M}, \\ \text{s.t.} \quad & g(\boldsymbol{p}) \le 0, \\ \text{and} \quad & h(\boldsymbol{p}) = 0, \end{aligned}\right. \tag{1.1}$$

where $\mathcal{M}$ is a smooth manifold. The objective $f\colon \mathcal{M} \to \mathbb{R}$ and the constraint functions $g\colon \mathcal{M} \to \mathbb{R}^m$ and $h\colon \mathcal{M} \to \mathbb{R}^q$ are assumed to be functions of class C^1. The main contribution of this paper is the development of first-order necessary optimality conditions in Karush-Kuhn-Tucker (KKT) form, well known when $\mathcal{M} = \mathbb{R}^n$, under appropriate constraint qualifications (CQs). Specifically, we introduce and discuss analogues of the linear independence, Mangasarian–Fromovitz, Abadie and Guignard CQ, abbreviated as LICQ, MFCQ, ACQ and GCQ, respectively; see for instance Solodov, 2010, Peterson, 1973 or Bazaraa, Sherali, Shetty, 2006, Ch. 5.

It is well known that KKT conditions are of paramount importance in nonlinear programming, both for theory and numerical algorithms. We refer the reader to Kjeldsen, 2000 for an account of the history of KKT condition in the Euclidean setting $\mathcal{M} = \mathbb{R}^n$. A variety of programming problems in numerous applications, however, are naturally given in a manifold setting. Well-known examples for smooth manifolds include spheres, tori, the general linear group $\mathrm{GL}(n)$ of non-singular matrices, the group of special orthogonal (rotation) matrices $\mathrm{SO}(n)$, the Grassmannian manifold

*Submitted to the editors April 17, 2018.

Funding: This work was partially supported by DFG grants BE 5888/2–1 and HE 6077/10–1, the latter within the Priority Program SPP 1962, which is gratefully acknowledged.

†Technische Universität Chemnitz, Faculty of Mathematics, Professorship Numerical Mathematics (Partial Differential Equations), D–09107 Chemnitz, Germany (RB: ronny.bergmann@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/bergmann; RH: roland.herzog@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/herzog).

of k-dimensional subspaces of a given vector space, and the orthogonal Stiefel manifold of orthonormal rectangular matrices of a certain size. We refer the reader to Absil, Mahony, Sepulchre, 2008 for an overview and specific examples. Recently optimization on manifolds has gained interest e.g., in image processing, where methods like the cyclic proximal point algorithm by Bačák, 2014, half-quadratic minimization by Bergmann, Chan, et al., 2016, and the parallel Douglas-Rachford algorithm by Bergmann, Persch, Steidl, 2016 have been introduced. They were then applied to variational models from imaging, i.e., optimization problems of the form (1.1), where the manifold is given by the power manifold $\mathcal{M}^N$ with N being the number of data items or pixel. We emphasize that all of the above consider *unconstrained* problems on manifolds.

In principle, inequality and equality constraints in (1.1) might be taken care of by considering a suitable submanifold of $\mathcal{M}$ (with boundary). This is much like in the case $\mathcal{M} = \mathbb{R}^n$, where one may choose not to include some of the constraints in the Lagrangian but rather treat them as abstract constraints. Often, however, there may be good reasons to consider constraints explicitly, one of them being that Lagrange multipliers carry sensitivity information for the optimal value function, although this is not addressed in the present paper.

To the best of our knowledge, a systematic discussion of constraint qualifications and KKT conditions for (1.1) is not available in the literature. We are aware of Udrişte, 1988 where KKT conditions are derived for convex inequality constrained problems and under a Slater constraint qualification on a complete Riemannian manifold. The work closest to ours is Yang, Zhang, Song, 2014, where KKT and also second-order optimality conditions are derived for (1.1) in the setting of a smooth Riemannian manifold, and under the assumption of LICQ. Other constraint qualifications are not considered. We also mention Ledyaev, Zhu, 2007 where a framework for generalized derivatives of non-smooth functions on smooth Riemannian manifolds is developed and Fritz-John type optimality conditions are derived as an application.

The novelty of the present paper is the formulation of analogues for a range of constraint qualifications (LICQ, MFCQ, ACQ, and GCQ) in the smooth manifold setting. We establish the classical "LICQ implies MFCQ implies ACQ implies GCQ" and prove that KKT conditions are necessary optimality conditions under any of these CQs. We also show that the classical connections between these constraint qualifications and the set of Lagrange multipliers continue to hold, e.g., Lagrange multipliers are generically unique if and only if LICQ holds. Finally, our work shows that the smooth structure on a manifold is a framework sufficient for the purpose of first-order optimality conditions. In particular, we do not need to introduce a Riemannian metric as in Yang, Zhang, Song, 2014.

We wish to point out that optimality conditions can also be derived by considering $\mathcal{M}$ to be embedded in a suitable ambient Euclidean space $\mathbb{R}^N$. This approach requires, however, to formulate additional, nonlinear constraints in order to ensure that only points in $\mathcal{M}$ are considered feasible. Another drawback of such an approach is that the number of variables grows since N is larger than the manifold dimension. In contrast to the embedding approach, we formulate KKT conditions and appropriate constraint qualifications (CQs) using *intrinsic* concepts on the manifold $\mathcal{M}$. This requires, in particular, the generalization of the notions of tangent and linearizing cones to the smooth manifold setting. The intrinsic point of view is also the basis of many optimization approaches for problems on manifolds; see for instance Absil, Mahony, Sepulchre, 2008; Absil, Baker, Gallivan, 2007; Boumal, 2015.

The material is organized as follows. In section 2 we review the necessary background material on smooth manifolds. Our main results are given in section 3, where KKT conditions are formulated and shown to hold for local minimizer under the Guignard constraint qualifications. We also formulate further constraint qualifications (CQs) and establish "LICQ implies MFCQ implies ACQ implies GCQ". Section 4 is devoted to the connections between CQs and the set of Lagrange multipliers. In section 5 we present an application of the theory.

Notation. Throughout the paper, ε is a positive number whose value may vary from occasion to occasion. We distinguish between column vectors (elements of $\mathbb{R}^n$) and row vectors (elements of $\mathbb{R}_n$).

2. Background Material.

In this section we review the required background material on smooth manifolds. We refer the reader to Spivak, 1979; Aubin, 2001; Lee, 2003; Tu, 2011; Jost, 2017 for a thorough introduction.

DEFINITION 2.1. *A Hausdorff, second-countable topological space $\mathcal{M}$ is said to be a smooth manifold of dimension $n \in \mathbb{N}$ if there exists an arbitrary index set A, a collection of open subsets $\{U_\alpha\}_{\alpha \in A}$ covering $\mathcal{M}$, together with a collection of homeomorphisms (continuous functions with continuous inverses) $\varphi_\alpha \colon U_\alpha \to \varphi_\alpha(U_\alpha) \subset \mathbb{R}^n$, such that the transition maps $\varphi_\beta \circ \varphi_\alpha^{-1} \colon \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$ are of class C^∞ for all $\alpha, \beta \in A$. A pair $(U_\alpha, \varphi_\alpha)$ is called a smooth chart, and the collection $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ is a smooth atlas.*

Well-known examples of smooth manifolds include $\mathbb{R}^n$, spheres, tori, $\mathrm{GL}(n)$, $\mathrm{SO}(n)$, the Grassmannian manifold of k-dimensional subspaces of a given vector space, and the orthogonal Stiefel manifold of orthonormal rectangular matrices of a certain size; see for instance Absil, Mahony, Sepulchre, 2008. From now on, a smooth manifold $\mathcal{M}$ will always be equipped with a given smooth atlas. In particular, $\mathbb{R}^n$ will be equipped with the standard atlas consisting of the single chart $(\mathbb{R}^n, \mathrm{id})$. Points on $\mathcal{M}$ will be denoted by bold-face letters such as $\boldsymbol{p}$ and $\boldsymbol{q}$.

Notions beyond continuity are defined by means of charts. In particular, the assumed C^1-property of the objective $f \colon \mathcal{M} \to \mathbb{R}$ means that $f \circ \varphi_\alpha^{-1}$, defined on the open subset $\varphi_\alpha(U_\alpha) \subset \mathbb{R}^n$ and mapping into $\mathbb{R}$, is of class C^1 for every chart $(U_\alpha, \varphi_\alpha)$ from the smooth atlas. The C^1-property of the constraint functions g and h is defined in the same way. Similarly, one may speak of C^1-functions which are defined only in an open subset $U \subset \mathcal{M}$, by replacing U_α by $U_\alpha \cap U$.

As is well known, tangential directions (to the feasible set) play a fundamental role in optimization. Tangential directions at a point can be viewed as derivatives of curves passing through that point. When $\mathcal{M} = \mathbb{R}^n$, these curves can be taken to be straight curves $t \mapsto \boldsymbol{p} + t\,\boldsymbol{v}$ of arbitrary velocity $\boldsymbol{v} \in \mathbb{R}^n$. This shows that $\mathbb{R}^n$ serves as its own tangent space. An adaptation to the setting of a smooth manifold leads to the following

DEFINITION 2.2 (Tangent space).

(a) *A function $\gamma \colon (-\varepsilon, \varepsilon) \to \mathcal{M}$ is called a C^1-curve about $\boldsymbol{p} \in \mathcal{M}$ if $\gamma(0) = \boldsymbol{p}$ holds and $\varphi_\alpha \circ \gamma$ is of class C^1 for some (equivalently, every) chart $(U_\alpha, \varphi_\alpha)$ about $\boldsymbol{p}$.*

(*b*) *Two C^1-curves γ and ζ about $\boldsymbol{p} \in \mathcal{M}$ are said to be equivalent if*

$$\frac{\mathrm{d}}{\mathrm{d}t}(\varphi_\alpha \circ \gamma)(t)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(\varphi_\alpha \circ \zeta)(t)\Big|_{t=0} \tag{2.1}$$

holds for some (equivalently, every) chart $(U_\alpha, \varphi_\alpha)$ about $\boldsymbol{p}$.

(*c*) *Suppose that γ is a C^1-curve about $\boldsymbol{p} \in \mathcal{M}$ and that $[\gamma]$ is its equivalence class. Then the following linear map, denoted by $[\dot\gamma(0)]$ or $[\frac{\mathrm{d}}{\mathrm{d}t}\gamma(0)]$ and defined as*

$$[\dot\gamma(0)](f) \coloneqq \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)\Big|_{t=0} \tag{2.2}$$

takes C^1-functions $f \colon U \to \mathbb{R}$ defined in some open neighborhood $U \subset \mathcal{M}$ of $\boldsymbol{p}$ into $\mathbb{R}$. It is called the tangent vector to $\mathcal{M}$ at $\boldsymbol{p}$ along (or generated by) the curve γ.

(*d*) *The collection of all tangent vectors at $\boldsymbol{p}$, i.e.,*

$$\mathcal{T}_{\mathcal{M}}(\boldsymbol{p}) \coloneqq \big\{[\dot\gamma(0)] \colon [\dot\gamma(0)] \text{ is generated by some } C^1\text{-curve } \gamma \text{ about } \boldsymbol{p}\big\}, \tag{2.3}$$

is termed the tangent space to $\mathcal{M}$ at $\boldsymbol{p}$.

Remark 2.3 (Tangent space).

1. We infer from (2.2) that the tangent vector $[\dot\gamma(0)]$ along the curve γ about $\boldsymbol{p}$ generalizes the notion of the directional derivative operator, acting on C^1-functions defined near $\boldsymbol{p}$.

2. It can be shown that the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ to $\mathcal{M}$ at $\boldsymbol{p}$ is a vector space of dimension n under the operations $\alpha \odot [\gamma]$ and $[\gamma] \oplus [\zeta]$, defined in terms of

$$\alpha \odot \gamma \colon t \mapsto \gamma(\alpha\, t) \in \mathcal{M} \quad \text{for } \alpha \in \mathbb{R}, \tag{2.4a}$$

$$\gamma \oplus \zeta \colon t \mapsto \varphi_\alpha^{-1}\big((\varphi_\alpha \circ \gamma)(t) + (\varphi_\alpha \circ \zeta)(t) - \varphi_\alpha(\boldsymbol{p})\big) \in \mathcal{M} \tag{2.4b}$$

for arbitrary representers of their respective equivalence classes. Here φ_α is an arbitrary chart about $\boldsymbol{p}$, and its choice does not affect the definition of $[\gamma] \oplus [\zeta]$.

Finally, we require the generalization of the notion of the derivative for functions $f \colon \mathcal{M} \to \mathbb{R}$.

Definition 2.4 (Differential). *Suppose that $f \colon \mathcal{M} \to \mathbb{R}$ is a C^1-function and $\boldsymbol{p} \in \mathcal{M}$. Then the following linear map, denoted by $(\mathrm{d}f)(\boldsymbol{p})$ and defined as*

$$(\mathrm{d}f)(\boldsymbol{p})\,[\dot\gamma(0)] \coloneqq [\dot\gamma(0)](f) \tag{2.5}$$

takes tangent vectors $[\dot\gamma(0)]$ into $\mathbb{R}$. It is called the differential of f at $\boldsymbol{p}$.

By definition, the differential $(\mathrm{d}f)(\boldsymbol{p})$ of a real-valued function is a cotangent vector, i.e., an element from the cotangent space $\mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$, the dual of the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$. In fact, every element of $\mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$ is the differential of a C^1-function s at $\boldsymbol{p}$. Therefore we denote, without loss of generality, generic elements of $\mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$ by $(\mathrm{d}s)(\boldsymbol{p})$.

Remark 2.5. In the literature on differential geometry the tangent space is usually denoted by $\mathcal{T}_{\boldsymbol{p}}\mathcal{M}$ and the cotangent space by $\mathcal{T}^*_{\boldsymbol{p}}\mathcal{M}$. Moreover the differential of a real-valued function s at $\boldsymbol{p}$ is written as $(\mathrm{d}s)_{\boldsymbol{p}}$. We hope that our slightly modified notation is more intuitive for readers familiar with nonlinear programming notation.

In the following two sections, we are going to derive the KKT theory for (1.1) and associated constraint qualifications on smooth manifolds. We wish to point out that the above notions from differential geometry are sufficient for these purposes. In particular, we do not need to introduce a Riemannian metric (a smoothly varying collection of inner products on the tangent spaces), nor do we need to consider embeddings of $\mathcal{M}$ into some $\mathbb{R}^N$ for some $N \ge n$. Moreover, we do not need to make further topological assumptions such as compactness, connectedness, or orientability of $\mathcal{M}$.

3. KKT Conditions and Constraint Qualifications.

In this section we develop first-order necessary optimality conditions in KKT form for (1.1). To begin with, we briefly recall the arguments when $\mathcal{M} = \mathbb{R}^n$; see for instance Nocedal, Wright, 2006, Chap. 12 or Forst, Hoffmann, 2010, Chap. 2.

3.1. KKT Conditions in $\mathbb{R}^n$.

We define $\Omega \coloneqq \big\{x \in \mathbb{R}^n : g(x) \le 0,\ h(x) = 0\big\}$ to be the feasible set and associate with (1.1) the Lagrangian

$$\mathcal{L}(x, \mu, \lambda) \coloneqq f(x) + \mu\, g(x) + \lambda\, h(x), \tag{3.1}$$

where $\mu \in \mathbb{R}_m$ and $\lambda \in \mathbb{R}_q$. Using Taylor's theorem, one easily shows that a local minimizer x^* satisfies the necessary optimality condition

$$f'(x^*)\, d \ge 0 \quad \text{for all } d \in \mathcal{T}_\Omega(x^*), \tag{3.2}$$

where $\mathcal{T}_\Omega(x^*)$ denotes the tangent cone,

$$\begin{aligned} \mathcal{T}_\Omega(x^*) \coloneqq \Big\{ d \in \mathbb{R}^n : {} & \text{there exist sequences } (x_k) \subset \Omega,\ x_k \to x^*,\ (t_k) \searrow 0, \\ & \text{such that } d = \lim_{k\to\infty} \frac{x_k - x^*}{t_k} \Big\}. \end{aligned} \tag{3.3}$$

This cone is also known as contingent cone or the Bouligand cone; compare Jiménez, Novo, 2006; Penot, 1985. Since $\mathcal{T}_\Omega(x^*)$ is inconvenient to work with, one introduces the linearizing cone

$$\begin{aligned} \mathcal{T}_\Omega^{\mathrm{lin}}(x^*) \coloneqq \big\{ d \in \mathbb{R}^n : {} & g_i'(x^*)\, d \le 0 \quad \text{for all } i \in \mathcal{A}(x^*), \\ & h_j'(x^*)\, d = 0 \quad \text{for all } j = 1, \ldots, q \big\}. \end{aligned} \tag{3.4}$$

Here $\mathcal{A}(x^*) \coloneqq \big\{1 \le i \le m : g_i(x^*) = 0\big\}$ is the index set of active inequalities at x^*. Moreover, $\mathcal{I}(x^*) \coloneqq \{1, \ldots, m\} \setminus \mathcal{A}(x^*)$ are the inactive inequalities. It is easy to see that $\mathcal{T}_\Omega(x^*)$ is a closed convex cone and that $\mathcal{T}_\Omega(x^*) \subset \mathcal{T}_\Omega^{\mathrm{lin}}(x^*)$ holds; see for instance Nocedal, Wright, 2006, Lem. 12.2.

Using the definition of the polar cone of a set $B \subset \mathbb{R}^n$,

$$B^\circ \coloneqq \{s \in \mathbb{R}_n : s\, d \le 0 \text{ for all } d \in B\}, \tag{3.5}$$

the first-order necessary optimality condition (3.2) can also be written as $-f'(x^*) \in \mathcal{T}_\Omega(x^*)^\circ$. Since the polar of the tangent cone is often not easily accessible, one prefers to work with $\mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ$ instead, which has the representation

$$\mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ = \Big\{ s = \sum_{i=1}^{m} \mu_i \, g_i'(x^*) + \sum_{j=1}^{q} \lambda_j \, h_j'(x^*), \quad \mu_i \ge 0 \text{ for } i \in \mathcal{A}(x^*),\ \mu_i = 0 \text{ for } i \in \mathcal{I}(x^*),\ \lambda_j \in \mathbb{R} \Big\} \subset \mathbb{R}_n, \tag{3.6}$$

as can be shown by means of the Farkas lemma; compare Nocedal, Wright, 2006, Lem. 12.4. We state it here in a slightly more general (yet equivalent) form than usual, where V is a finite dimensional vector space and $A \in \mathcal{L}(V, \mathbb{R}^q)$ is a linear map from V into $\mathbb{R}^q$ for some $q \in \mathbb{N}$. The adjoint of A, denoted by A^*, then belongs to $\mathcal{L}(\mathbb{R}_q, V^*)$, where V^* is the dual space of V.

LEMMA 3.1 (Farkas). *Suppose that V is a finite dimensional vector space, $A \in \mathcal{L}(V, \mathbb{R}^q)$ and $b \in V^*$. The following are equivalent:*

(a) *The system $A^* y = b$ has a solution $y \in \mathbb{R}_q$ which satisfies $y \ge 0$.*
(b) *For any $d \in \mathbb{R}^q$, $A\,d \ge 0$ implies $b\,d \ge 0$.*

Continuing our review, we notice that $\mathcal{T}_\Omega(x^*) \subset \mathcal{T}_\Omega^{\text{lin}}(x^*)$ entails $\mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ \subset \mathcal{T}_\Omega(x^*)^\circ$, hence (3.2) does *not* imply

$$-f'(x^*) \in \mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ. \tag{3.7}$$

Enter constraint qualifications, the weakest of which (the Guignard qualification, GCQ; see Guignard, 1969) requires the equality $\mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ = \mathcal{T}_\Omega(x^*)^\circ$. Realizing that (3.7) is nothing but the KKT conditions,

$$\mathcal{L}_x(x^*, \mu, \lambda) = f'(x^*) + \mu\, g'(x^*) + \lambda\, h'(x^*) = 0, \tag{3.8a}$$

$$h(x^*) = 0, \tag{3.8b}$$

$$\mu \ge 0, \quad g(x^*) \le 0, \quad \mu\, g(x^*) = 0, \tag{3.8c}$$

we obtain the well known

THEOREM 3.2. *Suppose that x^* is a local minimizer of (1.1) for $\mathcal{M} = \mathbb{R}^n$ and that the GCQ holds at x^*. Then there exist Lagrange multipliers $\mu \in \mathbb{R}_m$, $\lambda \in \mathbb{R}_q$, such that the KKT conditions (3.8) hold.*

In practice one of course often works with stronger constraint qualifications, which are easier to verify. We are going to consider in subsection 3.3 the analogue of the classical chain LICQ $\Rightarrow$ MFCQ $\Rightarrow$ ACQ $\Rightarrow$ GCQ on smooth manifolds.

3.2. KKT Conditions for Optimization Problems on Smooth Manifolds. In this section we adapt the argumentation sketched in subsection 3.1 to problem (1.1), where $\mathcal{M}$ is a smooth manifold. Our first result is the analogue of Theorem 3.2, showing that the GCQ renders the KKT conditions a system of first-order necessary optimality conditions for local minimizers. For convenience, we summarize in Table 1 how the relevant quantities need to be translated when moving from $\mathcal{M} = \mathbb{R}^n$ to manifolds.

$\mathcal{M} = \mathbb{R}^n$	$\mathcal{M}$ smooth manifold
tangent space $\mathbb{R}^n$	tangent space $\mathcal{T}_\mathcal{M}(\boldsymbol{p})$ (2.2)
tangent cone $\mathcal{T}_\Omega(x)$ (3.3)	tangent cone $\mathcal{T}_\mathcal{M}(\Omega;\boldsymbol{p})$ (3.12)
linearizing cone $\mathcal{T}_\Omega^{\text{lin}}(x)$ (3.4)	linearizing cone $\mathcal{T}_\mathcal{M}^{\text{lin}}(\Omega;\boldsymbol{p})$ (3.14)
cotangent space $\mathbb{R}_n$	cotangent space $\mathcal{T}_\mathcal{M}^*(\boldsymbol{p})$
derivative $f'(x) \in \mathbb{R}_n$	differential $(\mathrm{d}f)(\boldsymbol{p}) \in \mathcal{T}_\mathcal{M}^*(\boldsymbol{p})$ (2.5)
polar cone $\subset \mathbb{R}_n$ (3.6)	polar cone $\mathcal{T}_\mathcal{M}^{\text{lin}}(\Omega;\boldsymbol{p})^\circ \subset \mathcal{T}_\mathcal{M}^*(\boldsymbol{p})$ (3.16)
Lagrange multipliers $\mu \in \mathbb{R}_m$, $\lambda \in \mathbb{R}_q$	same as for $\mathcal{M} = \mathbb{R}^n$

Table 1: Summary of concepts related to KKT conditions and constraint qualifications.

Let us denote by

$$\Omega \coloneqq \big\{\boldsymbol{p} \in \mathcal{M} : g(\boldsymbol{p}) \le 0,\ h(\boldsymbol{p}) = 0\big\} \tag{3.9}$$

the feasible set of (1.1). As in $\mathbb{R}^n$, Ω is a closed subset of $\mathcal{M}$ due to the continuity of g and h.

A point $\boldsymbol{p}^* \in \Omega$ is a local minimizer of (1.1) if there exists a neighborhood U of $\boldsymbol{p}^*$ such that

$$f(\boldsymbol{p}^*) \le f(\boldsymbol{p}) \quad \text{for all } \boldsymbol{p} \in U \cap \Omega.$$

The first notion of interest is the tangent cone at a feasible point. In view of (2.2), it may be tempting to consider

$$\mathcal{T}_\mathcal{M}^{\text{classical}}(\Omega;\boldsymbol{p}) \coloneqq \big\{[\dot\gamma(0)] \in \mathcal{T}_\mathcal{M}(\boldsymbol{p}) : [\dot\gamma(0)] \text{ is generated by some } C^1\text{-curve } \gamma \text{ about } \boldsymbol{p} \text{ which satisfies } \gamma(t) \in \Omega \text{ for all } t \in [0,\varepsilon)\big\}. \tag{3.10}$$

In fact this is the analogue of what is known as the cone of attainable directions and it was used in the original works of Karush, 1939; Kuhn, Tucker, 1951. However, as is well known, this cone is, in general, strictly smaller than the Bouligand tangent cone (3.3) when $\mathcal{M} = \mathbb{R}^n$; see for instance Penot, 1985; Jiménez, Novo, 2006, Bazaraa, Shetty, 1976, Ch. 3.5 and Aubin, Frankowska, 2009, Ch. 4.1.

In order to properly generalize the Bouligand tangent cone (3.3) to the smooth manifold setting, we consider sequences rather than curves. This leads to the following

Definition 3.3 ((Bouligand) tangent cone). *Suppose that $\boldsymbol{p} \in \Omega$ holds, and let (U,φ) be a chart about $\boldsymbol{p}$.*

(a) *A sequence $(\Gamma_k) \coloneqq (\boldsymbol{p}_k, t_k) \subset (U \cap \Omega) \times \mathbb{R}$ is said to be a tangential sequence to Ω at $\boldsymbol{p}$ if $\boldsymbol{p}_k \to \boldsymbol{p}$, $t_k \searrow 0$, and $(\varphi(\boldsymbol{p}_k) - \varphi(\boldsymbol{p}))/t_k \to d$ for some $d \in \mathbb{R}^n$ holds.*

(b) *Two tangential sequences $(\boldsymbol{p}_k, t_k)$ and $(\boldsymbol{q}_k, s_k)$ to Ω at $\boldsymbol{p}$ are said to be equivalent if $\lim_{k\to\infty}(\varphi(\boldsymbol{p}_k) - \varphi(\boldsymbol{p}))/t_k = \lim_{k\to\infty}(\varphi(\boldsymbol{q}_k) - \varphi(\boldsymbol{p}))/s_k$ holds.*

(*c*) *Suppose that* (Γ_k) *is a tangential sequence to* Ω *at* $\boldsymbol{p}$ *and that* $[\Gamma]$ *is its equivalence class. Then the following linear map, denoted by* $[\dot\Gamma]$ *and defined as*

$$[\dot\Gamma](f) \coloneqq \lim_{k\to\infty} \frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p})}{t_k} = (f \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\, d \tag{3.11}$$

takes C^1*-functions* $f\colon U \to \mathbb{R}$ *defined in some open neighborhood* $U \subset \mathcal{M}$ *of* $\boldsymbol{p}$ *into* $\mathbb{R}$*. It is called the sequential tangent vector to* Ω *at* $\boldsymbol{p}$ *along (or generated by) the tangential sequence* (Γ_k)*.*

(*d*) *The collection of all sequential tangent vectors to* Ω *at* $\boldsymbol{p}$*, i.e.,*

$$\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p}) \coloneqq \big\{[\dot\Gamma] : [\dot\Gamma] \text{ is generated by some tangential sequence } (\boldsymbol{p}_k, t_k) \text{ to } \Omega \text{ at } \boldsymbol{p}\big\}, \tag{3.12}$$

is termed the (Bouligand) tangent cone to Ω *at* $\boldsymbol{p}$*.*

Let us confirm that the tangent cone is an intrinsic concept.

Lemma 3.4. *The tangent cone* (3.12) *is independent of the chart about* $\boldsymbol{p}$ *selected.*

Proof. Suppose that (U, φ) is a chart about $\boldsymbol{p}$ and that (Γ_k) is a tangential sequence to Ω at $\boldsymbol{p}$ w.r.t. φ, generating the sequential tangent vector $[\dot\Gamma]$. Moreover, let (V, ψ) be another chart about $\boldsymbol{p}$. Then by the chain rule,

$$\frac{\psi(\boldsymbol{p}_k) - \psi(\boldsymbol{p})}{t_k} \to (\psi \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\, d,$$

so (Γ_k) is a tangential sequence w.r.t. the chart ψ as well. Let us also observe that the action of $[\dot\Gamma]$ on a C^1-function f defined near $\boldsymbol{p}$ is independent of the chart; see (3.11). Indeed, the second equality in (3.11) amounts to

$$(f \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\, d = (f \circ \psi^{-1})'(\psi(\boldsymbol{p}))\, (\psi \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\, d,$$

which agree due to the chain rule. ∎

Remark 3.5 (Tangent cone).

1. Notice that although sequential tangent vectors are defined in terms of sequences, not curves, they can be understood as tangent vectors in the sense of Definition 2.2. Indeed, let $(\Gamma_k) = (\boldsymbol{p}_k, t_k)$ be a tangential sequence to Ω at $\boldsymbol{p}$. Suppose that φ is a chart about $\boldsymbol{p}$ and $(\varphi(\boldsymbol{p}_k) - \varphi(\boldsymbol{p}))/t_k \to d$ for some $d \in \mathbb{R}^n$. Define the curve

$$t \mapsto \gamma(t) \coloneqq \varphi^{-1}(\varphi(\boldsymbol{p}) + t\, d)$$

on a suitable open interval containing 0. Then it is easy to see that $[\dot\gamma(0)] = [\dot\Gamma]$, i.e., (Γ_k) can be understood as the representer of a tangent vector and thus as an element from the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$. Notice that $\gamma(t)$ is not necessarily feasible for some interval $[0, \varepsilon)$, which confirms that (3.12) indeed contains $\mathcal{T}_{\mathcal{M}}^{\text{classical}}(\Omega;\boldsymbol{p})$; see (3.10).

2. The tangent cone $\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ defined in (3.12) agrees with

$$[(\mathrm{d}\varphi)(\boldsymbol{p})]^{-1} \mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p})),$$

which is how it was introduced in Yang, Zhang, Song, 2014, eq. (3.7).

LEMMA 3.6 (Properties of the tangent cone). *For any $\boldsymbol{p} \in \Omega$, the tangent cone $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is a cone in the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$.*

Proof. Let $(\Gamma_k) = (\boldsymbol{p}_k, t_k)$ be a tangential sequence to Ω at $\boldsymbol{p}$. When t_k is replaced by t_k/α for some $\alpha > 0$, then it is easy to see that the resulting sequence is a tangential sequence generating the sequential tangent vector $\alpha\,[\dot\Gamma]$. This shows that $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is a cone. ∎

The analogue of (3.2) is the following

THEOREM 3.7 (First-order necessary optimality condition). *Suppose that $\boldsymbol{p}^* \in \Omega$ is a local minimizer of (1.1). Then we have*

$$[\dot\Gamma](f) \ge 0 \tag{3.13}$$

for all sequential tangent vectors $[\dot\Gamma] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)$.

Proof. Let $(\Gamma_k) = (\boldsymbol{p}_k, t_k)$ be a tangential sequence to Ω at $\boldsymbol{p}^*$ w.r.t. some chart φ about $\boldsymbol{p}^*$, generating the sequential tangent vector $[\dot\Gamma] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)$. Suppose $(\varphi(\boldsymbol{p}_k) - \varphi(\boldsymbol{p}^*))/t_k \to d$ for some $d \in \mathbb{R}^n$. Then, for some $\varepsilon > 0$, we have by local optimality of $\boldsymbol{p}^*$

$$\begin{aligned} & 0 \le \frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p}^*)}{t_k} && \text{for sufficiently large } k \\ \Rightarrow\quad & 0 \le [\dot\Gamma](f) && \text{by (3.11).} \end{aligned}$$

This concludes the proof. ∎

Next we introduce the concept of the linearizing cone (3.4) in the tangent space.

DEFINITION 3.8 (Linearizing cone). *For any $\boldsymbol{p} \in \Omega$, we define the linearizing cone to the feasible set Ω by*

$$\begin{aligned} \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p}) \coloneqq \big\{ [\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}) : {} & [\dot\gamma(0)](g^i) \le 0 \quad \textit{for all } i \in \mathcal{A}(\boldsymbol{p}), \\ & [\dot\gamma(0)](h^j) = 0 \quad \textit{for all } j = 1, \ldots, q \big\}. \end{aligned} \tag{3.14}$$

As in subsection 3.1, $\mathcal{A}(\boldsymbol{p}) \coloneqq \big\{ 1 \le i \le m : g^i(\boldsymbol{p}) = 0 \big\}$ is the index set of active inequalities at $\boldsymbol{p}$, and $\mathcal{I}(\boldsymbol{p}) \coloneqq \{1, \ldots, m\} \setminus \mathcal{A}(\boldsymbol{p})$ are the inactive inequalities. Notice that, as is customary in differential geometry, we denote the components of the vector-valued functions g and h by upper indices.

LEMMA 3.9 (Relation between the cones). *For any $\boldsymbol{p} \in \Omega$, $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is a convex cone, and $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}) \subset \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ holds.*

Proof. To show that $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is a convex cone, let γ_1 and γ_2 be two curves about $\boldsymbol{p}$, generating the elements $[\dot\gamma_1(0)]$ and $[\dot\gamma_2(0)]$ in $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})$, and let $\alpha_1, \alpha_2 > 0$. Since $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ is a vector space under $\odot$ and $\oplus$, we have

$$\begin{aligned} [(\alpha_1 \odot \gamma_1) \oplus (\alpha_2 \odot \gamma_2)](g^i) &= \alpha_1\,[\gamma_1](g^i) + \alpha_2\,[\gamma_2](g^i) \le 0 && \text{for } i \in \mathcal{A}(\boldsymbol{p}), \\ [(\alpha_1 \odot \gamma_1) \oplus (\alpha_2 \odot \gamma_2)](h^j) &= \alpha_1\,[\gamma_1](h^j) + \alpha_2\,[\gamma_2](h^j) = 0 && \text{for } j = 1, \ldots, q, \end{aligned}$$

hence $[(\alpha_1 \odot \gamma_1) \oplus (\alpha_2 \odot \gamma_2)]$ belongs to $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ as well.

Now let $[\dot\Gamma] \in \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ be generated by the tangential sequence $(\Gamma_k) = (\boldsymbol{p}_k, t_k)$ to Ω at $\boldsymbol{p}$. Recall that the points $\boldsymbol{p}_k$ are feasible. Consequently, for $i \in \mathcal{A}(\boldsymbol{p})$ and $k \in \mathbb{N}$ we have

$$0 \ge \frac{g^i(\boldsymbol{p}_k) - g^i(\boldsymbol{p})}{t_k} \quad \Rightarrow \quad [\dot\Gamma](g^i) \le 0.$$

Similarly, we get $[\dot\Gamma](h^j) = 0$ for $j = 1, \ldots, q$. This shows $[\dot\Gamma] \in \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})$. ◻

Similar to (3.5), the polar cone to a subset $B \subset \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ of the tangent space is defined as

$$B^\circ \coloneqq \big\{(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}^*_{\mathcal{M}}(\boldsymbol{p}) : (\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] \le 0 \text{ for all } [\dot\gamma(0)] \in B\big\}. \tag{3.15}$$

Let us calculate a representation of $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ$, similar to (3.6).

Lemma 3.10. *For any $\boldsymbol{p} \in \Omega$, we have*

$$\begin{aligned} \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ = \Big\{(\mathrm{d}s)(\boldsymbol{p}) = \sum_{i=1}^m \mu_i\,(\mathrm{d}g^i)(\boldsymbol{p}) + \sum_{j=1}^q \lambda_j\,(\mathrm{d}h^j)(\boldsymbol{p}), \\ \mu_i \ge 0 \textit{ for } i \in \mathcal{A}(\boldsymbol{p}),\ \mu_i = 0 \textit{ for } i \in \mathcal{I}(\boldsymbol{p}),\ \lambda_j \in \mathbb{R}\Big\} \subset \mathcal{T}^*_{\mathcal{M}}(\boldsymbol{p}), \end{aligned} \tag{3.16}$$

Proof. When $(\mathrm{d}s)(\boldsymbol{p})$ belongs to the set on the right hand side of (3.16) and $[\dot\gamma(0)] \in \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ is arbitrary, then

$$\begin{aligned} (\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] &= \sum_{i=1}^m \mu_i\,(\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] + \sum_{j=1}^q \lambda_j\,(\mathrm{d}h^j)(\boldsymbol{p})[\dot\gamma(0)] \\ &= \sum_{i=1}^m \mu_i\,[\dot\gamma(0)](g^i) + \sum_{j=1}^q \lambda_j\,[\dot\gamma(0)](h^j) \end{aligned}$$

by definition of the differential; see (2.5). Utilizing the sign conditions in (3.16) and the definition of $\mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ in (3.14) shows $(\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] \le 0$, i.e., $(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ$.

For the converse, consider the linear map

$$A \coloneqq \begin{pmatrix} -(\mathrm{d}g^i)(\boldsymbol{p})\big|_{i\in\mathcal{A}(\boldsymbol{p})} \\ -(\mathrm{d}h^j)(\boldsymbol{p})\big|_{j=1,\ldots,q} \\ (\mathrm{d}h^j)(\boldsymbol{p})\big|_{j=1,\ldots,q} \end{pmatrix}$$

which maps the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ into $\mathbb{R}^q$, where $q = |\mathcal{A}(\boldsymbol{p})| + 2\,q$. By (3.14), $[\dot\gamma(0)] \in \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ holds if and only if $A\,[\dot\gamma(0)] \ge 0$.

Now let $(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ$, i.e., $(\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] \le 0$ holds for all $[\dot\gamma(0)]$ such that $A\,[\dot\gamma(0)] \ge 0$. The Farkas Lemma 3.1 (with $V = \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ and $b = -(\mathrm{d}s)(\boldsymbol{p})$) shows that $A^* y = -(\mathrm{d}s)(\boldsymbol{p})$ has a solution $y \in \mathbb{R}_q$, $y \ge 0$. Now split $y =: (\mu_{|\mathcal{A}(\boldsymbol{p})}, \lambda^+, \lambda^-)$, set $\lambda \coloneqq \lambda^+ - \lambda^-$ and pad μ by setting $\mu_{|\mathcal{I}(\boldsymbol{p})} \coloneqq 0$. This shows that $(\mathrm{d}s)(\boldsymbol{p})$ indeed has the representation postulated in (3.16). ◻

We associate with (1.1) the Lagrangian

$$\mathcal{L}(\boldsymbol{p}, \mu, \lambda) \coloneqq f(\boldsymbol{p}) + \mu\,g(\boldsymbol{p}) + \lambda\,h(\boldsymbol{p}), \tag{3.17}$$

where $\mu \in \mathbb{R}_m$ and $\lambda \in \mathbb{R}_q$, and the KKT conditions

$$(\mathrm{d}\mathcal{L})(\boldsymbol{p},\mu,\lambda) = (\mathrm{d}f)(\boldsymbol{p}) + \mu\,(\mathrm{d}g)(\boldsymbol{p}) + \lambda\,(\mathrm{d}h)(\boldsymbol{p}) = 0, \tag{3.18a}$$

$$h(\boldsymbol{p}) = 0, \tag{3.18b}$$

$$\mu \ge 0, \quad g(\boldsymbol{p}) \le 0, \quad \mu\, g(\boldsymbol{p}) = 0. \tag{3.18c}$$

Here we introduced for convenience of notation the differential of the vector-valued functions $g = (g^1, \ldots, g^m)^{\mathrm{T}}$

$$(\mathrm{d}g)(\boldsymbol{p}) \coloneqq \begin{pmatrix} (\mathrm{d}g^1)(\boldsymbol{p}) \\ \vdots \\ (\mathrm{d}g^m)(\boldsymbol{p}) \end{pmatrix}$$

and similarly for h.

Just as in the case of $\mathcal{M} = \mathbb{R}^n$, it is easy to see by Lemma 3.10 that the KKT conditions (3.18) are equivalent to

$$-\,(\mathrm{d}f)(\boldsymbol{p}) \in \mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p})^\circ. \tag{3.19}$$

We thus obtain the analogue of Theorem 3.2:

THEOREM 3.11. *Suppose that $\boldsymbol{p}^*$ is a local minimizer of (1.1) and that the GCQ $\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p}^*)^\circ = \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p}^*)^\circ$ holds at $\boldsymbol{p}^*$. Then there exist Lagrange multipliers $\mu \in \mathbb{R}_m$, $\lambda \in \mathbb{R}_q$, such that the KKT conditions (3.18) hold.*

3.3. Constraint Qualifications for Optimization Problems on Smooth Manifolds. In this section we introduce the constraint qualifications (CQ) of linear independence (LICQ), Mangasarian–Fromovitz (MFCQ), Abadie (ACQ) and Guignard (GCQ) and show that the chain of implications

$$\text{LICQ} \;\Rightarrow\; \text{MFCQ} \;\Rightarrow\; \text{ACQ} \;\Rightarrow\; \text{GCQ} \tag{3.20}$$

continues to hold in the smooth manifold setting.

DEFINITION 3.12 (Constraint qualifications). *Suppose that $\boldsymbol{p} \in \Omega$ holds. We define the following constraint qualifications at $\boldsymbol{p}$.*

(*a*) *The LICQ holds at $\boldsymbol{p}$ if $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q \cup \{(\mathrm{d}g^i)(\boldsymbol{p})\}_{i\in\mathcal{A}(\boldsymbol{p})}$ is a linearly independent set in the cotangent space $\mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$.*

(*b*) *The MFCQ holds at $\boldsymbol{p}$ if $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q$ is a linearly independent set and if there exists a tangent vector $[\dot\gamma(0)]$ (termed an MFCQ vector) such that*

$$\begin{aligned} (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] &< 0 \quad \textit{for all } i \in \mathcal{A}(\boldsymbol{p}), \\ (\mathrm{d}h^j)(\boldsymbol{p})[\dot\gamma(0)] &= 0 \quad \textit{for all } j = 1,\ldots,q. \end{aligned} \tag{3.21}$$

(*c*) *The ACQ holds at $\boldsymbol{p}$ if $\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p}) = \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$.*

(*d*) *The GCQ holds at $\boldsymbol{p}$ if $\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p})^\circ = \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ$.*

PROPOSITION 3.13. *LICQ implies MFCQ.*

Proof. Consider the linear system

$$A\,[\dot\gamma(0)] \coloneqq \begin{pmatrix} (\mathrm{d}g^i)(\boldsymbol{p})\big|_{i\in\mathcal{A}(\boldsymbol{p})} \\ (\mathrm{d}h^j)(\boldsymbol{p})\big|_{j=1,\ldots,q} \end{pmatrix} [\dot\gamma(0)] = (-1,\ldots,-1,0,\ldots,0)^{\mathrm{T}}.$$

Since the linear map A is surjective by assumption, this system is solvable, and $[\dot\gamma(0)]$ satisfies the MFCQ conditions. ∎

In order to show that MFCQ implies ACQ, we first prove the following result; compare Geiger, Kanzow, 2002, Lem. 2.37.

PROPOSITION 3.14. *Suppose that $\boldsymbol{p}\in\Omega$ and that the MFCQ holds at $\boldsymbol{p}$ with the MFCQ vector $[\dot\gamma(0)]$. Then the curve γ about $\boldsymbol{p}$ which generates $[\dot\gamma(0)]$ can be chosen to satisfy the following:*

(a) $h^j(\gamma(t)) = 0$ *for all* $t\in(-\varepsilon,\varepsilon)$ *and all* $j=1,\ldots,q$.
(b) $\gamma(t)\in\Omega$ *for all* $t\in[0,\varepsilon)$ *and even* $g^i(\gamma(t))<0$ *for all* $t\in(0,\varepsilon)$ *and all* $i=1,\ldots,m$.

Proof. Choose a chart φ about $\boldsymbol{p}$ and set $x_0 \coloneqq \varphi(\boldsymbol{p})$. We start with an arbitrary C^1-curve ζ about $\boldsymbol{p}$ which generates the MFCQ vector $[\dot\gamma(0)]$. We are going to define, in the course of the proof, an alternative C^1-curve γ about $\boldsymbol{p}$ which generates the same tangent vector and which satisfies the conditions stipulated.

In the absence of equality constraints ($q=0$), we can simply take $\gamma=\zeta$. Suppose now that $q\ge 1$ holds. For some $\varepsilon>0$, $\zeta(t)$ belongs to the domain of φ whenever $t\in(-\varepsilon,\varepsilon)$. Define

$$H(y,t) \coloneqq (h\circ\varphi^{-1})\big((\varphi\circ\zeta)(t) + (h\circ\varphi^{-1})'(x_0)^{\mathrm{T}}y\big), \quad (y,t)\in\mathbb{R}^q\times(-\varepsilon,\varepsilon).$$

Then $H(0,0) = (h\circ\varphi^{-1})(x_0+0) = h(\boldsymbol{p}) = 0$ holds. Moreover, by the chain rule, the Jacobian of H w.r.t. y is

$$H_y(y,t) = (h\circ\varphi^{-1})'\big((\varphi\circ\zeta)(t) + (h\circ\varphi^{-1})'(x_0)^{\mathrm{T}}y\big)\,(h\circ\varphi^{-1})'(x_0)^{\mathrm{T}}$$

and in particular, $H_y(0,0) = (h\circ\varphi^{-1})'(x_0)\,(h\circ\varphi^{-1})'(x_0)^{\mathrm{T}}$. Since $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q$ is a linearly independent set of cotangent vectors, the $q\times n$-matrix $(h\circ\varphi^{-1})'(x_0)$ has rank q. To see this, consider the tangent vectors along the curves $t\mapsto\gamma_k(t)\coloneqq\varphi^{-1}(\varphi(\boldsymbol{p})+t\,e_k)$ for $k=1,\ldots,n$. The entry (j,k) of $(h\circ\varphi^{-1})'(x_0)$ equals $(\mathrm{d}h^j)(\boldsymbol{p})\,[\dot\gamma_k(0)] = \frac{\mathrm{d}}{\mathrm{d}t}(h^j\circ\gamma_k)(t)\big|_{t=0}$. Since the tangent vectors $\{[\dot\gamma_k]\}_{k=1}^n$ are linearly independent and the cotangent vectors $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q$ as well, the matrix $(h\circ\varphi^{-1})'(x_0)$ has full rank as claimed. This shows that $H_y(0,0)$ is symmetric positive definite. Moreover,

$$H_t(y,t) = (h\circ\varphi^{-1})'\big((\varphi\circ\zeta)(t) + (h\circ\varphi^{-1})'(x_0)^{\mathrm{T}}y\big)\,(\varphi\circ\zeta)'(t),$$

whence $H_t(0,0) = (h\circ\varphi^{-1})'(x_0)\,(\varphi\circ\zeta)'(0) = (h\circ\zeta)'(0)$. In particular, the j-th coordinate of $H_t(0,0)$ is equal to $[\dot\zeta(0)](h^j) = (\mathrm{d}h^j)(\boldsymbol{p})\,[\dot\zeta(0)] = 0$ by the properties of the MFCQ vector $[\dot\zeta(0)]$.

The implicit function theorem ensures that there exists a function $y\colon(-\varepsilon_0,\varepsilon_0)\to\mathbb{R}^q$ of class C^1 such that $H(y(t),t)=0$ and $y(0)=0$ holds, and moreover, $\dot y(0) = H_y(0,0)^{-1}H_t(0,0) = 0$.

Using $y(\cdot)$, we define, on a suitable open interval containing 0, the curve

$$\gamma(t) \coloneqq \varphi^{-1}\big((\varphi \circ \zeta)(t) + (h \circ \varphi^{-1})'(x_0)^{\mathrm{T}} y(t)\big) \in \mathcal{M}.$$

This curve is of class C^1 by construction, it satisfies $\gamma(0) = \varphi^{-1}(x_0 + 0) = \boldsymbol{p}$ and generates the same tangent vector as the original curve ζ. To see the latter, we consider an arbitrary C^1-function f defined near $\boldsymbol{p}$ and calculate

$$\begin{aligned}(f \circ \gamma)'(t) = (f \circ \varphi^{-1})'\big((\varphi \circ \zeta)(t) + (h \circ \varphi^{-1})'(x_0)^{\mathrm{T}} y(t)\big)\\ \cdot \big[(\varphi \circ \zeta)'(t) + (h \circ \varphi^{-1})'(x_0)^{\mathrm{T}} \dot{y}(t)\big].\end{aligned}$$

This implies

$$[\dot\gamma(0)](f) = (f \circ \gamma)'(0) = (f \circ \varphi^{-1})'(x_0)\,(\varphi \circ \zeta)'(0) = (f \circ \zeta)'(0) = [\dot\zeta(0)](f).$$

By construction, we have

$$h(\gamma(t)) = (h \circ \varphi^{-1})\big((\varphi \circ \zeta)(t) + (h \circ \varphi^{-1})'(x_0)^{\mathrm{T}} y(t)\big) = H(y(t), t) = 0$$

on a suitable interval $(-\varepsilon, \varepsilon)$. It remains to verify the conditions pertaining to the inequality constraints. When $i \in \mathcal{I}(\boldsymbol{p})$, then by continuity, $g^i(\gamma(t)) < 0$ for all $t \in (-\varepsilon_i, \varepsilon_i)$. When $i \in \mathcal{A}(\boldsymbol{p})$, consider the auxiliary function $\phi(t) \coloneqq g^i(\gamma(t))$, which satisfies $\phi(0) = g^i(\gamma(0)) = 0$ and $\dot\phi(0) = (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] = (\mathrm{d}g^i)(\boldsymbol{p})[\dot\zeta(0)] < 0$. An applications of Taylor's theorem now implies that there exists $\varepsilon_i > 0$ such that $\phi(t) < 0$ holds for $t \in (0, \varepsilon_i)$. Taking $\varepsilon = \min\{\varepsilon_i : i = 1, \ldots, m\}$ finishes the proof. ◻

PROPOSITION 3.15. *MFCQ implies ACQ.*

Proof. In view of Lemma 3.9, we only need to show $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}) \supset \mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega; \boldsymbol{p})$. To this end, suppose that $[\dot\gamma_0(0)]$ is an element of $\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega; \boldsymbol{p})$ defined in (3.14), generated by some C^1-curve about $\boldsymbol{p} = \gamma_0(0)$. Moreover, let γ be another C^1-curve about $\boldsymbol{p}$ such that $[\dot\gamma(0)]$ is an MFCQ vector, see (3.21). Finally, choose an arbitrary chart φ about $\boldsymbol{p}$.

For any $\tau \in (0, 1]$, consider the curve

$$\gamma_0 \oplus (\tau \odot \gamma) : t \mapsto \varphi^{-1}\big((\varphi \circ \gamma_0)(t) + (\varphi \circ \gamma)(\tau\, t) - \varphi(\boldsymbol{p})\big) \in \mathcal{M},$$

which is defined on an interval $(-\varepsilon, \varepsilon)$ where both γ and γ_0 are defined. Moreover by reducing ε if necessary we achieve that $\gamma(t)$ and $\gamma(\tau\, t)$ belong to the domain of the chosen chart φ when $t \in (-\varepsilon, \varepsilon)$.

We first show that $[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus (\tau \odot \gamma))(0)] \to [\dot\gamma_0(0)]$ as $\tau \searrow 0$. Indeed, for any C^1-function f defined near $\boldsymbol{p}$, we have

$$\begin{aligned}
&(\mathrm{d}f)(\boldsymbol{p})[\tfrac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus (\tau \odot \gamma))(0)]\\
&\quad = [\tfrac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus (\tau \odot \gamma))(0)](f) && \text{by definition of } (\mathrm{d}f)(\boldsymbol{p}), \text{ see (2.5)}\\
&\quad = \frac{\mathrm{d}}{\mathrm{d}t}\big[f \circ (\gamma_0 \oplus (\tau \odot \gamma))\big]\Big|_{t=0} && \text{by def. of tangent vectors, see (2.2)}\\
&\quad = (f \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\left[\frac{\mathrm{d}}{\mathrm{d}t}\big((\varphi \circ \gamma_0) + \tau\,(\varphi \circ \gamma)\big)\Big|_{t=0}\right] && \text{by the chain rule}
\end{aligned}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma_0)\Big|_{t=0} + \tau \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)\Big|_{t=0} \qquad \text{by the chain rule}$$
$$= (\mathrm{d}f)(\boldsymbol{p})[\dot\gamma_0(0)] + \tau\,(\mathrm{d}f)(\boldsymbol{p})[\dot\gamma(0)],$$

and the right hand side converges to $[\dot\gamma_0(0)](f)$ as $\tau \searrow 0$.

Next we show that the tangent vector along $\gamma_0 \oplus (\tau \odot \gamma)$ is an MFCQ vector for any $\tau \in (0,1]$. Similarly as above, we have

$$(\mathrm{d}g^i)(\boldsymbol{p})[\tfrac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus (\tau \odot \gamma))(0)] = (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma_0(0)] + \tau\,(\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)]$$

which is negative for any $i \in \mathcal{A}(\boldsymbol{p})$ since $\tau > 0$. Analogously, $(\mathrm{d}h^j)(\boldsymbol{p})[\tfrac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus (\tau \odot \gamma))(0)] = 0$ follows for all $j = 1, \ldots, q$. This confirms that $\gamma_0 \oplus (\tau \odot \gamma)$ is indeed an MFCQ vector.

Finally, by virtue of Proposition 3.14, we may assume, without loss of generality, that $\gamma_0 \oplus (\tau \odot \gamma)$ is feasible for $t \in [0, \varepsilon)$. In other words,

$$h\big((\gamma_0 \oplus (\tau \odot \gamma))(t)\big) = (h \circ \varphi^{-1}) \circ \big((\varphi \circ \gamma_0) + \tau\,(\varphi \circ \gamma)\big)(t) \equiv 0 \quad \text{for all } t \in [0, \varepsilon).$$

By continuity, we obtain in the limit $\tau \searrow 0$ that $h(\gamma_0(t)) = 0$ for $t \in [0, \varepsilon)$ holds as well. Similarly, $g(\gamma_0(t)) \le 0$ for $t \in [0, \varepsilon)$ follows. This shows that $[\dot\gamma_0(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ in the sense of Remark 3.5. □

Finally, the fact that ACQ implies GCQ is trivial, so (3.20) is proved.

4. Constraint Qualifications and the Polyhedron of Lagrange Multipliers. In this section we consider a number of results relating various constraint qualifications to the set of KKT multipliers at a local minimizer of (1.1). To this end, we fix an arbitrary feasible point $\boldsymbol{p} \in \Omega$ and consider the cone

$$\mathcal{F}(\boldsymbol{p}) \coloneqq \{f \in C^1(\mathcal{M}, \mathbb{R}) : \boldsymbol{p} \text{ is a local minimizer for (1.1)}\} \tag{4.1}$$

of objective functions of class C^1 attaining a local minimum at $\boldsymbol{p}$. For $f \in \mathcal{F}(\boldsymbol{p})$, we denote by

$$\Lambda(f; \boldsymbol{p}) \coloneqq \{(\mu, \lambda) \in \mathbb{R}_m \times \mathbb{R}_p : (3.18) \text{ holds}\} \tag{4.2}$$

the corresponding set of Lagrange multipliers. It is easy to see that $\Lambda(f; \boldsymbol{p})$ is a closed, convex (potentially empty) polyhedron.

The following theorem is known in the case $\mathcal{M} = \mathbb{R}^n$; see Gauvin, 1977; Gould, Tolle, 1971 and Wachsmuth, 2013, Thms. 1 and 2. It continues to hold verbatim for (1.1).

Theorem 4.1 (Connections between CQs and Lagrange Multipliers). *Suppose that $\boldsymbol{p} \in \Omega$.*

(*a*) *The set $\Lambda(f; \boldsymbol{p})$ is non-empty for all $f \in \mathcal{F}(\boldsymbol{p})$ if and only if (GCQ) holds at $\boldsymbol{p}$.*
(*b*) *Suppose (MFCQ) holds at $\boldsymbol{p}$. Then the set $\Lambda(f; \boldsymbol{p})$ is compact for all $f \in \mathcal{F}(\boldsymbol{p})$.*
(*c*) *If $\Lambda(f; \boldsymbol{p})$ is non-empty, compact for some $f \in \mathcal{F}(\boldsymbol{p})$, then (MFCQ) holds at $\boldsymbol{p}$.*
(*d*) *The set $\Lambda(f; \boldsymbol{p})$ is a singleton for all $f \in \mathcal{F}(\boldsymbol{p})$ if and only if (LICQ) holds at $\boldsymbol{p}$.*

In order to prove Theorem 4.1, we are going to work with some chart about $\boldsymbol{p}$ and apply the result in $\mathbb{R}^n$. Therefore, a preparatory step is required in order to confirm that this transformation leaves the notion of local minimum intact.

LEMMA 4.2 (compare Yang, Zhang, Song, 2014, Sec. 4.1). *Suppose that (U, φ) is a arbitrary chart about $\boldsymbol{p}^*$. The following are equivalent:*

(a) *$\boldsymbol{p}^*$ is a local minimizer of (1.1).*
(b) *$\varphi(\boldsymbol{p}^*)$ is a local minimizer of*

$$\text{(4.3)} \qquad \begin{cases} \textit{Minimize} & (f \circ \varphi^{-1})(x), \quad x \in \varphi(U) \subset \mathbb{R}^n \\ \textit{s.t.} & (g \circ \varphi^{-1})(x) \le 0 \\ \textit{and} & (h \circ \varphi^{-1})(x) = 0. \end{cases}$$

Proof. Suppose first that $\boldsymbol{p}^* \in \Omega$ is a local minimizer of (1.1), i.e., there exists an open neighborhood U_1 of $\boldsymbol{p}^*$ such that $f(\boldsymbol{p}^*) \le f(\boldsymbol{p})$ holds for all $\boldsymbol{p} \in U_1 \cap \Omega$. We can assume, by shrinking U_1 if necessary, that $U_1 \subset U$ holds. This implies $f(\varphi(\boldsymbol{p}^*)) \le f(\varphi(\boldsymbol{p}))$ for all $\boldsymbol{p} \in U_1 \cap \Omega$. Since $\varphi(U_1)$ is an open neighborhood of $\varphi(\boldsymbol{p}^*)$, $\varphi(\boldsymbol{p}^*)$ is a minimizer of (4.3). The converse is proved similarly. ∎

Proof of Theorem 4.1.

(a): Theorem 3.11 shows that (GCQ) implies $\Lambda(f; \boldsymbol{p}) \neq \emptyset$ for any $f \in \mathcal{F}(\boldsymbol{p})$. The converse is proved in Gould, Tolle, 1971, Sec. 4 for the case $\mathcal{M} = \mathbb{R}^n$; see also Bazaraa, Shetty, 1976, Thm. 6.3.2. We apply this result in the following way. Suppose that $(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})^\circ \subset \mathcal{T}^*_{\mathcal{M}}(\boldsymbol{p})$ holds. Fix an arbitrary chart (U, φ) about $\boldsymbol{p}$. Suppose that d is an arbitrary element from the tangent cone $\mathcal{T}_{\varphi(U \cap \Omega)}(\varphi(\boldsymbol{p}))$, i.e., there exist sequences $(x_k) \subset \varphi(U \cap \Omega)$ and $t_k \searrow 0$ such that $x_k \to x_0 \coloneqq \varphi(\boldsymbol{p})$ and $(x_k - x_0)/t_k \to d$. Define $\boldsymbol{p}_k \coloneqq \varphi(x_k)$. Then clearly, $(\Gamma_k) \coloneqq (\boldsymbol{p}_k, t_k)$ is a tangential sequence to Ω at $\boldsymbol{p}$ in the sense of Definition 3.3. When we denote the sequential tangent vector generated by (Γ_k) by $[\dot{\Gamma}]$, we have

$$(\mathrm{d}s)(\boldsymbol{p})\,[\dot{\Gamma}] = (s \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\, d \le 0.$$

This shows $(s \circ \varphi^{-1})'(\varphi(\boldsymbol{p})) \in \mathcal{T}_{\varphi(U \cap \Omega)}(\varphi(\boldsymbol{p}))^\circ$.

Using Bazaraa, Shetty, 1976, Thm. 6.3.2 we can construct a C^1-function $r \colon \mathbb{R}^n \to \mathbb{R}$ such that $r'(\varphi(\boldsymbol{p})) = -(s \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))$ holds and $\varphi(\boldsymbol{p})$ is a local minimizer of (4.3) but with the objective r in place of $(f \circ \varphi^{-1})$. By Lemma 4.2, $\boldsymbol{p}$ is a local minimizer of (1.1) with objective $r \circ \varphi$. By assumption, $\Lambda(r \circ \varphi, \boldsymbol{p})$ is non-empty, i.e., there exist Lagrange multipliers μ and λ such that

$$(\mathrm{d}(r \circ \varphi))(\boldsymbol{p}) + \mu\,(\mathrm{d}g)(\boldsymbol{p}) + \lambda\,(\mathrm{d}h)(\boldsymbol{p}) = 0$$

and (3.18b), (3.18c) hold. In other words, $-(\mathrm{d}(r \circ \varphi))(\boldsymbol{p}) \in \mathcal{T}^{\mathrm{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})^\circ$, see (3.19). Moreover, the differentials of $r \circ \varphi$ and $-s$ at $\boldsymbol{p}$ coincide since

$$\begin{aligned}
&(\mathrm{d}(r \circ \varphi))(\boldsymbol{p})\,[\dot{\gamma}(0)] \\
&\quad = [\dot{\gamma}(0)](r \circ \varphi) && \text{by definition (2.5) of the differential} \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}t}(r \circ \varphi \circ \gamma)(t)\Big|_{t=0} && \text{by definition (2.2) of a tangent vector} \\
&\quad = r'(x_0)\frac{\mathrm{d}}{\mathrm{d}t}(\varphi \circ \gamma)(t)\Big|_{t=0} && \text{by the chain rule}
\end{aligned}$$

$$
\begin{aligned}
&= -(s \circ \varphi^{-1})'(x_0) \frac{\mathrm{d}}{\mathrm{d}t}(\varphi \circ \gamma)(t)\Big|_{t=0} && \text{by construction of } r \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}(s \circ \gamma)(t)\Big|_{t=0} && \text{by the chain rule} \\
&= -(\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] && \text{by (2.2), (2.5)}
\end{aligned}
$$

holds for arbitrary tangent vectors $[\dot\gamma(0)]$ in $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$. This shows that $\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})^\circ \subset \mathcal{T}_{\mathcal{M}}^{\text{lin}}(\Omega;\boldsymbol{p})^\circ$ holds, i.e., the (GCQ) is satisfied.

(*b*) and (*c*): a possible proof of these results is based on linear programming arguments in the Lagrange multiplier space and thus it is directly applicable here as well. We sketch the proof following Burke, 2014 for the reader's convenience. One first observes that (MFCQ) is equivalent to the feasibility of the linear program

$$
\begin{aligned}
\text{Minimize}\quad & 0, \quad [\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}),\\
\text{s.t.}\quad & (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] \le -1 \quad \text{for all } i \in \mathcal{A}(\boldsymbol{p}),\\
\text{and}\quad & (\mathrm{d}h^j)(\boldsymbol{p})[\dot\gamma(0)] = 0 \quad \text{for all } j = 1, \dots, q.
\end{aligned}
\tag{4.4}
$$

Using strong duality, one shows that (MFCQ) is in turn equivalent to the system

$$
\begin{aligned}
&\mu\,(\mathrm{d}g)(\boldsymbol{p}) + \lambda\,(\mathrm{d}h)(\boldsymbol{p}) = 0,\\
&\mu_i \ge 0 \quad \text{for all } i \in \mathcal{A}(\boldsymbol{p}),\\
&\mu_i = 0 \quad \text{for all } i \in \mathcal{I}(\boldsymbol{p}),\\
&\lambda_j = 0 \quad \text{for all } j = 1, \dots, q
\end{aligned}
\tag{4.5}
$$

having the only solution $(\mu, \lambda) = 0$.

Now if $f \in \mathcal{F}(\boldsymbol{p})$ holds and $\Lambda(f;\boldsymbol{p})$ is not bounded, then there exists a non-zero direction (μ, λ) in $\Lambda(f;\boldsymbol{p})$ verifying (4.5), i.e., (MFCQ) does not hold. This shows (*b*). Conversely, if (MFCQ) does not hold, then there exists a non-zero vector (μ, λ) satisfying (4.5). When $(\mu_0, \lambda_0) \in \Lambda(f;\boldsymbol{p})$, then $(\mu_0, \lambda_0) + t\,(\mu, \lambda)$ belongs to $\Lambda(f;\boldsymbol{p})$ as well for any $t \ge 0$, hence $\Lambda(f;\boldsymbol{p})$ is not compact. This confirms (*c*).

(*d*): We have proved in section 3 that (LICQ) implies (GCQ), so $\Lambda(f;\boldsymbol{p})$ is non-empty. The uniqueness of the Lagrange multipliers then follows immediately from (3.18a). The converse statement is proved in Wachsmuth, 2013, Thm. 2, which applies without changes. ∎

5. Numerical Example. In this section we present a numerical example in which the fulfillment of the KKT conditions (3.18) is used as an algorithmic stopping criterion. While the framework of a smooth manifold was sufficient for the discussion of first-order optimality conditions, we require more structure for algorithmic purposes. Therefore we restrict the following discussion to complete Riemannian manifolds.

A manifold is Riemannian if its tangent spaces are equipped with a smoothly varying metric $\langle\cdot,\cdot\rangle_{\boldsymbol{p}}$. This allows the conversion of the differential of the objective f, $(\mathrm{d}f)(\boldsymbol{p}) \in \mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$, to the gradient $\nabla f(\boldsymbol{p}) \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$, which fulfills

$$
\langle [\dot\gamma(0)],\, \nabla f(\boldsymbol{p})\rangle_{\boldsymbol{p}}, = (\mathrm{d}f)(\boldsymbol{p})\,[\dot\gamma(0)] \quad \text{for all } [\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}).
$$

Completeness of a Riemannian manifold refers to the fact that there exists a geodesic between any two points $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}$.

The Riemannian center of mass, also known as (Riemannian) mean was introduced in Karcher, 1977 as a variational model. Given a set of points $\boldsymbol{d}_i$, $i = 1, \ldots, N$, their Riemannian center is defined as the minimizer of

$$f(\boldsymbol{p}) \coloneqq \frac{1}{N} \sum_{i=1}^{N} d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{d}_i), \tag{5.1}$$

where $d_{\mathcal{M}} \colon \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is the distance on the Riemannian manifold $\mathcal{M}$.

We extend this classical optimization problem on manifolds by adding the constraint that the minimizer should lie within a given ball of radius $r > 0$ and center $\boldsymbol{q} \in \mathcal{M}$. We obtain the following constrained minimization problem of the form (1.1),

$$\begin{cases} \text{Minimize} \quad f(\boldsymbol{p}), \quad \boldsymbol{p} \in \mathcal{M}, \\ \qquad\quad \text{s.t.} \quad d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{q}) - r^2 \le 0, \end{cases} \tag{5.2}$$

with associated Lagrangian

$$\mathcal{L}(\boldsymbol{p}, \mu) = \frac{1}{N} \sum_{i=1}^{N} d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{d}_i) + \mu \, (d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{q}) - r^2). \tag{5.3}$$

It can be shown, see for example Bačák, 2014; Afsari, Tron, Vidal, 2013, that the objective and the constraint are C^1-functions whose gradients are given by the tangent vectors

$$\nabla f(\boldsymbol{p}) = -\frac{2}{N} \sum_{i=1}^{N} \log_{\boldsymbol{p}} \boldsymbol{d}_i \quad \text{and} \quad \nabla g(\boldsymbol{p}) = -2 \log_{\boldsymbol{p}} \boldsymbol{q}. \tag{5.4}$$

Here log denotes the logarithmic (or inverse exponential) map on $\mathcal{M}$. In other words, the geodesic curve starting in $\boldsymbol{p}$ with velocity $\log_{\boldsymbol{p}} \boldsymbol{q} \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ reaches $\boldsymbol{q}$ at time 1.

In view of (5.4), the KKT conditions (3.18) become

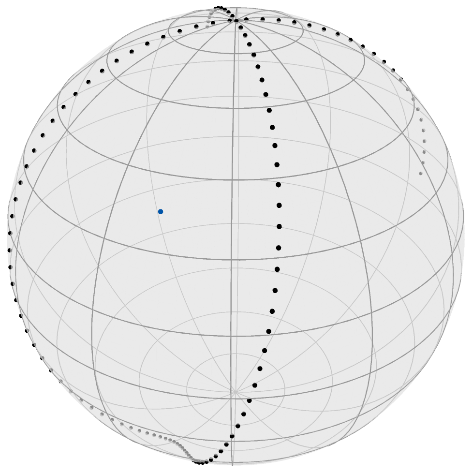
$$0 = (\mathrm{d}\mathcal{L})(\boldsymbol{p}, \mu)[\xi] = \frac{1}{N} \sum_{i=1}^{N} \langle \xi, \, -2 \log_{\boldsymbol{p}} \boldsymbol{d}_i \rangle_{\boldsymbol{p}} + \mu \, \langle \xi, \, -2 \log_{\boldsymbol{p}} \boldsymbol{q} \rangle_{\boldsymbol{p}} \quad \text{for all } \xi \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$$

$$\mu \ge 0, \quad d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{q}) \le r^2, \quad \mu \, (d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{q}) - r^2) = 0.$$

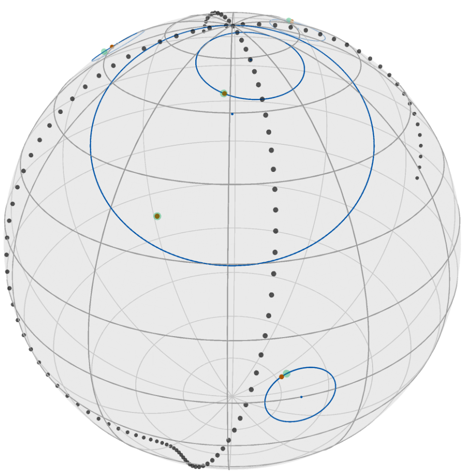
In our example we choose $\mathcal{M} = \mathbb{S}^2 \coloneqq \{\boldsymbol{p} \in \mathbb{R}^3 : |\boldsymbol{p}|_2 = 1\}$ the two-dimensional manifold of unit vectors in $\mathbb{R}^3$ or 2-sphere. The Riemannian metric is inherited from the ambient space $\mathbb{R}^3$. Since the feasible set

$$\Omega \coloneqq \{\boldsymbol{p} \in \mathbb{S}^2 : d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q}) \le r\} \tag{5.5}$$

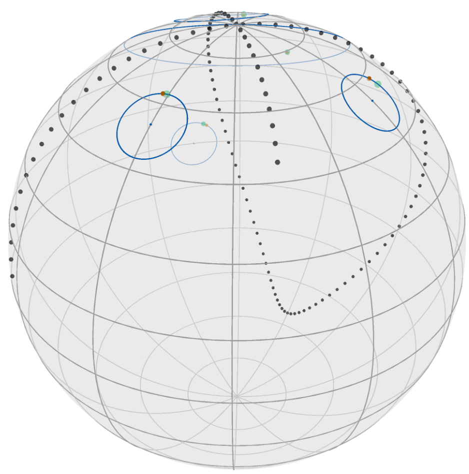
is compact, a global minimizer to (5.2) exists. Notice however, that unlike in the flat space $\mathbb{R}^2$, minimizers are not necessarily unique. Under the assumption of $r < \pi/4$, however, Ω is geodesically convex. In this case, there exists exactly one global (and no further local) solutions.



(a) Data points $\boldsymbol{d}_i$ and their mean $\bar{\boldsymbol{p}}$, the (unconstrained) Riemannian center of mass.



(b) Constrained solutions of (5.2) (light green) and projected unconstrained means $\operatorname{proj}_\Omega(\bar{\boldsymbol{p}})$ (orange) for five different feasible sets (blue).



(c) Same as Figure 1b, rotated by 180 degrees.

Fig. 1: Constrained centers of mass for five different feasible sets (centers and radii shown in blue). Unlike in $\mathbb{R}^2$, the minimizers $\boldsymbol{p}^*$ (light green) differ from the mean $\bar{\boldsymbol{p}}$ projected onto the feasible set (5.6) (orange).

Even in the absence of convexity, the LICQ is satisfied at every solution $\boldsymbol{p}^*$ unless $\boldsymbol{p}^* = \boldsymbol{q}$ holds, which is equivalent to the unconstrained mean $\bar{\boldsymbol{p}}$ coinciding with the center $\boldsymbol{q}$ of the feasible set. This does not happen for the data we use. Consequently, the Lagrange multiplier is unique by Theorem 4.1.

In our example, we choose a set of $N = 120$ data points $\boldsymbol{d}_i$ as shown in Figure 1a. Their unconstrained Riemannian center of mass $\bar{\boldsymbol{p}}$ is shown in blue. We then solve five

Algorithm 5.1 Projected gradient descent algorithm

Input: an objective function $f\colon \mathcal{M} \to \mathbb{R}$; a closed and convex set Ω; a fixed step size $s > 0$; and an initial value $\boldsymbol{p}^{(0)} \in \mathcal{M}$
$k \leftarrow 0$
repeat
  $\boldsymbol{p}^{(k+1)} \leftarrow \operatorname{proj}_\Omega\bigl(\exp_{\boldsymbol{p}^{(k)}}(s\nabla f(\boldsymbol{p}^{(k)}))\bigr)$
  $k \leftarrow k+1$
until a convergence criterion is reached
return $\boldsymbol{p}^* = \boldsymbol{p}^{(k)}$

variants of problem (5.2) which differ w.r.t. the centers $\boldsymbol{q}_i$ of the feasible set, and their radii r_i. The boundaries of the respective feasible sets, which are spherical caps, are displayed in blue in Figure 1b (front view) and Figure 1c (back view). For the choice $(\boldsymbol{q}_1, r_1)$, the distance constraint is inactive at the solution, while it is active in the other four cases. The constrained solutions $\boldsymbol{p}^*$ are shown in light green in Figures 1b and 1c.

Each instance of problem (5.2) was solved using a projected gradient descent method. Since it is a rather straightforward generalization of an unconstrained gradient algorithm, see for instance Absil, Mahony, Sepulchre, 2008, Ch. 4, Alg. 1, we only briefly sketch it here. We utilize the fact that the feasible set Ω is closed and geodesically convex when $r < \pi/4$, i.e., for any two points $\boldsymbol{p}, \boldsymbol{q} \in \Omega$, all (shortest) geodesics connecting these two points lie inside Ω. In this case the projection $\operatorname{proj}_\Omega\colon \mathcal{M} \to \Omega$ onto Ω is defined by

$$\operatorname{proj}_\Omega(\boldsymbol{p}) \coloneqq \operatorname*{arg\,min}_{\boldsymbol{q}\in\Omega} d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q}).$$

It can be computed in closed form, namely

$$\operatorname{proj}_\Omega(\boldsymbol{p}) = \exp_{\boldsymbol{q}}\bigl(b \log_{\boldsymbol{q}} \boldsymbol{p}\bigr), \quad \text{where } b = \min\Bigl\{\frac{r}{d_{\mathcal{M}}(\boldsymbol{p},\boldsymbol{q})}, 1\Bigr\}. \tag{5.6}$$

The projected gradient descent algorithm is given as pseudo code in Algorithm 5.1. The unconstrained problem with solution $\bar{\boldsymbol{p}}$ is solved similarly, omitting the projection step. This amounts to the classical gradient descent method on manifolds as given in Absil, Mahony, Sepulchre, 2008, Ch. 4, Alg. 1. In our experiments we set the step size to $s = \frac{1}{2}$ and used the first data point as initial data $\boldsymbol{p}^{(0)} = \boldsymbol{d}_1$, which is the 'bottom left' data point in Figure 1c, to solve the constrained instances. The algorithm was implemented within the Manifold-valued Image Restauration Toolbox (MVIRT)[1] Bergmann, 2017, providing a direct access to the necessary functions for the manifold of interest and the required algorithms.

Notice that in $\mathbb{R}^2$, the constrained mean of a set of points can simply be obtained by projecting the unconstrained mean $\bar{\boldsymbol{p}}$ onto the feasible disk. In $\mathbb{S}^2$, this would amount to $\operatorname{proj}_\Omega(\bar{\boldsymbol{p}})$, but this differs, in general, from the solution of (5.2) due to the curvature of $\mathbb{S}^2$. For comparison, we show the result of $\operatorname{proj}_\Omega(\bar{\boldsymbol{p}})$ in orange in Figures 1b and 1c.

[1] available open source at http://ronnybergmann.net/mvirt/.

By design, gradient type methods do not utilize Lagrange multiplier estimates. At an iterate $\boldsymbol{p}^{(k)}$, we therefore estimate the Lagrange multiplier $\mu^{(k)}$ by a least squares approach, which amounts to

$$\mu^{(k)} := -\frac{\langle \nabla g(\boldsymbol{p}^{(k)}),\, \nabla f(\boldsymbol{p}^{(k)}) \rangle_{\boldsymbol{p}^{(k)}}}{\langle \nabla g(\boldsymbol{p}^{(k)}),\, \nabla g(\boldsymbol{p}^{(k)}) \rangle_{\boldsymbol{p}^{(k)}}}. \tag{5.7}$$

We then evaluate the gradient of the Lagrangian,

$$\nabla_{\boldsymbol{p}} \mathcal{L}(\boldsymbol{p}^{(k)}, \mu^{(k)}) = -\frac{2}{N} \sum_{i=1}^{N} \log_{\boldsymbol{p}^{(k)}} \boldsymbol{d}_i - 2\, \mu^{(k)} \log_{\boldsymbol{p}^{(k)}} \boldsymbol{q} \tag{5.8}$$

and utilize its norm squared $n^{(k)} := \langle \nabla_{\boldsymbol{p}} \mathcal{L}(\boldsymbol{p}^{(k)}, \mu^{(k)}),\, \nabla_{\boldsymbol{p}} \mathcal{L}(\boldsymbol{p}^{(k)}, \mu^{(k)}) \rangle_{\boldsymbol{p}^{(k)}}$ as a stopping criterion.

For two of the five test cases we display the iteration history in Table 2. The first example is the large circle with center $\boldsymbol{q}_1 \approx (0.4319, 0.2592, 0.8639)^{\mathrm{T}}$ and radius $r_1 = \frac{\pi}{6}$. For this setup the constraint is inactive and $\bar{\boldsymbol{p}} = \boldsymbol{p}^*$ holds. The second example is shown to the right of Figure 1c and it is given by $\boldsymbol{q}_2 \approx (0, -0.5735, 0.8192)^{\mathrm{T}}$ and $r_2 = \frac{\pi}{24}$.

Since the unconstrained Riemannian mean is within the feasible set for the first example of $(\boldsymbol{q}_1, r_1)$, the projection is the identity after the first iteration. Hence for this case, the (projected) gradient descent algoriothm computes the unconstrained mean similar to Afsari, Tron, Vidal, 2013. We obtain $\boldsymbol{p}^* = \bar{\boldsymbol{p}} = \mathrm{proj}_{\Omega}(\bar{\boldsymbol{p}})$. Looking at the gradients ∇f and ∇g we see, cf. Figure 2a, that $\nabla f = 0$ while the constraint function g yields a gradient pointing towards the boundary $\partial\Omega$ of the feasible set. Clearly, the optimal Lagrange multplier is zero in this case. The iterates (green points) follow a typical gradient descent path of a Riemannian center of mass computation. Notice that the Lagrange multiplier approaches zero from below in this case. In view of (5.7), this is a result of the fact that the minimizer is approached from within the feasible set. While the objective decreases, the distance from $\boldsymbol{q}_1$ and thus g increases, leading to a negative multiplier estimate $\mu^{(k)}$.

For the second case, $(\boldsymbol{q}_2, r_2)$ the unconstrained mean lies outside the feasible set and the constraint g is strongly active, which in turn yields a nonzero value for μ. As we mentioned earlier, the optimal solution $\boldsymbol{p}^*$ is different from $\mathrm{proj}_{\Omega}(\bar{\boldsymbol{p}})$, their distance is 0.0409 , which is due to the curvature of the manifold.

References.

Absil, P.-A.; Baker, C. G.; Gallivan, K. A. (2007). "Trust-region methods on Riemannian manifolds". *Foundations of Computational Mathematics. The Journal of the Society for the Foundations of Computational Mathematics* 7.3, pp. 303–330. DOI: 10.1007/s10208-005-0179-9.

Absil, P.-A.; Mahony, R.; Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. DOI: 10.1515/9781400830244.

Afsari, B.; Tron, R.; Vidal, R. (2013). "On the convergence of gradient descent for finding the Riemannian center of mass". *SIAM Journal on Control and Optimization* 51.3, pp. 2230–2260. DOI: 10.1137/12086282X.

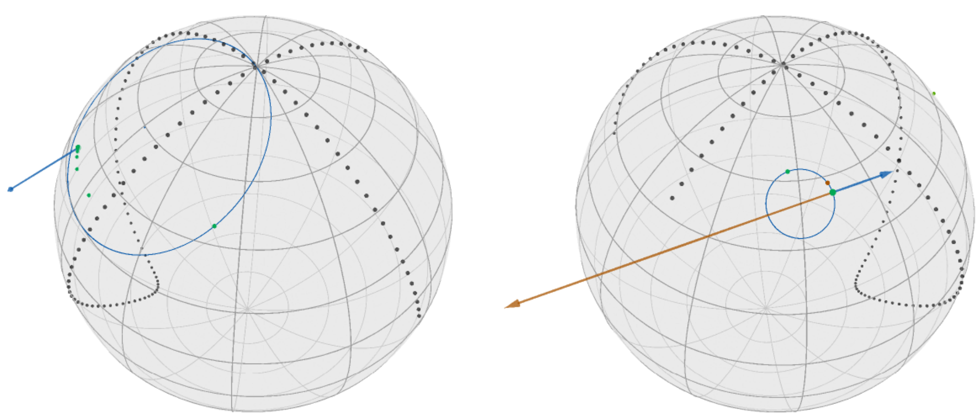
Table 2: Iteration history of Algorithm 5.1 for two instances of problem (5.2).

Results for $(\boldsymbol{q}_1, r_1)$.

k	$f(\boldsymbol{p}^{(k)})$	$n^{(k)}$	$\mu^{(k)}$
1	1.9129	0.6540	1.1722
2	1.4172	0.1243	0.2755
3	1.3754	0.0169	−0.0847
4	1.3695	0.0029	−0.0811
5	1.3684	0.0005	−0.0403
6	1.3682	0.0001	−0.0180
7	1.3682	1.18×10^{-5}	−0.0078
8	1.3682	3.26×10^{-6}	−0.0034
9	1.3682	6.02×10^{-7}	−0.0014
10	1.3682	1.11×10^{-7}	−0.0006
11	1.3682	2.05×10^{-8}	−0.0003
12	1.3682	3.79×10^{-9}	−0.0001
13	1.3682	6.99×10^{-10}	-4.94×10^{-5}
14	1.3682	1.29×10^{-10}	-2.12×10^{-5}
15	1.3682	2.38×10^{-11}	-9.13×10^{-6}
16	1.3682	4.40×10^{-12}	-3.93×10^{-6}
17	1.3682	8.13×10^{-13}	-1.69×10^{-6}
18	1.3682	1.50×10^{-13}	-7.25×10^{-7}
19	1.3682	2.77×10^{-14}	-3.11×10^{-7}
20	1.3682	5.12×10^{-15}	-1.34×10^{-7}
21	1.3682	9.45×10^{-16}	-5.75×10^{-8}
22	1.3682	1.74×10^{-16}	-2.47×10^{-8}

Results for $(\boldsymbol{q}_2, r_2)$.

k	$f(\boldsymbol{p}^{(k)})$	$n^{(k)}$	$\mu^{(k)}$
1	2.2190	2.1771	1.3833
2	2.0215	0.0011	1.2454
3	2.0214	5.04×10^{-6}	1.2475
4	2.0214	2.40×10^{-8}	1.2476
5	2.0214	1.15×10^{-10}	1.2477
6	2.0214	5.50×10^{-12}	1.2477
7	2.0214	2.63×10^{-15}	1.2477
8	2.0214	1.25×10^{-17}	1.2477



(a) Constraint data $(\boldsymbol{q}_1, r_1)$. (b) Constraint data $(\boldsymbol{q}_2, r_2)$.

Fig. 2: Iterates (green) of the projected gradient method and the final gradients of the objective f (orange) as well as the contraint g (blue).

Aubin, J.-P.; Frankowska, H. (2009). *Set-valued analysis*. Modern Birkhäuser Classics. Reprint of the 1990 edition. Boston, MA: Birkhäuser Boston Inc., pp. xx+461. DOI: 10.1007/978-0-8176-4848-0.

Aubin, T. (2001). *A course in differential geometry*. Vol. 27. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. xii+184.

Bačák, M. (2014). "Computing medians and means in Hadamard spaces". *SIAM Journal on Optimization* 24.3, pp. 1542–1566. DOI: 10.1137/140953393.

Bazaraa, M. S.; Shetty, C. M. (1976). *Foundations of optimization*. Lecture Notes in Economics and Mathematical Systems, Vol. 122. Springer-Verlag, Berlin-New York, pp. vi+193.

Bazaraa, M. S.; Sherali, H. D.; Shetty, C. M. (2006). *Nonlinear programming*. 3rd ed. Theory and algorithms. Wiley-Interscience, Hoboken, NJ, pp. xvi+853. DOI: 10.1002/0471787779.

Bergmann, R. (2017). "MVIRT, A toolbox for manifold-valued image registration". *IEEE International Conference on Image Processing, IEEE ICIP 2017, Beijing, China, September 17–20, 2017*. DOI: 10.1109/ICIP.2017.8296271.

Bergmann, R.; Chan, R. H.; Hielscher, R.; Persch, J.; Steidl, G. (2016). "Restoration of Manifold-Valued Images by Half-Quadratic Minimization". *Inverse Problems in Imaging* 10 (2), pp. 281–304. DOI: 10.3934/ipi.2016001.

Bergmann, R.; Persch, J.; Steidl, G. (2016). "A Parallel Douglas–Rachford Algorithm for Minimizing ROF-like Functionals on Images with Values in Symmetric Hadamard Manifolds". *SIAM Journal on Imaging Sciences* 9.4, pp. 901–937. DOI: 10.1137/15M1052858.

Boumal, N. (2015). "Riemannian trust regions with finite-difference Hessian approximations are globally convergent". *Geometric science of information*. Vol. 9389. Lecture Notes in Computer Science. Springer, Cham, pp. 467–475. DOI: 10.1007/978-3-319-25040-3_50.

Burke, J. V. (2014). *Nonlinear Optimization*. Lecture notes, Math 408, University of Washington, Seattle, WA.

Forst, W.; Hoffmann, D. (2010). *Optimization—theory and practice*. Springer Undergraduate Texts in Mathematics and Technology. Springer, New York, pp. xviii+402. DOI: 10.1007/978-0-387-78977-4.

Gauvin, J. (1977). "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming". *Mathematical Programming* 12.1, pp. 136–138. DOI: 10.1007/BF01593777.

Geiger, C.; Kanzow, C. (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. New York: Springer. DOI: 10.1007/978-3-642-56004-0.

Gould, F. J.; Tolle, J. W. (1971). "A necessary and sufficient qualification for constrained optimization". *SIAM Journal on Applied Mathematics* 20, pp. 164–172. DOI: 10.1137/0120021.

Guignard, M. (1969). "Generalized Kuhn-Tucker Conditions for Mathematical Programming in a Banach Space". *SIAM Journal on Control and Optimization* 7.2, pp. 232–241. DOI: 10.1137/0307016.

Jiménez, B.; Novo, V. (2006). "Characterization of the cone of attainable directions". *Journal of Optimization Theory and Applications* 131.3, pp. 493–499. DOI: 10.1007/s10957-006-9158-9.

Jost, J. (2017). *Riemannian geometry and geometric analysis*. 7th ed. Universitext. Springer, Cham, pp. xiv+697. DOI: 10.1007/978-3-319-61860-9.

Karcher, H. (1977). "Riemannian center of mass and mollifier smoothing". *Communications on Pure and Applied Mathematics* 30.5, pp. 509–541. DOI: 10.1002/cpa.3160300502.

Karush, W. (1939). "Minima of Functions of Several Variables with Inequalities as Side Constraints". M.Sc. Thesis. Department of Mathematics, University of Chicago.

Kjeldsen, T. H. (2000). "A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World War II". *Historia Mathematica* 27.4, pp. 331–361. DOI: 10.1006/hmat.2000.2289.

Kuhn, H. W.; Tucker, A. W. (1951). "Nonlinear programming". *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*. Berkeley and Los Angeles: University of California Press, pp. 481–492.

Ledyaev, Y. S.; Zhu, Q. J. (2007). "Nonsmooth analysis on smooth manifolds". *Transactions of the American Mathematical Society* 359.8, pp. 3687–3732. DOI: 10.1090/S0002-9947-07-04075-5.

Lee, J. M. (2003). *Introduction to smooth manifolds*. Vol. 218. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xviii+628. DOI: 10.1007/978-0-387-21752-9.

Nocedal, J.; Wright, S. (2006). *Numerical Optimization*. 2nd ed. New York: Springer. DOI: 10.1007/978-0-387-40065-5.

Penot, J.-P. (1985). "Variations on the theme of nonsmooth analysis: another subdifferential". *Nondifferentiable optimization: motivations and applications (Sopron, 1984)*. Vol. 255. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, pp. 41–54. DOI: 10.1007/978-3-662-12603-5_5.

Peterson, D. W. (1973). "A review of constraint qualifications in finite-dimensional spaces". *SIAM Review* 15, pp. 639–654. DOI: 10.1137/1015075.

Solodov, M. V. (2010). "Constraint Qualifications". *Wiley Encyclopedia of Operations Research and Management Science*. Ed. by Cochran, J. J.; Cox, L. A.; Keskinocak, P.; Kharoufeh, J. P.; Smith, J. C. John Wiley & Sons, Inc. DOI: 10.1002/9780470400531.eorms0978.

Spivak, M. (1979). *A comprehensive introduction to differential geometry. Vol. I.* 2nd ed. Publish or Perish, Inc., Wilmington, Del., pp. xiv+668.

Tu, L. W. (2011). *An introduction to manifolds*. 2nd ed. Universitext. Springer, New York, pp. xviii+411. DOI: 10.1007/978-1-4419-7400-6.

Udrişte, C. (1988). "Kuhn-Tucker theorem on Riemannian manifolds". *Topics in differential geometry, Vol. II (Debrecen, 1984)*. Vol. 46. Colloquia Mathematica Societatis János Bolyai. North-Holland, Amsterdam, pp. 1247–1259.

Wachsmuth, G. (2013). "On LICQ and the uniqueness of Lagrange multipliers". *Operations Research Letters* 41.1, pp. 78–80. DOI: 10.1016/j.orl.2012.11.009.

Yang, W. H.; Zhang, L.-H.; Song, R. (2014). "Optimality conditions for the nonlinear programming problems on Riemannian manifolds". *Pacific Journal of Optimization* 10.2, pp. 415–434.