

Bachelorarbeit

# Regression auf Mannigfaltigkeiten mittels optimaler Steuerung

Verena Brökelmann

Matr. 384046

Betreuer: Prof. Dr. Benedikt Wirth

Betreuender Assistent: Paul Striewski

Institut für numerische und angewandte Mathematik  
Westfälische Wilhelms - Universität Münster

30.09.2015



## **Kurzfassung**

Die vorliegende Arbeit behandelt Regression auf Mannigfaltigkeiten mittels optimaler Steuerung. Dabei wird von einem linearen Regressionsmodell im euklidischen Raum ausgegangen, welches auf eine Riemannsche Mannigfaltigkeit übertragen wird. Es wird der Frage nachgegangen, wie eine mögliche Verallgemeinerung der linearen Regression auf einer Riemannschen Mannigfaltigkeit aussehen kann. Ziel ist es, mittels eines Gradientenabstiegs das Optimum des Regressionsmodells zu ermitteln.

Im Verlauf der Arbeit wird zuerst ein Optimierungsproblem der optimalen Steuerung aufgestellt. Durch eine Parametrisierung der Mannigfaltigkeit wird es in den euklidischen Raum zurückgeführt, um Rechnungen wie gewohnt im euklidischen Raum aufstellen zu können. Dann werden, auf dem Optimierungsproblem aufbauend, Optimalitätsbedingungen erarbeitet. Schließlich wird mittels Gradientenabstieg das Minimum des Optimierungsproblems erreicht. Abschließend wird der Gradientenabstieg des Regressionsmodells auf einer zweidimensionalen Sphäre mit der Software MATLAB realisiert.



## Plagiatserklärung der Studierenden

Hiermit versichere ich, dass die vorliegende Arbeit über Regression auf Mannigfaltigkeiten mittels optimaler Steuerung selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

---

(Datum, Unterschrift)

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in eine Datenbank einverstanden.

---

(Datum, Unterschrift)



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Einführung in den theoretischen Hintergrund</b>	<b>2</b>
2.1. Theoretische Grundlagen . . . . .	2
2.2. Wie sieht Regression auf einer Mannigfaltigkeit aus? . . . . .	6
<b>3. Geodätische Differentialgleichung</b>	<b>8</b>
3.1. für eine allgemeine Riemannsche Mannigfaltigkeit . . . . .	8
3.2. am Beispiel der Sphäre . . . . .	10
3.3. Lösung der Differentialgleichung für die Sphäre . . . . .	13
<b>4. Optimalitätsbedingungen</b>	<b>16</b>
4.1. für eine allgemeine Riemannsche Mannigfaltigkeit . . . . .	16
4.2. am Beispiel der Sphäre . . . . .	23
<b>5. Gradientenabstieg</b>	<b>27</b>
5.1. Implementierung in Matlab . . . . .	29
<b>6. Fazit</b>	<b>31</b>
<b>A. Programmcode</b>	<b>32</b>
<b>B. Literaturverzeichnis</b>	<b>42</b>





# 1. Einleitung

Auf einer gekrümmten zweidimensionalen Fläche im dreidimensionalen reellen Raum werden Messwerte zu verschiedenen Zeitpunkten betrachtet. Ziel ist es, eine Kurve aufzustellen, deren Funktionswert zu jedem Zeitpunkt der Messung einen minimalen Abstand zum entsprechenden Messwert hat. Dieses Prinzip heißt Regression. Daraus ergibt sich ein Optimierungsproblem, welches mittels optimaler Steuerung gelöst werden kann. Dieses Prinzip lässt sich auch für eine Verallgemeinerung, welche einem gekrümmten  $n$ -dimensionalen Raum entspricht, durchführen. Damit beschäftigt sich diese Arbeit.

Die vorliegende Arbeit basiert auf einem Seminarvortrag aus dem Seminar „Optimale Steuerung“, das im Sommersemester 2015 von Prof. Dr. Benedikt Wirth gehalten wurde. Inhalt des Seminars war es, Optimalsteuerungsprobleme unter anderem mittels Gradientenabstieg oder Newton-Methode zu lösen. Zugrunde lag dabei die Lektüre „Optimization with PDE Constraints“ von M. Hinze, R. Pinnau, M. Ulbrich und S. Ulbrich [6].

Das Hauptziel dieser Arbeit ist, eine Regression auf einer Riemannschen Mannigfaltigkeit mittels eines Gradientenabstiegs zu optimieren.

Zunächst werden grundlegende theoretische Begriffe eingeführt, die zum besseren Verständnis der Arbeit beitragen. Die Begriffe beziehen sich hauptsächlich auf Themen der Differentialgeometrie und der optimalen Steuerung. Danach wird die Frage erläutert, wie eine Verallgemeinerung des linearen Regressionsmodells auf einer Mannigfaltigkeit aussehen kann. Dabei wird von einer linearen Regression im euklidischen Raum ausgegangen, welche dann auf eine Riemannsche Mannigfaltigkeit übertragen wird.

Die Verwirklichung des Hauptziels gliedert sich in drei große Kapitel. Ein Optimierungsproblem besteht aus einer Zielfunktion und einem System von Nebenbedingungen. Die im betrachteten Regressionsmodell auftretende Nebenbedingung ist, dass die Regressionskurve eine Geodätische ist. Diese Bedingung wird im ersten Teil zu einer gewöhnlichen Differentialgleichung umgeformt. Darauf aufbauend wird im nächsten Teil das zu optimierende Problem aufgestellt. Zusätzlich werden die zugehörigen Optimalitätsbedingungen ermittelt. Diese stellen die Grundlage des Gradientenabstiegs dar.

Im Verlauf der Arbeit werden die einzelnen Schritte, parallel zu der allgemeinen Riemannschen Mannigfaltigkeit, am Beispiel der Sphäre erläutert. Zum Schluss wird der Gradientenabstieg allgemein aufgestellt und für die Sphäre mittels der Software MATLAB realisiert.

## 2. Einführung in den theoretischen Hintergrund

Dieses Kapitel führt in den theoretischen Hintergrund der vorliegenden Arbeit ein. Dabei wird zuerst auf die theoretischen Grundlagen aus der Differentialgeometrie und der optimalen Steuerung eingegangen. Danach wird eine Motivation für das Thema der Arbeit gegeben. Sie beschäftigt sich mit der Frage, wie eine Regression auf einer Mannigfaltigkeit aussieht.

### 2.1. Theoretische Grundlagen

Um diese Arbeit gut verstehen zu können, benötigen wir Wissen aus der Differentialgeometrie, also sowohl über Riemannsche Mannigfaltigkeiten und Metriken als auch über Tangentialräume und Geodäten.

Wir wiederholen kurz vier grundlegende Begriffe, auf die wir später aufbauen:

1. Sei  $M$  eine Menge. Ein System von Teilmengen  $\mathcal{T} \subset \mathcal{P}(M)$ , wobei  $\mathcal{P}(M)$  die Potenzmenge von  $M$  bezeichnet, heißt **Topologie**, falls die folgenden drei Eigenschaften erfüllt sind:
  - (i)  $\emptyset, M \in \mathcal{T}$ .
  - (ii) Falls  $\{U_i\}_{i \in I} \in \mathcal{T}$ , so  $\bigcup_{i \in I} U_i \in \mathcal{T}$ .
  - (iii) Falls  $U_1, U_2 \in \mathcal{T}$ , dann auch  $U_1 \cap U_2 \in \mathcal{T}$ .
2. Sei  $M$  eine Menge und  $\mathcal{T}$  eine Topologie auf  $M$ . Eine Teilmenge  $U \subset M$  nennen wir **offen**, falls  $U \in \mathcal{T}$ . Dazu ergänzend heißt eine Teilmenge  $A \subset M$  **abgeschlossen**, wenn ihr Komplement ein Element der Topologie ist, also  $M \setminus A \in \mathcal{T}$ .
3. Seien  $U, V$  offene Teilmengen des  $\mathbb{R}^n$ . Falls die Abbildung  $f : U \rightarrow V$  bijektiv ist und sowohl  $f$  als auch  $f^{-1}$  stetig sind, so ist  $f$  ein **Homöomorphismus** ([8], S.17).
4. Sei  $f$  ein Homöomorphismus. Gilt zusätzlich, dass  $f$  und  $f^{-1}$  differenzierbar sind, dann heißt  $f$  **Diffeomorphismus** ([8], S.17).

Wir werden eine Regression auf einer Riemannschen Mannigfaltigkeit durchführen. Um eine Riemannsche Mannigfaltigkeit und ihre Metrik definieren zu können, müssen noch weitere Begriffe beschrieben werden. Dazu benötigen wir unter anderem eine differenzierbare Mannigfaltigkeit, die wiederum eine topologische Mannigfaltigkeit voraussetzt. Dieser thematische Abschnitt stützt hauptsächlich auf die Aussagen des Buches "Klassische Differentialgeometrie" von Wilhelm Klingenberg ([8], S. 83, 119-121).

Sei  $M$  eine Menge. Eine **Karte** ist ein Homöomorphismus  $x : U \rightarrow V$ . Dabei ist  $U$  eine offene und nichtleere Teilmenge von  $M$  und  $V \subset \mathbb{R}^n$  offen und zusammenhängend, wobei  $n$  fest gewählt sei. Ein **Atlas** ist eine Familie von Karten  $x_\alpha : U_\alpha \rightarrow V_\alpha$  für  $\alpha \in A$ , wobei

$\bigcup_{\alpha \in A} U_\alpha = M$  gilt. Dabei bezeichnet  $A$  eine Indexmenge. In anderen Worten heißt dies, dass die Familie  $(U_\alpha)_{\alpha \in A}$  ganz  $M$  überdeckt.

Nun definieren wir eine  $n$ -dimensionale **topologische Mannigfaltigkeit** als eine Menge  $M$ , für die die folgenden drei Eigenschaften erfüllt sind:

1.  $M$  ist Hausdorff. Dies bedeutet, dass für alle  $p, q \in M$  mit  $p \neq q$  offene Teilmengen  $p \in U \subset M$  und  $q \in V \subset M$  existieren, sodass  $U \cap V = \emptyset$  gilt.
2. Die Topologie von  $M$  hat eine abzählbare Basis. Wenn also  $\mathcal{T}$  diese Topologie ist, dann existiert eine abzählbare Teilmenge  $\mathcal{B} \subset \mathcal{T}$ , sodass es für jedes  $U \in \mathcal{T}$   $B_i \in \mathcal{B}$ ,  $i \in I$  gibt mit  $\bigcup_{i \in I} B_i = U$ .
3. Es gibt eine Familie  $(x_\alpha, U_\alpha)_{\alpha \in A}$  von Karten von  $M$ .

Sei  $M$  eine topologische Mannigfaltigkeit für die gilt: Zwei differenzierbare Atlanten sind genau dann äquivalent, wenn ihre Vereinigung wieder ein differenzierbarer Atlas ist. Wir sprechen von einem differenzierbaren Atlas, falls für alle  $(\alpha, \beta) \in A \times A$  der Homöomorphismus  $x_\beta \circ x_\alpha^{-1} : x_\alpha(U_\alpha \cap U_\beta) \rightarrow x_\beta(U_\alpha \cap U_\beta)$  ein Diffeomorphismus ist. Ist für  $M$  die Eigenschaft bezüglich der Äquivalenz erfüllt, so sprechen wir von einer **differenzierbaren Mannigfaltigkeit**.

Für den Begriff der Riemannschen Mannigfaltigkeit fehlt uns nur noch die Definition der Riemannschen Metrik. Diese baut wiederum auf dem Wissen über einen Tangentialraum auf.

Sei  $M$  eine differenzierbare Mannigfaltigkeit der Dimension  $n$  und  $p \in M$ . Für eine offene Umgebung  $U \subset M$  von  $p$  sei  $x : U \rightarrow \mathbb{R}^n$  eine beliebige Karte. Weiter sei für  $\epsilon > 0$   $\gamma : (-\epsilon, \epsilon) \rightarrow M$  mit  $\gamma(0) = p$  eine differenzierbare Kurve in  $M$ . Somit ist auch  $(x \circ \gamma) : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$  eine differenzierbare Kurve. Demnach existiert die Ableitung  $(x \circ \gamma)'(0)$ . Zwei Kurven  $\alpha, \beta : (-\epsilon, \epsilon) \rightarrow M$  mit  $\alpha(0) = \beta(0) = p$  heißen genau dann tangential äquivalent, wenn für eine Karte  $(x, U)$  um  $p$  gilt:  $(x \circ \alpha)'(0) = (x \circ \beta)'(0)$ . Diese Bedingung gilt unabhängig von der Wahl der Karte. Solche Kurven, die tangential äquivalent sind, bilden eine Äquivalenzklasse. Diese wird Tangentialvektor von  $M$  in  $p$  genannt. Zu jedem Punkt  $p \in M$  gibt es einen  $n$ -dimensionalen Vektorraum der Tangentialvektoren in  $p$ . Ihn bezeichnen wir als **Tangentialraum**  $T_p M$  ([7], S.29).

Für eine Karte  $x : U \rightarrow \mathbb{R}^n$  ist der Tangentialraum  $T_{x(p)} \mathbb{R}^n$  Repräsentant von  $T_p M$  bezüglich dieser Karte.

Sei  $M$  eine differenzierbare Mannigfaltigkeit und  $T_p M$  der zugehörige Tangentialraum im Punkt  $p$ . Zudem sei  $g = \{g_p\}_{p \in M}$  eine Familie von positiv definit symmetrischen Bilinearformen, welche wie folgt definiert sind:  $g_p : T_p M \times T_p M \rightarrow \mathbb{R}$  mit  $g_p(v, w) = \langle v | w \rangle_p$  für  $v, w \in T_p M$ .  $\langle \cdot | \cdot \rangle$  stellt das Skalarprodukt dar. Wenn  $x : U \subset M \rightarrow V \subset \mathbb{R}^n$  eine Karte von  $M$  ist, dann ist  $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$  die kanonische Basis von  $T_p M$  und  $dx_1, \dots, dx_n$  die duale

## 2. Einführung in den theoretischen Hintergrund

Basis des Kotangentialraumes  $T_p^*M$ . Die oben genannte Bilinearform habe die Eigenschaft, dass  $g_p(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}) =: g_{ij} : V \rightarrow \mathbb{R}$  differenzierbar ist. Somit gilt  $g_p(\frac{\partial}{\partial x_i}|_v, \frac{\partial}{\partial x_j}|_v) = g_{ij}(v) = g_{x^{-1}(v)}(\frac{\partial}{\partial x_i}|_{x^{-1}(v)}, \frac{\partial}{\partial x_j}|_{x^{-1}(v)})$  für ein beliebiges, aber festes  $v \in V$ . Solch eine Bilinearform  $g$  heißt **Riemannsche Metrik** ([7], S. 58-61, 214).

Eine **Riemannsche Mannigfaltigkeit**  $(M, g)$  ist eine differenzierbare Mannigfaltigkeit  $M$  mit einer Riemannschen Metrik  $g$ .

Nun werden wir uns mit der Theorie über eine kürzeste Verbindung zwischen zwei Punkten in einem gekrümmten Raum, das heißt auf einer Mannigfaltigkeit, auseinandersetzen. In einem euklidischen Raum ist dies eine Gerade. Eine detaillierte Ausführung dazu findet sich beispielsweise in "Elementare Differentialgeometrie" von Christian Bär ([1], S. 47, 170-172, 190-192, 197).

Dazu betrachten wir weiterhin eine  $n$ -dimensionale Riemannsche Mannigfaltigkeit  $M \subset \mathbb{R}^m$ . Ist  $c : I \rightarrow M$  eine Kurve, deren Tangentenvektorfeld  $\dot{c}(t)$  parallel ist, also

$$\frac{\nabla \dot{c}(t)}{dt} = 0$$

gilt, so nennen wir sie **Geodätische**. Sie stellt die lokal kürzeste Verbindung zwischen zwei Punkten dar.

Sei  $f : \mathbb{R}^n \rightarrow M$  eine Parametrisierung von  $M$ . Dann ist die Kurve  $u := (f^{-1} \circ c) : I \rightarrow \mathbb{R}^n$  ebenfalls eine Geodätische. Demnach ist für jede Komponente  $u_k$  die **geodätische Differentialgleichung** erfüllt:

$$\ddot{u}_k(t) + \sum_{i,j=1}^n \Gamma_{ij}^k(u(t)) \cdot \dot{u}_i(t) \cdot \dot{u}_j(t) = 0 \text{ für } k = 1, \dots, n.$$

Dabei bezeichnen  $\Gamma_{ij}^k$  die **Christoffel-Symbole**. Wenn wir für die zu  $M$  gehörige Riemannsche Metrik  $g$  die Matrix  $(g_{ij})_{i,j=1,\dots,n}$  und ihre Inverse  $(g^{ij})_{i,j=1,\dots,n}$  betrachten, dann gilt für die Christoffel-Symbole die Formel:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{m=1}^n \left( \frac{\partial g_{jm}}{\partial u_i} + \frac{\partial g_{im}}{\partial u_j} - \frac{\partial g_{ij}}{\partial u_m} \right) g^{mk} \text{ für } i, j, k = 1, \dots, n.$$

Wir betrachten eine Geodätische  $c : I \rightarrow M$  mit Startpunkt  $c(0) = p \in M$  und Startgeschwindigkeit  $\dot{c}(0) = v \in T_pM$ . Die zugehörige Abbildung  $\exp_p : \mathcal{D}_p \rightarrow M$  mit  $\exp_p(tv) = c(t)$  heißt **Exponentialabbildung**. Dabei ist  $\mathcal{D}_p$  eine bezüglich Null sternförmige Teilmenge von  $T_pM$ . Diese Eigenschaft von  $\mathcal{D}_p$  bedeutet, dass für jeden Tangentialvektor aus  $\mathcal{D}_p$  die ganze Geodätische zwischen dem Tangentialvektor und Null ganz in  $\mathcal{D}_p$  enthalten ist.

Der folgende Abschnitt orientiert sich an dem Buch "Optimization with PDE Constraints" von M. Hinze et al ([6], S. 50, 57-87). Wir widmen uns Themengebieten der optimalen

Steuerung. Diese benötigen wir, um ein Regressionsproblem zu optimieren.

Zunächst führen wir zwei Differentialbegriffe ein. Sei  $F : U \rightarrow Y$  eine Abbildung, wobei  $U$  eine nichtleere offene Teilmenge von  $X$  ist und sowohl  $X$  als auch  $Y$  Banachräume sind. Dann bezeichne  $D_h F(x)$  die Richtungsableitung von  $F$  in Richtung  $h$  im Punkt  $x \in U$ , falls  $dF(x, h) = \lim_{t \rightarrow 0^+} \frac{F(x+th) - F(x)}{t} \in Y$  für alle  $h \in X$  existiert. Falls zusätzlich  $D_h F(x) : X \rightarrow Y, h \mapsto dF(x, h)$  beschränkt und linear ist, so heißt  $F$  **Gâteaux differenzierbar** in  $x$ . Das Gâteaux-Differential wird auch oft mit  $\delta_h F(x)$  gekennzeichnet.  $F$  wird **Fréchet differenzierbar** genannt, falls  $F$  Gâteaux differenzierbar ist und folgende Bedingung erfüllt ist:  $\|F(x+h) - F(x) - D_h F(x)h\|_Y = o(\|h\|_X)$  für  $\|h\|_X \rightarrow 0$ .

Seien  $Y, U, Z$  Banachräume und  $W_{ad} \subset W := Y \times U$  eine nichtleere, abgeschlossene Teilmenge.  $W_{ad}$  beinhaltet alle zulässigen Tupel aus  $Y \times U$ . Weiter seien  $y \in Y$  ein **Zustand** und  $u \in U$  eine **Kontrolle**. Dann beschreiben  $J : Y \times U \rightarrow \mathbb{R}$  eine **Zielfunktion**,  $e : Y \times U \rightarrow Z$  einen Operator und  $U \ni u \mapsto y(u) \in Y$  einen Lösungsoperator. Schließlich sieht ein **Optimierungsproblem** wie folgt aus:

$$\min_{(y,u) \in Y \times U} J(y, u) \tag{1}$$

unter der Nebenbedingung  $e(y, u) = 0, (y, u) \in W_{ad}$ .

Wir nehmen an, dass sowohl die Zielfunktion  $J$ , als auch der Operator  $e$  stetig Fréchet differenzierbar sind. Es existiere zu jedem  $u \in U$  genau eine Lösung  $y(u) \in Y$ , sodass die **Zustandsgleichung**  $e(y, u) = 0$  erfüllt ist. Zudem sei  $e_y(y(u), u) \in \mathcal{L}(Y, Z)$  stetig invertierbar. (2)

Per Definition hänge das reduzierte Zielfunktional nur noch von der Kontrolle  $u$  ab. Wir bezeichnen es mit  $\hat{J}(u) := J(y(u), u)$ . So erhalten wir unter den genannten Annahmen (2) das **reduzierte Optimierungsproblem**:

$$\min_{u \in U} \hat{J}(u) \text{ unter der Nebenbedingung, dass } u \in U_{ad}.$$

Dabei enthält  $U_{ad} \subset U$ , als eine nichtleere und abgeschlossene Teilmenge, alle zulässigen Elemente aus  $U$ .

Seien die Annahmen (2) erfüllt. Zudem sei  $(\bar{y}, \bar{u})$  optimale Lösung des Optimierungsproblems (1). Dann existiert ein adjungierter Zustand  $\bar{p} \in Z^*$ , sodass folgende drei **Optimalitätsbedingungen** erfüllt sind:

$$\begin{aligned} e(\bar{y}, \bar{u}) &= 0 && \text{(Zustandsgleichung)} \\ e_y(\bar{y}, \bar{u})^* \bar{p} &= -J_y(\bar{y}, \bar{u}) && \text{(Adjungierte Gleichung)} \\ \bar{u} \in U_{ad}, \langle J_u(\bar{y}, \bar{u}) + e_u(\bar{y}, \bar{u})^* \bar{p}, u - \bar{u} \rangle_{U^*, U} &\geq 0 \quad \forall u \in U_{ad} && \text{(Adj. Gradientendarstellung)} \end{aligned}$$

Diese Optimalitätsbedingungen werden auch **Karush-Kuhn-Tucker (KKT)** -Bedingungen genannt.

## 2.2. Wie sieht Regression auf einer Mannigfaltigkeit aus?

Wir behandeln in dieser Arbeit die Regression auf einer Riemannschen Mannigfaltigkeit. Um uns dies besser vorstellen zu können, brechen wir das Ganze zunächst auf eine lineare Regression im euklidischen Raum herunter. Bei diesem Vorgehen folgen wir der Darstellung aus "Discrete geodesic regression in shape space" von B. Berkels et al ([3], S. 2-4).

Wir betrachten die lineare Regression in einem  $n$ -dimensionalen euklidischen Raum. Gegeben seien  $K$  Zeitpunkte  $t_1, \dots, t_K \in [0, 1]$ . Zu jedem Zeitpunkt  $t_l$  gibt es verschiedene Messwerte  $y_1^l, \dots, y_N^l \in \mathbb{R}^n$ . Wir wollen nun eine die Messwerte ausgleichende Kurve durch die Punktwolken legen. Gesucht ist eine lineare Kurve  $y : [0, 1] \rightarrow \mathbb{R}^n$ , die jedem Zeitpunkt  $t_l$  einen Funktionswert  $y(t_l)$  zuordnet.  $y(t_l)$  soll eine möglichst kleine quadratische Abweichung zu den jeweiligen Messwerten  $y_1^l, \dots, y_N^l$  haben. Wir sprechen auch von der Methode der kleinsten Quadrate. Das Ziel ist also, eine Gerade der Form  $y(t) = at + c$  zu finden, sodass  $\sum_{l=1}^K \sum_{i=1}^N \|y_i^l - y(t_l)\|_2^2$  minimal wird. Wir minimieren über die Koeffizienten  $\alpha$  und  $c$ . Dabei bezeichnen  $\alpha \in \mathbb{R}^n$  einen Richtungsvektor und  $c \in \mathbb{R}^n$  den Ordinatenabschnitt, wenn wir ein Koordinatensystem mit der Zeit  $t$  auf der Abzisse und der Funktion  $y$  auf der Ordinate betrachten. Die Abbildungen 1 und 2 illustrieren die lineare Regression am Beispiel der Ebene und des dreidimensionalen Raums.

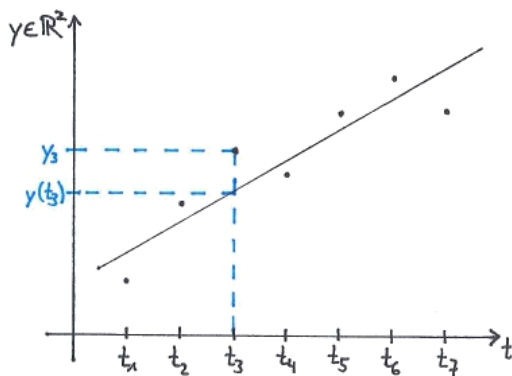


Abbildung 1: lineare Regression in der Ebene ( $K=7, N=1$ )

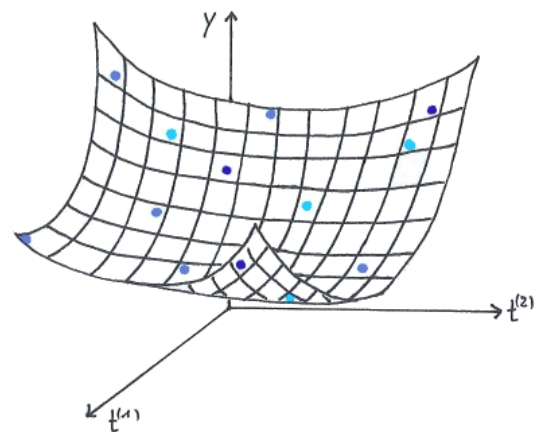


Abbildung 2: Regression im 3D-Raum ( $K=13, N=1$ ), Legende: hellblau: hinter Kurve, mittelblau: ganz nah bei Kurve, dunkelblau: vor Kurve

Eine Gerade im euklidischen Raum ist eindeutig festgelegt durch einen Punkt und einen Richtungsvektor. Analog dazu gilt für eine Geodäte auf einer Mannigfaltigkeit, dass sie eindeutig bestimmt ist durch ihren Startpunkt und ihre Startgeschwindigkeit. Mithilfe dieser Analogie übertragen wir nun das lineare Regressionsmodell im euklidischen Raum

## 2.2. Wie sieht Regression auf einer Mannigfaltigkeit aus?

auf eine Riemannsche Mannigfaltigkeit.

Sei  $M \subset \mathbb{R}^m$  eine  $n$ -dimensionale Riemannsche Mannigfaltigkeit und  $\phi : \mathbb{R}^n \rightarrow M$  ihre Parametrisierung. Mittels der Parametrisierung können wir die zu den Messwerten zugehörigen Eingabepunkte angeben. Die Eingabepunkte  $\{\phi(y_i^l)\}_{i=1, \dots, N}$  zu dem Zeitpunkt  $t_l$  liegen auf der Mannigfaltigkeit  $M$ . Des Weiteren wird der Ordinatenabschnitt  $c$  zum Startpunkt  $\hat{y}$  und der Richtungsvektor  $\alpha$  zur Startgeschwindigkeit  $\hat{v}$ . Somit entspricht die Gerade  $y(t) = \alpha t + c$  der Geodäte  $c : [0, 1] \rightarrow M$ ,  $c(t) = \exp_{\hat{y}}(t\hat{v})$ . Die Geodäte hat die Form einer Exponentialabbildung. Wir übertragen die Distanz zwischen den Messwerten und dem Funktionswert von der euklidischen Distanz  $\|y_i^l - y(t_l)\|_2$  auf die Riemannsche Distanz  $\text{dist}(\phi(y_i^l), c(t_l))$ . Schließlich erhalten wir das Optimierungsproblem in folgender Form: Minimiere  $\sum_{l=1}^K \sum_{i=1}^N \text{dist}^2(\phi(y_i^l), c(t_l))$  über  $\hat{y}$  und  $\hat{v}$  unter der Nebenbedingung, dass  $c$  eine Geodäte ist.

### 3. Geodätische Differentialgleichung

Wir haben ein Optimierungsproblem der optimalen Steuerung gegeben, welches eine Verallgemeinerung eines linearen Regressionsmodells auf einer Riemannschen Mannigfaltigkeit darstellt. Unser Ziel ist es, die entsprechenden Optimalitätsbedingungen aufzustellen.

Allgemein sieht ein Optimalsteuerungsproblem wie folgt aus: Es ist eine Zielfunktion  $J$  gegeben, die von einem Zustand  $y$  und einer Kontrolle  $u$  abhängt. Weiter gibt es einen Operator  $e$ , wieder abhängig von  $y$  und  $u$ . Dieser in der Form  $e(y, u) = 0$  stellt eine Nebenbedingung dar. Wir wollen die Zielfunktion über dem Zustand und der Kontrolle so minimieren, dass die gegebene Nebenbedingung erfüllt ist.

Um das Optimierungsproblem aufstellen zu können, brauchen wir sowohl die Zielfunktion als auch die entsprechende Nebenbedingung. Die Form der Zielfunktion haben wir schon in Kapitel 2.2 aufgestellt. Nun müssen wir noch die Nebenbedingung in Form einer gewöhnlichen Differentialgleichung herleiten. Damit beschäftigt sich dieses Kapitel.

#### 3.1. Geodätische Differentialgleichung für eine allgemeine Riemannsche Mannigfaltigkeit

Gegeben sei eine  $n$ -dimensionale riemannsche Mannigfaltigkeit  $M$  und die zugehörige riemannsche Metrik  $\tilde{g}$ . Um die gewöhnliche Differentialgleichung aufstellen zu können, gehen wir in vier Schritten vor:

1. Parametrisiere  $M$ : Es gibt eine Funktion  $\phi : \mathbb{R}^n \rightarrow M \subset \mathbb{R}^m$ ,  $(y_1, \dots, y_n) \mapsto \phi(y)$ , die ein Homoöomorphismus ist. Solch eine Funktion wird auch eine Karte von  $M$  genannt. Demnach ist  $\phi$  unsere gesuchte Parametrisierung von  $M$ .

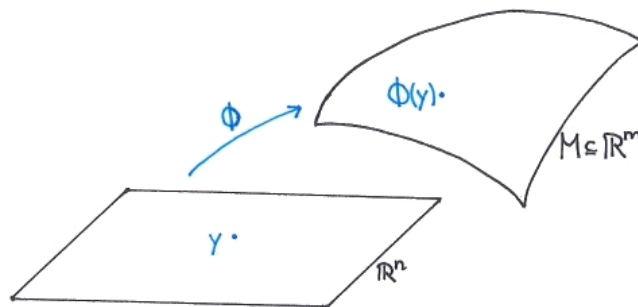


Abbildung 3: Parametrisierung einer Mannigfaltigkeit

2. Metrik  $g$  auf  $\mathbb{R}^n$ : Die Riemannsche Metrik  $\tilde{g}$  ist genauer gesagt eine Riemannsche Metrik auf dem Tangentialraum  $T_{\phi(y)}M$ . Wir wissen, dass die Metrik auf  $T_{\phi(y)}M$  die euklidische Metrik ist, also  $\tilde{g}_{\phi(y)}(v, w) = \langle v|w \rangle$ . Dies können wir auch in einer Form



### 3.1. für eine allgemeine Riemannsche Mannigfaltigkeit

mit Tensorprodukt schreiben:  $\tilde{g}_{\phi(y)} = \sum_{i,j=1}^n g_{ij}(y) \cdot dy_i \otimes dy_j$  ([2], S. 35 f.). Nun suchen wir eine riemannsche Metrik  $g$  auf  $\mathbb{R}^n$ , die zu der Metrik  $\tilde{g}$  auf  $T_{\phi(y)}M$  passt.

$\frac{\partial}{\partial y_1}|_{\phi(y)}, \dots, \frac{\partial}{\partial y_n}|_{\phi(y)}$  bezeichne die Basis des Tangentialraums  $T_{\phi(y)}M$  induziert durch die Karte  $\phi$  und  $dy_1|_{\phi(y)}, \dots, dy_n|_{\phi(y)}$  die duale Basis des Kotangentialraums  $(T_{\phi(y)}M)^* = T_{\phi(y)}^*M$ . Es gilt  $g_{ij} = g(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j}) = \langle \frac{\partial}{\partial y_i} | \frac{\partial}{\partial y_j} \rangle$ .  $g_{ij}$  sei ein Eintrag der Matrix  $(g_{ij})_{ij}$  für  $i, j = 1, \dots, n$ . Dann können wir wie folgt die Metrik  $g$  auf  $\mathbb{R}^n$  aufstellen:

$$\begin{aligned} g_y(v, w) &= \sum_{i,j=1}^n g_{ij} \cdot (dy_i|_{\phi(y)} \otimes dy_j|_{\phi(y)})(v, w) = \sum_{i,j=1}^n g_{ij} \cdot dy_i|_{\phi(y)}(v) \cdot dy_j|_{\phi(y)}(w) \\ &= \sum_{i,j=1}^n g_{ij} \cdot v_i \cdot w_j = v \cdot (g_{ij})_{ij} \cdot w \stackrel{(*)}{=} \langle D\phi|_y \cdot v | D\phi|_y \cdot w \rangle \end{aligned}$$

Herleitung von (\*):

$$\begin{aligned} v \cdot (g_{ij})_{ij} \cdot w &= \sum_{i,j=1}^n g_{ij} \cdot v_i \cdot w_j = \sum_{i,j=1}^n \langle \frac{\partial \phi(y)}{\partial y_i} | \frac{\partial \phi(y)}{\partial y_j} \rangle \cdot v_i \cdot w_j \\ &= \sum_{i,j=1}^n \langle \frac{\partial \phi(y)}{\partial y_i} \cdot v_i | \frac{\partial \phi(y)}{\partial y_j} \cdot w_j \rangle = \langle D\phi|_y \cdot v | D\phi|_y \cdot w \rangle \end{aligned}$$

Somit gilt für die gesuchte Metrik:

$$g_y(v, w) = \tilde{g}_{\phi(y)}(D\phi|_y \cdot v | D\phi|_y \cdot w) = \langle D\phi|_y \cdot v | D\phi|_y \cdot w \rangle.$$

3. Christoffel Symbole auf  $M$ : Die Christoffel Symbole  $\Gamma_{ij}^k$  auf  $M$  sind wie folgt definiert:  $\Gamma_{ij}^k := \frac{1}{2} \sum_{m=1}^n g^{mk} (\frac{\partial g_{im}}{\partial y_j} + \frac{\partial g_{jm}}{\partial y_i} - \frac{\partial g_{ij}}{\partial y_m})$ . Dabei sei  $(g^{ij})_{ij}$  die Inverse der vorher aufgeführten Matrix  $(g_{ij})_{ij}$  für  $i, j = 1, \dots, n$ . Es ist zu beachten, dass  $\Gamma_{ij} = \Gamma_{ji}$  für alle  $i, j = 1, \dots, n$  gilt.
4. Geodätische Differentialgleichung: Wir kennen nun schon die Parametrisierung  $\phi$  von  $M$  und die Riemannsche Metrik  $g$  auf  $\mathbb{R}^n$ . Weiter benötigen wir die Umkehrabbildung der Parametrisierung  $\phi^{-1} : M \rightarrow \mathbb{R}^n$ . Sei  $c : I \rightarrow M \subset \mathbb{R}^m$  mit  $t \mapsto (c_1(t), \dots, c_m(t))$  eine Geodäte. Dann ist  $u := (\phi^{-1} \circ c) : I \rightarrow \mathbb{R}^n$  mit  $t \mapsto (u_1(t), \dots, u_n(t))$  ebenfalls eine Geodäte. Schließlich können wir die lokale Darstellung der geodätischen Differentialgleichung aufstellen:  $\ddot{u}_k(t) + \sum_{i,j=1}^n \Gamma_{ij}^k(u_1(t), \dots, u_n(t)) \cdot \dot{u}_i(t) \cdot \dot{u}_j(t) = 0$  für  $k = 1, \dots, n$ .

Da eine Geodäte durch ihren Startpunkt und ihre Startgeschwindigkeit eindeutig festgelegt ist, kennen wir die Anfangsbedingungen  $c(0) = p$  und  $\dot{c}(0) = v$  der Geodäte  $c$ . Die entsprechenden Anfangsbedingungen für  $u$  definieren wir somit folgendermaßen:  $u(0) = \phi^{-1}(c(0)) = \phi^{-1}(p)$  und  $\dot{u}(0) = (D\phi|_{\phi^{-1}(p)})^{-1}(v) = D(\phi^{-1}|_p)(v)$ .

### 3. Geodätische Differentialgleichung

Zusammengefasst stellt nun die gewöhnliche Differentialgleichung zweiter Ordnung

$$0 = \left( \ddot{u}_k(t) + \sum_{i,j=1}^n \Gamma_{ij}^k(u_1(t), \dots, u_n(t)) \cdot \dot{u}_i(t) \cdot \dot{u}_j(t) \right)_{k=1, \dots, n}$$

mit den Anfangsbedingungen  $u(0) = \phi^{-1}(p)$ ,

$$\dot{u}(0) = D(\phi^{-1}|_p)(v)$$

die Nebenbedingung des Optimierungsproblems dar.

### 3.2. Geodätische Differentialgleichung am Beispiel der Sphäre

Zuvor haben wir die geodätische Differentialgleichung für eine allgemeine Riemannsche Mannigfaltigkeit aufgestellt. Jetzt wollen wir sie am Beispiel der  $S^2$ -Sphäre veranschaulichen. Die  $S^2$ -Sphäre ist die Oberfläche der Einheitskugel, definiert als  $S^2 := \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ . Vereinfacht können wir sie uns auch als Oberfläche der Weltkugel vorstellen.

1. Parametrisiere  $S^2$ : Eine Kurve auf einer gekrümmten Fläche verhält sich anders als eine in einem zweidimensionalen euklidischen Raum. In letzterem können wir leichter rechnen. Deswegen wollen wir ein Flächenstück der Kugel in der euklidischen Ebene darstellen. Dies machen wir mithilfe von Karten. In diesem Zusammenhang ist eine Karte eine Abbildung und nicht, wie angenommen werden könnte, das Bild der Abbildung, welches im allgemeinsprachlichen Gebrauch als Landkarte verstanden wird. Sie beschreibt wie die Kugel lokal auf den  $\mathbb{R}^2$  abgebildet wird. Die Umkehrabbildung der Karte ist ebenfalls eine Karte, welche unsere Parametrisierung der  $S^2$ -Sphäre darstellt.

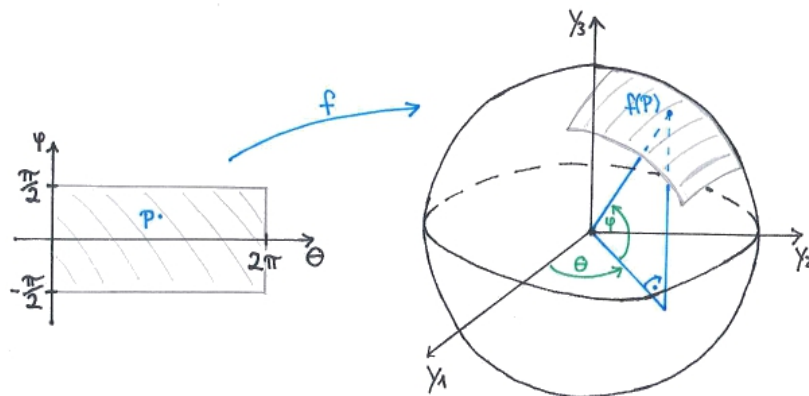


Abbildung 4: Parametrisierung der Sphäre

Wir betrachten die sphärischen Koordinaten  $y_1 = r \cdot \cos \varphi \cdot \cos \theta$ ,  $y_2 = r \cdot \cos \varphi \cdot \sin \theta$ ,  $y_3 = r \cdot \sin \varphi$ . Dabei beschreibt  $\varphi$  den Winkel zur  $y_1 - y_2$ -Ebene und  $\theta$  den Winkel

in der  $y_1 - y_2$ -Ebene. Die  $y_1 - y_2$ -Ebene können wir uns auch als Äquatorebene am Beispiel der Weltkugel vorstellen. Da wir von einer Einheitskugel ausgehen, gilt für den Radius  $r = 1$ . Die Abbildung 4 veranschaulicht diese Darstellung. Darauf aufbauend stellen wir die Parametrisierung  $f$  der  $S^2$ -Sphäre auf:

$$f : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times (0, 2\pi) \rightarrow S^2 \subset \mathbb{R}^3, \quad f(\varphi, \theta) := \begin{pmatrix} \cos \varphi \cdot \cos \theta \\ \cos \varphi \cdot \sin \theta \\ \sin \varphi \end{pmatrix}.$$

Da  $f$  ein Homöomorphismus ist, existiert eine Umkehrabbildung  $f^{-1}$ , die ebenfalls ein Homöomorphismus ist. Für die Umkehrabbildung wird

$$f(\varphi, \theta) = \begin{pmatrix} \cos \varphi \cdot \cos \theta \\ \cos \varphi \cdot \sin \theta \\ \sin \varphi \end{pmatrix} =: \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

nach  $\varphi$  und  $\theta$  umgeformt.

$$\begin{aligned} \sin \varphi = y_3 &\Leftrightarrow \varphi = \arcsin y_3 \\ \frac{y_2}{y_1} &= \frac{\cos \varphi \cdot \sin \theta}{\cos \varphi \cdot \cos \theta} = \tan \theta \end{aligned}$$

$Fkt\theta$  stellt die Umformung nach  $\theta$  dar.

$$Fkt\theta(y_1, y_2) = \begin{cases} \arctan\left(\frac{y_2}{y_1}\right) + \pi & , \text{ falls } y_1 > 0, \\ \arctan\left(\frac{y_2}{y_1}\right) + 2\pi & , \text{ falls } y_1 < 0, y_2 \geq 0, \\ \arctan\left(\frac{y_2}{y_1}\right) & , \text{ falls } y_1 < 0, y_2 \leq 0, \\ \frac{3\pi}{2} & , \text{ falls } y_1 = 0, y_2 > 0, \\ \frac{\pi}{2} & , \text{ falls } y_1 = 0, y_2 < 0, \\ \pi & , \text{ falls } y_1 = 0, y_2 = 0. \end{cases}$$

Wir erhalten

$$f^{-1} : S^2 \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times (0, 2\pi), \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \mapsto \begin{pmatrix} \arcsin y_3 \\ Fkt\theta(y_1, y_2) \end{pmatrix}.$$

2. Metrik  $g$  auf  $\mathbb{R}^2$ : Sei  $g|_{(\varphi, \theta)} : T_{(\varphi, \theta)}S^2 \times T_{(\varphi, \theta)}S^2 \rightarrow \mathbb{R}$  die Riemannsche Metrik auf  $\mathbb{R}^2$ . Für sie gilt  $g_{ij} := g|_{(\varphi, \theta)}(b_i, b_j) = \langle b_i | b_j \rangle$ , wobei  $b_1, \dots, b_n$  Basis von  $T_{(\varphi, \theta)}S^2$  ist.  $T_{(\varphi, \theta)}S^2$  hat genau zwei Basisvektoren, nämlich  $\frac{\partial f(\varphi, \theta)}{\partial x_i}$  für  $i = 1, 2$ . Diese sind

$$\frac{\partial f(\varphi, \theta)}{\partial \varphi} = \begin{pmatrix} -\sin \varphi \cos \theta \\ -\sin \varphi \sin \theta \\ \cos \varphi \end{pmatrix} \quad \text{und} \quad \frac{\partial f(\varphi, \theta)}{\partial \theta} = \begin{pmatrix} -\cos \varphi \sin \theta \\ \cos \varphi \cos \theta \\ 0 \end{pmatrix}.$$

### 3. Geodätische Differentialgleichung

$$\begin{aligned} g_{11} &= \left\langle \frac{\partial f(\varphi, \theta)}{\partial \varphi} \mid \frac{\partial f(\varphi, \theta)}{\partial \varphi} \right\rangle = \sin^2 \varphi \cos^2 \theta + \sin^2 \varphi \sin^2 \theta + \cos^2 \varphi \\ &= \sin^2 \varphi \underbrace{(\cos^2 \theta + \sin^2 \theta)}_{=1} + \cos^2 \varphi = \sin^2 \varphi + \cos^2 \varphi = 1. \end{aligned}$$

Analog zu  $g_{11}$  erhalten wir die anderen Matrixeinträge der Metrik

$$\begin{aligned} g_{22} &= \left\langle \frac{\partial f(\varphi, \theta)}{\partial \theta} \mid \frac{\partial f(\varphi, \theta)}{\partial \theta} \right\rangle = \cos^2 \varphi, \\ g_{12} = g_{21} &= \left\langle \frac{\partial f(\varphi, \theta)}{\partial \theta} \mid \frac{\partial f(\varphi, \theta)}{\partial \varphi} \right\rangle = 0. \end{aligned}$$

Wenn für die Matrixeinträge  $g_{ij} = \left\langle \frac{\partial f(\varphi, \theta)}{\partial x_i} \mid \frac{\partial f(\varphi, \theta)}{\partial x_j} \right\rangle$  gilt, dann sieht die gesamte Matrix, also die Riemannsche Metrik, so aus:  $g|_{(\varphi, \theta)}(v, w) = \langle Df|_{(\varphi, \theta)}(v) \mid Df|_{(\varphi, \theta)}(w) \rangle$ , wobei

$$\begin{aligned} Df|_{(\varphi, \theta)}(v) &= \begin{pmatrix} \frac{\partial f(\varphi, \theta)}{\partial \varphi} & \frac{\partial f(\varphi, \theta)}{\partial \theta} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -\sin \varphi \cdot \cos \theta \cdot v_1 - \cos \varphi \cdot \sin \theta \cdot v_2 \\ -\sin \varphi \cdot \sin \theta \cdot v_1 - \cos \varphi \cdot \cos \theta \cdot v_2 \\ \cos \varphi \cdot v_1 \end{pmatrix}, \\ \langle Df|_{(\varphi, \theta)}(v) \mid Df|_{(\varphi, \theta)}(w) \rangle &= v_1 \cdot w_1 + \cos^2 \varphi \cdot v_2 \cdot w_2 = v \cdot (g_{ij})_{i,j=1,2} \cdot w. \end{aligned}$$

Schließlich erhalten wir die Riemannsche Metrik  $g|_{(\varphi, \theta)}$  auf  $\mathbb{R}^2$ , die äquivalent zu der Metrik  $\tilde{g}|_{f(\varphi, \theta)}$  auf der  $S^2$ -Sphäre ist:

$$\begin{aligned} \tilde{g}|_{f(\varphi, \theta)}(v, w) &= \langle v \mid w \rangle \\ g|_{(\varphi, \theta)}(v, w) &= \langle Df|_{(\varphi, \theta)} \cdot v \mid Df|_{(\varphi, \theta)} \cdot w \rangle = v \cdot (g_{ij})_{i,j=1,2} \cdot w \end{aligned}$$

Der analytische Zusammenhang zwischen den beiden Metriken zeigt sich in der Gleichung

$$g|_{(\varphi, \theta)}(v, w) = \tilde{g}|_{f(\varphi, \theta)}(Df|_{(\varphi, \theta)} \cdot v \mid Df|_{(\varphi, \theta)} \cdot w).$$

### 3. Christoffel Symbole auf $S^2$ :

$$(g_{ij})_{i,j=1,2} = \begin{pmatrix} 1 & 0 \\ 0 & \cos^2 \varphi \end{pmatrix} \text{ ist, wie zuvor berechnet, die Matrix der Metrik.}$$

$$\text{Die Inverse dazu lautet } (g^{ij})_{i,j=1,2} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\cos^2 \varphi} \end{pmatrix}.$$

Wenn wir die partiellen Ableitungen der Matrixeinträge bilden, erhalten wir

$$\frac{\partial g_{22}}{\partial \varphi} = -2 \cos \varphi \cdot \sin \varphi.$$

Die restlichen partiellen Ableitungen sind Null. Die partiellen Ableitungen brauchen wir, um die Christoffel-Symbole zu berechnen. Wir erinnern uns an die Definition

### 3.3. Lösung der Differentialgleichung für die Sphäre

der Christoffel-Symbole:

$$\Gamma_{ij}^k := \frac{1}{2} \sum_{m=1}^n g^{mk} \left( \frac{\partial g_{im}}{\partial y_j} + \frac{\partial g_{jm}}{\partial y_i} - \frac{\partial g_{ij}}{\partial y_m} \right) \text{ für } i, j, k = 1, 2$$

und erhalten

$$\Gamma_{22}^1 = \cos \varphi \cdot \sin \varphi \quad \text{und} \quad \Gamma_{21}^2 = \Gamma_{12}^2 = -\tan \varphi.$$

Die übrigen Christoffel-Symbole sind jeweils Null.

4. Geodätische Differentialgleichung: Bisher haben wir für die  $S^2$ -Sphäre die Parametrisierung, die Metriken und die zugehörigen Christoffel Symbole aufgestellt. Mithilfe dieser drei Komponenten können wir nun die lokale Darstellung der geodätischen Differentialgleichung aufschreiben. Sie hat die Form  $\ddot{u}_k(t) + \sum_{i,j=1}^2 \Gamma_{ij}^k(u_1(t), u_2(t)) \cdot \dot{u}_i(t) \cdot \dot{u}_j(t) = 0$  für  $k = 1, 2$ . Dabei definieren wir  $u$  als die Verknüpfung der Parametrisierung mit der gesuchten Geodäte, also  $u := (f^{-1} \circ c) : I \rightarrow (-\frac{\pi}{2}, \frac{\pi}{2}) \times (0, 2\pi)$ . Angewendet auf die  $S^2$ -Sphäre gilt:

$$\begin{aligned} 0 &= \ddot{u}_1(t) + \cos u_1(t) \cdot \sin u_1(t) \cdot (\dot{u}_2(t))^2 \\ 0 &= \ddot{u}_2(t) - 2 \tan u_1(t) \cdot \dot{u}_1(t) \cdot \dot{u}_2(t). \end{aligned}$$

Die gesuchte Geodäte  $c$  ist eindeutig festgelegt durch ihren Startpunkt  $c(0) = p$  und ihre Startgeschwindigkeit  $\dot{c}(0) = v$ . Demnach sind  $u(0) = f^{-1}(c(0)) = f^{-1}(p)$  und  $\dot{u}(0) = (Df|_{f^{-1}(p)})^{-1}(v) = D(f^{-1}|_p)(v)$  die Anfangsbedingungen von  $u$ .

Die Nebenbedingung des Optimierungsproblems setzt sich zusammen aus einer gewöhnlichen Differentialgleichung zweiter Ordnung und zwei Anfangsbedingungen. Zusammengefasst ist dies nun:

$$\begin{aligned} 0 &= \ddot{u}_1(t) + \cos u_1(t) \cdot \sin u_1(t) \cdot (\dot{u}_2(t))^2 \\ 0 &= \ddot{u}_2(t) - 2 \tan u_1(t) \cdot \dot{u}_1(t) \cdot \dot{u}_2(t) \\ u(0) &= f^{-1}(p) \\ \dot{u}(0) &= (Df|_{f^{-1}(p)})^{-1}(v) \end{aligned} \tag{3}$$

### 3.3. Lösung der Differentialgleichung für die Sphäre

Im weiteren Verlauf werden wir zeigen, dass die Geodäte  $c$  dem Großkreis auf der  $S^2$ -Sphäre entspricht, welche das zuvor genannte Problem löst. Ein Großkreis ist ein Kreis auf der Oberfläche der Sphäre. Sein Mittelpunkt entspricht dem Schwerpunkt der Sphäre.

Wir richten das Koordinatensystem so aus, dass  $N = (1, 0, 0)^T$  den Nordpol der  $S^2$ -Sphäre beschreibt, also die  $x$ -Achse in die Höhe zeigt und  $y$ - und  $z$ -Achse die waagerechte Ebene aufspannen. Wir können annehmen, dass  $p = c(0)$  der Nordpol  $N$  ist (siehe Abb.

### 3. Geodätische Differentialgleichung

5). Somit gilt für die Anfangsbedingungen von  $u$ :

$$u(0) = f^{-1}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \pi \end{pmatrix},$$

$$\dot{u}(0) = (Df|_{(0,0)})^{-1}(v) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_3 \\ v_2 \end{pmatrix}.$$

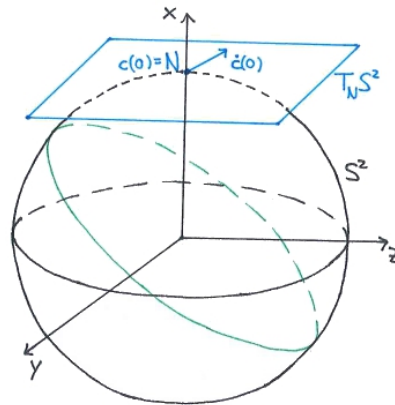


Abbildung 5: Zum Nordpol (blau) ausgerichtete Sphäre und Großkreis (grün)

Die Startgeschwindigkeit  $\dot{c}(0)$  der Geodäte  $c$  ist ein Tangentialvektor in ihrem Startpunkt  $c(0) = N$ . Demnach liegt  $\dot{c}(0)$  tangential an  $c(0)$ , also gilt  $\langle \dot{c}(0), N \rangle = 0$ . Damit die Gleichung erfüllt ist, muss die  $x$ -Komponente des Tangentialvektors Null sein. Dies gilt für alle Tangentialvektoren des Tangentialraums im Nordpol  $N$ . Daher definieren wir den Tangentialraum als

$$T_N S^2 := \{w \in \mathbb{R}^3 \mid \langle w, N \rangle = 0\} = \{(w_1, w_2, w_3) \in \mathbb{R}^3 \mid w_1 = 0\}.$$

Die folgende Rechnung zeigt uns, dass der Tangentialvektor  $\dot{c}(0)$  im Tangentialraum  $T_N S^2$  liegt.

$$\dot{c}(0) = v = (Df|_{(0,0)})((Df|_{(0,0)})^{-1}(v)) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_3 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ v_2 \\ v_3 \end{pmatrix},$$

$$\left\langle \begin{pmatrix} 0 \\ v_2 \\ v_3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = 0.$$

Der Tangentialvektor  $v$  ist fest im Nordpol  $N$  verankert. Durch Drehen der Sphäre können wir nur die  $y$ - und  $z$ -Komponenten von  $v$  beeinflussen. Wir richten  $v$  so aus, dass seine

### 3.3. Lösung der Differentialgleichung für die Sphäre

$z$ -Komponente verschwindet, das heißt  $v$  nur in  $y$ -Richtung zeigt. So gilt  $v_3 = 0$ . Somit sind die Startwerte für die Geodäte  $c$  auf der Sphäre

$$c(0) = N = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \dot{c}(0) = \begin{pmatrix} 0 \\ v_2 \\ 0 \end{pmatrix}$$

und entsprechend für die zugehörige Geodäte  $u$  im euklidischen Raum  $\mathbb{R}^2$

$$u(0) = \begin{pmatrix} 0 \\ \pi \end{pmatrix}, \quad \dot{u}(0) = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}.$$

Aus den Anfangswerten können wir schließen, dass  $u(t) = (0, v_2 \cdot t + \pi)$ . Die zugehörigen Ableitungen sind  $\dot{u}(t) = (0, v_2)$ ,  $\ddot{u}(t) = (0, 0)$ . Somit löst  $u(t) = (0, v_2 t)$  das Optimierungsproblem (3). Jetzt brauchen wir nur noch  $c(t)$  berechnen:

$$\begin{aligned} \|v\| &= \left\| \begin{pmatrix} 0 \\ v_2 \\ 0 \end{pmatrix} \right\| = \sqrt{v_2^2} = v_2 \\ c(t) &= (f \circ u)(t) = \begin{pmatrix} \cos(v_2 t) \\ \sin(v_2 t) \\ 0 \end{pmatrix} = \cos(v_2 t) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \sin(v_2 t) \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ &= \cos(v_2 t) \cdot N + \sin(v_2 t) \cdot \frac{1}{v_2} \begin{pmatrix} 0 \\ v_2 \\ 0 \end{pmatrix} = \cos(\|v\|t) \cdot p + \sin(\|v\|t) \cdot \frac{v}{\|v\|}. \end{aligned}$$

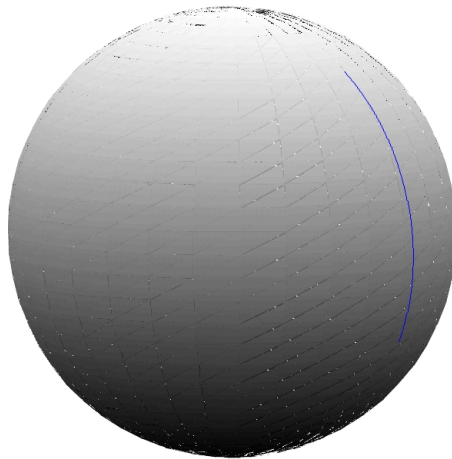


Abbildung 6: Die Kurve  $c(t)$  (blau) ist ein Teilstück des Großkreises (mit MATLAB erzeugt,  $K=300$ ,  $N=100$ )

## 4. Optimalitätsbedingungen

Dieses Kapitel widmet sich dem Aufstellen der Optimalitätsbedingungen sowohl für eine allgemeine Riemannsche Mannigfaltigkeit als auch für die  $S^2$ -Sphäre. Nach dem Aufstellen des Optimierungsproblems mit der Zustandsgleichung in Gestalt einer Differentialgleichung werden Gâteaux-Differentiale berechnet. Mit ihrer Hilfe werden dann die drei Bedingungen hergeleitet.

### 4.1. Optimalitätsbedingungen für eine allgemeine Riemannsche Mannigfaltigkeit

Zur Erinnerung sind hier die zentralen Abbildungen des bisher Erarbeiteten aufgeführt:

$\phi : \mathbb{R}^n \rightarrow M$  Parametrisierung,

$c : I \rightarrow M$  mit  $c(0) = p, \dot{c}(0) = v$  Geodäte,

$u := (\phi^{-1} \circ c) : I \rightarrow \mathbb{R}^n$  mit  $u(0) = \phi^{-1}(p), u'(0) = (D\phi|_{\phi^{-1}(p)})^{-1}(v)$  Geodäte.

Die Nebenbedingung haben wir in Form einer gewöhnlichen Differentialgleichung zweiter Ordnung mit zwei Anfangsbedingungen aufgestellt:

$$\begin{cases} \left( u''_k(t) + \sum_{i,j=1}^n \Gamma_{ij}^k(u_1(t), \dots, u_n(t)) \cdot u'_i(t) \cdot u'_j(t) \right)_{k=1, \dots, n} = 0 \\ u(0) =: \phi^{-1}(p) \\ u'(0) =: (D\phi|_{\phi^{-1}(p)})^{-1}(v) \end{cases}$$

Diese werden wir in ein System von gewöhnlichen Differentialgleichungen erster Ordnung überführen. Dazu benötigen wir die folgenden Definitionen:

$$\begin{aligned} u(0) &=: \hat{r} \\ u'(0) &=: \hat{s} \\ \tilde{r}(t) &:= u(t) - \hat{r}, \quad \tilde{r}(0) = 0 \\ \tilde{s}(t) &:= u'(t) - \hat{s}, \quad \tilde{s}(0) = 0 \end{aligned} \tag{4}$$

$$g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) := - \left( \sum_{i,j=1}^n \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j) \right)_{k=1, \dots, n} .$$

Wir wählen  $\tilde{r}(t), \tilde{s}(t)$  aus  $\tilde{C}_0^1([0, 1], \mathbb{R}^n) := \{f \in C^1(I, \mathbb{R}^n) | f(0) = 0\}$ .  $\hat{r}, \hat{s}$  seien reellwertige  $n$ -dimensionale Vektoren. Dann stellen  $(\tilde{r}(t), \tilde{s}(t))$  den Zustand und  $(\hat{r}, \hat{s})$  die Kontrolle dar.

Wenn wir die neu definierten Werte und Funktionen (4) in die obige Differentialgleichung



4.1. für eine allgemeine Riemannsche Mannigfaltigkeit

einsetzen, erhalten wir die Nebenbedingung

$$e : \tilde{C}_0^1([0, 1], \mathbb{R}^{2n}) \times \mathbb{R}^{2n} \rightarrow C^0([0, 1], \mathbb{R}^{2n}),$$

$$e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) = \begin{pmatrix} (\tilde{r}'_k(t) - \tilde{s}_k(t) - \hat{s}_k)_{k=1, \dots, n} \\ (\tilde{s}'_k(t) + \sum_{i,j=1}^n \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j))_{k=1, \dots, n} \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{r}(t) \\ \tilde{s}(t) \end{pmatrix}' - \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix}.$$

Aus Kapitel 3 wissen wir, dass die Zielfunktion auf der Mannigfaltigkeit die Gestalt  $J(c(t), (p, v)) = \sum_{l=1}^K \sum_{i=1}^N \text{dist}^2(\phi(y_i^l), c(t_l))$  hat. Wir formen sie zu einem Funktional im euklidischen Raum um und erhalten  $J(u(t), (\phi^{-1}(p), D(\phi^{-1}|_p))(v)) = \sum_{l=1}^K \sum_{i=1}^N \text{dist}^2(y_i^l, u(t_l))$ . Jetzt setzen wir die obigen Definitionen (4) ein. Daraus folgt die Zielfunktion

$$J : \tilde{C}_0^1([0, 1], \mathbb{R}^{2n}) \times \mathbb{R}^{2n} \rightarrow \mathbb{R}, ((\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})) \mapsto \sum_{l=1}^K \sum_{i=1}^N \text{dist}^2(y_i^l, \tilde{r}(t_l) + \hat{r}).$$

Damit ergibt sich das **zu optimierende Problem**:

$$\min_{(\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})} \sum_{l=1}^K \sum_{i=1}^N \text{dist}^2(y_i^l, \tilde{r}(t_l) + \hat{r}),$$

$$\text{sodass } \begin{pmatrix} \tilde{r}(t) \\ \tilde{s}(t) \end{pmatrix}' - \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix} = 0 \text{ für } \tilde{r}(t), \tilde{s}(t) \in \tilde{C}_0^1([0, 1], \mathbb{R}^{2n}), \hat{r}, \hat{s} \in \mathbb{R}^{2n} \text{ gilt.}$$

Es seien alle Elemente aus  $\tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$  und  $\mathbb{R}^{2n}$  zulässig. Somit gilt offensichtlich, dass  $\mathbb{R}^{2n}$  eine nichtleere, abgeschlossene und konvexe Teilmenge von  $\mathbb{R}^{2n}$  ist. Sowohl  $\mathbb{R}^{2n}$ ,  $\tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$  als auch  $C^0([0, 1], \mathbb{R}^{2n})$  sind Banachräume.

Wir nehmen an, dass Voraussetzungen aus (2) für die Optimalitätsbedingungen zutreffen:

- $J : \tilde{C}_0^1([0, 1], \mathbb{R}^{2n}) \times \mathbb{R}^{2n} \rightarrow \mathbb{R}$ ,  $e : \tilde{C}_0^1([0, 1], \mathbb{R}^{2n}) \times \mathbb{R}^{2n} \rightarrow C^0([0, 1], \mathbb{R}^{2n})$  sind stetig Fréchet-differenzierbar.
- Für alle  $u$  in einer Umgebung  $V \subset \mathbb{R}^{2n}$  von  $\mathbb{R}^{2n}$  existiert genau eine Lösung  $y(u)$  aus  $\tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$ , sodass die Zustandsgleichung  $e(y(u), u) = 0$  erfüllt ist.
- $e_y(y(u), u) \in \mathcal{L}(\tilde{C}_0^1([0, 1], \mathbb{R}^{2n}), C^0([0, 1], \mathbb{R}^{2n}))$  hat eine beschränkte Inverse.

Dann existiert ein adjungierter Zustand  $p \in (C^0([0, 1], \mathbb{R}^{2n}))^*$ , sodass die Optimalitätsbedingungen erfüllt sind:

1. Zustandsgleichung:  $e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) = 0$
2. Adjungierte Gleichung:  $e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t) + J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) = 0$
3. Adj. Gradientendarstellung:  $\langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) + e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t), \hat{u} - (\hat{r}, \hat{s})^T \rangle \geq 0$   
für alle  $\hat{u} \in \mathbb{R}^{2n}$ .

#### 4. Optimalitätsbedingungen

Um diese Bedingungen aufstellen zu können, bilden wir zuerst die Gâteaux-Ableitungen der Zielfunktion und des Operators  $e$ . Seien  $(\varphi_r(t), \varphi_s(t)) \in \tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$  und  $(\omega_r, \omega_s) \in \mathbb{R}^{2n}$ . Dann gilt:

$$\begin{aligned} \delta_{\tilde{r}(t)} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_r(t)) &= \frac{d}{d\tau} (J(\tilde{r}(t) + \tau\varphi_r, \tilde{s}(t), \hat{r}, \hat{s})(\varphi_r(t))) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N \frac{d}{d\tau} \left( \text{dist}^2(y_i^l, \tilde{r}(t_l) + \tau\varphi_r(t_l) + \hat{r}) \right) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N \left( 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \tau\varphi_r(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \tau\varphi_r(t_l) + \hat{r}) \cdot \varphi_r(t_l) \right) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot \varphi_r(t_l) \end{aligned}$$

$$\delta_{\tilde{s}(t)} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_s(t)) = 0$$

$$\begin{aligned} \delta_{\hat{r}} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_r) &= \frac{d}{d\tau} (J(\tilde{r}(t), \tilde{s}(t), \hat{r} + \tau\omega_r, \hat{s})(\omega_r)) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N \frac{d}{d\tau} \left( \text{dist}^2(y_i^l, \tilde{r}(t_l) + \hat{r} + \tau\omega_r) \right) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N \left( 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r} + \tau\omega_r) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r} + \tau\omega_r) \cdot \omega_r \right) \Big|_{\tau=0} \\ &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot \omega_r \end{aligned}$$

$$\delta_{\hat{s}} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_s) = 0$$

$$\begin{aligned} \delta_{(\tilde{r}(t), \tilde{s}(t))} e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_r(t), \varphi_s(t)) &= \frac{d}{d\tau} (e(\tilde{r}(t) + \tau\varphi_r(t), \tilde{s}(t) + \tau\varphi_s(t), \hat{r}, \hat{s})) \Big|_{\tau=0} \\ &= \frac{d}{d\tau} \left( \left( \begin{array}{c} \tilde{r}(t) + \tau\varphi_r(t) \\ \tilde{s}(t) + \tau\varphi_s(t) \end{array} \right)' - \left( g(\tilde{r}(t) + \tau\varphi_r(t), \tilde{s}(t) + \tau\varphi_s(t), \hat{r}, \hat{s}) \right) \right) \Big|_{\tau=0} \\ &= \left( \begin{array}{c} \varphi_r'(t) \\ \varphi_s'(t) \end{array} \right) - \left( \begin{array}{c} \varphi_s(t) \\ \frac{\partial}{\partial \tilde{r}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \varphi_r(t) + \frac{\partial}{\partial \tilde{s}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \varphi_s(t) \end{array} \right) \\ \delta_{(\hat{r}, \hat{s})} e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_r, \omega_s) &= \frac{d}{d\tau} (e(\tilde{r}(t), \tilde{s}(t), \hat{r} + \tau\omega_r, \hat{s} + \tau\omega_s)) \Big|_{\tau=0} \\ &= \frac{d}{d\tau} \left( \left( \begin{array}{c} \tilde{r}(t) \\ \tilde{s}(t) \end{array} \right)' - \left( g(\tilde{r}(t), \tilde{s}(t), \hat{r} + \tau\omega_r, \hat{s} + \tau\omega_s) \right) \right) \Big|_{\tau=0} \\ &= - \left( \begin{array}{c} \omega_s \\ \frac{\partial}{\partial \tilde{r}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \omega_r + \frac{\partial}{\partial \tilde{s}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \omega_s \end{array} \right) \end{aligned}$$

Die in der Gâteaux-Ableitung von  $e$  auftauchenden partiellen Ableitungen von  $g$  lassen sich wie folgt berechnen:

$$\begin{aligned}
 \frac{\partial}{\partial \tilde{r}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &= \frac{\partial}{\partial \tilde{r}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \\
 &= \left( \frac{\partial}{\partial \tilde{r}_l(t)} \left( \sum_{i,j=1}^n \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j) \right) \right)_{l,k=1, \dots, n} \\
 &= \left( \sum_{i,j=1}^n \frac{\partial}{\partial \tilde{r}_l(t)} \left( \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \right) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j) \right)_{l,k=1, \dots, n} \\
 \frac{\partial}{\partial \tilde{s}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &= \frac{\partial}{\partial \tilde{s}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \\
 &= \left( \frac{\partial}{\partial \tilde{s}_l(t)} \left( \sum_{i,j=1}^n \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j) \right) \right)_{l,k=1, \dots, n} \\
 &= \left( \sum_{i,j=1}^n \Gamma_{ij}^k(\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot \frac{\partial}{\partial \tilde{s}_l(t)} ((\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j)) \right)_{l,k=1, \dots, n}
 \end{aligned}$$

Die **2. Karush-Kuhn-Tucker-Bedingung** lautet

$$e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p = -J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}),$$

das heißt es gilt

$$\langle \xi, e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^*(p) \rangle_{Y, Y^*} = \langle \xi, -J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \rangle_{Y, Y^*}$$

für alle  $\xi \in Y$ . Dabei sind  $Y := \tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$  und  $Z := C^0([0, 1], \mathbb{R}^{2n})$  definiert. Mit den Eigenschaften der dualen Paarung folgt:

$$\begin{aligned}
 - \langle \xi, e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^*(p) \rangle_{Y, Y^*} &= \langle e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi), p \rangle_{Z, Z^*}, \\
 - \langle \xi, -J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \rangle_{Y, Y^*} &= -J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi).
 \end{aligned}$$

Die Elemente aus  $(C^0([0, 1], \mathbb{R}^{2n}))^*$  sind Radon-Maße. Sei  $M([0, 1])$  die Menge aller Radon-Maße mit  $\mu \in M([0, 1])$ . Mittels

$$\mu \mapsto \left( \phi \in C^0([0, 1], \mathbb{R}^{2n}) \mapsto \int_0^1 \phi(x) d\mu(x) \right)$$

#### 4. Optimalitätsbedingungen

lässt sich jedem  $\mu$  eine Distribution  $T_\mu(\xi)$  zuordnen. Also erzeugt das Radon-Maß  $\mu$  die Distribution

$$T_\mu(\phi) = \int_0^1 \phi(x) d\mu(x) = \langle \mu, \phi \rangle_{Z^*, Z}$$

für Testfunktionen  $\phi \in Z$  ([4], S. 344; [5], S. 203f., 217). Mithilfe dieser Distributionen formen wir

$$\langle e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi), p \rangle_{Z, Z^*}$$

für ein Radon-Maße  $p \in Z^*$  und  $e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi) \in Z$  zu

$$\int_0^1 e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi(t)) dp(t)$$

um. Die kommenden Rechnungen sind für Maße formal nicht richtig. Wir nehmen an, dass das Radon-Maß  $p$  eine differenzierbare Funktion ist. Später werden wir sehen, dass wir  $p$  tatsächlich so auffassen können.

$$\begin{aligned} & \int_0^1 e_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi(t)) \cdot p(t) dt \\ &= \int_0^1 \left[ \begin{pmatrix} \xi_r'(t) \\ \xi_s'(t) \end{pmatrix} - \begin{pmatrix} \xi_s(t) \\ \frac{\partial}{\partial \tilde{r}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \xi_r(t) + \frac{\partial}{\partial \tilde{s}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot \xi_s(t) \end{pmatrix} \right] \begin{pmatrix} p_r(t) \\ p_r(t) \end{pmatrix} dt \\ &= \int_0^1 \begin{pmatrix} \xi_r'(t) \\ \xi_s'(t) \end{pmatrix}^T \begin{pmatrix} p_r(t) \\ p_r(t) \end{pmatrix} dt - \int_0^1 \begin{pmatrix} p_r(t) \\ p_r(t) \end{pmatrix}^T \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}(t)} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}(t)} \end{pmatrix} \begin{pmatrix} \xi_r(t) \\ \xi_s(t) \end{pmatrix} dt \end{aligned}$$

Definiere  $M := \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}(t)} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}(t)} \end{pmatrix}$  und wende partielle Integration auf das erste Integral an.

$$[\xi(t) \cdot p(t)]_0^1 - \int_0^1 \xi(t) \cdot p'(t) dt + \int_0^1 M^T(t)p(t) \cdot \xi(t) dt$$

Da  $\xi$  aus  $\tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$ , verschwindet  $\xi$  im Punkt Null.

$$\xi(1)p(1) - \int_0^1 [p'(t) + M^T(t)p(t)] \cdot \xi(t) dt$$

Für das Zielfunktional gilt:

$$\begin{aligned} J_{(\tilde{r}(t), \tilde{s}(t))}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\xi(t)) &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot \xi_r(t_l) \\ &= \sum_{l=1}^K \begin{pmatrix} \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} \xi_r(t_l) \\ \xi_s(t_l) \end{pmatrix} \end{aligned}$$

4.1. für eine allgemeine Riemannsche Mannigfaltigkeit

Die Delta-Distribution ist ein stetiges lineares Funktional  $\hat{\delta} : C^0([0, 1], \mathbb{R}^{2n}) \rightarrow \mathbb{R}$  mit  $\hat{\delta}(f) = f(0)$ . Sie hat die Faltungseigenschaft  $f(a) = \int_0^1 \delta(x-a)f(x)dx$ . Weiter gelten  $\delta(x-a) = \infty$  für  $x = a$  und  $\delta(x-a) = 0$  für  $x \neq a$  ([5], S. 204). Da  $\xi \in C^0([0, 1], \mathbb{R}^{2n})$ , sind sowohl  $\xi(t_l) = \int_0^1 \delta(t-t_l)\xi(t)dt$  als auch  $\delta(t-t_l) = \infty$  für  $t = t_l$  und  $\delta(t-t_l) = 0$  für  $t \neq t_l$  erfüllt. Zusätzlich definieren wir  $DZ(t_l) := \begin{pmatrix} \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \\ 0 \end{pmatrix}$ .

Also können wir schreiben:

$$\int_0^1 \sum_{l=1}^K DZ(t_l) \cdot \xi(t) \cdot \delta(t-t_l) dt$$

Fassen wir linke und rechte Seite der Bedingung zusammen, ergibt sich:

$$\xi(1)p(1) - \int_0^1 [p'(t) + M^T(t)p(t)] \cdot \xi(t) dt = - \int_0^1 \sum_{l=1}^K DZ(t_l) \cdot \xi(t) \cdot \delta(t-t_l) dt$$

für jedes  $\xi \in \tilde{C}_0^1([0, 1], \mathbb{R}^{2n})$ .

Daraus folgt:

$$p'(t) + M^T(t)p(t) - \sum_{l=1}^K DZ(t_l) \cdot \delta(t-t_l) = 0,$$

$$p(1) = 0.$$

Dies bedeutet, dass die adjungierte Gleichung für  $t \neq t_l$  von der Form

$$p'(t) + M^T(t)p(t) = 0,$$

$$p(1) = 0$$

ist. Für  $t = t_l$  wird sie zu

$$p'(t_l) + M^T(t_l)p(t_l) - \sum_{l=1}^K DZ(t_l) \cdot \delta(0) = 0,$$

$$p(1) = 0.$$

In anderen Worten heißt dies, dass die Lösung  $p(t)$  der Differentialgleichung an den Stellen  $t_l$  für  $l = 1, \dots, K$  einen Sprung macht.

Die **3. Karush-Kuhn-Tucker-Bedingung** ist gegeben durch

$$\langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) + e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t), \begin{pmatrix} u_r \\ u_s \end{pmatrix} - \begin{pmatrix} \hat{r} \\ \hat{s} \end{pmatrix} \rangle \geq 0 \quad \forall \begin{pmatrix} u_r \\ u_s \end{pmatrix} \in \mathbb{R}^{2n}.$$

Da sowohl  $(u_r, u_s)^T \in \mathbb{R}^{2n}$  als auch  $(\hat{r}, \hat{s})^T \in \mathbb{R}^{2n}$  können wir  $(u_r, u_s)^T$  als zwei entgegengesetzte Richtungen schreiben:

#### 4. Optimalitätsbedingungen

- Für  $(u_r, u_s)^T := (\hat{r}, \hat{s})^T + z$  gilt  
 $\langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) + e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t), z) \rangle \geq 0$ .
- Für  $(u_r, u_s)^T := (\hat{r}, \hat{s})^T - z$  gilt  
 $\langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) + e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t), z) \rangle \leq 0$ .

Zusammengefasst folgt daraus:

$$\langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) + e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^*(p(t)), z)_{U^*, U} = 0 \text{ für alle } z \in U = \mathbb{R}^{2n}.$$

Nun werden Eigenschaften der dualen Paarung angewandt:

$$\begin{aligned} \langle J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}), z \rangle &= J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(z) \text{ bildet das Gâteaux-Differential von } J. \\ \langle e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^*(p), z)_{U^*, U} &= \langle p, e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(z) \rangle_{Z^*, Z} \end{aligned}$$

Indem wir die Radon-Maße als differenzierbare Funktionen auffassen, formen wir die letzte Gleichung, analog zu der Adjungierten Gleichung, um zu:

$$\begin{aligned} &\langle e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(z), p(t) \rangle_{Z, Z^*} \\ &= - \int_0^1 \left( \frac{\partial}{\partial \tilde{r}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot z_r + \frac{\partial}{\partial \tilde{s}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \cdot z_s \right)^T \begin{pmatrix} p_r(t) \\ p_s(t) \end{pmatrix} dt \\ &= - \int_0^1 \begin{pmatrix} p_r(t) \\ p_s(t) \end{pmatrix}^T \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}} \end{pmatrix} \begin{pmatrix} z_r \\ z_s \end{pmatrix} dt \end{aligned}$$

Wenn wir nun  $N := \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}} \end{pmatrix}$  definieren, dann wissen wir, dass  
 $\langle z, e_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})^* p(t) \rangle_{U, U^*} = - \int_0^1 N^T(t) p(t) \cdot z dt$  gilt.

Für das Gâteaux-Differential der Zielfunktion angewendet auf  $z$  erhalten wir:

$$\begin{aligned} J_{(\hat{r}, \hat{s})}(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(z) &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot z_r \\ &= \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \right) \cdot \begin{pmatrix} z_r \\ z_s \end{pmatrix}. \end{aligned}$$

Letztendlich fassen wir beide Terme der Gleichung zusammen und erhalten für alle  $z \in \mathbb{R}^{2n}$ :

$$- \int_0^1 N^T(t) p(t) dt \cdot z + \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \right) dt \cdot z = 0.$$

Demnach hat die 3. KKT-Bedingung die Form

$$\int_0^1 N^T(t)p(t) dt - \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \cdot D_2 \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \right) dt = 0.$$

Zusammengefasst haben wir die **Optimalitätsbedingungen** in folgender Gestalt aufgestellt:

1. 
$$\begin{pmatrix} \tilde{r}(t) \\ \tilde{s}(t) \end{pmatrix}' - \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix} = 0$$
2. 
$$p'(t) + M^T(t)p(t) - \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \cdot D_2 \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \right) \cdot \delta(t - t_l) = 0$$
  
$$p(1) = 0$$
3. 
$$\int_0^1 N^T(t)p(t) dt - \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \cdot D_2 \underset{0}{\text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r})} \right) dt = 0$$

Dazu sind  $M := \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}(t)} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}(t)} \end{pmatrix}$ ,  $N := \begin{pmatrix} 0 & I_n \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{r}} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{s}} \end{pmatrix}$

und  $g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) := - \left( \sum_{i,j=1}^n \Gamma_{ij}^k (\tilde{r}_1(t) + \hat{r}_1, \dots, \tilde{r}_n(t) + \hat{r}_n) \cdot (\tilde{s}_i(t) + \hat{s}_i) \cdot (\tilde{s}_j(t) + \hat{s}_j) \right)_{k=1, \dots, n}$  definiert.

## 4.2. Optimalitätsbedingungen am Beispiel der Sphäre

Wir werden, analog zu der allgemeinen Riemannschen Mannigfaltigkeit, das Optimalitätsproblem und die zugehörigen Karush-Kuhn-Tucker-Bedingungen für die  $S^2$ -Sphäre aufstellen. Gegeben haben wir die gewöhnliche Differentialgleichung zweiter Ordnung

$$\begin{aligned} 0 &= u_1''(t) + \cos u_1(t) \cdot \sin u_1(t) \cdot (u_2'(t))^2 \\ 0 &= u_2''(t) - 2 \tan u_1(t) \cdot u_1'(t) \cdot u_2'(t) \end{aligned}$$

mit den Anfangsbedingungen  $u(0) = f^{-1}(p)$  und  $u'(0) = f^{-1}(p)$ . Für  $u : [0, 1] \rightarrow \mathbb{R}^2$  definieren wir:

$$\begin{aligned} u(0) &=: \hat{r} \\ u'(0) &=: \hat{s} \\ \tilde{r}(t) &:= u(t) - \hat{r}, \quad \tilde{r}(0) = 0 \\ \tilde{s}(t) &:= u'(t) - \hat{s}, \quad \tilde{s}(0) = 0 \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &:= \begin{cases} -\cos(\tilde{r}_1(t) + \hat{r}_1) \cdot \sin(\tilde{r}_1(t) + \hat{r}_1) \cdot (\tilde{s}_2(t) + \hat{s}_2)^2 \\ 2 \tan(\tilde{r}_1(t) + \hat{r}_1) \cdot (\tilde{s}_1(t) + \hat{s}_1) \cdot (\tilde{s}_2(t) + \hat{s}_2) \end{cases} \end{aligned}$$

#### 4. Optimalitätsbedingungen

Weiter seien  $\tilde{r}(t), \tilde{s}(t) \in \tilde{C}_0^1([0, 1], \mathbb{R}^2)$  und  $\hat{r}, \hat{s} \in \mathbb{R}^2$ . Mithilfe des gerade Aufgestellten formen wir die Nebenbedingung um und erhalten Folgendes:

$$\begin{aligned} e : \tilde{C}_0^1([0, 1], \mathbb{R}^4) \times \mathbb{R}^4 &\rightarrow C^0([0, 1], \mathbb{R}^4) \\ e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &= \begin{pmatrix} \tilde{r}'_1(t) - \tilde{s}_1(t) - \hat{s}_1 \\ \tilde{r}'_2(t) - \tilde{s}_2(t) - \hat{s}_2 \\ \tilde{s}'_1(t) + \cos(\tilde{r}_1(t) + \hat{r}_1) \cdot \sin(\tilde{r}_1(t) + \hat{r}_1) \cdot (\tilde{s}_2(t) + \hat{s}_2)^2 \\ \tilde{s}_2(t) - 2 \tan(\tilde{r}_1(t) + \hat{r}_1) \cdot (\tilde{s}_1(t) + \hat{s}_1) \cdot (\tilde{s}_2(t) + \hat{s}_2) \end{pmatrix} \\ &= \begin{pmatrix} \tilde{r}(t) \\ \tilde{s}(t) \end{pmatrix}' - \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix}. \end{aligned}$$

Für die Riemannsche Distanz auf der Sphäre setzen wir

$$\begin{aligned} dist(x, y) &:= \sqrt{g_x(x - y, x - y)} = \sqrt{\tilde{g}_{f(x)}(Df|_x \cdot (x - y), Df|_x \cdot (x - y))} \\ &= \|Df|_x \cdot (x - y)\|_2, \end{aligned}$$

wobei  $Df|_x$  die Jakobi-Matrix der Parametrisierung  $f$  an der Stelle  $x$  bezeichnet und  $x, y$  zweidimensionale Vektoren sind. Mithilfe der Gleichung  $\sin^2 \alpha + \cos^2 \alpha = 1$  für ein beliebiges  $\alpha$  aus  $\mathbb{R}$  folgt

$$\begin{aligned} dist(x, y) &= \left\| \begin{pmatrix} -\sin x_1 \cos x_2 & -\cos x_1 \sin x_2 \\ -\sin x_1 \sin x_2 & \cos x_1 \cos x_2 \\ \cos x_1 \end{pmatrix} \cdot \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix} \right\|_2 \\ &= \left\| \begin{pmatrix} -(\sin x_1 \cos x_2(x_1 - y_1) + \cos x_1 \sin x_2(x_2 - y_2)) \\ -\sin x_1 \sin x_2(x_1 - y_1) + \cos x_1 \cos x_2(x_2 - y_2) \\ \cos x_1(x_1 - y_1) \end{pmatrix} \right\|_2 \\ &= \sqrt{(x_1 - y_1)^2 + \cos^2(x_1 - y_1) \cdot (x_2 - y_2)^2}. \end{aligned}$$

Setzen wir die Messwerte  $y_i^l$  mit  $y_i^l = (y_{i,1}^l, y_{i,2}^l)^T \in \mathbb{R}^2$  für  $x$  und  $\tilde{r}(t_l) + \hat{r}$  für  $y$  ein, so hat die Zielfunktion die Form

$$\begin{aligned} J : \tilde{C}_0^1([0, 1], \mathbb{R}^4) \times \mathbb{R}^4 &\rightarrow \mathbb{R} \\ J((\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})) &= \sum_{l=1}^K \sum_{i=1}^N \left( (y_{i,1}^l - \tilde{r}_1(t_l) - \hat{r}_1)^2 + \cos^2(y_{i,1}^l - \tilde{r}_1(t_l) - \hat{r}_1) \cdot (y_{i,2}^l - \tilde{r}_2(t_l) - \hat{r}_2)^2 \right). \end{aligned}$$

Unter den oben aufgestellten Formeln für Zielfunktion und Nebenbedingung, ist das **Optimalitätsproblem**:

$$\begin{aligned} \min_{(\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})} & J((\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})), \quad \text{sodass } e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) = 0 \\ & \text{für } \tilde{r}(t), \tilde{s}(t) \in C_0^1([0, 1], \mathbb{R}^2) \text{ und } \hat{r}, \hat{s} \in \mathbb{R}^2 \text{ gilt.} \end{aligned}$$



Wir differenzieren die Funktion  $g$ , welche Teil der Nebenbedingung ist.

$$\begin{aligned} \frac{\partial}{\partial \tilde{r}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &= \frac{\partial}{\partial \hat{r}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \\ &= \begin{pmatrix} [\sin^2(\tilde{r}_1(t) + \hat{r}_1) - \cos^2(\tilde{r}_1(t) + \hat{r}_1)](\tilde{s}_2 + \hat{s}_2)^2 & 0 \\ 2 \cos^{-2}(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_1 + \hat{s}_1)(\tilde{s}_2 + \hat{s}_2) & 0 \end{pmatrix} \\ \frac{\partial}{\partial \tilde{s}(t)} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) &= \frac{\partial}{\partial \hat{s}} g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \\ &= \begin{pmatrix} 0 & -2 \cos(\tilde{r}_1(t) + \hat{r}_1) \sin(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_2 + \hat{s}_2) \\ 2 \tan(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_2 + \hat{s}_2) & 2 \tan(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_1 + \hat{s}_1) \end{pmatrix} \end{aligned}$$

Die Gâteaux-Ableitungen der Nebenbedingung sind demnach

$$\begin{aligned} \delta_{(\tilde{r}(t), \tilde{s}(t))} e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_r(t), \varphi_s(t)) &= \begin{pmatrix} \varphi_r'(t) \\ \varphi_s'(t) \end{pmatrix} \begin{pmatrix} 0 & I_2 \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}(t)} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}(t)} \end{pmatrix} \begin{pmatrix} \varphi_r(t) \\ \varphi_s(t) \end{pmatrix}, \\ \delta_{(\hat{r}, \hat{s})} e(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_r, \omega_s) &= - \begin{pmatrix} 0 & I_2 \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{r}} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{s}} \end{pmatrix} \begin{pmatrix} \omega_r \\ \omega_s \end{pmatrix}, \end{aligned}$$

wobei  $(\varphi_r(t), \varphi_s(t)) \in \tilde{C}_0^1([0, 1], \mathbb{R}^4)$  und  $(\omega_r, \omega_s) \in \mathbb{R}^4$ . Um die Gâteaux-Differentiale für die Zielfunktion aufstellen zu können, differenzieren wir zuerst die Distanzfunktion nach ihrer zweiten Komponente. Es gilt

$$\begin{aligned} D_2 \text{dist}(x, y) &= \begin{pmatrix} \frac{\partial \text{dist}(x, y)}{\partial y_1} & \frac{\partial \text{dist}(x, y)}{\partial y_2} \end{pmatrix}^T \\ &= \frac{-1}{\sqrt{(x_1 - y_1)^2 + \cos^2(x_1 - y_1) \cdot (x_2 - y_2)^2}} \cdot \begin{pmatrix} (x_1 - y_1) - \cos(x_1 - y_1) \sin(x_1 - y_1)(x_2 - y_2)^2 \\ \cos^2(x_1 - y_1)(x_2 - y_2) \end{pmatrix} \end{aligned}$$

für die zweidimensionalen Vektoren  $x$  und  $y$ . Somit folgen schließlich die Gâteaux-Ableitungen:

$$\begin{aligned} \delta_{\tilde{r}(t)} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_r(t)) &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot \varphi_r(t_l) \\ &= -2 \sum_{l=1}^K \sum_{i=1}^N \begin{pmatrix} (y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1 - \cos(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1) \sin(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1)(y_{i,2}^l - \tilde{r}_2(t) - \hat{r}_2)^2 \\ \cos^2(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1)(y_{i,2}^l - \tilde{r}_2(t) - \hat{r}_2) \end{pmatrix} \cdot \varphi_r(t), \\ \delta_{\tilde{s}(t)} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\varphi_s(t)) &= 0, \\ \delta_{\hat{r}} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_r) &= \sum_{l=1}^K \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot \varphi_r(t_l) \\ &= -2 \sum_{l=1}^K \sum_{i=1}^N \begin{pmatrix} (y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1 - \cos(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1) \sin(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1)(y_{i,2}^l - \tilde{r}_2(t) - \hat{r}_2)^2 \\ \cos^2(y_{i,1}^l - \tilde{r}_1(t) - \hat{r}_1)(y_{i,2}^l - \tilde{r}_2(t) - \hat{r}_2) \end{pmatrix} \cdot \omega_r, \\ \delta_{\hat{s}} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})(\omega_s) &= 0. \end{aligned}$$

#### 4. Optimalitätsbedingungen

Mit den berechneten Gâteaux-Ableitungen der Zielfunktion und der Nebenbedingung stellen wir nun, analog zu der allgemeinen Riemannschen Mannigfaltigkeit, die Karush-Kuhn-Tucker-Bedingungen für die  $S^2$ -Sphäre auf.

Wir haben für die Sphäre

$$M = \begin{pmatrix} 0 & I_2 \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{r}(t)} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \tilde{s}(t)} \end{pmatrix}$$

und

$$N = \begin{pmatrix} 0 & I_2 \\ \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{r}} & \frac{\partial g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})}{\partial \hat{s}} \end{pmatrix}$$

berechnet. Die zugehörigen partiellen Ableitungen von  $g$  sind

$$\frac{\partial}{\partial \tilde{r}(t)} g = \frac{\partial}{\partial \hat{r}} g = \begin{pmatrix} [\sin^2(\tilde{r}_1(t) + \hat{r}_1) - \cos^2(\tilde{r}_1(t) + \hat{r}_1)](\tilde{s}_2 + \hat{s}_2)^2 & 0 \\ 2 \cos^{-2}(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_1 + \hat{s}_1)(\tilde{s}_2 + \hat{s}_2) & 0 \end{pmatrix}$$

und

$$\frac{\partial}{\partial \tilde{s}(t)} g = \frac{\partial}{\partial \hat{s}} g = \begin{pmatrix} 0 & -2 \cos(\tilde{r}_1(t) + \hat{r}_1) \sin(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_2 + \hat{s}_2) \\ 2 \tan(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_2 + \hat{s}_2) & 2 \tan(\tilde{r}_1(t) + \hat{r}_1)(\tilde{s}_1 + \hat{s}_1) \end{pmatrix}.$$

Vereinfacht durch diese Definitionen sind die **Optimalitätsbedingungen** die Folgenden:

1. Zustandsgleichung:

$$\begin{pmatrix} \tilde{r}(t) \\ \tilde{s}(t) \end{pmatrix}' - \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix} = 0$$

2. Adjungierte Gleichung:

$$p'(t) + M^T(t)p(t) - \sum_{l=1}^K \begin{pmatrix} \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \\ 0 \end{pmatrix} \cdot \delta(t - t_l) = 0$$

$$p(1) = 0$$

3. Adjungierte Gradientendarstellung:

$$\int_0^1 N^T(t)p(t) dt - \sum_{l=1}^K \begin{pmatrix} \sum_{i=1}^N 2 \cdot \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \text{dist}(y_i^l, \tilde{r}(t_l) + \hat{r}) \\ 0 \end{pmatrix} = 0$$

## 5. Gradientenabstieg

Bisher haben wir die Karush-Kuhn-Tucker-Bedingungen für ein Problem der optimalen Steuerung aufgestellt. Diese werden wir nutzen, um das Minimum des Problems zu finden. Dazu wählen wir ein Abstiegsverfahren, welches als Abstiegsrichtung den Gradienten des reduzierten Zielfunktional hat. Dieses Verfahren heißt Gradientenabstieg. Im Folgenden orientieren wir uns an dem Buch "Optimization with PDE Constraints" von M. Hinze et al ([6], S. 59, 99, 101, 105).

Wir nutzen die Adjungierte Methode, um den Gradienten des reduzierten Zielfunktional aufzustellen. Dabei gehen wir wie folgt vor:

Nach dem Aufstellen der Zustandsgleichung in Form einer gewöhnlichen Differentialgleichung in Kapitel 3 haben wir ein vollständiges Optimierungsproblem, nämlich:

$$\min_{(\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}), \text{ sodass } \begin{cases} \tilde{r}'(t) = \tilde{s}(t) + \hat{s} \\ \tilde{s}'(t) = g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{cases}$$

für  $(\tilde{r}(t), \tilde{s}(t)) \in \tilde{C}_0^1([0, 1], \mathbb{R}^4)$  und  $(\hat{r}, \hat{s}) \in \mathbb{R}^4$ .

Lösen wir die Zustandgleichung nach dem Zustand  $(\tilde{r}(t), \tilde{s}(t))$  auf, so erhalten wir für den Zustand eine Funktion  $F$ , die nur von der Kontrolle  $(\hat{r}, \hat{s})$  abhängt. Durch Einsetzen von  $(\tilde{r}(t), \tilde{s}(t)) = F(\hat{r}, \hat{s})$  wird die reduzierte Zielfunktion aufgestellt:

$$J(F(\tilde{r}(t), \tilde{s}(t)), (\hat{r}, \hat{s})) = \hat{J}(\hat{r}, \hat{s}).$$

Damit hängt das neue Problem, das sogenannte reduzierte Optimierungsproblem, nur von der Kontrolle ab:

$$\min_{(\hat{r}, \hat{s})} J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \text{ für } (\hat{r}, \hat{s}) \in \mathbb{R}^4.$$

Nun wird  $F(\hat{r}, \hat{s})$  in die Adjungierte Gleichung eingesetzt:

$$p'(t) + M^T(\hat{r}, \hat{s}, t)p(t) - \sum_{l=1}^K \left( \sum_{i=1}^N 2 \cdot \underset{0}{\text{dist}}(y_i^l, F_1(\hat{r}) + \hat{r}) \cdot D_2 \underset{0}{\text{dist}}(y_i^l, F_1(\hat{r}) + \hat{r}) \right) \cdot \delta(t - t_l) = 0$$

$$p(1) = 0$$

Diese Differentialgleichung erster Ordnung mit Endwert wird nach  $p$  aufgelöst, indem sie abschnittsweise rückwärts gelöst wird. Wir haben

$$DZ(t_l) = \left( \sum_{i=1}^N 2 \cdot \underset{0}{\text{dist}}(y_i^l, \tilde{r}(t_l) + \hat{r}) \cdot D_2 \underset{0}{\text{dist}}(y_i^l, \tilde{r}(t_l) + \hat{r}) \right)$$

## 5. Gradientenabstieg

gesetzt. Es gilt:

$$\begin{aligned}
 p(t_K) &= p(1) = 0 \\
 p(t_{K-1}^+) &= p(1) - \int_1^{t_{K-1}^+} (-M^T(t)p(t)) dt \\
 p(t_{K-1}) &= \infty \\
 p(t_{K-1}^-) &= p(t_{K-1}^+) - \int_{t_{K-1}^+}^{t_{K-1}^-} DZ(t_{K-1})\delta(t - t_{K-1}^-) dt \\
 &= p(t_{K-1}^+) - DZ(t_{K-1}^-) \int_{t_{K-1}^+}^{t_{K-1}^-} \delta(t - t_{K-1}^-) dt = p(t_{K-1}^+) - DZ(t_{K-1}^-) \\
 p(t_{K-2}^+) &= p(t_{K-1}^-) + \int_{t_{K-1}^-}^{t_{K-2}^+} M^T(t)p(t) dt
 \end{aligned}$$

und so weiter bis  $p(t_1^+)$ . Daraus ergibt sich die folgende abschnittsweise Darstellung für die Differentialgleichung: Es gilt für  $l = K, \dots, 1$ :

$$\begin{aligned}
 p'(x) &= -M^T(x) \cdot p(x) \text{ für } t_{l-1}^+ \leq x \leq t_l^-, \\
 p(t_l^-) &= p(t_l^+) - DZ(t_l^-), \text{ wobei } p(t_K) = 0.
 \end{aligned}$$

Nach dem Lösen der Differentialgleichung nach dem Adjungierten Zustand  $p$  erhalten wir ihn als Funktion in Abhängigkeit von der Kontrolle:  $p = G(\hat{r}, \hat{s})$ . Schließlich werden der Zustand  $(\tilde{r}(t), \tilde{s}(t)) = F(\hat{r}, \hat{s})$  und der adjungierte Zustand  $p = G(\hat{r}, \hat{s})$  in die Adjungierte Gradientendarstellung eingesetzt. Diese Gleichung stellt den Gradienten des reduzierten Zielfunktional dar:

$$\nabla \hat{J}(\hat{r}, \hat{s}) = \int_0^1 N^T(t)p(t) dt - \sum_{l=1}^K DZ(t_l)$$

Dabei hängen sowohl die Matrizen  $N$  und  $DZ$  als auch der adjungierte Zustand  $p$  nur von der Kontrolle und Zeit ab.

Der Gradientenabstieg wird auch Verfahren des steilsten Abstiegs genannt, da in einem Punkt auf dem Graphen der negative Gradient dem steilsten Abstieg entspricht. Das Verfahren löst numerisch das Optimierungsproblem. Um es übersichtlich darzustellen, definieren wir  $u := (\hat{r}, \hat{s}) \in \mathbb{R}^{2n}$ . Die Abbildung 7 veranschaulicht das Vorgehen.

Der **Gradientenabstieg** beginnt in einem Startpunkt  $u_0 \in \mathbb{R}^{2n}$ . Jede Iteration ist in vier Schritte unterteilt. Für  $k = 0, 1, 2, 3, \dots$  gilt:

1. Falls  $\|\nabla \hat{J}(u_k)\| < \epsilon$  für  $\epsilon > 0$ , stoppe das Verfahren.
2. Setze  $s_k := -\nabla \hat{J}(u_k)$  als Abstiegsrichtung.

3. Berechne eine Schrittweite  $\sigma_k > 0$ , für die  $\hat{J}(u_k + \sigma_k \cdot s_k) < \hat{J}(u_k)$  erfüllt ist.
4. Setze  $u_{k+1} := u_k + \sigma_k \cdot s_k$ .

Zur Berechnung der Schrittweite wählen wir die **Armijo-Regel**. Für den Fall des Gradientenverfahrens sieht sie wie folgt aus:

Seien die Abstiegsrichtung  $\nabla \hat{J}(u_k)$  in  $u_k$  und eine Konstante  $\gamma \in (0, 1)$  gegeben. Dann wähle  $\sigma_k \in \{(\frac{1}{2})^n\}_{n \in \mathbb{N}_0}$  so maximal, dass  $\hat{J}(u_k + \sigma_k \cdot s_k) - \hat{J}(u_k) \leq \gamma \sigma_k \nabla \hat{J}(u_k)^T \nabla \hat{J}(u_k)$  gilt.

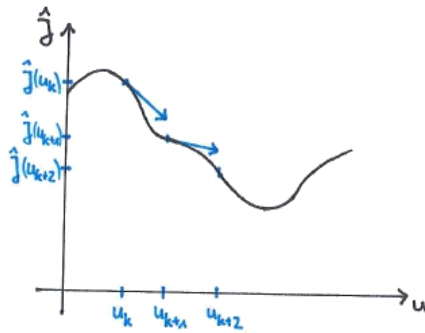


Abbildung 7: Gradientenabstieg

Letztendlich erreichen wir mittels Gradientenabstieg mit Armijo-Schrittweite das Minimum der Zielfunktion unter gegebener Nebenbedingung.

## 5.1. Implementierung in Matlab

Dieser Teilabschnitt erläutert, wie der Gradientenabstieg am Beispiel der  $S^2$ -Sphäre in der Software MATLAB implementiert wird. Der MATLAB-Code selbst ist Anhang A zu entnehmen.

Indem wir die Ergebnisse bezüglich der  $S^2$ -Sphäre aus den vorigen Kapiteln aufgreifen, werden die für den Gradientenabstieg benötigten Funktionen aufgestellt.

Für das Optimierungsproblem werden sowohl die Zielfunktion  $J(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s})$  (siehe Listing 14) als auch die Nebenbedingung in Form einer gewöhnlichen Differentialgleichung (siehe L. 7) benötigt:

$$\begin{pmatrix} \tilde{r}'(t) \\ \tilde{s}'(t) \end{pmatrix} = \begin{pmatrix} \tilde{s}(t) + \hat{s} \\ g(\tilde{r}(t), \tilde{s}(t), \hat{r}, \hat{s}) \end{pmatrix}.$$

Mittels eines MATLAB-Lösers für Differentialgleichungen erster Ordnung (ode45) wird die Nebenbedingung nach  $(\tilde{r}(t), \tilde{s}(t)) =: Y$  aufgelöst. Setzen wir  $Y$  in die Zielfunktion

## 5. Gradientenabstieg

$J$  ein (siehe L. 14), so erhalten wir das reduzierte Zielfunktion  $\hat{J}(\hat{r}, \hat{s})$  (siehe L. 12). Von diesem werden wir den Gradienten bilden (siehe L. 8). Um dies zu erreichen, lösen wir zuerst abschnittsweise die zweite Karush-Kuhn-Tucker-Bedingung nach dem adjungierten Zustand  $P$  auf. Dies erfolgt wieder mit dem ode45-Löser (siehe L. 8, Z. 24-37). Schließlich werden sowohl  $Y$  als auch  $P$  in die dritte Optimalitätsbedingung eingesetzt (siehe L. 8, Z. 61-78). Diese neue Formel entspricht dann dem Gradienten des reduzierten Zielfunktional und hängt nur noch von  $(\hat{r}, \hat{s})$  ab. Nun bilden wir eine Hilfsfunktion, die die reduzierte Zielfunktion und ihren Gradienten ausgibt (siehe L. 9). Damit sind alle für den Gradientenabstieg relevanten Funktionen aufgestellt.

An mehreren Stellen tritt das Problem auf, dass wir eine Kurve mit Werten zu verschiedenen Zeitpunkten haben und sie aufgrund der Anzahl der Zeitpunkte nicht mit einer Formel kompatibel ist. Die Formel benötigt dann eine Kurve zu weniger Zeitpunkten. Ziel ist es, eine neue Kurve zu den für die Formel benötigten Zeitpunkten zu erstellen. Dabei sollen die neuen Kurvenwerte einzelnen Werten der ursprünglichen Kurve entsprechen. Dafür nutzen wir die Nearest-Neighbor-Interpolation. Diese Methode weist einem Zeitpunkt der neuen Kurve den Wert der ursprünglichen Kurve zum nächstgelegenen Zeitpunkt zu.

Für das Programm (siehe L. 1) müssen nur Messwerte und ein Startpunkt eingegeben werden. Mit der vom mathematischen Institut vorgegebenen Funktion "descendSteepest" (siehe L. 3), die einen Gradientenabstieg darstellt, wird dann das Minimum ermittelt.

## 6. Fazit

Ziel der vorliegenden Arbeit war es, eine Verallgemeinerung der linearen Regression auf einer Riemannschen Mannigfaltigkeit durchzuführen. Zu diesem Zweck wurden anhand des Optimierungsproblems die Optimalitätsbedingungen aufgestellt. Dabei ergab sich, dass Radon-Maße als Funktionen aufgefasst wurden. Damit die Rechnungen formal richtig sind, ist es zu empfehlen, sich an dieser Stelle noch eingehender mit der Theorie über Radon-Maße zu beschäftigen.

Zur numerischen Ermittlung des Minimums wurde das Verfahren des Gradientenabstiegs zur Hilfe genommen. Schauen wir uns die Ergebnisse des Programms in MATLAB an, so stellen wir fest, dass sie nicht korrekt sein können. Dies liegt vermutlich an Fehlern im Code des Gradienten der reduzierten Zielfunktion  $\nabla \hat{J}$  (siehe L. 8). Dabei sind besonders die Zeilen 39-78 zu beachten, da MATLAB sowohl für den Zustand  $Y$  als auch für den adjungierten Zustand  $P$  richtige Ergebnisse liefert. Der Zustand wurde getestet, indem die Kurve  $\tilde{r}(t)$  um  $\hat{r}$  verschoben und von dieser neuen Kurve der Gradient gebildet wurde. Der Gradient, welcher der Geschwindigkeit der Geodätischen entspricht, ist im Verlauf der Zeitpunkte konstant. Des Weiteren wurden Tests für die Funktionen 2 und 4 durchgeführt, welche grundlegend für die Berechnung des adjungierten Zustands  $P$  sind. Zur Überprüfung der Funktion 8 kann mit einem Ableitungstest, der auf finiten Differenzen basiert, gearbeitet werden.

In Fällen, in denen eine Interpolation von Kurvendaten erforderlich war, wurde Nearest-Neighbor-Interpolation verwendet. Allerdings gehen bei dieser Interpolationsmethode Daten verloren. Da alle Rechnungen in euklidischen Räumen durchgeführt wurden, ließe sich die Nearest-Neighbor-Interpolation durch lineare Interpolation oder Interpolation höherer Ordnung ersetzen. Beide Methoden sind effizientere Alternativen.

In dieser Arbeit wurde eine lineare Regression auf eine Riemannsche Mannigfaltigkeit übertragen, indem Geraden durch Geodätische ersetzt wurden. Interessant wäre es, zu untersuchen, ob es noch andere Möglichkeiten gibt, die lineare Regression auf eine Mannigfaltigkeit zu übertragen, und ob dies vergleichbare Ergebnisse liefert.

Abschließend lässt sich sagen, dass die Arbeit für die gewählte Übertragungsart einen guten Überblick über lineare Regression auf Riemannschen Mannigfaltigkeiten liefert.

## A. Programmcode

Es folgen das für den Gradientenabstieg verwendete Programm und die dazugehörigen Funktionen. Sie wurden mit der Software MATLAB realisiert und durch Kommentare (mit % gekennzeichnet) ausführlich dokumentiert.

Das Programm sieht wie folgt aus:

Listing 1: Programm

```
1 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2 % Dieses Programm ist ein Regressionproblem auf der Sphre mittels
3 % Gradientenabstieg. Zu gegebenen Messwerten A und einem Startwert u wird
4 % das Minimum x ermittelt.
5 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
6 clear all
7 clc
8
9 % Geben Sie die Anzahl der quidistanten Zeitpunkte im Intervall [0,1] an,
10 % zu denen die Messungen durchgeführt werden.
11 K=10;
12
13 % Geben Sie die Anzahl der Messungen pro Zeitpunkt an.
14 N=3;
15
16 % Geben Sie die Messwerte in Form einer (N x K x 2)-Matrix ein.
17 A=randn(N,K,2);
18 % Nach Eingabe der Matrix A in geeigneter Form kann man Anzahl der
19 % Zeitpunkte K und Anzahl der Messungen N zu jedem Zeitpunkt auch durch
20 % [N,K,~]=size(A) ermitteln.
21
22 % Geben Sie den Startvektor u ein. Dabei mssen der 1. Eintrag aus dem
23 % Intervall (-pi/2,pi/2) und der 2. Eintrag aus (0,2pi) sein.
24 u=randn(1,4);
25 while u(2) == 0 %u(2)=phi aus der Parametrisierung f ist aus (0,2pi).
26     u = rand(1,4);
27 end
28
29 % Das Minimum x der reduzierten Zielfunktion Jhut wird numerisch mittels
30 % Gradientenabstieg berechnet. gradJhut stellt den Gradienten des
31 % reduzierten Zielfunktionals dar.
32 [x,e,de] = descendSteepest( @(x) Hilfsfunktion( u, A ), u, 50, true );
```



Die Funktionen sind alphabetisch nach dem Dateinamen aufgeführt:

Listing 2: Ableitung der Distanzfunktion

```
1 function z = D2dist( a,b )
2 % D2dist entspricht der Ableitung der Distanzfunktion dist nach dem 2.
3 % Argument.
4
5 z= [0,0];
6 if dist(a,b)~=0 %Falls dist=0, so ist D2dist nicht definiert.
7     z(1) = (-1/dist(a,b))*( (a(1)-b(1)) - cos((a(1)-b(1)))*sin((a(1)-b(1)))
8         *(a(2)-b(2))^2);
9     z(2) = (-1/dist(a,b))*(cos(a(1)-b(1)))^2*(a(2)-b(2));
10
11 end
```

Listing 3: ]Gradientenabstieg [9]

```
1 function [x,e,de] = descendSteepest( f, x, maxNumSteps, display )
2 % gradient descent iteration to minimise f(x)
3 % f needs to be a matlab function that takes a Matlab vector x and returns
4 % the scalar function value and its gradient as a Matlab vector,
5 % [functionValue,functionGradient] = f(x)
6 % the Matlab vector x is the starting point
7 % maxNumSteps is maximum number of iterations
8 % display=true displays information about current iterates
9 stepsize = 1;
10 [e,de] = f( x );
11 %% gradient descent iteration
12 for ii = 1:maxNumSteps
13     if display
14         fprintf( 'step %d, stepsize %f, energy %f, gradient-norm %e\n', ii -1,
15                 stepsize ,e ,norm(de,2) )
16     end
17     %% do descent step
18     normSqr = sum( de(:).^2 );
19     stepsize = min( stepsize , e / normSqr );
20     trialX = x - stepsize * de;
21     [trialE ,trialDE] = f( trialX );
22     %% Armijo linesearch
23     while ( trialE > e - .5 * stepsize * normSqr )
24         stepsize = stepsize / 2;
25         trialX = x - stepsize * de;
26         [trialE ,trialDE] = f( trialX );
27     end
28     x = trialX;
29     e = trialE;
30     de = trialDE;
31 end
```

## A. Programmcode

Listing 4: Ableitung von  $g$  aus  $gDgl$

```
1 function [ dy ] = Dg_gDgl( a,b,c,d )
2 %partielle Ableitung von g nach r^, s^
3 %a=r~, b=s~, c=r~, d=s^
4 %Dabei sind a, b KurvenPUNKTE zu einem bestimmten Zeitpunkt, nicht zu allen
5 %Zeitpunkten.
6 %a(1)+c(1)=r~_1+r^-1=u-1 in (-pi/2,pi/2) (offenes Intervall!)
7
8 dy=zeros(2,4);
9 %dg nach dr^
10 dy(1,1) = ( sin(a(1)+c(1))^2 - cos(a(1)+c(1))^2 ) * (b(2)+d(2))^2;
11 dy(1,2) = 0;
12 dy(2,1) = ( 2/(cos(a(1)+c(1))^2) ) * (b(1)+d(1)) * (b(2)+d(2));
13 dy(2,2) = 0;
14
15 %dg nach ds^
16 dy(1,3) = 0;
17 dy(1,4) = -2 * cos(a(1)+c(1)) * sin(a(1)+c(1)) * (b(2)+d(2));
18 dy(2,3) = 2 * tan(a(1)+c(1)) * (b(2)+d(2));
19 dy(2,4) = 2 * tan(a(1)+c(1)) * (b(1)+d(1));
20
21 end
```

Listing 5: Distanzfunktion

```
1 function [ res ] = dist( a,b )
2 % Die Riemannsche Distanz wurde wie folgt gewhlt:
3 % dist(a,b) = 2-Norm( Df|_a (a-b) )
4 % Dabei steht Df|_a fr die Ableitung der Parametrisierung f (der
5 % Mannigfaltigkeit) ausgewertet in a.
6 % size(a)=size(b)=[1, 2]
7
8 [~,n]=size(a);
9 [~,q]=size(b);
10 if n~=q
11     disp('Groesse_der_Vektoren_a_und_b_sind_nicht_kompatibel. ');
12 end
13
14 res = sqrt((a(1)-b(1))^2 + (cos(a(1)-b(1)))^2 * (a(2)-b(2))^2);
15
16 end
```

Listing 6: Funktion g aus gDgl

```

1 function [ z ] = g_gDgl( a,b,c,d )
2 % g_gDl beschreibt die Funktion g aus der Zustandsgleichung. Dabei
3 % entspricht g einer Summe, die unter anderem aus Christoffel-Symbolen
4 % besteht.
5
6 % [a, b]=[r~, s~]
7 % [c, d]=[r^, s^]
8 % Dabei sind a, b Kurvenpunkte zu einem bestimmten Zeitpunkt, nicht zu
9 % allen
10 % Zeitpunkten.
11
12 z=zeros(1,2);
13 z(1) = -cos(a(1)+c(1))*sin(a(1)+c(1))*(b(2)+d(2))^2;
14 z(2) = 2*tan(a(1)+c(1))*(b(1)+d(1))*(b(2)+d(2));
15
16 end

```

Listing 7: gewöhnliche Differentialgleichung

```

1 function [ dy ] = gDgl( t, y, rshut)
2 % gDgl entspricht der Zustandsgleichung e(y,u)=0 in Form einer
3 % gewöhnlichen Differentialgleichung.
4 %
5 % Fr den Zustand y gilt: y=[a, b]=[r~, s~].
6 % Fr die Kontrolle u gilt: u=[c, d]=[r^, s^].
7 % Dabei entspricht y einem Kurvenpunkt zu dem Zeitpunkt t.
8
9 a=y(1:2);
10 b=y(3:4);
11 c=rshut(1:2);
12 d=rshut(3:4);
13
14 % g_gDl beschreibt die Funktion g aus der Zustandsgleichung. Dabei
15 % entspricht g einer Summe, die unter anderem aus Christoffel-Symbolen
16 % besteht.
17 z = g_gDgl(a,b,c,d);
18
19 dy = zeros(4,1);
20 dy(1) = b(1)+d(1);
21 dy(2) = b(2)+d(2);
22 dy(3) = z(1);
23 dy(4) = z(2);
24
25 end

```

Listing 8: Gradient des reduzierten Zielfunktionals

```

1  function [ res ] = grad_RedZielfkt( A,Y,u )
2  % Der Gradient des reduzierten Zielfunktionals Jhut entspricht der
   Adjungierten
3  % Gradientendarstellung (3. KKT-Bedingung):
4  % grad_Jhut = int_0^1 N'(t) * p(t) dt - sum_l [sum_i 2*dist()*D2dist(); 0]
5  %
6  % A sind die Messwerte y_i^l zu den Zeitpunkten l=1,...,K.
7  % Y ist der Zustand in Abhngigkeit von der Kontrolle u zu den
8  % jeweiligen Zeitpunkten.
9  % u=[r^,s^].
10
11 % K ist die Anzahl der Zeitpunkte , zu denen es Messwerte gibt.
12 [~, K, ~] = size(A);
13
14 % F gibt die Feinheit der Diskretisierung der Zustandsgleichung e(y,u)=0 an
15 F = 4*K;
16
17 % S ist die Anzahl der Zeitpunkte des Teilintervalls (t_{l-1},t_l) von dem
18 % Intervall (0,1) mit der Unterteilung in K Zeitpunkte.
19 S=F/K;
20
21 % DZ = [sum_i 2*dist()*D2dist(); 0]. DZ hngt noch von der Zeit ab.
22 DZ=[ZielfktAbl(A,Y(:,1:2),u(1:2)), zeros(K,2) ];
23
24 % Der Adjungierte Zustand P aus der 2. KKT-Bedingung wird berechnet:
25 p0=zeros(K,4);
26 p0(1,:) =0;
27 [Z, P] = ode45(@(t,p) KKT2(t,p,Y,u,1,S), linspace(0,1/K,S), p0(1,:) ');
28 size(P);
29 GesamtP = zeros(S*K, 4);
30 GesamtP(1:S,:) = P;
31
32 for l=2:K
33     p0(l,:)=P(end,:)+DZ(l,:);
34     [Z, P] = ode45(@(t,p) KKT2(t,p,Y,u,l,S), linspace((l-1)/K,l/K,S), p0(l
        ,:)) ');
35     GesamtP((l-1)*S+1: l*S,:) = P;
36 end
37 GesamtP = flipud(GesamtP);
38
39 % Produkt aus N und P wird berechnet:
40 % In den Rechnungen entspricht die Matrix N der Matrix M aus der 2.
41 % KKT-Bedingung.
42 Prod = zeros(F,4);
43 for i=1:F
44     Prod(i,:) = MatrixM(Y(i,1:2), Y(i,3:4), u(1:2), u(3:4)) * GesamtP(i,:)
        ';
45 end
46
47 % Sucht passenden Index fr das Produkt N*P mittels Nchster-Nachbar
48 % Interpolation:

```

```

49 [m,~]=size(Prod);
50 ProdNeu = zeros(K,4);
51 for j=1:K
52     if ceil(m*j/K) == 0
53         ProdNeu(j,:) = Prod(1,:);
54     elseif ceil(m*j/K) > m
55         ProdNeu(j,:) = Prod(m,:);
56     else
57         ProdNeu(j,:) = Prod(ceil(m*j/K),:);
58     end
59 end
60
61 % 3. KKT-Bedingung wird aufgestellt, also der Gradient des reduzierten
62 % Zielfunktional.
63 % Das Integral  $\int_0^1 N'(t) * p(t) dt$  wird mittels Trapezregel
64 % approximiert:
65 Int = zeros(1,4);
66 for s=1:4
67     Int(s) = Trapezregel(ProdNeu(:,s),K,1,K);
68 end
69
70 % DZNeu =  $\sum_l [\sum_i 2*dist()*D2dist(); 0] = \sum_l DZ(t_l)$ . DZNeu ist
71 % unabhngig von der Zeit und hngt nur noch von der Kontrolle u ab.
72 DZNeu = zeros(1,4);
73 for l=1:K
74     DZNeu = DZ(l,:);
75 end
76
77 % Gradient des reduzierten Zielfunktional:
78 res = Int(:, :) - DZNeu(:, :);
79
80 end

```

Listing 9: Hilfsfunktion

```

1 function [fx, gradfx ] = Hilfsfunktion( x, A )
2 % Die Hilfsfunktion besteht aus dem reduziertem Zielfunktional Jhut und
3 % ihrem Gradienten.
4 %
5 % A sind die Messwerte  $y_i^l$  zu den Zeitpunkten  $l=1, \dots, K$ .
6 %  $x=[r^s]$  mit  $\text{size}(x) = [1,4]$ .
7
8 [fx,Y] = RedZielfkt( A, x ); % size(fx)=[1,1]
9 gradfx = grad_RedZielfkt( A, Y, x ); % size(gradfx)=[1,4]
10
11 end

```

Listing 10: 2. Karush-Kuhn-Tucker-Bedingung

```

1 function [ dp ] = KKT2( t, p, Y, u, ind, S )
2 % Gibt einen Teil aus der abschnittsweise definierten Differentialgleichung
3 % der 2. KKT-Bedingung an:
4 %  $p' = -M^T * p$ 
5 %
6 % Die 2. KKT ist die Dgl
7 %  $p' = -M^T * p + \sum_l (DZ(t_l) ; 0) * \delta(t-t_l)$  mit
8 % mit Endwert  $p(1)=0$ , wobei  $DZ(t_l) = \sum_i^2 \text{dist}() D2\text{dist}()$ .
9
10 % Nchster Nachbar Interpolation:
11 [m, ~] = size(Y);
12 offset = (ind -1) * S;
13 if ceil(t*m) == 0 && ind > 1
14     nearestIndex = offset;
15 elseif ceil(t*m) == 0 && ind == 1
16     nearestIndex = 1;
17 elseif offset + ceil(t * m) > m
18     nearestIndex = m;
19 else
20     nearestIndex = offset + ceil(t * m);
21 end
22
23 %  $dp = -M*p$ :
24 M=MatrixM(Y(nearestIndex ,1:2) , Y(nearestIndex ,3:4) , u(1:2) , u(3:4));
25 dp=zeros(4,1);
26 dp(1) = -( M(1)*p(1)+M(5)*p(2)+M(9)*p(3)+M(13)*p(4) );
27 dp(2) = -( M(2)*p(1)+M(6)*p(2)+M(10)*p(3)+M(14)*p(4) );
28 dp(3) = -( M(3)*p(1)+M(7)*p(2)+M(11)*p(3)+M(15)*p(4) );
29 dp(4) = -( M(4)*p(1)+M(8)*p(2)+M(12)*p(3)+M(16)*p(4) );
30
31 end

```

Listing 11: Matrix M

```

1 function [ M ] = MatrixM( a,b,c,d )
2 % MatrixM ist die transponierte Matrix  $M^T$  aus der 2. KKT-Bdeingung.
3 % Die Matrix N aus 3. KKT-Bedingung entspricht der Matrix M.
4
5 g=Dg_gDgl(a,b,c,d);
6 M=zeros(4,4);
7 M(1:2,1:2) = zeros(2,2);
8 M(1:2,3:4) = eye(2,2);
9 M(3:4,1:2) = g(:,1:2);
10 M(3:4,3:4) = g(:,3:4);
11 M = M';
12
13 end

```

Listing 12: Reduzierte Zielfunktion

```

1 function [ res , Y ] = RedZielfkt( A, u )
2 % Das reduziertes Zielfunktional Jhut(u) entspricht dem Zielfunktional J
3 % mit der Eigenschaft , dass es nur noch von der Kontrolle u abhngt.
4 %
5 % A sind die Messwerte  $y_i^l$  zu den Zeitpunkten  $l=1,\dots,K$ .
6 %  $u=[r^{\wedge},s^{\wedge}]$  mit  $\text{size}(u) = [1,4]$ .
7
8 % K ist die Anzahl der Zeitpunkte , zu denen es Messwerte gibt.
9 [~, K, ~] = size(A);
10
11 % F gibt die Feinheit der Diskretisierung der Zustandsgleichung  $e(y,u)=0$  an
12 F = 4*K;
13
14 % Die Zustandsgleichung (1. KKT-Bedingung)  $e(y,u)=0$  wird nach dem Zustand
15 %  $y=[r^{\wedge},s^{\wedge}]$  aufgelst. T gibt die Zeitpunkte an, zu denen die Gleichung
16 % gelst wird. Y ist der Zustand in Abhngigkeit von der Kontrolle u zu den
17 % jeweiligen Zeitpunkten. gDgl entspricht der Zustandsgleichung in Form
18 % einer gewhnlichen Differentialgleichung.
19 [T,Y] = ode45(@(t,y) gDgl(t,y,u), linspace(0,1,F), zeros(4,1));
20
21 % Der berechnete Zustand Y, der von der Kontrolle u abhngt, wird in die
22 % Zielfunktion J eingesetzt.
23 res = Zielfkt(A, Y(:,1:2), u(1:2));
24
25 end

```

Listing 13: Trapezregel

```

1 function [ res ] = Trapezregel( data, numT, a, b )
2 % Die Trapezregel approximiert ein Integral. Das Integral geht von a bis b.
3 %
4 % (Matlab beginnt immer bei 1 an zu zhlen , nicht schon bei 0.)
5 % data enthlt Daten der Kurve c, dabei sollte  $\text{data}(k,:) = c(t_k)$  sein.
6 % numT entspricht der Anzahl an Zeitpunkten t (d.h. c ist in  $a=t_1, \dots,$ 
7 %  $t_{\text{numT}}=b$  gegeben).
8
9 res = 0;
10 if numT > 1
11     S = 0;
12     for i=2:numT-1
13         S = S+data(i, :);
14     end
15     S = S+ (data(1, :)/2) + (data(end, :)/2);
16     res = S* ((b-a)/(numT-1));
17 end
18 end

```

Listing 14: Zielfunktion

```

1 function [ res ] = Zielfkt(A,B,d)
2 % Zielfkt entspricht der Zielfunktion J:
3 %  $J(y_{i^l}, r^{\sim}, r^{\wedge}) = \sum_l \sum_i \text{dist}(y(i,l), r^{\sim}(l) + r^{\wedge})^2$ .
4 %
5 % A sind die Messwerte  $y_{i^l}$  zu den Zeitpunkten  $l=1, \dots, K$ .
6 % Die Kurve  $B=r^{\sim}$  ist die erste Komponente aus dem Zustand y.
7 % Der Vektor  $d=r^{\wedge}$  ist die erste Komponente aus der Kontrolle u.
8
9 % n ~ K Anzahl der Zeitpunkte, zu denen gemessen wird.
10 % m ~ N Anzahl der Messungen pro Zeitpunkt.
11 % q ~ F Feinheit der Diskretisierung der Zustandsgleichung  $e(y,u)=0$ .
12 [m,n,~] = size(A);
13 [q,~] = size(B);
14
15 % Naechster Nachbar Interpolation:
16 % Produziert einen Vektor, welcher  $(r^{\sim}(t_0), r^{\sim}(t_1), \dots, r^{\sim}(t_n))$ 
17 % approximiert.
18 BNeu = zeros(n,2);
19 for k=1:n
20     if B(ceil(q*k/n),:) == 0
21         BNeu(k,:) = B(1+ceil(q*k/n),:);
22     else
23         BNeu(k,:) = B(ceil(q*k/n),:);
24     end
25
26 % Die Zielfunktion  $J(y_{i^l}, r^{\sim}, r^{\wedge})$  wird aufgestellt:
27 S=0;
28 for l=1:n
29     for i=1:m
30         S=S+(dist(A(i,l,:), BNeu(l,:)+d))^2;
31     end
32 end
33 res=S;
34 end

```

Listing 15: Ableitung der Zielfunktion

```

1 function [ res ] = ZielfktAbl( A,B,d )
2 % ZielfktAbl ist ein Teil aus dem Gteaux-Differential der Zielfunktion J:
3 %  $\sum_l \sum_i 2 * \text{dist}() * D2\text{dist}()$ 
4 %
5 % A sind die Messwerte  $y_{i^l}$  zu den Zeitpunkten  $l=1, \dots, K$ .
6 % Die Kurve  $B=r^{\sim}$  ist die erste Komponente aus dem Zustand y.
7 % Der Vektor  $d=r^{\wedge}$  ist die erste Komponente aus der Kontrolle u.
8
9 % n ~ K Anzahl der Zeitpunkte, zu denen gemessen wird.
10 % m ~ N Anzahl der Messungen pro Zeitpunkt.
11 % q ~ F Feinheit der Diskretisierung der Zustandsgleichung  $e(y,u)=0$ .
12 [m,n,~] = size(A);
13 [q,~] = size(B);
14
15 % Naechster Nachbar Interpolation:

```



```

16 % Produziert einen Vektor, welcher ( $\tilde{r}(t_0), \tilde{r}(t_1), \dots, \tilde{r}(t_n)$ )
17 % approximiert.
18 BNeu = zeros(n,2);
19 for k=1:n
20     if B(ceil(q*k/n),:) == 0
21         BNeu(k,:) = B(1+ceil(q*k/n),:);
22     else
23         BNeu(k,:) = B(ceil(q*k/n),:);
24     end
25 end
26
27 % Das Ergebnis der ZielfktAbl wird aufgestellt:
28 S=zeros(n,2);
29 for l=1:n
30     for i=1:m
31         S(l,1) = S(l,1) + (A(i,l,1)-BNeu(l,1)-d(1)) - cos(A(i,l,1)-BNeu(l
32             ,1)-d(1)) ...
33             * sin(A(i,l,1)-BNeu(l,1)-d(1)) * (A(i,l,2)-BNeu(l,2)-d(2))^2;
34         S(l,2) = S(l,2) + cos(A(i,l,1)-BNeu(l,1)-d(1))^2 * (A(i,l,2)-BNeu(l
35             ,2)-d(2));
36     end
37 end
38 res = -2 * S;
39
40 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
41 %Alternative Form ist die Folgende. Dabei ist der Definitionsbereich
42 %von D2dist zu beachten.
43 % for l=1:n
44 %     for i=1:m
45 %         S=S+(dist(A(i,l,:),B(l,:)+d) * D2dist(A(i,l,:),B(l,:)+d));
46 %     end
47 % end
48 % res = 2*S;
49 end

```

## B. Literaturverzeichnis

1. Bär, Christian (2001): Elementare Differentialgeometrie, Berlin/New York: Walter de Gruyter.
2. Bär, Christian (2013): Differential Geometry, Potsdam: Universität Potsdam, URL: <http://geometrie.math.uni-potsdam.de/documents/baer/skripte/skript-DiffGeo-engl.pdf> (Zugriff: 29.09.2015).
3. Berkels, B., Fletcher, P. T., Heeren, B., Rumpf, M., Wirth, B. (2013): 'Discrete geodesic regression in shape space'. In: Heyden, A. u. a. (Hrsg.): Energy Minimization Methods in Computer Vision and Pattern Recognition, Lund: Springer, S. 108-122.
4. Elstrodt, Jürgen (2009): Maß- und Integrationstheorie, 6. Auflage, Berlin/Heidelberg: Springer.
5. Forster, Otto (2012): Analysis 3, Maß- und Integrationstheorie, Integralsätze im  $\mathbb{R}^n$  und Anwendungen, 7. Auflage, Wiesbaden: Springer.
6. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S. (2009): Optimization with PDE Constraints, Cham: Springer.
7. Jänich, K. (2005): Vektoranalysis, 5. Auflage, Berlin/Heidelberg/New York: Springer.
8. Klingenberg, Wilhelm P. A. (2004): Klassische Differentialgeometrie, Eine Einführung in die Riemannsche Geometrie, Leipzig: Edition am Gutenbergplatz.
9. Abstiegsverfahren zur Optimierung, Münster: Institut für Numerische und Angewandte Mathematik, WWU Münster, URL: [http://dune-project.uni-muenster.de/git/projects/public-wiki/wiki/\\_Abstiegsverfahren\\_zur\\_Optimierung](http://dune-project.uni-muenster.de/git/projects/public-wiki/wiki/_Abstiegsverfahren_zur_Optimierung) (Zugriff: 29.09.2015).