SIAM J. OPTIM.
Vol. 29, No. 4, pp. 2423–2444

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

DOI. 10.1137/18M1181602

1. Introduction.

We consider constrained, nonlinear optimization problems

$$\begin{cases} \text{minimize} & f(\boldsymbol{p}) \text{ w.r.t. } \boldsymbol{p} \in \mathcal{M} \\ \text{s.t.} & g(\boldsymbol{p}) \le 0 \\ \text{and} & h(\boldsymbol{p}) = 0, \end{cases} \tag{1.1}$$

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 CQs, abbreviated as LICQ, MFCQ, ACQ, and GCQ, respectively; see for instance Solodov (2010), Peterson (1973), or Bazaraa, Sherali, and 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 conditions 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 nonsingular matrices, the group of special orthogonal (rotation) matrices $\mathrm{SO}(n)$, 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. We refer the reader

*Received by the editors April 18, 2018; accepted for publication (in revised form) June 11, 2019; published electronically October 3, 2019.
https://doi.org/10.1137/18M1181602
Funding: The work of the authors was partially supported by DFG grants BE 5888/2-1 and HE 6077/10-1, the latter within the Priority Program SPP 1962.
†Faculty of Mathematics, Technische Universität Chemnitz, 09107 Chemnitz, Germany (roland.herzog@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/bergmann, ronny.bergmann@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/herzog).

to Absil, Mahony, and 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 et al. (2016), and the parallel Douglas–Rachford algorithm by Bergmann, Persch, and Steidl (2016) have been introduced. These 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 pixels. 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 $\mathcal{M} = \mathbb{R}^n$ case, 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. To be precise, the objective is convex along geodesics, and the feasible set is described by a finite collection of inequality constraints which are likewise geodesically convex. The work closest to ours is Yang, Zhang, and 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. The emphasis of the present paper is on constraint qualifications and first-order necessary conditions of KKT type, but in contrast to Yang, Zhang, and Song (2014) we do not discuss second-order optimality conditions. We also mention Ledyaev and Zhu (2007), where a framework for generalized derivatives of nonsmooth functions on smooth Riemannian manifolds is developed and Fritz John-type optimality conditions are derived as an application. Recently, a discussion of algorithms for equality and inequality constrained problems on Riemannian manifolds was performed in Liu and Boumal (2019)

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 CQs 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, and 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, the formulation of 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 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, and Sepulchre (2008), Absil, Baker, and Gallivan (2007), and Boumal (2015).

We also mention that since CQs and KKT conditions are local concepts, the results of this paper can be stated and derived in a different way: one can transcribe (1.1) locally into an optimization problem in Euclidean space and subsequently apply the theory of CQs and KKT in $\mathbb{R}^n$. This leads to equivalent definitions and results. However we find it more instructive to formulate CQs and KKT conditions using the language of differential geometry and to minimize the explicit use of charts.

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 minimizers under the Guignard constraint qualifications. We also formulate further 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 $\mathbb{R}_n$). Moreover, we recall that a subset K of a vector space V is said to be a *cone* if $\alpha K \subseteq K$ for all $\alpha > 0$. A cone K may or may not be convex.

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), and Jost (2017) for a thorough introduction.

DEFINITION 2.1. *Suppose that $\mathcal{M}$ is a Hausdorff, second-countable topological space. One says that $\mathcal{M}$ can be endowed with a smooth structure 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) $\varphi_\alpha \colon U_\alpha \to \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^n$, 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 $\mathcal{A} \coloneqq \{(U_\alpha, \varphi_\alpha)\}_{\alpha\in A}$ is a smooth atlas. Then the pair $(\mathcal{M}, \mathcal{A})$ is called a smooth manifold.*

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, and Sepulchre (2008). From now on, a smooth manifold $\mathcal{M}$ will always be equipped with a given smooth atlas $\mathcal{A}$. 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 $\varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^n$ 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 $U \subseteq \mathcal{M}$ 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.

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*

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

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 linear map, denoted by $[\dot\gamma(0)]$ or $[\frac{\mathrm{d}}{\mathrm{d}t}\gamma(0)]$, defined as*

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

takes C^1-functions $f\colon U \to \mathbb{R}$ defined in some open neighborhood $U \subseteq \mathcal{M}$ 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 $\alpha \odot [\gamma] = [\alpha \odot \gamma]$ and $[\gamma] \oplus [\zeta] = [\gamma \oplus_\varphi \zeta]$, defined in terms of

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

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

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

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 linear map, denoted by $(\mathrm{d}f)(\boldsymbol{p})$, 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. We also remark that Definition 2.4 easily generalizes to vector-valued functions $g\colon \mathcal{M} \to \mathbb{R}^m$ by applying (2.5) component by component.

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 $N \geq n$. Moreover, we do not need to make further topological assumptions such as compactness, connectedness, or orientability of $\mathcal{M}$.

As we mentioned in the introduction, the subsequent results could be derived by transcribing (1.1) locally into a problem in Euclidean space, using a chart. This is due to the fact that this transformation leaves the notion of local minimum intact, as shown by the following lemma.

LEMMA 2.6 (compare Yang, Zhang, and Song (2014, sect. 4.1)). *Suppose that (U, φ) is an 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*

$$\left\{\begin{array}{rl} \textit{minimize} & (f \circ \varphi^{-1})(x) \ \textit{w.r.t. } x \in \varphi(U) \subseteq \mathbb{R}^n \\ \textit{s.t.} & (g \circ \varphi^{-1})(x) \leq 0 \\ \textit{and} & (h \circ \varphi^{-1})(x) = 0. \end{array}\right. \tag{2.6}$$

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 $f(\boldsymbol{p}^*) \leq f(\boldsymbol{p})$ holds for all $\boldsymbol{p} \in U_1 \cap \Omega$. We can assume, by shrinking U_1 if necessary, that $U_1 \subseteq U$ holds. This implies $f(\varphi(\boldsymbol{p}^*)) \leq f(\varphi(\boldsymbol{p}))$ 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 (2.6). The converse is proved similarly. ∎

However, we are going to prefer working directly with (1.1) using the language of differential geometry and minimize the explicit use of charts.

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 and Wright (2006, Chap. 12) or Forst and Hoffmann (2010, Chap. 2).

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

We define $\Omega \coloneqq \{x \in \mathbb{R}^n : g(x) \leq 0,\ h(x) = 0\}$ 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,

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

This cone is also known as the contingent cone or the Bouligand cone; cf. Jiménez and Novo (2006) and Penot (1985). It is closed but not necessarily convex. Since $\mathcal{T}_\Omega(x^*)$ is inconvenient to work with, one introduces the linearizing cone

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

Here $\mathcal{A}(x^*) \coloneqq \{1 \le i \le m : g_i(x^*) = 0\}$ 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 $\mathcal{T}_\Omega^{\text{lin}}(x^*)$ is a closed convex cone and that $\mathcal{T}_\Omega(x^*) \subseteq \mathcal{T}_\Omega^{\text{lin}}(x^*)$ holds; see for instance Nocedal and Wright (2006, Lem. 12.2).

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

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

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

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

as can be shown by means of the Farkas lemma; cf. Nocedal and Wright (2006, Lem. 12.4).

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

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

This is where constraint qualifications come into play. The weakest, 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 following well-known theorem.

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

$\mathcal{M} = \mathbb{R}^n$	$\mathcal{M}$ 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 $\mathcal{T}_{\mathcal{M}}^{\text{lin}}(\Omega; \boldsymbol{p})$ (3.15)
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 $\subseteq \mathbb{R}_n$ (3.6)	Polar cone $\mathcal{T}_{\mathcal{M}}^{\text{lin}}(\Omega; \boldsymbol{p})^\circ \subseteq \mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p})$ (3.18)
Lagrange multipliers $\mu \in \mathbb{R}_m$, $\lambda \in \mathbb{R}_q$	Same as for $\mathcal{M} = \mathbb{R}^n$

THEOREM 3.1. *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 3.1, 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.

Let us denote by

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

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

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

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) and Kuhn and 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 and Novo (2006), Bazaraa and Shetty (1976, Ch. 3.5), and Aubin and 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.

DEFINITION 3.2 ((Bouligand) tangent cone). *Suppose that $\boldsymbol{p} \in \Omega$ holds.*

(a) *A tangent vector $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ is called a tangent vector to Ω at $\boldsymbol{p}$ if there exist sequences $(\boldsymbol{p}_k) \subseteq \Omega$ and $t_k \searrow 0$ such that for all C^1-functions f defined near $\boldsymbol{p}$, we have*

$$[\dot\gamma(0)](f) = \lim_{k\to\infty} \frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p})}{t_k}. \tag{3.11}$$

We refer to the sequence $(\boldsymbol{p}_k, t_k)$ as a tangential sequence to Ω at $\boldsymbol{p}$. Notice that necessarily $\boldsymbol{p}_k \to \boldsymbol{p}$ holds.

(b) *The collection of all tangent vectors to Ω at $\boldsymbol{p}$ is termed the (Bouligand) tangent cone to Ω at $\boldsymbol{p}$ and denoted by*

$$\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p}) := \{[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}) : [\dot\gamma(0)] \text{ is a tangent vector to } \Omega \text{ at } \boldsymbol{p}\}. \tag{3.12}$$

The following proposition shows that (3.12) could also have been defined as a lifting via the chart differential of the classical tangent cone to the chart image of the feasible set near $\boldsymbol{p}$. This was in fact used as the definition of the tangent cone in Yang, Zhang, and Song (2014, eq. (3.7)).

PROPOSITION 3.3. *Suppose that $\boldsymbol{p} \in \Omega$, and let (U, φ) be a chart about $\boldsymbol{p}$. Then*

$$\big((\mathrm{d}\varphi)(\boldsymbol{p})\big)\, \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p}) = \mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p})). \tag{3.13}$$

Proof. We divide the proof into two parts and first prove "$\supseteq$" in (3.13). To this end, suppose that $d \in \mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p}))$, i.e., there exist sequences $(x_k) \subseteq \varphi(U\cap\Omega)$, $x_k \to \varphi(\boldsymbol{p}) =: x$ and $t_k \searrow 0$ such that $d = \lim_{k\to\infty}(x_k - x)/t_k$; see (3.3). Define $\boldsymbol{p}_k := \varphi^{-1}(x_k) \in U \cap \Omega$ and $\boldsymbol{p} := \varphi^{-1}(x) \in U\cap\Omega$. Then $\boldsymbol{p}_k \to \boldsymbol{p}$ since φ^{-1} is continuous. Further, define a curve γ via $\gamma(t) := \varphi^{-1}(\varphi(\boldsymbol{p}) + t\,d)$ for $|t|$ sufficiently small. We show that $[\dot\gamma(0)]$ belongs to $\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ by verifying (3.11). To this end, let f be an arbitrary C^1-function defined near $\boldsymbol{p}$. Then we have

$$[\dot\gamma(0)](f) = \frac{\mathrm{d}}{\mathrm{d}t}(f\circ\gamma)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\big((f\circ\varphi^{-1})(\varphi(\boldsymbol{p}) + t\,d)\big)\Big|_{t=0} = (f\circ\varphi^{-1})'(\varphi(\boldsymbol{p}))\,d$$

by the definition of γ and the chain rule. On the other hand,

$$\lim_{k\to\infty}\frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p})}{t_k} = \lim_{k\to\infty}\frac{(f\circ\varphi^{-1})(x_k) - (f\circ\varphi^{-1})(x)}{t_k} = (f\circ\varphi^{-1})'(\varphi(\boldsymbol{p}))\,d$$

holds, which proves (3.11) and thus $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$. By Definition 2.4, Remark 2.5, (2.2), and the definition of γ, we have

$$(\mathrm{d}\varphi)(\boldsymbol{p})\,[\dot\gamma(0)] = [\dot\gamma(0)](\varphi) = \frac{\mathrm{d}}{\mathrm{d}t}(\varphi\circ\gamma)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\big((\varphi\circ\varphi^{-1})(\varphi(\boldsymbol{p}) + t\,d)\big)\Big|_{t=0} = d.$$

This confirms $d \in \big((\mathrm{d}\varphi)(\boldsymbol{p})\big)\,\mathcal{T}_{\mathcal{M}}(\Omega;\boldsymbol{p})$ and thus the first part of the proof.

For the reverse inclusion "$\subseteq$", we begin with an element $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ and an associated tangential sequence $(\boldsymbol{p}_k, t_k)$ as in (3.11). Again by Definition 2.4 and Remark 2.5, we obtain

$$\big((\mathrm{d}\varphi)(\boldsymbol{p})\big)\, \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}) = [\dot\gamma(0)](\varphi) = \lim_{k\to\infty} \frac{\varphi(\boldsymbol{p}_k) - \varphi(\boldsymbol{p})}{t_k},$$

and the limit exists by (3.11). The sequence $\varphi(\boldsymbol{p}_k, t_k)$ satisfies all the requirements to generate an element of $\mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p}))$; cf. (3.3). ∎

Remark 3.4 (tangent cone). The notion of tangent vectors to subsets of smooth manifolds can be traced back to Motreanu and Pavel (1982, Def. 2.1), where they were called *quasitangent* vectors and introduced, in our notation, as vectors $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ satisfying

$$\lim_{h\to 0} \frac{1}{h} \operatorname{dist}\big(\varphi(\boldsymbol{p}) + h\,(D\varphi)(\boldsymbol{p})\,[\dot\gamma(0)],\ \varphi(U\cap\Omega)\big) = 0.$$

Here (U, φ) is a chart about $\boldsymbol{p}$, $(D\varphi)(\boldsymbol{p})$ is the derivative (push-forward) of φ at $\boldsymbol{p}$, and dist denotes the (Euclidean) distance between a point and a set in $\mathbb{R}^n$. It is straightforward to show that this definition is equivalent to (3.12).

LEMMA 3.5 (properties of the tangent cone). *For any $\boldsymbol{p} \in \Omega$, the tangent cone $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is a closed cone in the tangent space $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$.*

Proof. The result follows immediately from Proposition 3.3 since $\mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p}))$ is a closed cone in $\mathbb{R}^n$ and $(\mathrm{d}\varphi)(\boldsymbol{p})$ is a bijective, linear map between the vector spaces $\mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$ and $\mathbb{R}^n$. ∎

The analogue of (3.2) is the following theorem.

THEOREM 3.6 (first-order necessary optimality condition). *Suppose that $\boldsymbol{p}^* \in \Omega$ is a local minimizer of* (1.1). *Then we have*

$$[\dot\gamma(0)](f) \ge 0 \tag{3.14}$$

for all tangent vectors $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)$.

Proof. Suppose that $[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)$ and that $(\boldsymbol{p}_k, t_k)$ is an associated tangential sequence. Then we have, by local optimality of $\boldsymbol{p}^*$,

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

This concludes the proof. ∎

Next we introduce the concept of the linearizing cone (3.4) in the tangent space, similar to Yang, Zhang, and Song (2014, Def. 4.1).

DEFINITION 3.7 (linearizing cone). *For any $\boldsymbol{p} \in \Omega$, we define the linearizing cone to the feasible set Ω by*

$$\begin{aligned} \mathcal{T}^{\text{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p}) \coloneqq \big\{[\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}) : {} & [\dot\gamma(0)](g^i) \le 0 \textit{ for all } i \in \mathcal{A}(\boldsymbol{p}), \\ & [\dot\gamma(0)](h^j) = 0 \textit{ for all } j = 1, \ldots, q\big\}. \end{aligned} \tag{3.15}$$

As in subsection 3.1, $\mathcal{A}(\boldsymbol{p}) \coloneqq \{1 \le i \le m : g^i(\boldsymbol{p}) = 0\}$ 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.

Remark 3.8. The linearizing cone could be defined equivalently as

$$\big((\mathrm{d}\varphi)(\boldsymbol{p})\big)\, \mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p}) = \mathcal{T}_{\varphi(U\cap\Omega)}^{\mathrm{lin}}(\varphi(\boldsymbol{p})); \tag{3.16}$$

cf. Proposition 3.3 for the parallel result for the tangent cone.

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

Proof. The result follows immediately from (3.16) and the corresponding result in $\mathbb{R}^n$. ∎

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

$$B^\circ \coloneqq \big\{(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p}) : (\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] \le 0 \text{ for all } [\dot\gamma(0)] \in B\big\}. \tag{3.17}$$

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

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

$$\begin{aligned}\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p})^\circ = \Big\{(\mathrm{d}s)(\boldsymbol{p}) = \sum_{i=1}^m \mu_i\,(\mathrm{d}g^i)(\boldsymbol{p}) + \sum_{j=1}^q \lambda_j\,(\mathrm{d}h^j)(\boldsymbol{p}),&\\ \mu_i \ge 0 \textit{ for } i \in \mathcal{A}(\boldsymbol{p}),\ \mu_i = 0 \textit{ for } i \in \mathcal{I}(\boldsymbol{p}),\ \lambda_j \in \mathbb{R}\Big\} \subseteq \mathcal{T}_{\mathcal{M}}^*(\boldsymbol{p}).&\end{aligned} \tag{3.18}$$

Proof. It is easy to see that for vector spaces V and W of finite dimension and bijective, linear $A : V \to W$, we have $(A^{-1}K)^\circ = A^*K^\circ$ in V^* for all $K \subseteq W$. Here V^* and W^* are the dual spaces of V and W and $A^* : W^* \to V^*$ is the adjoint map. We apply this with $K = \mathcal{T}_{\varphi(U\cap\Omega)}^{\mathrm{lin}}(\varphi(\boldsymbol{p})) \subseteq W = \mathbb{R}^n$, $V = \mathcal{T}_{\mathcal{M}}(\boldsymbol{p})$, and $A = (\mathrm{d}\varphi)(\boldsymbol{p})$ to obtain

$$\begin{aligned}
&\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega;\boldsymbol{p})^\circ\\
&\quad= \Big(\big((\mathrm{d}\varphi)(\boldsymbol{p})\big)^{-1}\mathcal{T}_{\varphi(U\cap\Omega)}^{\mathrm{lin}}(\varphi(\boldsymbol{p}))\Big)^\circ && \text{by (3.16)}\\
&\quad= \big((\mathrm{d}\varphi)(\boldsymbol{p})\big)^*\mathcal{T}_{\varphi(U\cap\Omega)}^{\mathrm{lin}}(\varphi(\boldsymbol{p}))^\circ\\
&\quad= \big((\mathrm{d}\varphi)(\boldsymbol{p})\big)^*\Big\{\sum_{i=1}^m \mu_i\,(g^i\circ\varphi^{-1})'(\varphi(\boldsymbol{p})) + \sum_{j=1}^q \lambda_j\,(h^j\circ\varphi^{-1})'(\varphi(\boldsymbol{p})),\\
&\qquad\qquad \mu_i \ge 0 \text{ for } i \in \mathcal{A}(\varphi(\boldsymbol{p})),\ \mu_i = 0 \text{ for } i \in \mathcal{I}(\varphi(\boldsymbol{p})),\ \lambda_j \in \mathbb{R}\Big\} && \text{by (3.6)}\\
&\quad= \Big\{\sum_{i=1}^m \mu_i\,(\mathrm{d}g^i)(\boldsymbol{p}) + \sum_{j=1}^q \lambda_j\,(\mathrm{d}h^j)(\boldsymbol{p}),\\
&\qquad\qquad \mu_i \ge 0 \text{ for } i \in \mathcal{A}(\boldsymbol{p}),\ \mu_i = 0 \text{ for } i \in \mathcal{I}(\boldsymbol{p}),\ \lambda_j \in \mathbb{R}\Big\}.
\end{aligned}$$

The last equality follows from the chain rule applied to $(g^i \circ \varphi^{-1}) \circ \varphi$. ∎

We associate with (1.1) the Lagrangian

$$\mathcal{L}(\boldsymbol{p},\mu,\lambda) \coloneqq f(\boldsymbol{p}) + \mu\, g(\boldsymbol{p}) + \lambda\, h(\boldsymbol{p}), \tag{3.19}$$

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

$$(\mathrm{d}\mathcal{L})(\boldsymbol{p}, \mu, \lambda) = (\mathrm{d}f)(\boldsymbol{p}) + \mu\,(\mathrm{d}g)(\boldsymbol{p}) + \lambda\,(\mathrm{d}h)(\boldsymbol{p}) = 0, \tag{3.20a}$$

$$h(\boldsymbol{p}) = 0, \tag{3.20b}$$

$$\mu \ge 0, \quad g(\boldsymbol{p}) \le 0, \quad \mu\, g(\boldsymbol{p}) = 0. \tag{3.20c}$$

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

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

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 (3.20) are equivalent to

$$-\,(\mathrm{d}f)(\boldsymbol{p}) \in \mathcal{T}^{\mathrm{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p})^{\circ}. \tag{3.21}$$

We thus obtain the analogue of Theorem 3.1.

THEOREM 3.11. *Suppose that $\boldsymbol{p}^*$ is a local minimizer of* (1.1) *and that the GCQ $\mathcal{T}^{\mathrm{lin}}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)^{\circ} = \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}^*)^{\circ}$ holds at $\boldsymbol{p}^*$. Then there exist Lagrange multipliers $\mu \in \mathbb{R}_m$, $\lambda \in \mathbb{R}_q$ such that the KKT conditions* (3.20) *hold.*

3.3. Constraint qualifications for optimization problems on smooth manifolds. In this section we introduce the constraint qualifications (CQs) of linear independence (LICQ), Mangasarian–Fromovitz (MFCQ), Abadie (ACQ), and Guignard (GCQ) and show that the chain of implications

$$\text{LICQ} \;\Rightarrow\; \text{MFCQ} \;\Rightarrow\; \text{ACQ} \;\Rightarrow\; \text{GCQ} \tag{3.22}$$

continues to hold in the smooth manifold setting. Except for the LICQ, which has been used in Yang, Zhang, and Song (2014), this is the first time these conditions have been formulated and utilized on smooth manifolds.

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*

$$\begin{aligned} (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] &< 0 \quad \textit{for all } i \in \mathcal{A}(\boldsymbol{p}), \\ (\mathrm{d}h^j)(\boldsymbol{p})[\dot\gamma(0)] &= 0 \quad \textit{for all } j = 1, \ldots, q. \end{aligned} \tag{3.23}$$

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

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

Remark 3.13. The constraint qualifications in Definition 3.12 are equivalent to their respective counterparts for the local transcription of (1.1) into an optimization problem in Euclidean space; see (2.6). For instance, when φ is a chart about $\boldsymbol{p} \in \Omega$, then the LICQ is equivalent to the linear independence of the derivatives

$\{(h^j \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\}_{j=1}^q \cup \{(g^i \circ \varphi^{-1})'(\varphi(\boldsymbol{p}))\}_{i\in\mathcal{A}(\varphi(\boldsymbol{p}))}$. A similar statement holds for the MFCQ, ACQ, and GCQ. The result (3.22) can therefore be shown by invoking the corresponding statement for (2.6). However, we also provide direct proofs in Propositions 3.14 and 3.16.

Proposition 3.14. *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; cf. Geiger and Kanzow (2002, Lem. 2.37).

Proposition 3.15. *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

$$t \mapsto \gamma_k(t) \coloneqq \varphi^{-1}(\varphi(\boldsymbol{p}) + t\,e_k) \quad \text{for } k = 1,\ldots,n.$$

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 $\{[\dot\gamma_k(0)]\}_{k=1}^n$ are linearly independent and the cotangent vectors $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q$ are 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 $H_t(0,0) = (h\circ\varphi^{-1})'(x_0)\,(\varphi\circ\zeta)'(0) = (h\circ\zeta)'(0)$. Notice that the jth 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 $[\dot\zeta(0)]$ for any $j = 1,\ldots,q$. Thus we conclude $H_t(0,0) = 0$.

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, and 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 $i \in \mathcal{I}(\boldsymbol{p})$, 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 application 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 3.16. *MFCQ implies ACQ.*

Proof. In view of Lemma 3.9, we only need to show $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p}) \supseteq \mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega; \boldsymbol{p})$. To this end, suppose that $[\dot{\gamma}_0(0)]$ is an element of $\mathcal{T}_{\mathcal{M}}^{\mathrm{lin}}(\Omega; \boldsymbol{p})$ defined in (3.15), 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 (3.23). Finally, choose an arbitrary chart φ about $\boldsymbol{p}$.

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

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

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 and that $(\varphi \circ \gamma_0)(t) + (\varphi \circ \gamma)(\tau\, t) - \varphi(\boldsymbol{p})$ belongs to the image of φ so that $\gamma_0 \oplus_\varphi (\tau \odot \gamma)$ is well defined for $t \in (-\varepsilon, \varepsilon)$.

We first show that $[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)] \to [\dot{\gamma}_0(0)]$ as $\tau \searrow 0$. Indeed, for any C^1-function f defined near $\boldsymbol{p}$, we have

$$\begin{aligned}&(\mathrm{d}f)(\boldsymbol{p})\left[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)\right]\\ &\quad = \left[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)\right](f) \qquad \text{by definition of } (\mathrm{d}f)(\boldsymbol{p})\text{; see (2.5)}\end{aligned}$$

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

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

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

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

which is negative for any $i \in \mathcal{A}(\boldsymbol{p})$ since $\tau > 0$. Analogously,

$$(\mathrm{d}h^j)(\boldsymbol{p})[\tfrac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)] = 0$$

follows for all $j = 1, \ldots, q$. This confirms that $\gamma_0 \oplus_\varphi (\tau \odot \gamma)$ is indeed an MFCQ vector.

Fix $\tau \in (0, 1]$. While $\gamma_0 \oplus_\varphi (\tau \odot \gamma)$ itself may not be feasible near $t = 0$, Proposition 3.15 shows that we can replace it by an equivalent C^1-curve which is feasible for $t \in [0, \varepsilon_\tau)$. In other words, the equivalence class $[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)]$ belongs to the tangent cone $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$. We showed above that $[\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_0 \oplus_\varphi (\tau \odot \gamma))(0)] \to [\dot\gamma_0(0)]$ holds as $\tau \searrow 0$. Since $\mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ is closed, we conclude that $[\dot\gamma_0(0)] \in \mathcal{T}_{\mathcal{M}}(\Omega; \boldsymbol{p})$ holds. ◻

Finally, the fact that ACQ implies GCQ is trivial, so (3.22) 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 of objective functions of class C^1 attaining a local minimum at $\boldsymbol{p}$,

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

For $f \in \mathcal{F}(\boldsymbol{p})$, we denote by

$$\Lambda(f; \boldsymbol{p}) \coloneqq \{(\mu, \lambda) \in \mathbb{R}_m \times \mathbb{R}_p : \text{(3.20) holds}\} \tag{4.2}$$

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 $\mathcal{M} = \mathbb{R}^n$ case; see Gauvin (1977), Gould and 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 nonempty 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 nonempty, 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 $\boldsymbol{p}$.*

Proof. (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 and Tolle (1971, sect. 4) for the $\mathcal{M} = \mathbb{R}^n$ case; see also Bazaraa and Shetty (1976, Thm. 6.3.2). In order to utilize this result directly and to avoid stating an analogous one on $\mathcal{M}$, we temporarily depart from our standing principle of minimizing the use of charts. Suppose that $(\mathrm{d}s)(\boldsymbol{p}) \in \mathcal{T}_\mathcal{M}(\Omega;\boldsymbol{p})^\circ \subseteq \mathcal{T}^*_\mathcal{M}(\boldsymbol{p})^\circ$ holds. Fix an arbitrary chart (U,φ) about $\boldsymbol{p}$. Suppose that d is an arbitrary element from the tangent cone $\mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p}))$. Then we can construct, as in the proof of Proposition 3.3, the curve $\gamma(t) \coloneqq \varphi^{-1}(\varphi(\boldsymbol{p}) + t\,d)$ so that $[\dot\gamma(0)] \in \mathcal{T}_\mathcal{M}(\Omega;\boldsymbol{p})$ and $d = \big((\mathrm{d}\varphi)(\boldsymbol{p})\big)[\dot\gamma(0)]$ holds. We obtain

$$0 \ge (\mathrm{d}s)(\boldsymbol{p})\,[\dot\gamma(0)] = \frac{\mathrm{d}}{\mathrm{d}t}(s\circ\varphi^{-1}\circ\varphi\circ\gamma)\Big|_{t=0} = (s\circ\varphi^{-1})'(\varphi(\boldsymbol{p}))\,d$$

from $\mathcal{T}_\mathcal{M}(\Omega;\boldsymbol{p})^\circ$, Definition 2.4, the chain rule, and the definition of γ. This shows $(s\circ\varphi^{-1})'(\varphi(\boldsymbol{p})) \in \mathcal{T}_{\varphi(U\cap\Omega)}(\varphi(\boldsymbol{p}))^\circ$.

Using Bazaraa and 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 (2.6) but with the objective r in place of $(f\circ\varphi^{-1})$. By Lemma 2.6, $\boldsymbol{p}$ is a local minimizer of (1.1) with objective $r\circ\varphi$. By assumption, $\Lambda(r\circ\varphi,\boldsymbol{p})$ is nonempty, 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 (3.20b), (3.20c) hold. In other words, $-(\mathrm{d}(r\circ\varphi))(\boldsymbol{p}) \in \mathcal{T}^{\mathrm{lin}}_\mathcal{M}(\Omega;\boldsymbol{p})^\circ$; see (3.21). Moreover, the differentials of $r\circ\varphi$ and $-s$ at $\boldsymbol{p}$ coincide since

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

holds for arbitrary tangent vectors $[\dot\gamma(0)]$ in $\mathcal{T}_\mathcal{M}(\boldsymbol{p})$. This shows that $\mathcal{T}_\mathcal{M}(\Omega;\boldsymbol{p})^\circ \subseteq \mathcal{T}^{\mathrm{lin}}_\mathcal{M}(\Omega;\boldsymbol{p})^\circ$ 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 the existence of an MFCQ vector in (MFCQ) is equivalent to the feasibility of the linear program

$$\left\{\begin{array}{rll} \text{minimize} & 0 \text{ w.r.t. } [\dot\gamma(0)] \in \mathcal{T}_{\mathcal{M}}(\boldsymbol{p}) & \\ \text{s.t.} & (\mathrm{d}g^i)(\boldsymbol{p})[\dot\gamma(0)] \le -1 & \text{for all } i \in \mathcal{A}(\boldsymbol{p}) \\ \text{and} & (\mathrm{d}h^j)(\boldsymbol{p})[\dot\gamma(0)] = 0 & \text{for all } j = 1, \ldots, q. \end{array}\right. \tag{4.3}$$

Using strong duality, one shows that (MFCQ) holds if and only if $\{(\mathrm{d}h^j)(\boldsymbol{p})\}_{j=1}^q$ is linearly independent and

$$\begin{aligned} &\mu\,(\mathrm{d}g)(\boldsymbol{p}) + \lambda\,(\mathrm{d}h)(\boldsymbol{p}) = 0, \\ &\mu_i \ge 0 \quad \text{for all } i \in \mathcal{A}(\boldsymbol{p}), \\ &\mu_i = 0 \quad \text{for all } i \in \mathcal{I}(\boldsymbol{p}), \end{aligned} \tag{4.4}$$

has 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 sequence of Lagrange multipliers $(\mu^{(k)}, \lambda^{(k)})$ whose Euclidean norm $|(\mu^{(k)}, \lambda^{(k)})|_2$ diverges to ∞. Consequently, there exists a subsequence (which we do not relabel) such that $(\mu^{(k)}, \lambda^{(k)})/|(\mu^{(k)}, \lambda^{(k)})|_2$ converges to some $(\mu, \lambda) \neq 0$. Exploiting the KKT conditions (3.20) for $(\mu^{(k)}, \lambda^{(k)})$ it follows that (4.4) holds. Consequently, (MFCQ) is violated. This shows (b).

Conversely, if (MFCQ) does not hold, then there exists a nonzero vector (μ, λ) satisfying (4.4). When $(\mu_0, \lambda_0) \in \Lambda(f;\boldsymbol{p})$, $(\mu_0, \lambda_0) + t\,(\mu, \lambda)$ belongs to $\Lambda(f;\boldsymbol{p})$ as well for any $t \ge 0$, and hence $\Lambda(f;\boldsymbol{p})$ is not compact. This confirms (c).

(d) We proved in section 3 that (LICQ) implies (GCQ), so $\Lambda(f;\boldsymbol{p})$ is nonempty. The uniqueness of the Lagrange multipliers then follows immediately from (3.20a). 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 (3.20) 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. In this section we denote tangent vectors by the symbol ξ instead of $[\dot\gamma(0)]$.

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

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

Completeness of a Riemannian manifold refers to the fact that geodesics emanating from any point $\boldsymbol{p} \in \mathcal{M}$ in the direction of an arbitrary tangent vector ξ exist for all time $t \in \mathbb{R}$.

The Riemannian center of mass, also known as the (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

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

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 $\boldsymbol{c} \in \mathcal{M}$. We obtain the following constrained minimization problem of the form (1.1):

$$\left\{ \begin{aligned} &\text{minimize} \quad f(\boldsymbol{p}) \text{ w.r.t. } \boldsymbol{p} \in \mathcal{M} \\ &\qquad\text{s.t.} \quad d^2_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{c}) - r^2 \le 0, \end{aligned} \right. \tag{5.2}$$

with associated Lagrangian

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

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

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

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

In view of (5.4), the KKT conditions (3.20) become

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

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. We further have to restrict the data to not include antipodal points, i.e. the case in which for some $i, j \in \{1, \dots, N\}$ it holds that $\boldsymbol{d}_i = -\boldsymbol{d}_j$ is excluded. The Riemannian metric is inherited from the ambient space $\mathbb{R}^3$. Since the feasible set

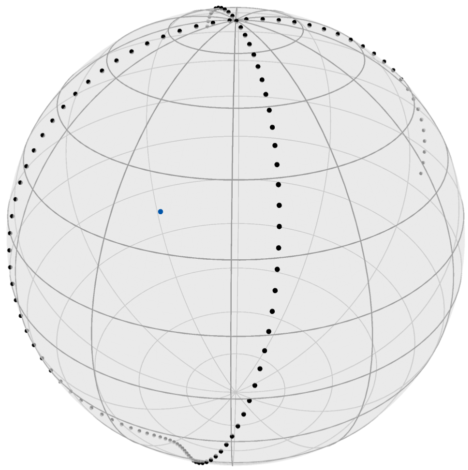
$$\Omega \coloneqq \{\boldsymbol{p} \in \mathbb{S}^2 : d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{c}) \le r\} \tag{5.5}$$

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.

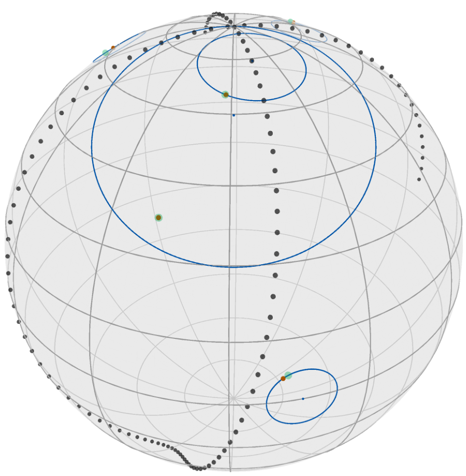
Even in the absence of convexity, the LICQ is satisfied at every solution $\boldsymbol{p}^*$ unless $\boldsymbol{p}^* = \boldsymbol{c}$ holds, which is equivalent to the unconstrained mean $\bar{\boldsymbol{p}}$ coinciding with the center $\boldsymbol{c}$ of the feasible set. This does not happen for 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 1(a).[1] Their unconstrained Riemannian center of mass $\bar{\boldsymbol{p}}$ is shown in blue. We solve five variants of problem (5.2) which differ w.r.t. the centers $\boldsymbol{c}_i$ and radii r_i of the feasible sets Ω_i. The boundaries $\partial\Omega_i$ of the feasible sets, which are spherical caps, are displayed in blue in Figure 1(b) (front view) and Figure 1(c) (back view). The constrained solutions $\boldsymbol{p}_i^*$ are shown in light green in Figures 1(b) and 1(c).

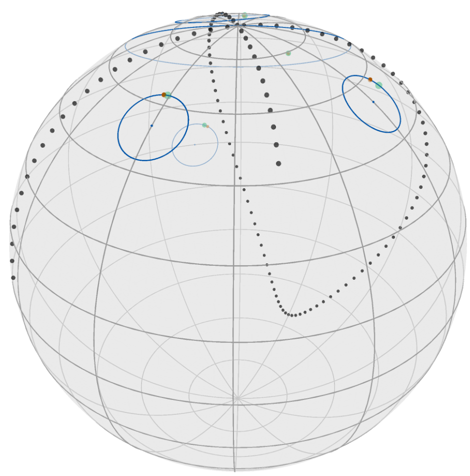
[1]See online version for color figures.



(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* 1(b)*, rotated by* 180 *degrees.*

Fig. 1. *Constrained centers of mass for five different feasible sets (centers and boundaries of the feasible sets 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).*

Each instance of (5.2) is solved using a projected gradient descent method. Since it is a rather straightforward generalization of an unconstrained gradient algorithm (see for instance Absil, Mahony, and 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 $r < \pi/2$, 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

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

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
$\quad \boldsymbol{p}^{(k+1)} \leftarrow \operatorname{proj}_\Omega\bigl(\exp_{\boldsymbol{p}^{(k)}}(-s\nabla f(\boldsymbol{p}^{(k)}))\bigr)$
$\quad k \leftarrow k+1$
until a convergence criterion is reached
return $\boldsymbol{p}^* = \boldsymbol{p}^{(k)}$

It can be computed in closed form, namely

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

whenever the logarithmic map is uniquely determined. This in turn holds whenever $\boldsymbol{p} \neq -\boldsymbol{c}$.

The projected gradient descent algorithm is given as pseudocode 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, and 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 1(c), to solve the constrained instances. The algorithm was implemented within the Manifold-valued Image Restoration Toolbox (MVIRT)[2] (see Bergmann (2017)), providing 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 1(b) and 1(c).

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,

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

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 $\boldsymbol{c}_1 \approx (0.4319, 0.2592, 0.8639)^{\mathrm{T}}$ and radius $r_1 = \frac{\pi}{6}$. For this setup the constraint is inactive and $\boldsymbol{p}_1^* = \bar{\boldsymbol{p}}$ holds. The second example is shown to the right of Figure 1(c), and it is given by $\boldsymbol{c}_2 \approx (0, -0.5735, 0.8192)^{\mathrm{T}}$ and $r_2 = \frac{\pi}{24}$. For this and remaining three cases the constraint is active.

[2]Available open source at http://ronnybergmann.net/mvirt/.

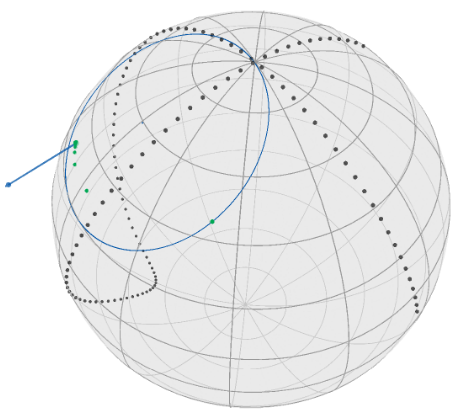
Table 2
Iteration history of Algorithm 5.1 *for two instances of problem* (5.2).

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

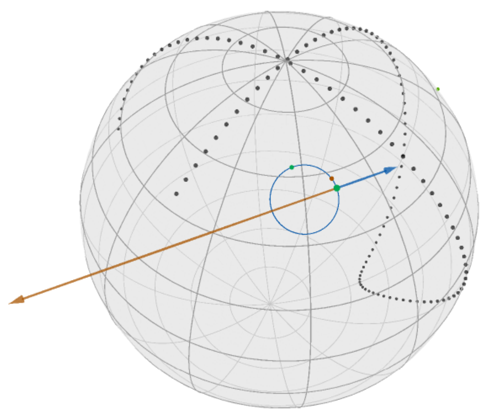
k	$f(\boldsymbol{p}^{(k)})$	$n^{(k)}$	$\mu^{(k)}$
1	1.9129	0.6540	1.1722
2	1.4172	0.1243	0.2755
3	1.3754	0.0169	−0.0847
4	1.3695	0.0029	−0.0811
5	1.3684	0.0005	−0.0403
6	1.3682	0.0001	−0.0180
7	1.3682	1.18×10^{-5}	−0.0078
8	1.3682	3.26×10^{-6}	−0.0034
9	1.3682	6.02×10^{-7}	−0.0014
10	1.3682	1.11×10^{-7}	−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 $(\boldsymbol{c}_2, r_2)$.

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* $(\boldsymbol{c}_1, r_1)$.



(b) *Constraint data* $(\boldsymbol{c}_2, r_2)$.

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

Since the unconstrained Riemannian mean is within the feasible set for the first example of $(\boldsymbol{c}_1, r_1)$, the projection is the identity after the first iteration. Hence for this case, the (projected) gradient descent algorithm computes the unconstrained mean similarly to in Afsari, Tron, and Vidal (2013). We obtain $\boldsymbol{p}_1^* = \bar{\boldsymbol{p}} = \operatorname{proj}_{\Omega_1}(\bar{\boldsymbol{p}})$. Looking at the gradients ∇f and ∇g we see (cf. Figure 2(a)) that $\nabla f = 0$ while the constraint function g yields a gradient pointing towards the boundary $\partial\Omega_1$ of the feasible set. Clearly, the optimal Lagrange multiplier 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 happens to approach zero from below in this case. While the objective decreases, the distance from $\boldsymbol{c}_1$, and thus g, increases, leading to a negative multiplier estimate $\mu^{(k)}$.

For the second case, $(\boldsymbol{c}_2, r_2)$, the unconstrained mean lies outside the feasible set, and the constraint g is strongly active, i.e., the multiplier μ is strictly positive. As we mentioned earlier, the optimal solution $\boldsymbol{p}_2^*$ is different from $\operatorname{proj}_{\Omega_2}(\bar{\boldsymbol{p}})$ (their distance is 0.0409), which is due to the curvature of the manifold.

Acknowledgments. We would like to thank two anonymous reviewers for their constructive comments, which helped to significantly improve the manuscript.

REFERENCES

P.-A. Absil, C. G. Baker, and K. A. Gallivan (2007), *Trust-region methods on Riemannian manifolds*, Found. Comput. Math., 7, pp. 303–330.

P.-A. Absil, R. Mahony, and R. Sepulchre (2008), *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ.

B. Afsari, R. Tron, and R. Vidal (2013), *On the convergence of gradient descent for finding the Riemannian center of mass*, SIAM J. Control Optim., 51, pp. 2230–2260.

J.-P. Aubin and H. Frankowska (2009), *Set-valued Analysis*, Mod. Birkhäuser Class., Birkhäuser, Boston, MA.

T. Aubin (2001), *A Course in Differential Geometry*, Grad. Stud. Math. 27, American Mathematical Society, Providence, RI.

M. Bačák (2014), *Computing medians and means in Hadamard spaces*, SIAM J. Optim., 24, pp. 1542–1566.

M. S. Bazaraa, H. D. Sherali, and C. M. Shetty (2006), *Nonlinear Programming*, 3rd ed., Wiley-Interscience, Hoboken, NJ.

M. S. Bazaraa and C. M. Shetty (1976), *Foundations of Optimization*, Lecture Notes in Econom. and Math. Systems 122, Springer, Berlin.

R. Bergmann (2017), *MVIRT, a toolbox for manifold-valued image restoration*, in Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), IEEE Press, Piscataway, NJ, pp. 201–205.

R. Bergmann, R. H. Chan, R. Hielscher, J. Persch, and G. Steidl (2016), *Restoration of manifold-valued images by half-quadratic minimization*, Inverse Probl. Imaging, pp. 281–304.

R. Bergmann, J. Persch, and G. Steidl (2016), *A parallel Douglas–Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds*, SIAM J. Imaging Sci., 9, pp. 901–937.

N. Boumal (2015), *Riemannian trust regions with finite-difference Hessian approximations are globally convergent*, in Geometric Science of Information, Lecture Notes in Comput. Sci. 9389, Springer, Cham, pp. 467–475.

J. V. Burke (2014), *Nonlinear Optimization*, Lecture Notes, Math 408, University of Washington, Seattle, WA; available at https://sites.math.washington.edu/~burke/crs/408/notes/Math408_Spring2014/.

W. Forst and D. Hoffmann (2010), *Optimization—Theory and Practice*, Springer Undergrad. Texts Math. Technol., Springer, New York.

J. Gauvin (1977), *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Program., 12, pp. 136–138.

C. Geiger and C. Kanzow (2002), *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer-Lehrbuch Masterclass, Springer, New York.

F. J. Gould and J. W. Tolle (1971), *A necessary and sufficient qualification for constrained optimization*, SIAM J. Appl. Math., 20, pp. 164–172.

M. Guignard (1969), *Generalized Kuhn–Tucker conditions for mathematical programming in a Banach space*, SIAM J. Control Optim., 7, pp. 232–241.

B. Jiménez and V. Novo (2006), *Characterization of the cone of attainable directions*, J. Optim. Theory Appl., 131, pp. 493–499.

J. Jost (2017), *Riemannian Geometry and Geometric Analysis*, 7th ed., Universitext, Springer, Cham.

H. Karcher (1977), *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math., 30, pp. 509–541.

W. Karush (1939), *Minima of Functions of Several Variables with Inequalities as Side Constraints*,

M.Sc. thesis, University of Chicago, Chicago, IL.
T. H. Kjeldsen (2000), *A contextualized historical analysis of the Kuhn–Tucker theorem in nonlinear programming: The impact of World War* II, Historia Math., 27, pp. 331–361.
H. W. Kuhn and A. W. Tucker (1951), *Nonlinear programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, pp. 481–492.
Y. S. Ledyaev and Q. J. Zhu (2007), *Nonsmooth analysis on smooth manifolds*, Trans. Amer. Math. Soc., 359, pp. 3687–3732.
J. M. Lee (2003), *Introduction to Smooth Manifolds*, Grad. Texts in Math. 218, Springer, New York.
C. Liu and N. Boumal (2019), *Simple Algorithms for Optimization on Riemannian Manifolds with Constraints*, preprint, https://arxiv.org/abs/1901.10000.
D. Motreanu and N. H. Pavel (1982), *Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds*, J. Math. Anal. Appl., 88, pp. 116–132.
J. Nocedal and S. Wright (2006), *Numerical Optimization*, 2nd ed., Springer, New York.
J.-P. Penot (1985), *Variations on the theme of nonsmooth analysis: Another subdifferential*, Nondifferentiable Optimization: Motivations and Applications, Lecture Notes in Econom. and Math. Systems 255, Springer, Berlin, pp. 41–54.
D. W. Peterson (1973), *A review of constraint qualifications in finite-dimensional spaces*, SIAM Rev., 15, pp. 639–654.
M. V. Solodov (2010), *Constraint Qualifications*, Wiley Encyclopedia Oper. Res. Management Sci., J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, eds., John Wiley & Sons.
M. Spivak (1979), *A Comprehensive Introduction to Differential Geometry*, Vol. I, 2nd ed., Publish or Perish, Wilmington, DE.
L. W. Tu (2011), *An Introduction to Manifolds*, 2nd ed., Universitext, Springer, New York.
C. Udrişte (1988), *Kuhn–Tucker Theorem on Riemannian Manifolds*, in Topics in Differential Geometry, Vol. II, Colloquia Mathematica Societatis János Bolyai 46, North-Holland, Amsterdam, pp. 1247–1259.
G. Wachsmuth (2013), *On LICQ and the uniqueness of Lagrange multipliers*, Oper. Res. Lett., 41, pp. 78–80.
W. H. Yang, L.-H. Zhang, and R. Song (2014), *Optimality conditions for the nonlinear programming problems on Riemannian manifolds*, Pac. J. Optim., 10, pp. 415–434.