Intrinsic Formulation of KKT Conditions and Constraint Qualifications on Smooth Manifolds.[a]

Ronny Bergmann Roland Herzog

Technische Universität Chemnitz

S16: Optimization, GAMM 2019

Vienna, February 20, 2019.

[a] supported by DFG Grant BE 5888/2-1

Introduction

A lot of theory and algorithms exist for unconstrained problems

$$\text{Minimize} \quad f(\boldsymbol{p}) \in \mathcal{M}$$

where $\mathcal{M}$ is a smooth (Riemannian) manifold.

However: little work on the theory of constrained problems

$$\begin{cases} \text{Minimize} & f(\boldsymbol{p}), \quad \boldsymbol{p} \in \mathcal{M}, \\ \text{s.t.} & g(\boldsymbol{p}) \leq 0, \\ \text{and} & h(\boldsymbol{p}) = 0. \end{cases}$$

This talk: functions $g\colon \mathcal{M} \to \mathbb{R}^m$ and $h\colon \mathcal{M} \to \mathbb{R}^q$.

[Absil, Mahony, Sepulchre, 2008; Udrişte, 1988; Yang, Zhang, Song, 2014; Liu, Boumal, 2019]

First-Order Optimality Conditions on $\mathbb{R}^n$

For the Euclidean case $\mathcal{M} = \mathbb{R}^n$ using the feasible set

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

A local minimizer x^* necessarily satisfies

$$f'(x^*)\, d \geq 0 \quad \text{for all } d \in \mathcal{T}_\Omega(x^*) \quad \Leftrightarrow \quad -f'(x^*) \in \mathcal{T}_\Omega(x^*)^\circ$$

where the (Bouligand) tangent cone is defined as

$$\mathcal{T}_\Omega(x^*) \coloneqq \Bigl\{d \in \mathbb{R}^n : \exists \text{ sequences } (x_k) \subset \Omega,\ x_k \to x^*,\ (t_k) \searrow 0, \\ \text{such that } d = \lim_{k\to\infty} \frac{x_k - x^*}{t_k}\Bigr\}$$

and B° denotes the polar cone of B.

KKT Conditions and Constraint Qualifications

Easier to work with the linearizing cone

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

$$\supset \mathcal{T}_\Omega(x^*)$$

Then the KKT conditions

$$\begin{cases} \mathcal{L}_x(x^*, \mu, \lambda) = f'(x^*) + \mu\, g'(x^*) + \lambda\, h'(x^*) = 0 & , \\ h(x^*) = 0, \quad \mu \ge 0, \quad g(x^*) \le 0, \quad \mu\, g(x^*) = 0 \end{cases}$$

are nothing but the statement

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

But: A local minimizer x^* is not necessarily a KKT point:

$$-f'(x^*) \in \mathcal{T}_\Omega(x^*)^\circ \quad \not\Rightarrow \quad -f'(x^*) \in \mathcal{T}_\Omega^{\text{lin}}(x^*)^\circ$$

Solution: Constraint Qualifications to close this gap.

(Smooth) Manifold & Charts

A topological manifold $\mathcal{M}$ is a

- second countable Hausdorff topological space
- locally homeomorphic to $\mathbb{R}^n$
- local homeomorphisms:
  charts $\varphi_\alpha \colon \mathcal{M} \supset U_\alpha \to \varphi(U_\alpha) \subset \mathbb{R}^n$

A manifold $\mathcal{M}$ is smooth if the transition maps $\varphi_\beta \circ \varphi_\alpha^{-1}$, $\alpha, \beta \in A$, are smooth.

The collection $\mathcal{A} \coloneqq \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$ of such charts "covering" $\mathcal{M}$ is a smooth atlas.

(Smooth) Manifold & Charts

A topological manifold $\mathcal{M}$ is a

- second countable Hausdorff topological space
- locally homeomorphic to $\mathbb{R}^n$
- local homeomorphisms: charts $\varphi_\alpha \colon \mathcal{M} \supset U_\alpha \to \varphi(U_\alpha) \subset \mathbb{R}^n$

A manifold $\mathcal{M}$ is smooth if the transition maps $\varphi_\beta \circ \varphi_\alpha^{-1}$, $\alpha, \beta \in A$, are smooth.

The collection $\mathcal{A} \coloneqq \bigl\{(U_\alpha, \varphi_\alpha)\bigr\}_{\alpha \in A}$ of such charts "covering" $\mathcal{M}$ is a smooth atlas.

Examples

- a sphere $\mathbb{S}^n$
- symmetric positive definite matrices $\mathcal{P}(n)$
- special orthogonal group $\mathrm{SO}(n)$

Tangent Space: Vectors and Covectors

- curve $\gamma\colon (-\varepsilon, \varepsilon) \to \mathcal{M}$ is C^1 about $\boldsymbol{p}$ if
$$\gamma(0) = p \text{ and } \varphi_\alpha \circ \gamma \text{ is } C^1$$
- two C^1-curves γ, ζ are equivalent if
$$\frac{\mathrm{d}}{\mathrm{d}t}(\varphi_\alpha \circ \gamma)(t)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(\varphi_\alpha \circ \zeta)(t)\Big|_{t=0}$$
- we introduce the linear map $[\dot\gamma(0)]$ on the equivalence classes $[\gamma]$ as $[\dot\gamma(0)]f := \frac{\mathrm{d}}{\mathrm{d}t}(\varphi_\alpha \circ f)\Big|_{t=0}$ for all C^1 functions $f\colon U \to \mathbb{R}$, $U \subset \mathcal{M}$ about $\boldsymbol{p}$.

The tangent space is defined as

$$\mathcal{T}_{\boldsymbol{p}}\mathcal{M} := \left\{[\dot\gamma(0)] \colon [\dot\gamma(0)] \text{ is generated by some } C^1\text{-curve } \gamma \text{ about } \boldsymbol{p}\right\}.$$

and is a vector space.

It's dual space $\mathcal{T}^*_{\boldsymbol{p}}\mathcal{M}$ is called cotangent space, its elements are called covectors.

The Tangent Cone in $\mathbb{R}^n$

1. A tangent vector $d \in \mathbb{R}^n$ is called tangent vector to Ω at x if sequences $x_k \to x$, $t_k \searrow 0$ exist such that

$$d = \lim_{k\to\infty} \frac{x_k - x}{t_k} \quad \text{for all } C^1\text{-functions } f \text{ near } \boldsymbol{p}.$$

[Bergmann, Herzog, 2018; Motreanu, Pavel, 1982; Yang, Zhang, Song, 2014]

The Tangent Cone on $\mathcal{M}$

1. A tangent vector $d \in \mathbb{R}^n$ is called tangent vector to Ω at $\boldsymbol{p}$ if sequences $\boldsymbol{p}_k \to \boldsymbol{p}$, $t_k \searrow 0$ exist such that

$$d = \lim_{k\to\infty} \frac{\boldsymbol{p}_k - \boldsymbol{p}}{t_k}$$ for all C^1-functions f near $\boldsymbol{p}$.

[Bergmann, Herzog, 2018; Motreanu, Pavel, 1982; Yang, Zhang, Song, 2014]

The Tangent Cone on $\mathcal{M}$

1. A tangent vector $[\dot\gamma(0)] \in \mathcal{T}_p\mathcal{M}$ is called tangent vector to Ω at $\boldsymbol{p}$ if sequences $\boldsymbol{p}_k \to \boldsymbol{p}$, $t_k \searrow 0$ exist such that

$$[\dot\gamma(0)] = \lim_{k\to\infty} \frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p})}{t_k} \quad \text{for all } C^1\text{-functions} f \text{ near } \boldsymbol{p}.$$

[Bergmann, Herzog, 2018; Motreanu, Pavel, 1982; Yang, Zhang, Song, 2014]

The Tangent Cone on $\mathcal{M}$

1. A tangent vector $[\dot{\gamma}(0)] \in \mathcal{T}_p\mathcal{M}$ is called tangent vector to Ω at $\boldsymbol{p}$ if sequences $\boldsymbol{p}_k \to \boldsymbol{p}$, $t_k \searrow 0$ exist such that

$$[\dot{\gamma}(0)] = \lim_{k\to\infty} \frac{f(\boldsymbol{p}_k) - f(\boldsymbol{p})}{t_k} \quad \text{for all } C^1\text{-functions } f \text{ near } \boldsymbol{p}.$$

2. The collection of all tangent vectors to Ω at $\boldsymbol{p}$,

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

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

[Bergmann, Herzog, 2018; Motreanu, Pavel, 1982; Yang, Zhang, Song, 2014]

The Linearizing Cone

The linearizing cone to Ω at p is defined as

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

We can show the following results parallel to $\mathbb{R}^n$:

1. For any $\boldsymbol{p} \in \Omega$, $\mathcal{T}^{\mathrm{lin}}_{\Omega;\boldsymbol{p}}\mathcal{M}$ is a closed convex cone, and
$$\mathcal{T}_{\Omega;\boldsymbol{p}}\mathcal{M} \subset \mathcal{T}^{\mathrm{lin}}_{\Omega;\boldsymbol{p}}\mathcal{M}$$
holds.
2. For any $\boldsymbol{p} \in \Omega$, we have (by Farkas lemma)
$$\mathcal{T}^{\mathrm{lin}}_{\Omega;\boldsymbol{p}}\mathcal{M}^\circ = \Big\{\sum_{i=1}^{m} \mu_i \, (\mathrm{d}g^i)_{\boldsymbol{p}} + \sum_{j=1}^{q} \lambda_j \, (\mathrm{d}h^j)_{\boldsymbol{p}}, \\ \mu_i \ge 0 \text{ for } i \in \mathcal{A}(\boldsymbol{p}),\ \mu_i = 0 \text{ for } i \in \mathcal{I}(\boldsymbol{p}),\ \lambda_j \in \mathbb{R}\Big\} \subset \mathcal{T}^*_{\boldsymbol{p}}\mathcal{M}$$

Formulation of Constraint Qualifications

We define the following constraint qualifications at $\boldsymbol{p} \in \Omega$.

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

$$X(g^i) < 0 \quad \text{for all } i \in \mathcal{A}(\boldsymbol{p}),$$
$$X(h^j) = 0 \quad \text{for all } j = 1, \ldots, q.$$

3. The ACQ holds at p if $\mathcal{T}_{\Omega;\boldsymbol{p}}^{\mathrm{lin}}\mathcal{M} = \mathcal{T}_{\Omega;\boldsymbol{p}}\mathcal{M}$.
4. The GCQ holds at p if $\mathcal{T}_{\Omega;\boldsymbol{p}}^{\mathrm{lin}}\mathcal{M}^\circ = \mathcal{T}_{\Omega;\boldsymbol{p}}\mathcal{M}^\circ$.

As in $\mathbb{R}^n$, we can show

$$\text{LICQ} \;\Rightarrow\; \text{MFCQ} \;\Rightarrow\; \text{ACQ} \;\Rightarrow\; \text{GCQ}.$$

KKT Theorem on Manifolds

Theorem
Suppose that $\boldsymbol{p} \in \Omega$ is a local minimizer of our problem and that one of the constraint qualifications holds at p.

Then there exist Lagrange multipliers $\mu \in \mathbb{R}_m$ and $\lambda \in \mathbb{R}_q$ such that the KKT conditions

$$
\begin{aligned}
&(\mathrm{d}f)_{\boldsymbol{p}} + \mu\,(\mathrm{d}g)_{\boldsymbol{p}} + \lambda\,(\mathrm{d}h)_{\boldsymbol{p}} = 0 \quad \text{in } \mathcal{T}^*_{\boldsymbol{p}}\mathcal{M},\\
&h(\boldsymbol{p}) = 0,\\
&\mu \geq 0, \quad g(\boldsymbol{p}) \leq 0, \quad \mu\, g(\boldsymbol{p}) = 0
\end{aligned}
$$

hold.

Note: All these properties are stated independent of the choice of chart(s).

A Numerical Example

The Constrained Karcher Mean

$\mathbb{R}^n$: average $x^* = \frac{1}{N}\sum_{i=1}^{N} d_i$ of data points $d_i \in \mathbb{R}^n$ is the unique solution of

$$\text{Minimize} \quad \frac{1}{N}\sum_{i=1}^{N} \left|x - d_i\right|_2^2, \quad x \in \mathbb{R}^n.$$

On $\mathcal{M}$: Karcher mean (Riemannian center of mass) with constraints

$$\text{Minimize} \quad \frac{1}{N}\sum_{i=1}^{N} d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{d}_i), \quad p \in \mathcal{M},$$

$$\text{s.t.} \quad d_{\mathcal{M}}^2(\boldsymbol{p}, \boldsymbol{p}_0) - r^2 \leq 0.$$

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

Constrained Karcher Mean: Analysis

Since the feasible set $\Omega = \{\boldsymbol{p} \in \mathcal{M} : d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{p}_0) \leq r\}$ is compact, a global minimizer to the constrained Karcher mean problem exists. Unlike in the case $\mathcal{M} = \mathbb{R}^n$, there may exist additional local minimizers on manifolds with positive sectional curvature.

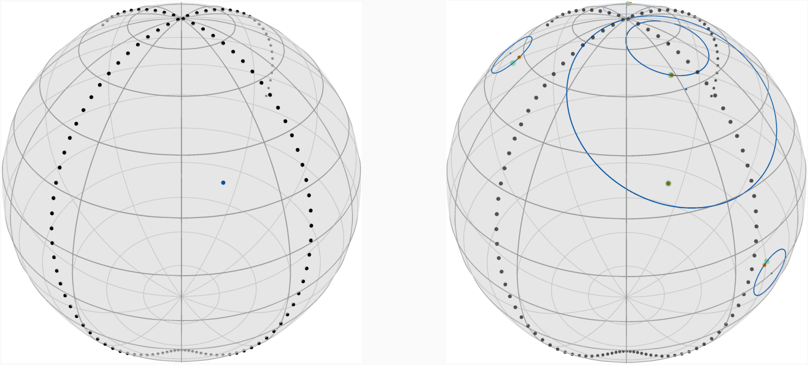
Since the gradient (the Riesz representer of the derivative) of $d^2_{\mathcal{M}}(p, q)$ is equal to $-2 \log_p q$, we can express the KKT conditions as

$$0 = \frac{1}{N} \sum_{i=1}^{N} (-2 \log_{\boldsymbol{p}} \boldsymbol{d}_i, \cdot)_g + \mu \, (-2 \log_{\boldsymbol{p}} \boldsymbol{p}_0, \cdot)_g \quad \text{in } \mathcal{T}^*_{\boldsymbol{p}} \mathcal{M}$$

$$\mu \geq 0, \quad d^2_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{p}_0) \leq r^2, \quad \mu \, (d^2_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{p}_0) - r^2) = 0.$$

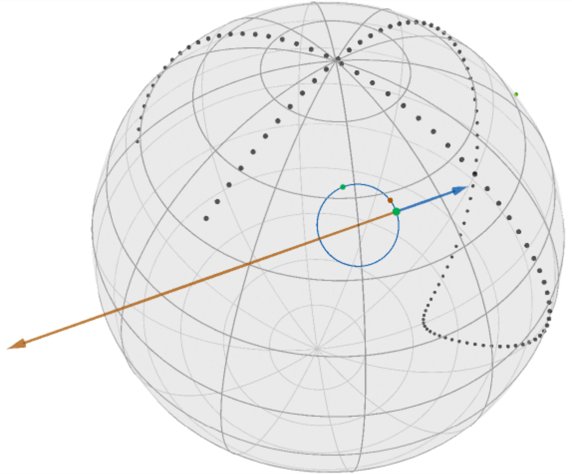
Constrained Karcher Mean: Solution

We consider the problem on the 2-sphere $\mathcal{M} = \mathbb{S}^2$.



Solution (light green) and projected unconstrained solutions (orange) for five different feasible sets (blue). The solution was computed using a projected gradient descent method.

Constraint Karcher Mean: Gradients



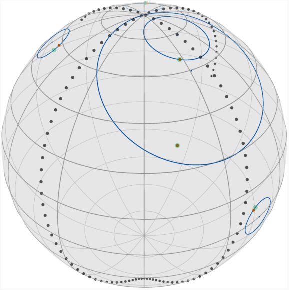
For one of the sets: gradient of the objective f (orange) and the constraint g (blue)

Summary

- KKT conditions for constrained optimization problems on smooth manifolds.
- generalized the notion of tangent cone, linearizing cone and their polars to manifolds.
- constrained Karcher mean problem as an example.

Future Work

- manifold-valued constraints.
- second-order optimality conditions



Thank you for your attention.

References

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

Bergmann, R.; Herzog, R. (2018). *Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds*. SIAM Journal on Optimization, under revision. arXiv: 1804.06214.

Liu, C.; Boumal, N. (2019). “Simple algorithms for optimization on Riemannian manifolds with constraints”. arXiv: 1901.10000.

Motreanu, D.; Pavel, N. H. (1982). “Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds”. *Journal of Mathematical Analysis and Applications* 88.1, pp. 116–132. DOI: 10.1016/0022-247X(82)90180-9.

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

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

https://ronnybergmann.net/talks/2019-GAMM-KKT-on-manifolds.pdf