Constrained Optimization on Smooth Manifolds

1.1. Introduction

There have been some recent contributions towards the generalization of results from the constrained optimization in euclidean space $\mathbb{R}^n$ to constrained optimization on general smooth manifolds. That is, instead of working with the space $\mathbb{R}^n$, we work with a smooth manifold $\mathcal{M}$ and generalize the results to this new setting. The goal of this write up is to conduct a survey of some of these recent advances and the results that form the core of this survey are due to Bergmann and Herzog [2] and Liu and Boumal [3]. Bergman and Herzog have proposed an intrinsic formulation of the KKT conditions and constraint qualification on smooth manifolds, whereas Liu and Boumal have generalized the penalty and Augmented Lagrangian algorithms to the Riemannian manifold setting. The constrained optimization problem can be formulated as

$$\begin{aligned}
\text{minimize} \quad & f(p) \\
\text{subject to} \quad & g_j(p) \leq 0,\ j = 1, 2, ..., l \\
& h_i(p) = 0,\ i = 1, 2, ..., m \qquad\qquad (\mathcal{P}_1) \\
& p \in \mathcal{M}
\end{aligned}$$

where $\mathcal{M}$ is a smooth manifold and the functions f, g and h are assumed to be in C^1. The C^1- property of objective (and constraint functions) means that $f \circ \phi_\alpha^{-1}$ defined on $\phi_\alpha(\mathcal{U}_\alpha)$ and mapping into $\mathbb{R}$ is of class C^1 for every chart $(\mathcal{U}_\alpha, \phi_\alpha)$ from the smooth atlas for the manifold $\mathcal{M}$. The set of such functions is denoted by

$$C^1(\mathcal{M}) = \{f : \mathcal{M} \to \mathbb{R}\} \tag{1.1}$$

Structure of write up: Section 1 consists of a discussion of some standard concepts and terminology from differential geometry and constrained optimization in the context of the problem $\mathcal{P}_1$. Section 2 consists of a detailed description of the derivation of KKT necessary condition for the optimal solution to the problem $(\mathcal{P}_1)$ using intrinsic concepts on the manifold. Section 3 includes a discussion of constraint qualifications. As before, constraint qualifications are conditions on the constraint functions that ensure that the optimal solution satisfies the KKT conditions. As shown in [2], the linear independence constraint qualification, Mangasarian Fromovitz constraint qualification, Abadie's constraint qualification and the Guinard constraint qualification (GCQ) can be generalized on smooth manifolds. Further, it is shown that the implication $LICQ \Rightarrow MFCQ \Rightarrow ACQ \Rightarrow GCQ$ continues to hold on smooth manifolds and GCQ is necessary and sufficient for local optimal to be a KKT point.

1.2. Preliminaries

Topological Manifold. A hausdorff, second countable topological space $(\mathcal{M}, \mathcal{O})$ is called a d-dimensional topological manifold if for every point $p \in \mathcal{M}$, there exists $\mathcal{U} \in \mathcal{O}$ containing p and is locally homeomorphic to $\mathbb{R}^d$. A homeomorphism is a map $\phi : \mathcal{U} \to \mathbb{R}^d$ such that ϕ is continuous, invertible and ϕ^{-1} is continuous. Some terminology is as follows.

1. $(\mathcal{U}, \phi)$ is called a chart on $(M, \mathcal{O})$

2. $\phi : \mathcal{U} \to \mathbb{R}^n$ is called the chart map.

3. $\mathcal{A} = \mathcal{U}_\alpha, \phi_\alpha | \alpha \in A$, where A is an index set, is called an atlas of the topological space $((M, \mathcal{O}))$.

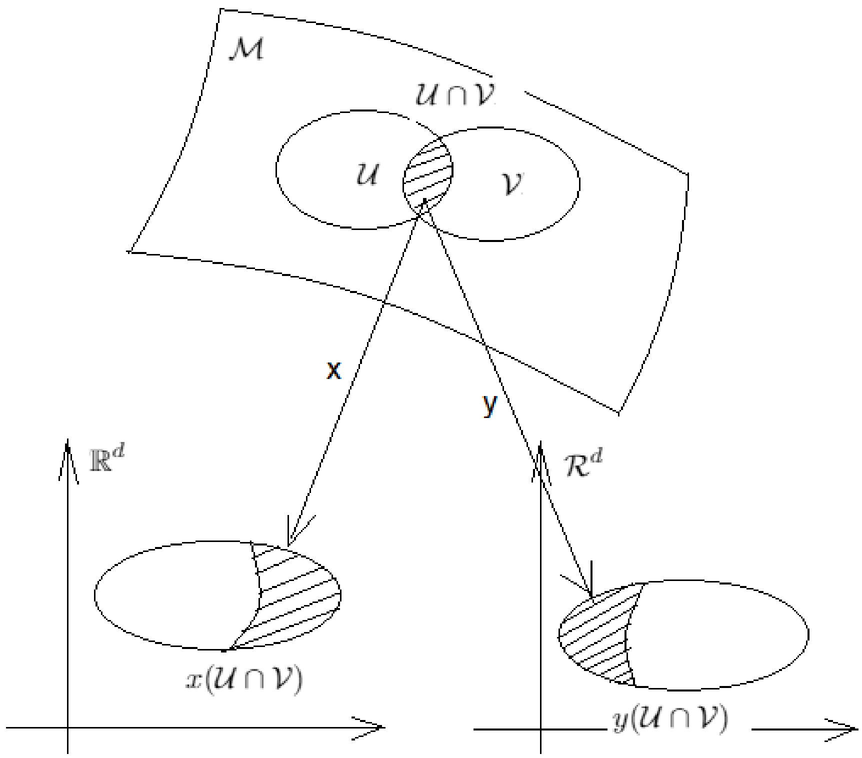

Figure 1.1. Chart Transition

Chart transition maps. Given two charts $(\mathcal{U}_1, \phi_1)$ and $(\mathcal{U}_1, \phi_1)$ with overlapping regions, that is $\mathcal{U}_1 \cap \mathcal{U}_2 \neq \phi$, the map $\phi_2 \circ \phi_1^{-1} : \phi_1(\mathcal{U}_1 \cap \mathcal{U}_2) \to \phi_2(\mathcal{U}_1 \cap \mathcal{U}_2)$ is called the chart transition map. Note that $\phi_1(\mathcal{U}_1 \cap \mathcal{U}_2)$ and $\phi_2(\mathcal{U}_1 \cap \mathcal{U}_2)$ are subsets of $\mathbb{R}^d$ and hence the usual notion of smoothness applies.

Smooth Manifold. A $d-$ dimensional smooth manifold is a $d-$dimensional topological manifold such that the transition maps are of class C^∞. The chart $(\mathcal{U}_\alpha, \phi_\alpha)$ is called a smooth chart and the atlas $\mathcal{A}$ is called a smooth atlas.

Some examples of smooth manifolds include the sphere, the general linear group $GL(n)$ of non singular

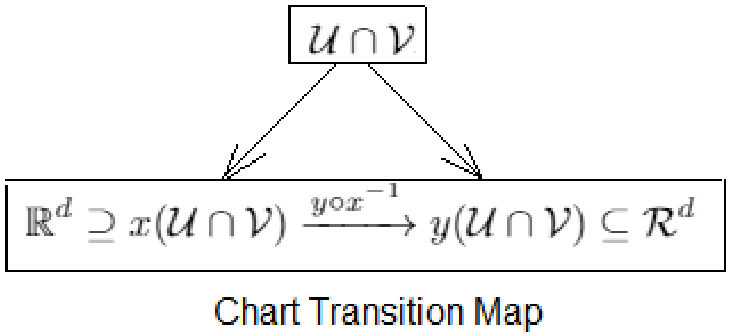

Figure 1.2

matrices, the group of special orthogonal matrices $SO(n)$, the orthogonal Stiefel manifold of orthonormal rectangular matrices etc [2].

We now discuss the crucial notion of the tangent space to a manifold. In order to generalize optimization algorithms and the KKT conditions to the smooth manifolds, we need the notions of gradients and hessians. Note that what we refer to as the euclidean space $\mathbb{R}^n$ is endowed with a linear vector space structure and an inner product. This structure allows us to define gradients and hessians. In general for a smooth manifold, we do not have the linear vector space structure to begin with. The spehere serves as a good example. However we can locally linearize a smooth manifold at every point. This is done by defining a tangent space to the manifold at each point. This tangent space has a vector space structure with the same dimension as the dimension of the smooth manifold. As we will point out later, this tangent space can be endowed with an inner product which leads to the notion of the Riemannian manifold.

Tangent Space. To a differentiable manifold, it is possible to attach at every point x, a tangent space, which is a real vector space that intuitively contains every possible direction one can tangentially pass through x. For example for a 2-sphere as shown, the tangent space at a point is the plane that touches the sphere at that point and is perpendicular to the radius through that point []. This informal description of the tangent space relies on the manifold's ability to be embedded in an ambient euclidean space, in this case the 2-sphere is embedded in $\mathbb{R}^3$ so that tangent vectors can be thought of as directed line segments sticking out of the manifold into the ambient space. However the intrinsic approach used in differential geometry makes use of the manifold itself to define the tangent space and minimizes the reference to an ambient euclidean space. To define a tangent space using only the manifold itself is to let the vectors survive as the directional derivatives they induce. To this extent, first consider the following notions.

C^1 **Curve**: A function $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is called a C^1-curve about $p \in \mathcal{M}$ if $\gamma(0) = p$ holds and $\phi_\alpha \circ \gamma : \mathbb{R} \to \mathbb{R}^d$ is C^1 for some chart (U_α, ϕ_α) about p for some and hence every chart (U_α, ϕ_α) about p. The reason as to why C^1 smoothness in some chart implies C^1 smoothness every chart is skipped in the current discussion and can be found in []. However it is worth mentioning that this can be attributed to the definition of manifold itself which requires that chart transition maps be smooth.

Equivalent Curves. Two curves γ and η about $p \in \mathcal{M}$ are equivalent if

$$\frac{d}{dt}(\phi_\alpha \circ \gamma)(t)|_{t=0} = \frac{d}{dt}(\phi_\alpha \circ \eta)(t)|_{t=0} \tag{1.2}$$

Tangent Vector. Given a C^1 curve γ about $p \in \mathcal{M}$ and $[\gamma]$ be its equivalence class. Then the linear map denoted as $[\dot{\gamma}(0)] : C^1(M) \to \mathbb{R}$ is defined as

$$[\dot{\gamma}(0)](f) = \frac{d}{dt}(f \circ \gamma)(t)|_{t=0}. \tag{1.3}$$

It is also called the tangent vector to $\mathcal{M}$ at p generated by the curve γ.

Tangent Space. Finally the tangent space can now be defined at a point $p \in \mathcal{M}$, denoted as $T_{\mathcal{M}}(p)$, as the collection of all tangent vectors at p,

$$T_{\mathcal{M}}(p) = \{[\dot{\gamma}(0)] : [\dot{\gamma}(0)] \text{ is generated by some } C^1 - \text{curve } \gamma \text{ about p}\} \tag{1.4}$$

The tangent space as defined forms a vector space of the same dimension as the manifold. This important property will be used in further discussions. As the purpose of this write up is to be largely self contained, this fact is stated in the following theorem along with its proof.

Theorem 1.2.1. The tangent space $T_{\mathcal{M}}(p)$ of a d-dimensional manifold $\mathcal{M}$ is a vector space.

Proof. We first define the addition and scaler multiplication operations on $T_{\mathcal{M}}(p)$ as

$$([\gamma] \oplus [\delta])(f) = [\gamma](f) +_{\mathbb{R}} [\delta](f) \tag{1.5}$$

$$(\alpha \otimes [\gamma])(f) = \alpha \times_{\mathbb{R}} [\gamma](f) \tag{1.6}$$

It should be noted that this proof only involves the proof of the fact that the addition and scaler multiplication operations defined are closed. To show that we must show that there exist curves σ and τ such that

$$[\gamma] \oplus [\delta] = [\sigma] \tag{1.7}$$

$$\alpha \otimes [\gamma] = [\tau] \tag{1.8}$$

Claim (1). $\tau : \mathbb{R} \to \mathcal{M}$ defined as $\tau(\lambda) = \gamma(\alpha\lambda)$ is the required curve. Note that $\lambda \in R$ and also denote $\mu_\alpha : \mathbb{R} \to \mathbb{R}$ as $\mu_\alpha(\lambda) = \alpha\lambda$. We have $\tau(0) = \gamma(0) = p \in \mathcal{M}$, so $\tau(0) = p$ holds true. Also $\mu_\alpha(0) = 0$ and $\mu_\alpha^{'}(0) = \alpha$. Now we have

$$[\tau](f) = (f \circ \tau)^{'}(0) \tag{1.9}$$

$$= (f \circ \gamma \circ \mu_\alpha)^{'}(0) \tag{1.10}$$

$$= (f \circ \gamma)^{'}(\mu_\alpha(0))\mu_\alpha^{'}(0) \text{ (chain rule)} \tag{1.11}$$

$$= \alpha(f \circ \gamma)^{'}(0) = \alpha \otimes [\gamma] \tag{1.12}$$

To show the existence of curve σ we make use of a chart $(\mathcal{U}, x)$ and make the following claim.

Claim (2). The curve σ is given as

$$\sigma_x(\lambda) = x^{-1}\Big((x \circ \lambda)(\lambda) +_{\mathbb{R}^d} (x \circ \delta)(\lambda) -_{\mathbb{R}^d} (x \circ \lambda)(0)\Big) \tag{1.13}$$

To begin with we have

$$\sigma_x(0) = x^{-1}\Big((x \circ \gamma)(0) +_{\mathbb{R}^d} (x \circ \delta)(0) -_{\mathbb{R}^d} (x \circ \gamma)(0)\Big) \tag{1.14}$$

$$= x^{-1}\Big((x \circ \delta)(0)\Big) \tag{1.15}$$

$$\delta(0) = p \tag{1.16}$$

We can now write

$$[\delta_x](f) = (f \circ \delta_x)^{'}(0) \\ = (f \circ x^{-1} \circ x \circ \delta_x)^{'}(0) \tag{1.17}$$

Now $f \circ x^{-1}$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$ whereas $x \circ \delta_x$ is a function from $\mathbb{R}$ to $\mathbb{R}^d$. The derivative can be evaluated using the chain rule and is simply the dot product of derivative vectors of the two functions. Using the Einstein's summation notation the dot product can be written as

$$(f \circ \sigma_x)^{'}(0) = \Big((x \circ \sigma_x)^{i^{'}}(0)\Big)\Big(\partial_i(f \circ x^{-1})(x(\sigma_x(0))\Big) \tag{1.18}$$

$$= \Big((x \circ \gamma)^{i^{'}}(0) + (x \circ \delta)^{i^{'}}(0)\Big)\Big(\partial_i(f \circ x^{-1})(x(\sigma_x(0)))\Big) \\ = \Big((x \circ \gamma)^{i^{'}}(0)\Big)\Big(\partial_i(f \circ x^{-1})(x(\sigma_x(0)))\Big)(x(p)) + \Big((x \circ \delta)^{i^{'}}(0)\Big)\Big(\partial_i(f \circ x^{-1})(x(\sigma_x(0)))\Big)(x(p)) \tag{1.19}$$

Using the chain rule backwards can be written as

$$= (f \circ x^{-1} \circ x \circ \gamma)^{'}(0) + (f \circ x^{-1} \circ x \circ \delta)^{'}(0) \tag{1.20}$$

$$= (f \circ \gamma)^{'}(0) + (f \circ \delta)^{'}(0) \tag{1.21}$$

$$= [\gamma](f) + [\delta](f) \tag{1.22}$$

$\square$

Differential. The generalization of the notion of the derivative for C^1 function $f : \mathcal{M} \to \mathbb{R}$ and $p \in \mathcal{M}$ is the linear map denoted as $(df)(p)$ where, $(df)(p) : T_{\mathcal{M}}(p) \to \mathbb{R}$ and defined as

$$(df)(p)[\dot{\gamma}(0)] = [\dot{\gamma}(0)](f) \tag{1.23}$$

It is important to point out that this definition is a natural generalization of the notion of derivative for a function f from $\mathbb{R}^n$ to $\mathbb{R}$. The derivative vector, that is, the gradient ∇f acts on a vector $v \in T_p(\mathbb{R})^n$ and gives the directional derivative of f at p in the direction of v. The equation **??** captures exactly this action of the derivative $(df)(p)$ on $[\dot{\gamma}(0)]$. The cone of linearized feasible directions (definition in the first write up on KKT conditions) is defined as collection of vectors (tangent vectors) in whose directions the constraint functions admit non positive or zero directional derivatives. Thus the definition of the differential as the generalization of function derivative plays a central role in defining the cone of linearized directions for the constraint set on a smooth manifold.

Further, it should be noted that since $T_{\mathcal{M}}(p)$ is a vector space, $(df)(p)$ maps this vector space to $\mathbb{R}$, it belongs to the dual space of $T_{\mathcal{M}}(p)$ and we denote the dual space $T^*_{\mathcal{M}}(p)$ also called as the cotangent space.

We get back to the optimization problem $\mathcal{P}_1$ and discuss the generalization of the KKT conditions to the smooth manifold setting.

Lemma 1.1. Suppose that $(\mathcal{U}, \phi)$ is an arbitrary chart about p^*. Then the following are equivalent.
(a)p^* is a local minimizer of $\mathcal{P}_1$
(b)$\phi(p^*)$ is a local minimzer for the problem,

$$\begin{aligned} \text{minimize} \quad & f \circ \phi^{-1}(x) \text{ w.r.t.} x \in \phi(\mathcal{U}) \subseteq \mathbb{R}^n \\ \text{subject to} \quad & g \circ \phi^{-1}(x) \leq 0 \\ & h \circ \phi^{-1}(x) = 0 \end{aligned}$$

Proof. Suppose that is a local minimizer of $\mathcal{P}_1$ then there exists an open neighborhood of $\mathcal{U}_1$ of $p^*)$ such that $f(p^*)) \leq f(p)$ holds for all $p \in \mathcal{U}_1 \cap \Omega$. Assume by shrinking if necessary that $\mathcal{U}_1 \subseteq \mathcal{U}$. This can be rewritten as $(f \circ \phi^{-1})(\phi(p^*)) \leq (f \circ \phi^{-1})(\phi(p))$ for all $p \in \mathcal{U}_1 \cap \Omega$. Since ϕ is a homeomorphism $\phi(\mathcal{U}_1)$ is an open set containing $\phi(p^*)$ □

However the goal is to use intrinsic concepts from differential geometry and minimize the use of charts. To this extent we begin by defining a tangent vector to the constraint set

Definition 1.2.1. Feasible set. The feasible set is denoted by Ω and is given by

$$\Omega = \{p \in \mathcal{M} : g(p) \leq 0,\ h(p) = 0\} \tag{1.24}$$

Definition 1.2.2. Local minimizer. A point $p^* \in \Omega$ is a local minimizer of $\mathcal{P}_1$ if there exists U of p^* such that

$$f(p^*) \leq f(p)\, \forall p \in U \cap \Omega \tag{1.25}$$

Definition 1.2.3. Active constraints. Similar to the definition seen in the case when $\mathcal{M} = \mathbb{R}^n$ any point $\bar{p}$ in the feasible set, active constraints $\mathcal{A}$ are the constraints which are satisfied with equality instead of strict inequality. We denote the set of indices corresponding to the active constraints at a point p by $\mathcal{A}(p)$ and is defined as;

$$\mathcal{A}(p) = \{j : g_j(p) = 0\} \tag{1.26}$$

The next crucial generalization is that of the Bouligand tangent cone. To that end, the definition of tangent vector to the constraint set Ω is now given.

Definition 1.2.4. Tangent vector to the constraint set Ω. $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(p)$ is called a tangent vector to Ω at p if there exist sequences $(p_k) \subseteq \Omega$ and $t_k \to 0$ such that for all C^1 functions f defined near p, we have

$$[\dot{\gamma}(0)](f) = \lim_{k \to \infty} \frac{f(p_k) - f(p)}{t_k} \tag{1.27}$$

Definition 1.2.5. Bouligand Tangent Cone. The collection of all tangent vectors to Ω at p is termed as the Tangent cone to Ω at p and denoted by

$$T_{\mathcal{M}}(\Omega, p) = \{[\dot{\gamma}(0)] \in T_{\mathcal{M}}(p) : [\dot{\gamma}(0)] \text{ is a tangent vector to } \Omega \text{ at } p\} \tag{1.28}$$

Finally the linearizing cone to the feasible set Ω can be defined as follows. This is the generalization of the cone of linearized feasible directions from chapter 1.

Definition 1.2.6. Cone of linearized feasible directions. For any $p \in \Omega$, we define the linearizing cone to the feasible set Ω by

$$T^{\text{lin}}_{\mathcal{M}}(\Omega, p) = \{[\dot{\gamma}(0)] \in T_{\mathcal{M}}(p) : [\dot{\gamma}(0)](g^j) \leq 0\ \forall j \in \mathcal{A}(p) \text{ and } [\dot{\gamma}(0)](h^i)(p) = 0\ \forall\ i\} \tag{1.29}$$

It was seen in lemma in the write up on KKT conditions that the Bouligand Tangent cone is a subset of the cone of linearized feasible directions. A similar result holds here.

Lemma 1.2. For a feasible point p, $T_{\mathcal{M}}(\Omega, p) \subseteq T^{\text{lin}}_{\mathcal{M}}(\Omega, p)$.

Proof. Let $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(\Omega, p)$ and let the associated tangential sequence be (p_k, t_k) to Ω at p. Note that all the points in the sequence p_k are feasible. Now for all $j \in \mathcal{A}(p)$ we have

$$\frac{g^j(p_k) - g^j(p)}{t_k} \leq 0 \tag{1.30}$$

since $g^j(p_k) \leq 0$ and $g^j(p) = 0$ as $j \in \mathcal{A}(p)$. Since 1.31 holds true for all $k \in \mathbb{N}$, we have $[\dot{\gamma}(0)](g^j) \leq 0$. Similarly we have

$$\frac{h^i(p_k) - h^i(p)}{t_k} = 0 \tag{1.31}$$

for all $k \in \mathbb{N}$ as p_k and p are feasible. Thus $[\dot{\gamma}(0)](h^i) = 0$ and $[\dot{\gamma}(0)] \in T^{\text{lin}}_{\mathcal{M}}(\Omega, p)$. □

We now have the following generalization of the necessary condition for local optimality as stated in theorem (1.3.3) in the write up on KKT conditions.

Theorem 1.2.2. First order necessary optimality condition. Suppose $p^* \in \Omega$ is a local minimizer of $\mathcal{P}_1$. Then we have

$$[\dot{\gamma}(0)](f) \geq 0 \tag{1.32}$$

for all tangent vectors $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(\Omega, p)$

Proof. Suppose [] and p_k, t_k is the associated tangential sequence. Then since p^* is a local minimum, we have for sufficiently large $k \in \mathbb{N}$

$$\frac{f(p_k) - f(p^*)}{t_k} \geq 0 \tag{1.33}$$

which gives $[\dot{\gamma}(0)](f) \geq 0$. □

As seen earlier, the Farkas' lemma plays a crucial role in stating the KKT conditions. Bergman and Herzog use a slightly more general formulation of the Farkas' lemma. The lemma is stated for a linear transformation between a finite dimensional vector space V and $\mathbb{R}^n$. The form of Farkas' lemma used is exactly the same as used in the write up on KKT conditions for $\mathcal{M} = \mathbb{R}^n$. There the lemma was stated with two statements with exactly one of them being true. Instead we can negate one of the two statements and write the lemma as two equivalent statements. So if has a solution then the negation of the statement should hold true. That is, should imply .

Theorem 1.2.3. Farkas' Lemma. For a finite dimensional vector space V, let $\mathbb{A} \in \mathcal{L}(V, R^q)$ and $b \in V^*$. Also denote adjoint of A as $A^* \in \mathcal{L}(V, \mathbb{R}^*)$ Then the following are equivalent:
a. The system $A^*y = b$ has a solution in $y \in \mathbb{R}_{\shortparallel}$, which satisfies $y \geq 0$
b. For any $d \in V$, $Ad \geq 0$ implies $bd \geq 0$.

It is important to point out the following calculations which clarify the operations discussed above. Suppose $A \in \mathcal{L}(V, R^q)$ and $A^* \in \mathcal{L}(R_q, V^*)$. Let $r \in R_q$ and $v \in V$. Then A^* takes in a vector $r \in R_q$ to yield an element $A^*(r) \in V^*$. Since V^* is the dual space of the fdvs V, an element in that space will act on an element of V to yield a real number. That is $A^*(r)(v) \in \mathbb{R}$ and this is evaluated as

$$A^*(r)(v) = r(Av). \tag{1.34}$$

Note that the expression on the right is just a matrix multiplication of a row and column vector or equal size. Now consider a matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \end{pmatrix}$ in $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$. Now the dual for the matrix A is the matrix $A^* = \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}$ and the dual for $\mathbb{R}^3$ is $\mathbb{R}_3$ and $\mathbb{R}^2$ is $\mathbb{R}_2$ where the subscript describes row vector of appropriate dimensions. Now we can think of the matrix A^* as belonging to $\mathcal{L}(\mathbb{R}_2, \mathbb{R}_3)$ in the following sense. Matrix A^* applied to an element of $\mathbb{R}_2$ say $r = (2, 2)$ should yield an element in $\mathbb{R}_3$ that is row vector of size 1×3. This row vector can be found by evaluating the linear combination of rows of A or equivalently the columns of A^* with elements of the row vector $(2, 2)$ as the weights. So we have $A^*r = 2(1 \quad 2 \quad 3) + 2(3 \quad 4 \quad 2) = (8 \quad 12 \quad 6)$. This row vector as an element of $\mathbb{R}_3$ can act on any element of $\mathbb{R}_3$ say $(1, 1, 1)^T$ as a dot product to yield $(8, 12, 6)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 26$. Now we must check whether $A^*(r)(v) = r(Av)$. Clearly we have $r(Av) = (2, 2)\begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = (2, 2)\begin{pmatrix} 4 \\ 9 \end{pmatrix} = 26.$

With this background the KKT conditions for the problem $\mathcal{P}_1$ can now be derived subject to certain constraint qualifications. Derivation of KKT conditions subject to ACQ as well as GCQ is provided in the subsequent discussion.

1.3. KKT conditions

The derivation of the KKT conditions subject to Abadie's constraint qualification is discussed in this section. The Abadie's constraint qualification has been defined by Bergman and Herzog in [2] and is defined exactly as the ACQ in the case for $\mathcal{M} = \mathbb{R}^n$. That is, the set of linearized feasible directions and the set of tangents to the constraint set are equal.

Abadie Constraint Qualification. The ACQ states that the sets defined as the linearizing cone and the tangent cone should be equal, that is

$$T_{\mathcal{M}}^{\text{lin}}(\Omega, p) = T_{\mathcal{M}}(\Omega, p) \tag{1.35}$$

Theorem 1.3.1. Karush-Kuhn-Tucker Necessary conditions. For the problem, let p be a local optimal solution. Assume the Abadie's constraint qualification hold true. Then there exist vectors $\mu \in \mathbb{R}^l$ and $\lambda \in \mathbb{R}^m$

$$\mathrm{g}_j(p) \geq 0 \text{ and } h_i(p) = 0 \ \forall i, j \text{ (Feasibility)}$$
$$\text{(df)(p)} + \sum_{j=1}^{l} \mu_j (dg_j)(p) + \sum_{i=1}^{m} \lambda_i (dh_i)(p) = 0 \text{ (Stationarity)}$$
$$\mu_j g_j(p) = 0, \ j = 1, \ ..., \ l \ , \mu \geq 0 \text{ (Complimentary Slackness)}$$

Proof. Since ACQ holds at $p^* \in \mathcal{M}$ we have $T_{\mathcal{M}}^{\text{lin}}(\Omega, p^*) = T_{\mathcal{M}}(\Omega, p^*)$. Given that p^* is a local minimum for the problem as a result for all $[\dot\gamma(0)] \in T_{\mathcal{M}}(\Omega, p^*) = T_{\mathcal{M}}^{\text{lin}}(\Omega, p^*)$ the following is true for all j and i,

$$[\dot\gamma(0)](f) \geq 0 \tag{1.36}$$
$$[\dot\gamma(0)](g^j) \leq 0 \tag{1.37}$$
$$[\dot\gamma(0)](h^i) = 0 \tag{1.38}$$

Now the equality 1.36 can also be split into two inequalities as $[\dot\gamma(0)](h^i) \leq 0$ and $-[\dot\gamma(0)](h^i) \leq 0$. We now rewrite the less than inequalities into greater than inequalities and for all $[\dot\gamma(0)] \in T_{\mathcal{M}}(\Omega, p^*) = T_{\mathcal{M}}^{\text{lin}}(\Omega, p^*)$ we have,

$$[\dot\gamma(0)](f) \geq 0 \tag{1.39}$$
$$-[\dot\gamma(0)](g^j) \geq 0 \tag{1.40}$$
$$-[\dot\gamma(0)](h^i) \geq 0 \tag{1.41}$$
$$[\dot\gamma(0)](h^i) \geq 0. \tag{1.42}$$

Now since $[\dot\gamma(0)](f) = (df)(p)[\dot\gamma(0)]$, $[\dot\gamma(0)](g^j) = (dg^j)(p)[\dot\gamma(0)]$ and $[\dot\gamma(0)](h^i) = (dh^i)(p)[\dot\gamma(0)]$, $3.19 - 3.22$ can equivalently be written as

$$(df)(p)[\dot\gamma(0)] \geq 0 \tag{1.43}$$
$$-(dg^j)(p)[\dot\gamma(0)] \geq 0 \tag{1.44}$$
$$-(dh^i)(p)[\dot\gamma(0)] \geq 0 \tag{1.45}$$
$$(dh^i)(p)[\dot\gamma(0)] \geq 0. \tag{1.46}$$

Now consider the linear map $A = \begin{pmatrix} (-dg^j)(p) \\ -(dh^i)(p) \\ (dh^i)(p) \end{pmatrix}$, $j \in \mathcal{A}$ and $i = 1, ..., m$ which maps $T_{\mathcal{M}}(\Omega, p)$ to R^p, where $p = \#(\mathcal{A}) + 2m$ such that $A[\dot\gamma(0)] \geq 0$. Now let $b = (df)(p) \in T_{\mathcal{M}}^*(\Omega, p)$ $\big($note that $T_{\mathcal{M}}(\Omega, p) \subseteq T_{\mathcal{M}}(p)$ and so $T_{\mathcal{M}}^*(p) \subseteq T_{\mathcal{M}}^*(\Omega, p)$, thus $b = (df)(p) \in T_{\mathcal{M}}^*(p) \Rightarrow b = (df)(p) \in T_{\mathcal{M}}^*(\Omega, p)\big)$. For all $[\dot\gamma(0)] \in T_{\mathcal{M}}(\Omega, p)$, $(df)(p)[\dot\gamma(0)] \geq 0$, that is, $b\gamma(0) \geq 0$. Thus for $V = T_{\mathcal{M}}(\Omega, p)$ and $\mathbb{R}^p$ the second statement of the Farkas' lemma holds true and hence is equivalent to the first statement of the lemma. As a result there exist scalers $\mu_j \geq 0, \forall j$ and $\lambda_i \geq 0$ such that $A^* y = b$ holds true for some $y \geq 0$. This gives

$$(df)(p^*) = -\sum_{j=1}^{l} \mu_j (dg_j)(p^*) + \sum_{i=1}^{m} \lambda_i (dh_i)(p^*) \tag{1.47}$$

and finally we have

$$(df)(p^*) + \sum_{j=1}^{l} \mu_j (dg_j)(p^*) + \sum_{i=1}^{m} \lambda_i (dh_i)(p^*) = 0 \tag{1.48}$$

□

1.4. Constraint Qualifications

This section is dedicated to the study of constraint qualifications in the context of constrained optimization problem $\mathcal{P}_1$. Bergmann and Herzog [] have shown that the constraint qualifications LICQ, MFCQ, ACQ and GCQ can be generalized to the smooth manifold setting and the chain of implications LICQ⇒MFCQ⇒ACQ⇒ GCQ continues to hold. Further it is shown that GCQ is the weakest constraint qualification as in theorem in the write up on KKT conditions.

Since the generalization of gradients is the differential, for a point $p \in \Omega$, the constraint qualifications at p are defined as follows.

Definition 1.4.1. LICQ. The LICQ is satisfied at $p \in \Omega$ if the set $\{(dh^i)(p)\} \cup \{(dg^j)(p)\}$, where $i \in \mathcal{E}$ and $j \in \mathcal{A}(p)$ is a linearly independent set in the cotangent space $T^*_{\mathcal{M}}(\Omega, p)$.

Definition 1.4.2. MFCQ. The MFCQ is satisfied at $p \in \Omega$ if the set is linearly independent and there exists a tangent vector $[\dot{\gamma}(0)]$ such that

$$(dg^j)(p)[\dot{\gamma}(0)] < 0 \text{ for all } j \in \mathcal{A}(p) \tag{1.49}$$

$$(dh^i)(p)[\dot{\gamma}(0)] = 0 \text{ for all } i \in \mathcal{E} \tag{1.50}$$

Definition 1.4.3. ACQ. The ACQ holds at p if the linearized cone of tangents is equal to the Bouligand tangent cone, that is

$$T^{\text{lin}}_{\mathcal{M}}(\Omega, p) = T_{\mathcal{M}}(\Omega, p) \tag{1.51}$$

Definition 1.4.4. GCQ. The GCQ holds at the point p if the polar to the set of linearized cone of tangents is equal to the polar to the Bouligand tangent cone, that is

$$T^{\text{lin}}_{\mathcal{M}}(\Omega, p)^\circ = T_{\mathcal{M}}(\Omega, p)^\circ \tag{1.52}$$

The relationship between the constraint qualifications is stated in the following theorem

Theorem 1.4.1. LICQ ⇒ MFCQ ⇒ ACQ ⇒ GCQ

Proof. The following proof does not consist of the proof of $MFCQ \Rightarrow ACQ$. The proof can be found in [2]. (LICQ ⇒ MFCQ) The proof is similar to the counterpart in euclidean setting. Consider the linear system

$$A[\dot{\gamma}(0)] = \begin{pmatrix} (dg^j)(p)|_{j\in\mathcal{A}} \\ (dh^i)(p)|_{i=1\ldots m} \end{pmatrix} [\dot{\gamma}(0)] = (-1, \ldots, -1, 0, \ldots, 0)^T \tag{1.53}$$

Since LICQ is assumed, the linear map A is surjective and the system has a unique solution and thus $\dot{\gamma}(0)$ solves the system and satisfies the MFCQ. □

Remark

1.5. KKT Conditions for Riemannian Manifolds

In order to state the KKT conditions for the case when the manifold $\mathcal{M}$ is a Riemannian manifold, we need the notion of gradient on the Riemannian manifold. But before that we define the Riemannian manifold. We need the following definitions. These definitions have been taken from [1]

The notion of tangent bundle is needed first in order to define smooth vector fields on smooth manifolds. A tangent bundle is really the collection of all tangent spaces at every point on the manifold. It is defined as follows

Definition 1.5.1. Tangent Bundle. The tangent bundle of a manifold $\mathcal{M}$ is the disjoint union of tangent spaces of $\mathcal{M}$,

$$T\mathcal{M} = \{(x, v) : \; x \in \mathcal{M} \,\text{and}\, v \in T_x\mathcal{M}\} \tag{1.54}$$

A more detailed description of tangent bundles can be found in the handwritten notes written as part of a systematic study of differential geometry [**?**]. The discussion in the notes also includes a discussion of defining a topology on the tangent bundle to create a smooth bundle out of the tangent bundle. This allows for defining smooth maps between the manifold and the tangent bundle.

The vector field on a manifold can be defined as follows.

Definition 1.5.2. Vector Field. A vector field on a manifold $\mathcal{M}$ is a map $V : \mathcal{M} \to T\mathcal{M}$ such that $V(x)$ is in $T_x\mathcal{M}$ for all $x \in \mathcal{M}$. If V is a smooth map, it is called a smooth vector field.

The description of a vector field as a section of the tangent bundle has been included in the hand written notes [].

Bibliography

[1] Nicholas Boumal.
`An Introduction to Optimization on Smooth Manifolds`. Springer.

[2] Ronny Bergmann and Roland Herzog.
`Intrinsic Formulation of KKT Conditions and Constraint Qualifications On Smooth Manifolds`.

[3] Changshuo Liu and Nicholas Boumal.
`Simple Algorithms for Optimization on Riemannian Manifolds with Constraints`

[4] F. J. Gould and Jon W. Tolle
`A necessary and sufficient qualification for constrained optimization,` SIAM Journal on Applied Mathematics, 20:164-172, 1971