

---

# Arbeitsgruppe Mathematische Optimierung

Bachelorarbeit

## Höherdimensionaler optimaler Transport mit Anwendung für die Klassifizierung von Daten

Lucas Plagwitz

Matr. 418502

Betreuer: Dr. Bernhard Schmitzer

Zweitgutachter: Prof Dr. Benedikt Wirth

# Inhaltsverzeichnis

<b>1. Optimaler Transport</b>	<b>3</b>
1.1. Grundlagen . . . . .	3
1.2. Weiterführende Definitionen . . . . .	7
1.3. Numerik . . . . .	10
<b>2. Erweiterung des optimalen Transport auf höhere Dimensionen</b>	<b>15</b>
2.1. Der Produktraum . . . . .	16
2.2. Diskretisierung . . . . .	22
2.3. Anwendungsbeispiele . . . . .	23
<b>3. Höherdimensionale Betrachtung des MNIST Datensatzes</b>	<b>26</b>
3.1. Vorbereitung . . . . .	26
3.2. Ergebnisse . . . . .	29
<b>4. Zusammenfassung und Ausblick</b>	<b>33</b>
<b>Anhang</b>	<b>34</b>
<b>Anhang A. Anhang</b>	<b>34</b>
A.1. Kullback-Leibler Divergenz . . . . .	34

# Einleitung

Die Suche des optimalen Transports, der auch diese Arbeit gewidmet ist, ist auf den französischen Mathematiker Gaspard Monge (1746-1818) zurückzuführen. Monge beschrieb 1783 eine mathematische Formulierung eines Transportproblems. Damals stellte sich die Frage, wie man gewonnenes Eisenerz aus verschiedenen Eisenminen am effizientesten auf verschiedene Fabriken verteilt. Nach Monges Arbeiten über dieses Problem gab es über die Jahre viele Mathematiker, die neben neuen Lösungsansätzen auch weitere Anwendungen fanden. So kann man neben logistischen Probleme auch Themen im Bereich der Bildverarbeitung, Bilderkennung, Interpolation sowie Machine-Learning betrachten.

Das Ziel dieser Arbeit besteht darin, den Leser in das Gebiet des höherdimensionalen optimalen Transports einzuführen, seine vielseitigen Eigenschaften zu demonstrieren und zum skalaren Fall zu vergleichen. Hierbei sei angemerkt, dass die Höherdimensionalität zusätzliche Information zu der skalaren Version betrachtet und keine einfache Erweiterung der Raumdimensionen darstellt. Unter Datenlifting (kurz Lifting), zu deutsch wohl das Anheben von Daten, versteht man das Erweitern der skalaren Version um weitere Dimensionen, die vorher nicht explizit betrachtet wurden. Ein Anwendungsbeispiel des optimalen Transports ist die Bildinterpolation zwischen zwei gegebenen Bildern  $A$  und  $B$ . Normalerweise würde man nun nur die zweidimensionale Raumstruktur betrachten, jedem Pixel ein Grauwert von 0 bis 1 (0=weiß, 1=schwarz) zuordnen und den optimalen Transport von Bild  $A$  zu Bild  $B$  bestimmen. Dies muss nicht immer den natürlichsten Transport beschreiben, welchen wir durch das Betrachten weiterer Dimensionen besser approximieren können, wie wir im Verlauf dieser Arbeit noch sehen werden.

Um dieses und weitere Ergebnisse zu zeigen, werden zuerst einige Definitionen und Erkenntnisse aus grundlegenden Theorien und Anwendungen des optimalen Transports benötigt, welche in Kapitel 1 beschrieben werden. In Kapitel 2 wird gezeigt, wie die Theorie

## Einleitung

---

auf höherdimensionale Daten erweitert werden kann und kurze Fallbeispiele betrachten. In Kapitel 3 werden die Erkenntnisse auf den berühmten MNIST Datensatz angewendet.

# 1 | Optimaler Transport

Bevor wir die Theorie auf zusätzliche Dimensionen erweitern, betrachten wir zunächst grundlegende Erkenntnisse rund um den optimalen Transport. Hierzu leiten wir aus der Monge Formulierung ähnliche Optimierungsansätze her, welche wir gegeneinander vergleichen werden, bevor wir sie in Kapitel 2 für geliftete Daten anpassen.

## 1.1 Grundlagen

In dieser Sektion führen wir grundlegende Begriffe wie den Pushforward eines Maßes unter einer Funktion, die Monge sowie die Kantorovich Formulierung ein, die essentiell für jede tiefere Auseinandersetzung mit der Theorie des optimalen Transport sind:

### 1.1.1. Monge-Formulierung

Wie bereits in der Einleitung erwähnt, geht die erste Formulierung der Optimierung eines Transports auf Monge im Jahr 1783 zurück. Zu ihrem Verständnis benötigen wir zuvor folgende Definition.

**Definition 1.1.1** (messbare Mengen, Pushforward). Seien  $(\Omega_0, P_0, \mu_0)$  und  $(\Omega_1, P_1, \mu_1)$  zwei Maßräume, so ist

1.  $T : \Omega_0 \rightarrow \Omega_1$  messbar, wenn  $T^{-1}(A_1) \in P_0 \ \forall A_1 \in P_1$
2. der Pushforward von  $\mu_0$  unter  $T$  definiert als

$$T_{\#}\mu_0 : P_0 \rightarrow (-\infty, \infty], (T_{\#}\mu_0)(A) := \mu_0(T^{-1}(A)) \text{ mit } A \in P_0$$

Im Folgenden werden wir mit  $\text{rca}(\Omega_0)$  die Menge aller Radonmaße auf dem messbaren Raum  $\Omega_0$ , sowie mit  $\text{rca}_1(\Omega_0) \subset \text{rca}(\Omega_0)$  die Menge aller Wahrscheinlichkeitsmaße auf  $\Omega_0$  definieren.

Als erstes sei angemerkt, dass wir  $T$  als Transportfunktion zwischen den Mengen  $\Omega_0$  und  $\Omega_1$  bezeichnen. Für die maßtheroretischen Definitionen, die wir benötigen sei für nähere Erläuterungen auf [FO12] hingewiesen. Wir konzentrieren uns auf den Pushforward und wollen diesen anhand eines Beispiels erläutern: Nehmen wir an, wir wollen aus einem Haufen Sand eine gewisse Form gestalten. In diesem Fall stellt die Dichtefunktion  $\rho_{\mu_0}$  die Umriss des Ausgangshaufens und  $\rho_{\mu_1}$  die Umriss der Zielform dar. Nun betrachtet der Pushforward das Maß des Urbilds einer Teilmenge unserer Zielverteilung unter der Transportfunktion.

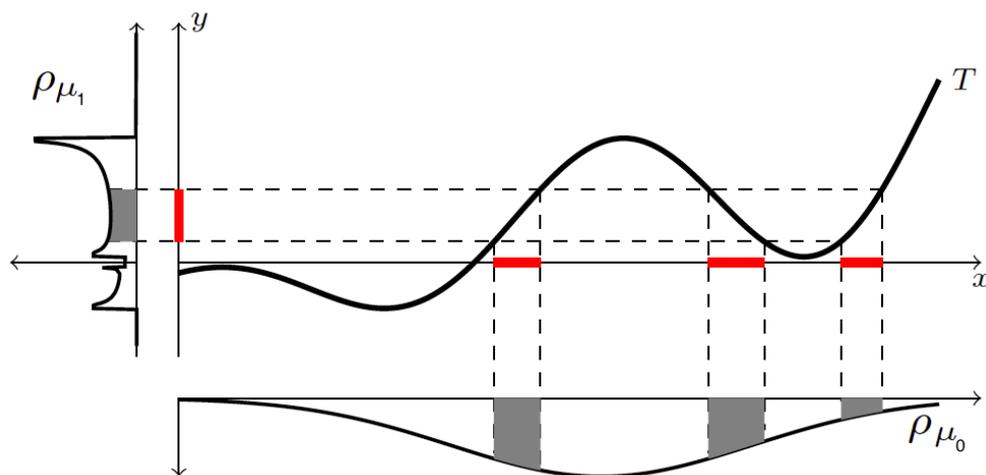


Abbildung 1.1.: Veranschaulichung des Pushforwards [TH16]

Links neben der  $y$ -Achse ist die Zielverteilung in Form der Dichtefunktion  $\rho_{\mu_1}$  des Maßes  $\mu_1$  gegeben. Unter der  $x$ -Achse ist die Ausgangsverteilung in Form des Haufens (Dichtefunktion  $\rho_{\mu_0}$  des Maßes  $\mu_0$ ) zu sehen. Die Transportfunktion  $T$  gibt uns die Information, wohin die Masse von  $\mu_0$  zu  $\mu_1$  transportiert wird (vgl. gestrichelte Linien). Der graue Bereich des  $\mu_1$  zeigt die Teilmenge, von der wir das Urbild unter der Funktion  $T$  betrachten (die roten Linien auf der  $x$ -Achse). Die graue Teilmenge unter  $\rho_{\mu_1}$  befand sich somit vorher in der grauen Fläche unter  $\rho_{\mu_0}$ .

Nach der Einführung des Pushforwards können wir so Monges Formulierung aufstellen (vgl. [SA15]):

**Definition 1.1.2** (Formulierung nach Monge). Sei eine Transportfunktion  $T$  wie in Definition 1.1.1, zwei Maße  $\mu_0, \mu_1$  mit  $\mu_1 = T_{\#}\mu_0$  und eine Kostenfunktion  $c : \Omega_0 \times \Omega_1 \rightarrow [0, \infty]$  gegeben, so ist das Problem nach Monge formuliert durch:

$$OT_M(\mu_0, \mu_1) := \inf \{M(T)\} \quad M(T) := \int_{\Omega_0} c(x, T(x))\mu_0(x)dx$$

Das Optimierungsproblem nach Monge sucht somit nach der günstigsten Transportabbildung  $T$  um bei Kosten  $c(x, T(x))$  von  $\mu_0$  nach  $\mu_1$  abzubilden, bzw. die Mongekosten  $M(T)$  zu minimieren. Die Bedingung  $\mu_1 = T_{\#}\mu_0$  garantiert uns, dass keine Masse beim Abbilden durch  $T$  verloren geht. Als Kostenfunktion wäre beispielsweise die euklidische Norm  $\|T(x) - x\|_2$  denkbar. Zusammenfassend integrieren wir also über jeden Massepartikel in  $\mu_0$  und seine Kosten auf dem Weg nach  $\mu_1$  und versuchen diese zu minimieren.

Monge gibt uns hier eine recht verständliche Formulierung an die Hand. Diese hat allerdings zwei Nachteile:

1. Massensplitting. Da wir mit  $T$  eine Abbildung definiert haben, können keine Massepartikel aufgeteilt werden, da sonst die Wohldefiniertheit unserer Abbildung verloren ginge. Schon bei Monges Ausgangsproblem, der Suche nach einer optimaler Verteilung von Erz zwischen Minen und Fabriken, erhalten wir so eine harte Einschränkung, da im Falle eines hohen Ertrags einer Mine trotzdem nur eine Fabrik beliefert werden darf.
2. Existenz. Es ist nicht möglich die Existenz einer solchen Transportfunktion  $T$  zu garantieren. (vgl.[SA15])

### 1.1.2. Kantorovich-Formulierung

Durch diese Nachteile motiviert, kommen wir nun zu der Formulierung nach Kantorovich, eine deutlich jüngere Theorie. Kantorovich erhielt 1975 den Nobelpreis für Wirtschaftswissenschaften für seine Verdienste rund um die optimale Zuordnung und den damit verbundenen optimalen Transport, der auf seine grundlegende Formulierung zurückgeht (vgl. [BH00]). In diesem Abschnitt beschreiben wir seine Herangehensweise.

**Definition 1.1.3** (Menge der Transportpläne). Gegeben seien zwei Maßräume  $(\Omega_0, \mathcal{P}_0, \mu_0)$  und  $(\Omega_1, \mathcal{P}_1, \mu_1)$ . Die Menge der zulässigen Transportpläne zwischen  $\mu_0$  und  $\mu_1$  definieren wir durch:

$$\Pi(\mu_0, \mu_1) = \{\pi \in \text{rca}(\Omega_0 \times \Omega_1) : \text{pr}_{i\#} \pi = \mu_i \text{ für } i \in \{0, 1\}\}$$

mit  $\text{pr}_i: (\Omega_0 \times \Omega_1) \rightarrow \Omega_i, (\mu_0, \mu_1) \mapsto \mu_i$  für  $i \in \{0, 1\}$

In der Kantorovich-Formulierung ist der Transportplan das Pendant zu der Transportabbildung von Monge. Ein Transportplan gibt an, wie viel Masse von jedem Punkt in  $\mu_0$  zu jedem Punkt in  $\mu_1$  wandert. Die Nebenbedingung, in der der Pushforward des Maßes  $\pi$  unter der jeweiligen Projektion betrachtet wird, erhält uns wieder die Masse zwischen den beiden Maßen  $\mu_0$  und  $\mu_1$ .

Mit Hilfe der Menge der Transportpläne können wir nun die Kantorovich-Formulierung aufstellen (vgl. [SA15]).

**Definition 1.1.4** (Formulierung nach Kantorovich). Seien zwei Maßräume  $(\Omega_0, \mathcal{P}_0, \mu_0)$  und  $(\Omega_1, \mathcal{P}_1, \mu_1)$  sowie  $\Pi(\mu_0, \mu_1)$ , die Menge der zulässigen Transportpläne von  $\mu_0$  nach  $\mu_1$ , gegeben, so ist der optimale Transport nach Kantorovich für eine Kostenfunktion  $c$  gegeben durch:

$$OT_K(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\Omega_0 \times \Omega_1} c(x, y) d\pi(x, y)$$

Hier suchen wir den optimalen Transportplan, der die Kosten jeder Verbindung aller Punkte in  $\mu_0$  nach  $\mu_1$  minimiert. Dafür minimieren wir das Integral über das Maß  $\pi$  jedes Paares von Massepunkten und die dazugehörigen Kosten.

Für die oben aufgezeigten beiden Nachteile der Monge Formulierung gilt für die Formulierung nach Kantorovich:

1. Massensplitting. Da wir nun durch den Transportplan angeben wie viel Masse von jedem Punkt in  $\Omega_0$  zu jedem Punkt in  $\Omega_1$  zugeordnet wird, ist das Massensplittingproblem nicht mehr vorhanden.

2. Existenz. Man kann zeigen, dass wenn  $\Omega_0$  und  $\Omega_1$  kompakt sind ein optimales  $\pi$  existiert, sodass  $OT_K(\mu_0, \mu_1)$  eindeutig gelöst werden kann. (vgl.[SA15])

## 1.2 Weiterführende Definitionen

Wir haben durch die Formulierungen nach Monge und Kantorovich alle wichtigen Grundlagen des optimalen Transports kennen gelernt. Da wir im späteren Verlauf aber etwas speziellere Formulierungen auf unsere höherdimensionalen Daten anwenden wollen, führen wir diese in den folgenden Abschnitten ein. Hierbei werden die Wasserstein Distanz und der Hellinger-Kantorovich Abstand dargestellt und gegeneinander verglichen.

### 1.2.1. Wasserstein Distanz

Der Name verrät schon, dass es sich hier um eine Metrik (zwischen zwei Maßen) handelt. So werden wir im weiteren Verlauf nur noch den optimalen Transport von zwei Maßen innerhalb des selben Maßraums betrachten, um so einen metrischen Raum zu erhalten. Die Wasserstein Distanz verallgemeinert die Kantorovich-Formulierung zu einer Metrik. Wir tauschen deshalb die Formulierung der Kosten  $c(x, y)$  durch die einer Metrik  $d(x, y)$  um später einheitliche Notationen gewährleisten zu können.

**Definition 1.2.1 (Wasserstein-Distanz).** Sei ein Maßraum  $(\Omega, P)$  mit zwei Maßen  $\mu_0, \mu_1$  sowie  $\Pi(\mu_0, \mu_1)$ , die Menge der zulässigen Transportpläne von  $\mu_0$  nach  $\mu_1$ , gegeben, so ist die  $p$ -Wasserstein Distanz definiert durch:

$$d_{W,\Omega}^p(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\Omega^2} d_{\Omega}^p(x, y) d\pi(x, y)$$

Auch hier suchen wir den günstigsten Plan  $\pi$ , mit gegebenen Kosten zu jedem Paar  $(x, y) \in \Omega^2$  von  $d_{\Omega}^p(x, y)$ . Wie schon erwähnt erhalten wir mit dieser Definition eine Metrik auf  $rca(\Omega)$  (siehe [SA15]). Im Folgenden werden wir uns auf die Betrachtung der 2-Wasserstein Distanz beschränken. Es ist anzumerken, dass die  $p$ -Wasserstein Distanz nur die Distanz zwischen zwei normalisierten Maßen angeben kann, da sonst  $\Pi(\mu_0, \mu_1) = \emptyset$

und  $d_{W,\Omega}(x, y) = +\infty$  gilt.

### 1.2.2. Hellinger-Kantorovich Abstand

Durch diesen Umstand kommen wir nun zur letzten Formulierung dieses Kapitels, der die Notwendigkeit der Normalisierung der Maße behebt, dem Hellinger-Kantorovich Abstand. Für den Hellinger-Kantorovich Abstand benötigen wir die Kullback-Leibler Divergenz.

**Definition 1.2.2** (Kullback-Leibler Divergenz). Es seien  $\mu_0$  und  $\mu_1$  zwei Maße auf einem Maßraum, sowie  $\rho_{\mu_0}$  und  $\rho_{\mu_1}$  deren Dichtefunktionen, so ist die Kullback-Leibler Divergenz von  $\mu_0$  und  $\mu_1$  definiert durch:

$$KL(\mu_0 || \mu_1) = \int_{-\infty}^{\infty} \rho_{\mu_0}(x) \cdot \log \left( \frac{\rho_{\mu_0}(x)}{\rho_{\mu_1}(x)} \right) dx$$

Die Kullback-Leibler Divergenz ist ein Maß, welches den Grad der Abweichung zweier Maße gegeneinander beschreibt. Für eine kurze Einführung wird auf Anhang A verwiesen.

Die Betrachtung des Hellinger-Kantorovich Abstands ist nun möglich.

**Definition 1.2.3** (Hellinger-Kantorovich Abstand). Sei  $\kappa > 0$ ,  $\mu_0$  und  $\mu_1$  wie zuvor zwei Maße eines Maßraums, so definiert sich der Hellinger-Kantorovich Abstand wie folgt:

$$d_{HK,\Omega}^2(\mu_0, \mu_1) = \kappa^2 \inf_{\pi \in rca(\Omega^2)} \left\{ \sum_{i=0}^1 KL(\text{pr}_{\Omega, i\#} \pi | \mu_i) + \int_{\Omega^2} c_{\Omega,\kappa}(x, y) d\pi(x, y) \right\}$$

mit Abstandsfunktion:

$$c_{X,\kappa}(x, y) := \begin{cases} -\log([\cos(\frac{d_{\Omega}(x,y)}{\kappa})]^2) & \text{für } d_{\Omega}(x, y) < \kappa \cdot \frac{\pi}{2} \\ +\infty & \text{sonst} \end{cases}$$

Dieser zuerst unhandlich erscheinende Abstand hat doch vieles mit der 2-Wasserstein-Distanz gemein. Der größte Unterschied zwischen beiden liegt darin, dass nun nicht mehr die Menge der Transportpläne betrachtet wird, bei der es wichtig war die Masse zu erhalten, sondern allgemeinere Radonmaße  $rca(\Omega^2)$ , die ein Ungleichgewicht zwischen den

Massen erlauben. Zwar kann so jedes nicht-negative Maß  $\pi \in \text{rca}(\Omega^2)$  genutzt werden, doch wird die Abweichung eingeschränkt durch die Kullback-Leibler Divergenz, die zu dem Abstand addiert wird. Für ein  $\pi \in \text{rca}_1(\Omega)$  würden die beiden KL-Terme verschwinden. Um ein Gefühl davon zu bekommen, wie unterschiedlich die Wasserstein-Distanz gegenüber dem Hellinger-Kantorovich Abstand arbeitet, betrachten wir die Geodäten, die die kürzeste Verbindung zwischen  $\mu_0$  und  $\mu_1$  im metrischen Raum  $(\text{rca}_1(\Omega), d_{W,\Omega})$  und  $(\text{rca}(\Omega), d_{HK,\Omega})$  beschreibt.

1. Wasserstein-Geodäte. Seien  $\mu_0, \mu_1 \in \text{rca}_1(\Omega)$  und  $\pi \in \text{rca}_1(\Omega^2)$  der optimale Transportplan für die Wasserstein Distanz. So gewinnen wir aus  $\pi$  eine Geodäte, die ein Massepartikel mit konstanter Masse und Geschwindigkeit von  $\mu_0$  nach  $\mu_1$  transportiert.
2. Hellinger-Kantorovich-Geodäte. Seien  $\mu_0, \mu_1 \in \text{rca}(\Omega)$  und  $\pi \in \text{rca}(\Omega^2)$  das optimale Transportmaß für den Hellinger-Kantorovich Abstand. Dann kann ein Massepartikel auf der Geodäte von  $\mu_0$  zu  $\mu_1$  sowohl seine Masse, als auch seine Geschwindigkeit ändern. Aufgrund dieser Tatsache sprechen wir beim Hellinger-Kantorovich Abstand auch vom unbalancierten Transport.

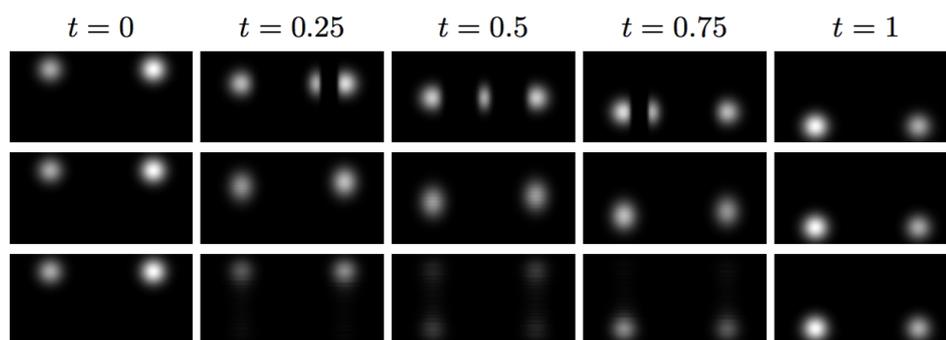


Abbildung 1.2.: Veranschaulichung der unterschiedlichen Geodäten [Fl17]

In der ersten Zeile wird mittels Wasserstein Geodäten interpoliert. In der zweiten Zeile ist die Interpolation mittels Hellinger-Kantorovich Geodäten für  $\kappa = 60$  abgebildet. Im Gegensatz zur ersten Zeile ist eine Veränderung der Masse sichtbar. Man achte auch auf eine leicht elliptische Form der interpolativen Verteilungen. In der dritten Zeile wird mit Hellinger-Kantorovich Geodäten für  $\kappa = 16$  interpoliert. Hier wird nur durch wachsen und schrumpfen der Verteilungen interpoliert.

### 1.3 Numerik

Nach den Formulierungen der verschiedenen Optimierungsprobleme versuchen wir diese zu lösen. Dabei suchen wir aber keine analytische sondern eine numerische Lösung, für die eine Raumdiskretisierung notwendig ist. Hierunter versteht man in der numerischen Mathematik das Approximieren eines Ausgangsgebiets durch eine endliche Anzahl von Daten.

#### 1.3.1. Diskretisierung

Wir versuchen nun Schritt für Schritt die bisherigen Definitionen des Kapitels zu diskretisieren. Als erstes werden die beiden Maße  $\mu_0, \mu_1$ , zwischen denen wir den optimalen Transport suchen, durch  $r, c \in \Sigma_d := \{x \in \mathbb{R}_+^d : x^T \mathbf{1}_d = 1\}$  beschrieben mit  $\mathbf{1}_d = \underbrace{(1, \dots, 1)^T}_{d\text{-mal}}$ . Hierbei sei angemerkt, dass wir ein Pixelgitter  $X = \{x_1, \dots, x_d\} \subset \Omega$  einführen, um ein beliebiges Maß eindimensional darzustellen und so die Kosten bestimmen zu können. So nennen wir im folgenden  $r$  und  $c$  Marginale. Die diskrete Version von  $\Pi$ , der Menge der Transportpläne, bezeichnen wir nun mit  $U$  und definieren wir durch:

$$U(r, c) := \{P \in \mathbb{R}_+^{d \times d} \mid P \mathbf{1}_d = r, P^T \mathbf{1}_d = c\}$$

Jeder Eintrag  $p_{ij}$  der  $(d \times d)$ -Matrix  $P$  beschreibt die Menge der Masse, die von  $x_i$  nach  $x_j$  transportiert wird. Die Nebenbedingungen garantieren uns, dass jeweils die gesamte Masse unserer diskreten Marginale  $r$  und  $c$  erhalten bleibt. Wir bezeichnen mit  $\langle A, B \rangle := \sum_{i,j=1}^d a_{ij} b_{ij}$  für zwei Matrizen  $A, B \in \mathbb{R}^{d \times d}$  das Frobenius-Skalarprodukt. So können wir den optimalen Transport zwischen den Marginalen  $r$  und  $c$  mit einer geeigneter Kostenmatrix  $M$  beschreiben durch:

$$d_M(r, c) := \min_{P \in U(r, c)} \langle P, M \rangle \quad (1.1)$$

So würde die Kostenmatrix  $M$  für die diskrete Version der 2-Wasserstein Distanz Einträge  $m_{ij} = \|x_i - x_j\|_2$  mit  $x_i, x_j \in X$  beinhalten. Analog hierzu können wir für eine allgemeine

Transportmatrix auch die diskrete Version des Hellinger-Kantorovich Abstands angeben (vgl. [Fl17]):

$$d_{HK, \mathbb{R}^d}^2(r, c) = \kappa^2 \inf_{P \in \mathbb{R}_+^{d \times d}} \text{KL}(P \mathbf{1}_d | r) + \text{KL}(P^T \mathbf{1}_d | c) + \langle P, M \rangle$$

mit  $p_{ij} = c_{HK, \kappa}(x_i, x_j)$  als Kostenmatrix, wobei die  $c_{HK, \kappa}$  die in Definition 1.2.3 definierte Kostenfunktion ist. Die diskrete Kullback-Leibler Divergenz erhalten wir durch:

$$\text{KL}(c || r) = \begin{cases} \sum_{i=1}^d c_i \cdot \log\left(\frac{c_i}{r_i}\right) & \text{wenn } [r_i = 0] \Rightarrow [c_i = 0] \\ +\infty & \text{sonst} \end{cases} \quad (1.2)$$

### 1.3.2. Sinkhorn

Wir haben die Diskretisierung des optimalen Transport Problems so weit beschrieben, dass wir es nun konkret Lösen wollen. Hierzu konzentrieren wir uns direkt auf das ursprüngliche Transportproblem (1.1) und regulieren dieses mit einem entropischen Term. Die Regularisierung wird unser Problem in ein streng konvexes Problem umformen, so dass wir schlussendlich die Fixpunktiteration von Sinkhorn (1967) anwenden werden. (vgl. [CU17])

Für die entropische Regularisierung des Problems (1.1) definieren wir die diskrete Entropie  $h$ , sowie die diskrete Kullback-Leibler Divergenz  $\text{KL}$  zweier Transportpläne  $P_0, P_1$ .

$$h(r) = - \sum_{i=1}^d r_i \log r_i, \quad h(P_0) = - \sum_{i,j=1}^d p_{ij} \log p_{ij}, \quad \text{KL}(P_0 || P_1) = \sum_{i,j=1}^d p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Zum Verständnis der beiden Begriffe, sowie zur Demonstration der Äquivalenz der nächsten Formeln, sei auch hier nochmal auf den Anhang A hingewiesen. So stellen wir die regularisierte Version von (1.1) auf.

$$U_\alpha(r, c) := \{P \in U(r, c) | \text{KL}(P || rc^T) \leq \alpha\} = \{P \in U(r, c) | h(P) \geq h(r) + h(c) - \alpha\} \subset U(r, c) \quad (1.3)$$

Also ist  $P_\alpha$  die Menge von Transportplänen die sich im Kullback-Leibler Sinne maximal um den Betrag  $\alpha$  von dem Transportplan  $rc^T$  unterscheidet.



Transportplan  $P^*$ . Mit  $P_\alpha$  in dem regulierten Gebiet erhalten wir eine gute Annäherung, die wir durch  $P^\lambda$  bis auf Maschinengenauigkeit optimieren können. Durch das Hinzufügen eines simplen Regularisierungsterms  $-\frac{1}{\lambda}h(P)$  ergibt sich eine bestmögliche Annäherung an  $P^*$  für großes  $\lambda$ , da  $P^\lambda$  gegen  $P^*$  konvergiert. Jetzt stellt sich die Frage, wie wir zu dem  $P^\lambda$  gelangen.

**Satz 1.3.2.** Für  $\lambda \geq 0$  ist die Lösung  $P^\lambda$  eindeutig und hat die Form:

$$P^\lambda = \text{diag}(u) \cdot K \cdot \text{diag}(v) \quad K := e^{-\lambda M}$$

mit zwei bis auf Vorfaktoren eindeutigen  $u, v \in \mathbb{R}_{\geq 0}^d$  und  $e^{-\lambda M}$  als die elementenweise Exponentialfunktion von  $-\lambda M$ .

*Beweis.* Die Eindeutigkeit und Existenz der Lösung  $P^\lambda$  erhalten wir durch die Beschränktheit von  $U(r, c)$  und der strikten Konvexität des  $-h(P)$  (vgl. [CU15]). Wir überführen das Problem der dualen Sinkhorn Distanz in dessen Lagrange Dualität  $L(P, \alpha, \beta)$  ( $\alpha, \beta$  duale Variablen) mit den Nebenbedingungen  $P\mathbf{1}_d = r$  und  $P^T\mathbf{1}_d = c$ .

$$L(P, \alpha, \beta) = \sum_{i,j} \frac{1}{\lambda} p_{ij} \log p_{ij} + p_{ij} m_{ij} + \alpha(P\mathbf{1}_d - r) + \beta(P^T\mathbf{1}_d - c)$$

Wir suchen also das Maximum von  $L(P, \alpha, \beta)$  und bestimmen dazu die erste Ableitung. So ergibt sich für jedes Paar  $(i, j)$ :

$$\left( \frac{\delta L}{\delta p_{ij}} = 0 \right) \Rightarrow 0 = \frac{1}{\lambda} (1 + \log p_{ij}) + m_{ij} + \alpha + \beta \Rightarrow p_{ij} = \underbrace{e^{-1/2 - \lambda \alpha_i}}_{\text{diag}(u)} \underbrace{e^{-\lambda m_{ij}}}_K \underbrace{e^{-1/2 - \lambda \beta_j}}_{\text{diag}(v)}$$

Nur dank der strikten Positivität von  $K$  konnten wir  $u$  und  $v$  wie gesehen gestalten. Mit einer simplen Fixpunktiteration  $(u, v) \leftarrow (\text{diag}(1/r) \cdot K \cdot v, \text{diag}(1/c) \cdot K^T \cdot u)$  können wir diese dann ermitteln und  $P^*$  mittels  $P^\lambda$  approximieren.  $\square$

Es ergibt sich der folgende Algorithmus, der auf Sinkhorn zurückgeht (vgl. [CU15]):

**Algorithmus 1** Sinkhorns Fixpunktiteration

- 
- 1: Input:  $M, \lambda, r, c$
  - 2:  $I_r = (r > 0); r = r(I_r); I_c = (c > 0); c = c(I_c); M = M(I_r, I_c); K = \exp(-\lambda M)$
  - 3: **while**  $u$  ändert sich kaum noch **do**
  - 4:      $u = r./Kv$
  - 5:      $v = c./K^T u$
  - 6: **end while**
- 

Mithilfe Sinkhorns Fixpunktiteration können wir somit leicht für eine beliebige Kostenmatrix  $M$  die optimale Transportmatrix und damit die minimalen Gesamtkosten des ausgeglichenen Problems bestimmen.

**Bemerkung.** Durch leichte Anpassung der Fixpunktiteration erhalten wir eine Berechnung für den unausgeglichenen Transport, bei uns in Form des Hellinger-Kantorovich Abstands (vgl. [CH17]):

$$u_{i+1} = (\text{diag}(1/r) \cdot K \cdot v_i)^{\frac{\kappa}{\kappa+\epsilon}} \quad v_{i+1} = (\text{diag}(1/c) \cdot K^T \cdot u_{i+1})^{\frac{\kappa}{\kappa+\epsilon}} \quad \text{mit } \epsilon := \frac{1}{\lambda}$$

## 2 | Erweiterung des optimalen Transport auf höhere Dimensionen

Als nächstes erweitern wir die Erkenntnisse aus Kapitel 1, um höherdimensionale Daten betrachten zu können. Hierbei sei wie in der Einleitung angemerkt, dass wir unter Höherdimensionalität die Existenz weiterer Information jedes Massepartikels verstehen. Unser Ziel ist es, die Wasserstein und Hellinger-Kantorovich Distanzen so zu erweitern, dass diese neben dem Abstand im metrischen Ausgangsraum, welchen wir im Folgenden als Bildraum bezeichnen, ebenfalls den Abstand in einem weiteren Raum, im folgenden Signalraum genannt, berücksichtigen. Um unser Vorhaben zu motivieren, greifen wir das Beispiel aus Kapitel 1 (Abbildung 1.2) auf:

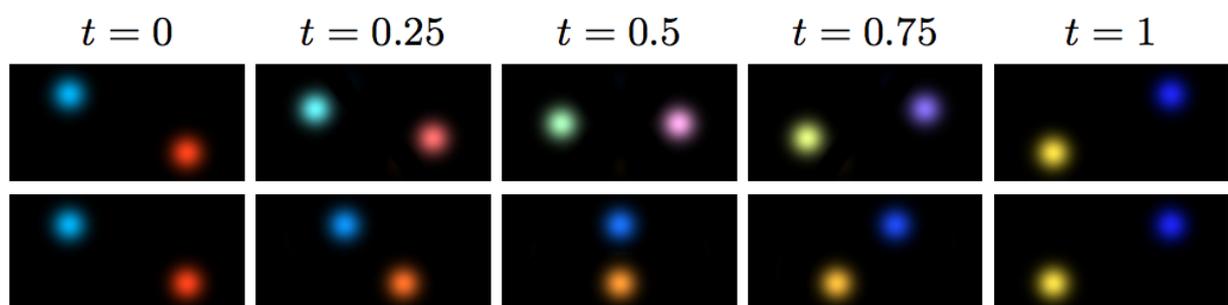


Abbildung 2.1.: Optimaler Transport skalar und höherdimensional [F17]

In diesem Fall beschreibt der Bildraum die Position und Konzentration der Masse, der Signalraum einen Farbraum. Die obere Zeile zeigt den optimalen Transport der Masse, die in Form von Kugeln dargestellt wird, ohne Betrachtung des Farbraums. Die untere Zeile gewichtet den Farbraum stark.

## 2.1 Der Produktraum

Dieser Abschnitt beschäftigt sich mit der Einführung eines Produktraumes, der das Produkt zwischen Bildraum und Signalraum bildet. So beschäftigen wir uns zuerst mit den Eigenschaften dieses Raumes und zeigen, dass es sich hierbei um einen metrischen Raum handelt, bevor auf ihm eine Ordnung von Massepartikeln zu Signalen hergestellt wird. Nach dieser Vorarbeit können wir im darauf folgenden Abschnitt die in Kapitel 1 erläuterten Distanzen auf den Produktraum erweitern.

### 2.1.1. Eigenschaften

Betrachten wir die beiden metrischen Räume, den Bildraum  $(\Omega \subset \mathbb{R}^d, d_\Omega)$  und den Signalraum  $(S, d_S)$ . Unser Ziel ist es nun beide Räume zu einem metrischen Raum zu vereinen.

**Satz 2.1.1.** Seien  $\Omega$  und  $S$  zwei metrische Räume mit Metriken  $d_\Omega(x_0, x_1)$  und  $d_S(s_0, s_1)$  für  $x_0, x_1 \in \Omega$ ,  $s_0, s_1 \in S$ . So ist eine Metrik auf dem Produktraum  $Y = \Omega \times S$  gegeben durch:

$$d_Y((x_0, s_0), (x_1, s_1)) = \sqrt{d_\Omega^2(x_0, x_1) + \lambda^2 \cdot d_S^2(s_0, s_1)}$$

mit  $\lambda > 0$  als relativ gewichtender Parameter.

*Beweis.* Wir zeigen, dass  $d_Y((x_0, s_0), (x_1, s_1))$  wie oben eine Metrik definiert.

1. Positive Definitheit. Für  $(x_0, s_0), (x_1, s_1) \in Y$  gilt:

$$0 = d_Y((x_0, s_0), (x_1, s_1)) \Leftrightarrow 0 = d_\Omega^2(x_0, x_1) + \lambda^2 \cdot d_S^2(s_0, s_1)$$

$$\Leftrightarrow 0 = d_\Omega(x_0, x_1) = d_S(s_0, s_1) \Leftrightarrow (x_0, s_0) = (x_1, s_1)$$

wird erfüllt, da  $d_\Omega$  und  $d_S$  Metriken sind, genauso wie:

$$0 \leq d_Y((x_0, s_0), (x_1, s_1)) \Leftrightarrow 0 \leq \underbrace{d_\Omega^2(x_0, x_1)}_{\geq 0} + \underbrace{\lambda^2 \cdot d_S^2(s_0, s_1)}_{\geq 0}$$

2. Symmetrie. Auch hier können wir wie bei der Positiven Definitheit die Symmetrie aus den Metriken  $d_\Omega$  und  $d_S$  gewinnen.
3. Dreiecksungleichung.

$$\begin{aligned}
& (d_Y((x_0, s_0), (x_2, s_2)) + d_Y((x_1, s_1), (x_2, s_2)))^2 \\
&= \left( \left\| \begin{pmatrix} d_\Omega(x_0, x_2) \\ d_S(s_0, s_2) \end{pmatrix} \right\| + \left\| \begin{pmatrix} d_\Omega(x_1, x_2) \\ d_S(s_1, s_2) \end{pmatrix} \right\| \right)^2 \\
&= \left\| \begin{pmatrix} d_\Omega(x_0, x_2) \\ d_S(s_0, s_2) \end{pmatrix} \right\|^2 + 2 \left\| \begin{pmatrix} d_\Omega(x_0, x_2) \\ d_S(s_0, s_2) \end{pmatrix} \right\| \left\| \begin{pmatrix} d_\Omega(x_1, x_2) \\ d_S(s_1, s_2) \end{pmatrix} \right\| + \left\| \begin{pmatrix} d_\Omega(x_1, x_2) \\ d_S(s_1, s_2) \end{pmatrix} \right\|^2 \\
&\stackrel{\geq}{\geq} \left\| \begin{pmatrix} d_\Omega(x_0, x_2) \\ d_S(s_0, s_2) \end{pmatrix} \right\|^2 + 2 \left\langle \begin{pmatrix} d_\Omega(x_0, x_2) \\ d_S(s_0, s_2) \end{pmatrix}, \begin{pmatrix} d_\Omega(x_1, x_2) \\ d_S(s_1, s_2) \end{pmatrix} \right\rangle + \left\| \begin{pmatrix} d_\Omega(x_1, x_2) \\ d_S(s_1, s_2) \end{pmatrix} \right\|^2 \\
&= \left( \left\| \begin{pmatrix} d_\Omega(x_0, x_2) + d_\Omega(x_1, x_2) \\ d_S(s_0, s_2) + d_S(s_1, s_2) \end{pmatrix} \right\| \right)^2 \\
&\geq d_Y^2((x_0, s_0), (x_2, s_2))
\end{aligned}$$

□

Da wir in Kapitel 1 einen allgemeinen metrischen Raum betrachtet haben, sei an dieser Stelle angemerkt, dass wir alle Erkenntnisse auf den Bildraum, den Signalraum und sogar auf unseren metrischen Produktraum  $(Y, d_Y)$  anwenden können. So erhalten wir im ersten Schritt für  $\nu_0, \nu_1 \in \text{rca}_1(Y)$  (bzw.  $\nu_0, \nu_1 \in \text{rca}(Y)$  in (2.3)):

1. Menge der Transportpläne

$$\Pi_Y(\nu_0, \nu_1) = \{\pi \in \text{rca}_1(Y^2) : \text{pr}_{X_i} \pi = \nu_i \text{ für } i \in \{0, 1\}\} \quad (2.1)$$

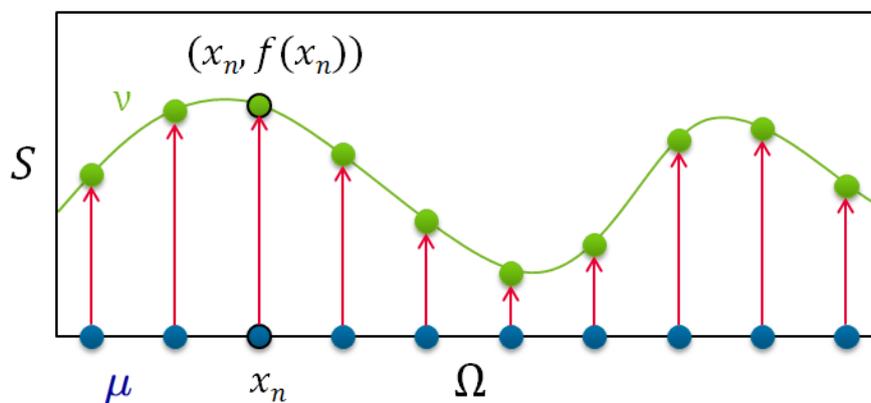
2. Wasserstein Distanz

$$d_{W,Y}^2(\nu_0, \nu_1) = \inf_{\pi \in \Pi_Y(\nu_0, \nu_1)} \int_{Y^2} d_Y^2((x_0, s_0), (x_1, s_1)) d\pi((x_0, s_0), (x_1, s_1)) \quad (2.2)$$

## 3. Hellinger-Kantorovich Abstand

$$d_{HK,X}^2(\nu_0, \nu_1) = \kappa^2 \inf_{\pi \in P(Y^2)} \left\{ \sum_{i=0}^1 KL(\text{pr}_{Y,i\#} \pi | \nu_i) + \int_{Y^2} c_{Y,\kappa}(x, y) d\pi((x_0, s_0), (x_1, s_1)) \right\} \quad (2.3)$$

Wichtig ist nun die Zuordnung des Signalraums zum Bildraum (engl. „the lifting of  $S$  over  $\Omega$ “) in unserem Produktraum  $Y$ . Nur wenn das Zuordnen realisiert wurde, können wir später Verteilungen auf dem Produktraum mittels  $d_{W,Y}$  oder  $d_{HK,Y}$  sinnvoll vergleichen.

Abbildung 2.2.: Lifting von  $S$  über  $\Omega$ 

Die blauen Punkte stellen die diskretisierten Punkte in  $\Omega$  dar, denen jeweils ein Signalwert zugeordnet wird.

Um die metrischen Räume einander zuzuordnen, definieren wir zwei messbare  $S$ -wertige Funktionen  $f_i : \Omega \rightarrow S$  und  $\mu_i \in rca(\Omega)$  für  $i \in \{0, 1\}$ . Beide Funktionen ordnen jedem Punkt des Ausgangs- und Zielmaßes einen Signalwert zu (siehe Abbildung 2.2).

So erhalten wir für das Paar  $(f_i, \mu_i)$  zwei Maße auf dem Produktraum  $Y$  für das jeweilige  $i \in \{0, 1\}$  durch:

$$\nu_i = F_{i\#} \mu_i, \text{ mit } F_i : \Omega \rightarrow Y, x \mapsto (x, f_i(x))$$

Insgesamt bilden die beiden Funktionen  $F_0$  und  $F_1$  einen Punkt von  $\Omega$  auf das Paar aus Bildpunkt und Signalpunkt ab. Wir bezeichnen die Zusammensetzung  $F$  der beiden Funk-

tion als Liftingfunktion:

$$F : (x_0, x_1) \mapsto (F_0(x_0), F_1(x_1)) = ((x_0, f_0(x_0)), (x_1, f_1(x_1)))$$

Somit codieren wir das nicht-skalare Signal als Teil der Position der (skalaren) Masse, um so die Höherdimensionalität zu modellieren. Mit Hilfe dieser Erkenntnisse können wir den Pushforward aller Transportpläne von  $\mu_0, \mu_1 \in \Omega$  unter der Liftingfunktion  $F$  betrachten für den gilt:

**Lemma 2.1.2.** Für den Bildraum  $\Omega$ , den Produktraum  $Y$ , die Maße  $\mu_0, \mu_1$  und die Liftingfunktion  $F : \Omega^2 \rightarrow Y^2, (x_0, x_1) \mapsto ((x_0, f_0(x_0)), (x_1, f_1(x_1)))$  wie zuvor gilt:

$$F_{\#} \Pi_{\Omega}(\mu_0, \mu_1) = \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$$

*Beweis.* Vorarbeit: Für die Maßräume  $(A, \Sigma_A), (B, \Sigma_B), (C, \Sigma_C)$ , zwei messbare Funktionen  $f : A \rightarrow B, g : B \rightarrow C$  und ein Maß  $\mu$  auf  $A$  besteht die Gleichung  $g_{\#} f_{\#} \mu = (g \circ f)_{\#} \mu$ , denn mit einem beliebigen  $\tilde{C} \in \Sigma_C$  gilt:

$$(g_{\#} f_{\#} \mu)(\tilde{C}) = f_{\#} \mu(g^{-1}(\tilde{C})) = \mu(f^{-1}(g^{-1}(\tilde{C}))) = \mu((g \circ f)^{-1}(\tilde{C})) = ((g \circ f)_{\#} \mu)(\tilde{C})$$

Für den Beweis des Lemmas betrachten wir die beiden getrennten Inklusion:

$$1. F_{\#} \Pi_{\Omega}(\mu_0, \mu_1) \subset \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$$

Hierfür ist zu zeigen: Für alle  $\pi \in \Pi_{\Omega}(\mu_0, \mu_1)$  gilt  $F_{\#} \pi \in \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$

Hierzu können wir dank der Vorarbeit und der Gleichung  $\text{pr}_{Y,i} \circ F = F_i \circ \text{pr}_{\Omega,i}$  zeigen, dass das  $i$ -te Marginal des gelifteten Paares das Ergebnis des Liftings des  $i$ -ten Marginals ist:

$$\text{pr}_{Y,i\#} F_{\#} \pi = (\text{pr}_{Y,i} \circ F)_{\#} \pi = (F_i \circ \text{pr}_{\Omega,i})_{\#} \pi = F_{i\#} \mu_0$$

Damit folgt die Inklusion.

$$2. F_{\#} \Pi_{\Omega}(\mu_0, \mu_1) \supset \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$$

Es ist zu zeigen: Für alle  $\pi \in \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$  existiert ein  $\tilde{\pi} \in \Pi_{\Omega}(\mu_0, \mu_1)$  mit

$$F_{\#}\tilde{\pi} = \pi.$$

Sei  $P : Y \rightarrow \Omega, (x, s) \mapsto x$  und  $(P, P) : Y^2 \rightarrow \Omega^2, ((x_0, s_0), (x_1, s_1)) \mapsto (x_0, x_1)$

Wir gehen davon aus, dass  $\tilde{\pi} = (P, P)_{\#}\pi$  gilt. Wir zeigen in Anlehnung zu 1., dass  $\tilde{\pi} \in \Pi_Y(\mu_0, \mu_1)$ , indem wir  $pr_{\Omega, i\#}\tilde{\pi} = \mu_i$  nachweisen.

$$pr_{\Omega, i\#}\tilde{\pi} = pr_{\Omega, i\#}(P, P)_{\#}\pi = (pr_{\Omega, i} \circ (P, P))_{\#}\pi = P_{\#}pr_{Y, i\#}\pi$$

$$P_{\#}F_{i\#}\mu_i = (P \circ F_i)_{\#}\mu_i = \mu_i$$

Hiefür haben wir die Formeln  $pr_{\Omega, i\#} \circ (P, P) = P \circ pr_{Y, i}$  und  $P \circ F_i = id$  genutzt. Es kann gezeigt werden, dass  $F_{\#}(P, P)_{\#}\pi = \pi$  mit  $\pi \in \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)$  (vgl. [SC17]).

□

Mit Hilfe dieses Lemmas haben wir gezeigt, wie sich die Menge der Transportpläne beim Lifting verhält. Dies war nötig um im nächsten Abschnitt  $d_{W, Y}$  (2.2) und  $d_{HK, Y}$  (2.3) vollständig ausformulieren zu können.

## 2.1.2. Optimierungsprobleme auf dem Produktraum

Mit Hilfe des Lemmas 2.1.3 können wir nun die Wasserstein Distanz von gelifteten Maßen aufstellen:

**Proposition 2.1.3** (Geliftete Wasserstein-Distanz). *Sei  $(Y, d_Y)$  der Produktraum wie zuvor. So können wir  $\nu_i = F_{i\#}\mu_i$  als geliftetes Maß schreiben und erhalten die geliftete Wasserstein Distanz durch:*

$$d_{W, Y}^2(\nu_0, \nu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\Omega^2} (d_{\Omega}^2(x_0, x_1) + \lambda^2 d_S^2(f_0(x_0), f_1(x_1))) d\pi(x_0, x_1)$$

*Beweis.*

$$d_{W, Y}^2(\nu_0, \nu_1) = \inf_{\pi \in \Pi_Y(\nu_0, \nu_1)} \int_{Y^2} d_Y^2(y_0, y_1) d\pi(y_0, y_1) = \inf_{\pi \in \Pi_Y(F_{0\#}\mu_0, F_{1\#}\mu_1)} \int_{Y^2} d_Y^2(y_0, y_1) d\pi(y_0, y_1)$$

$$\begin{aligned}
& \stackrel{2.1.2}{=} \inf_{\pi \in F_{\#} \Pi_{\Omega}(\mu_0, \mu_1)} \int_{Y^2} d_Y^2(y_0, y_1) d\pi(y_0, y_1) = \inf_{\pi \in \Pi_{\Omega}(\mu_0, \mu_1)} \int_{Y^2} d_Y^2(y_0, y_1) d\pi(F^{-1}(y_0, y_1)) \\
& = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\Omega^2} (d_{\Omega}^2(x_0, x_1) + \lambda^2 d_S^2(f_0(x_0), f_1(x_1))) d\pi(x_0, x_1)
\end{aligned}$$

□

Diese Formulierung ist nach unserer Zuordnung in Abschnitt 2.1.2 sowohl hilfreich, als auch eingängig. Dies liegt daran, dass wir den Abstand zwar nun im Produktraum betrachten, diesen aber als Problem über dem Bildraum beschreiben. Der einzige Unterschied zwischen dieser und der uns schon bekannten Wasserstein-Distanz liegt in der Kostenfunktion, die nicht mehr rein von  $d_{\Omega}(x_0, x_1)$ , sondern noch zu einem gewichteten Teil von  $d_S(f_0(x_0), f_1(x_1))$  abhängt.

Als vergleichendes Beispiel zwischen der gelifteten Version der Wasserstein Distanz und der skalaren Version sei auf die Abbildung 2.1 hingewiesen.

Die gleiche Herangehensweise wenden wir nun auf den Hellinger-Kantorovich Abstand an (vgl. [F17]):

**Proposition 2.1.4** (Gelifteter Hellinger-Kantorovich Abstand). Sei  $(Y, d_Y)$  der Produktraum wie zuvor. Dann lässt sich der geliftete Hellinger-Kantorovich Abstand für zwei geliftete Maße  $\nu_i = F_{i\#}\mu_i$  schreiben durch:

$$d_{HK,Y}^2(\nu_0, \nu_1) = \kappa^2 \inf_{\pi \in \text{rca}(\Omega^2)} \left\{ \sum_{i=0}^1 KL(\text{pr}_{Y,i\#} \pi | \nu_i) + \int_{\Omega^2} c_{Y,\kappa}((x, f_0(x)), (y, f_1(y))) \pi(x, y) \right\}$$

mit

$$c_{Y,\kappa}((x, f_0(x)), (y, f_1(y))) := \begin{cases} -\log([\cos(\frac{d_{\Omega}^2(x,y) + \lambda^2 d_S^2(f_0(x), f_1(y))}{\kappa})]^2) & \text{für } d_{\Omega}(x, y) < \kappa \cdot \frac{\pi}{2} \\ +\infty & \text{sonst} \end{cases}$$

Der Beweis ist nach einer Verallgemeinerung des Lemmas 2.1.2 ähnlich zu dem der gelifteten Wasserstein-Distanz zu führen.

So haben wir nun beiden Distanzen auf höherdimensionale Daten erweitert.

Um später Interpolationen zwischen zwei Verteilungen  $\nu_0, \nu_1$  zu beobachten, betrachten

wir nun wie schon in Kapitel 1 die Geodäten  $t \mapsto \nu_t$  für  $d_{HK,Y}$  und  $d_{W,Y}$ . Der größte Unterschied zwischen den Geodäten in Kapitel 1 und denen in diesem Kapitel liegt darin, dass wir nun nicht nur bewegende Masse in räumlicher Richtung betrachten, sondern auch noch Veränderungen der zugeordneten Signalwerte. Hierbei kann sich im Gegensatz zur einfachen räumlichen Betrachtung ein Problem ergeben. Es kann ein Zeitpunkt  $t_0$  existieren, zu dem zwei Massepartikel den gleichen räumlichen Ort haben, aber nicht den gleichen Signalwert. Damit würde keine Eindeutigkeit des Paares  $(f_{t_0}, \mu_{t_0})$  vorliegen. Dieser Uneindeutigkeit werden wir mithilfe eines Baryzentrums im Signalraum zu jedem Ortspunkt entgegenwirken. Für ein Maß  $\rho_x \in \text{rca}_1(S)$ , welches die Information über Vorkommen und Häufigkeit eines Signals am Ortspunkt  $x \in \Omega$  beschreibt, definiert sich das Baryzentrum durch:

$$\text{Bar}(\rho_x) = \operatorname{argmin}_{\bar{s} \in S} \int_S d_S^2(\bar{s}, s) d\rho_x(s) \quad (2.4)$$

Nun kann man zeigen, dass  $f(x) = \text{Bar}(\rho_x)$  wohldefiniert ist (siehe [F17]) und wir so wie in Kapitel 1 die verschiedenen Geodäten betrachten können, die die gleichen Eigenschaften wie in Kapitel 1 besitzen.

## 2.2 Diskretisierung

Wie schon am Ende des ersten Kapitels wollen wir nun die theoretischen Ergebnisse auf den diskreten Fall anwenden um damit die Brücke zu konkreten Implementierungen zu schlagen und die Vorarbeit für Kapitel 3 zu komplettieren. Im Wesentlichen werden wir durch die Vorarbeit in Kapitel 1 lediglich eine kurze Erweiterung benötigen.

Wir betrachten dazu zu unserem Pixelgitter  $X = \{x_1, \dots, x_d\} \subset \Omega$  und den diskreten Maßen  $r, c \in \mathbb{R}^d$  die Signalfunktion  $f_0$  bzw.  $f_1$ , die den Signalraum über das Maß  $r$  bzw.  $c$  legt. Des Weiteren erinnern wir uns an die diskrete Menge der Transportpläne  $U(r, c) := \{P \in \mathbb{R}^{d \times d} : P \mathbf{1}_d = r, P^T \mathbf{1}_d = c\}$ . Mit ihr erhalten wir analog zu Kapitel 1 für eine Transportmatrix  $P \in U(r, c)$  das Problem:

$$d_{W,Y}(F_{0\#}\mu_0, F_{1\#}\mu_1) = \inf_{P \in U(r,c)} \langle P, M \rangle \quad \text{mit} \quad \langle P, M \rangle = \sum_{j,k=1}^d m_{jk} p_{jk}$$

## 2. Erweiterung des optimalen Transport auf höhere Dimensionen *Anwendungsbeispiele*

Wie wir schon festgestellt haben ändert sich im Vergleich zum skalaren Fall nur die Kostenmatrix  $M$  wie folgt:

$$m_{i,j} = \sqrt{d_{\Omega}^2(x_i, x_j) + \lambda^2 d_{\xi}^2(f_0(x_i), f_1(x_j))}$$

Für alle  $x_i \in X$  gilt wie zuvor  $r_i = r(x_i)$  und  $c_i = c(x_i)$ .

Auf die gleiche Weise finden wir auch den unbalancierten Transport, bei uns in Form der Hellinger-Kantorovich Formulierung:

$$d_{HK,Y}^2(r, c) = \kappa^2 \inf_{P \in \mathbb{R}_+^{d \times d}} \text{KL}(P \mathbf{1}_d | r) + \text{KL}(P^T \mathbf{1}_d | c) + \langle P, M \rangle$$

mit Kostenmatrix  $M$  für die gilt:  $m_{ij} = c_{HK,Y}((x_i, f_0(x_i)), (x_j, f_1(x_j)))$  und diskreter Kullback-Leibler Divergenz (siehe 1.2).

Man stellt fest, dass wir für beide Versionen keine Diskretisierung des Signalraum benötigen. Im nächsten Schritt könnte man nun mittels Sinkhorn Algorithmus (siehe Kapitel 1) den optimalen Transport explizit berechnen.

### 2.3 Anwendungsbeispiele

In diesem Abschnitt wollen wir einen kurzen Überblick über Anwendungsmöglichkeiten des höherdimensionalen Transports geben. Wir stellen kurz Ergebnisse aus dem Bereich der Bildinterpolation und Zeichenerkennung vor.

#### 2.3.1. Farbraum

Wir knüpfen nun an das Beispiel am Anfang dieses Kapitels an. Dabei handelt es sich um das Lifting eines Bildraumes mittels eines geeigneten Farbraumes. Dabei müssen verschiedene Liftingansätze in Betracht gezogen werden, um das optimale Ergebnis zu erhalten.

Hierfür nehmen wir die Räume wie in Abbildung (2.1). So ist  $X \subset \mathbb{R}^2$  der diskrete Bildraum, sowie  $f : X \rightarrow \mathbb{R}_+$  ein Graubild und  $g : X \rightarrow S$  ein Farbbild. Als Farbräume betrachten

## 2. Erweiterung des optimalen Transport auf höhere Dimensionen *Anwendungsbeispiele*

wir  $S_{RGB} = [0, 1]^3$ , den RGB-, und  $S_{HSV} = H \times [0, 1]^2$ , den HSV-Farbraum. Für eine genaue Beschreibung des Liftings bzw. der Farbräume wird auf [F17] bzw. [RGo9] verwiesen. Ein Vorteil des HSV-Farbraums besteht darin, dass durch geschicktes Lifting (bei dem man die Masse mit der Sättigung verknüpft) jedem Schwarzwert keinerlei Masse zugeordnet wird, sodass Bildinterpolation von Bildern mit schwarzem Hintergrund hier deutlich realistischer erscheinen (siehe Abbildung 2.3).



Abbildung 2.3.: Lifting des Bildraums durch Farbsignale [F17]

Die ersten beiden Zeilen zeigen Transport zwischen einem Baum in verschiedenen Jahreszeiten. Hier interpolieren wir einmal, indem wir das den Bild- und Signalraum betrachten (erste) und nur den Farbraum (zweite). Da wir zwischen den Jahreszeiten interpolieren, gibt das Ergebnis auf dem alleinigen Farbraum bessere Ergebnisse. Im Gegensatz dazu in Zeile drei ein Ergebnis der Interpolation zweier Bilder, in der Signal und Bildraum betrachtet wurden.

### 2.3.2. Zeichenerkennung

Zuletzt schließen wir das Kapitel mit dem maschinellen Erkennen der Zeichensprache AUSLAN (Australian Sign Language). Der gleichnamige Datensatz enthält 95 verschiedenen Klassen, die die verschiedenen Worte der Sprache repräsentieren, sowie pro Wort 22 Signale zwischen  $-1$  und  $1$ .

## 2. Erweiterung des optimalen Transport auf höhere Dimensionen *Anwendungsbeispiele*

x-Position	y-Position	z-Position	Ballung	Neigung	Abweichung
-0.097816	-0.127761	-0.029051	0.537977	0.249162	0.508908
Daumen	Zeigefinger	Mittelfinger	Ringfinger	kleiner Finger	
0.239445	0.039819	0.000000	0.007009	0.019041	

Tabelle 2.1.: Beispieldaten eines Zeichens einer Hand zu einem Zeitpunkt

Um ein ganzes Zeichen zu repräsentieren betrachtet man die zu messenen Daten zu verschiedenen Zeitpunkten  $T := \{t_1, \dots, t_N\}$ . Am Ende erhalten wir eine Funktion von  $f : T \rightarrow \mathbb{R}^{22}$ .

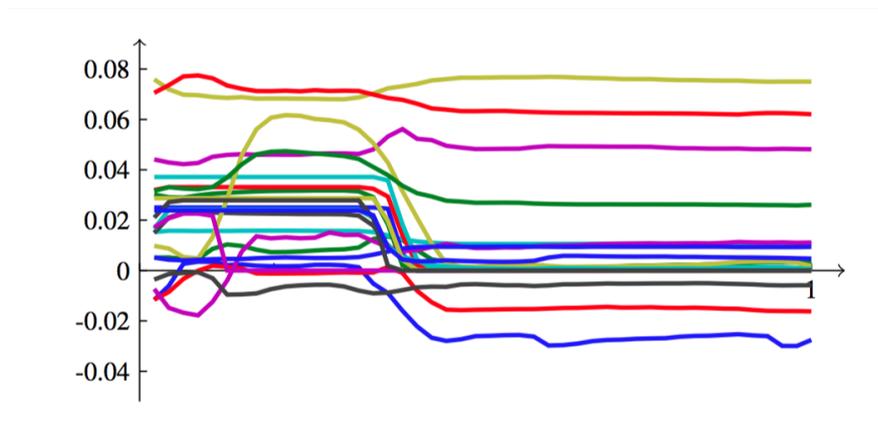


Abbildung 2.4.: Beispiel Signal des AUSLAN Datensatzes [TH16]

Für zwei Signale  $f$  und  $g$  betrachten wir die skalare und geliftete Wasserstein Distanz, die neben dem Abstand  $f(t_i) - g(t_j)$  auch den Abstand  $t_i - t_j$  wertet für  $t_i, t_j \in T$ . Damit ist  $\Omega = \mathbb{R}^{22}$  und  $S = T$ . Zudem vergleichen wir diese noch mit der Wasserstein Distanz der Ableitungen der Signale.

Datensatz	$d_{W,\Omega}(f, g)$	$d_{W,\gamma}(f, g)$	$d_{W,\Omega}(f', g')$	$d_{W,\gamma}(f', g')$
AUSLAN	10.4	9.5	14,7	2,4

Tabelle 2.2.: Fehlerrate in Prozent für eine „1-nearest-neighbor“ Klassifizierung

Für eine Beschreibung des Klassifizierers wird auf [GE17] verwiesen. Für die Berechnung wurden 25 der 95 Klassen und insgesamt 675 Signale betrachtet. Man erkennt, dass die geliftete Version bessere Ergebnisse erzielt. Die geliftete Version der Ableitung sticht durch ihre deutlich niedrigere Fehlerquote hervor.

## 3 | Höherdimensionale Betrachtung des MNIST Datensatzes

Legen wir nun den Fokus auf eine zu 2.3.2 vergleichbare Anwendung des effizienteren Klassifizierens von Daten mittels Betrachtung weiterer Dimensionen. Im Gegensatz zu Abschnitt 2.3.2 werden wir einen etwas anderen Datensatz betrachten und die Ergebnisse dieser Arbeit detailliert veranschaulichen. Hierzu wurde für diese Arbeit eine Applikation in der Programmiersprache Python implementiert, die sowohl als Applikation, als auch als Jupyter-Notebook auf der beiliegenden DVD zu finden ist. Es sei darauf hingewiesen, dass das Jupyter-Notebook einen Beispielcode für jeder Grafik in Kapitel 3 bereitstellt.

### 3.1 Vorbereitung

In diesem Abschnitt widmen wir uns neben der näheren Betrachtung des Datensatzes und der Formulierung unserer Ziele auch der Vorarbeit zur späteren Analyse und Interpretation.

#### 3.1.1. Der MNIST Datensatz

Der Modified National Institute of Standards and Technology (kurz MNIST) Datensatz besteht aus 70.000 ( $28 \times 28$ ) Pixelbilder mit handgeschriebenen Ziffern von Null bis Neun und deren zugehörigen Klassen. Im Bereich Klassifizierung des Machine-Learnings ist er einer der bekanntesten Datensätze.

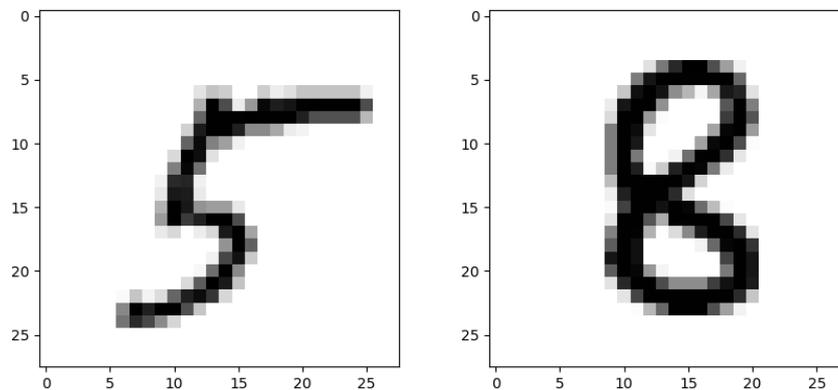


Abbildung 3.1.: Ziffern des MNIST Datensatzes

Beim Klassifizieren dieses Datensatzes wird versucht jede Ziffer der richtigen Klasse zuzuordnen. Wir versuchen nun eine Klassifizierung über den optimalen Transport zu ermöglichen. Die Masse wird hierbei durch den Grauwert eines Pixels beschreiben. Die Grundidee dieses Klassifizierens ist es, dass die Kosten des Transport der Masse, die von einer Ziffer zur anderen transportiert wird, zwischen zwei Ziffern einer Klasse geringer ist als zwischen zweier aus verschiedenen Klassen.

Zur Umsetzung dieser Idee können wir unsere Erkenntnisse aus den beiden ersten Kapiteln verwenden. So werden wir die diskreten Formulierungen auf den Datensatz Schritt für Schritt anwenden. Wir beginnen zunächst damit, das ungeliftete Problem zu betrachten, die Daten für den Sinkhorn Algorithmus anzureichern und mittels Hellinger-Kantorovich bzw. Wasserstein Distanz den Abstand zweier Ziffern zu berechnen.

Im nächsten Schritt greifen wir die Thematik des Kapitels 2 auf. So wollen wir zusätzlich die Struktur jeder Ziffer an jedem Punkt mittels lokaler Hessematrix betrachten (siehe Abbildung 3.2). Auf den gelifteten Daten arbeiten wir dann wie in Kapitel 2 beschreiben. Zum Schluss vergleichen wir die Ergebnisse der gelifteten gegen die ungelifteten Daten. Hierzu werden wir neben der multidimensionalen Skalierung auch die Fehlerrate eines Klassifizierers betrachten.

### 3.1.2. Datenvorbereitung

Um unsere Erkenntnisse auf die MNIST Daten anwenden zu können, müssen wir etwas Vorarbeit leisten, die in diesem Abschnitt beschrieben wird. Sie besteht darin, die Daten für den Sinkhorn Algorithmus vorzubereiten. Hierfür stellen wir unsere Bilder als normalisierte Marginale dar, indem wir die Zeilen im Bild sequenziell hintereinander zu einer einzigen Zeile umformen. Um spätere Kosten ableiten zu können, erstellen wir eine Zuordnungsmatrix, die jedem Wert in unserem Marginal seine Bildposition zuordnet.

$$r, c \in \mathbb{R}^{784} \quad X \in \mathbb{N}^{2 \times 784} \text{ mit } X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \cdots & 28 & 28 & 28 & 28 \\ 1 & 2 & 3 & 4 & 5 & \cdots & 25 & 26 & 27 & 28 \end{bmatrix}$$

Um sich später mit lokalen Strukturen beschäftigen zu können, betrachten wir als  $S$ -wertiges Signal eine geglättete lokale Hessematrix mit  $d_S$  als Frobeniusnorm. Damit erhalten wir für die Notationen aus Kapitel 2  $\Omega \subset \mathbb{R}^2$  als diskreten Bildraum und  $S = \{H \in \mathbb{R}^{2 \times 2} | h_{21} = h_{12}\}$  als die Menge der Hessematrizen.

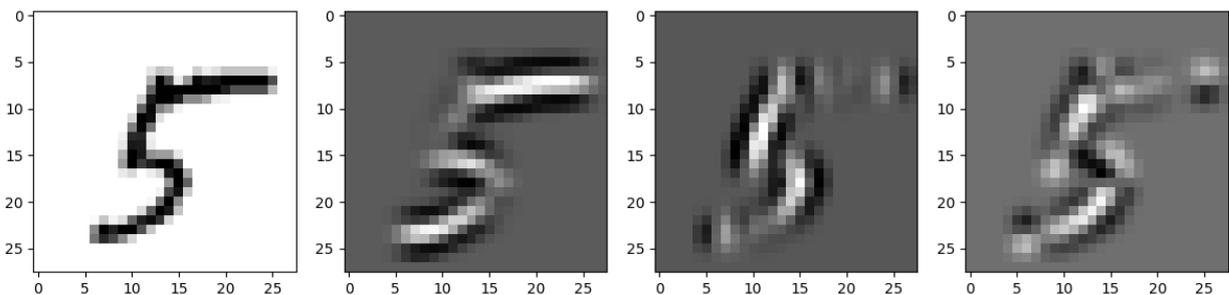


Abbildung 3.2.: Einträge der Hessematrix zu jedem Pixel

Von links nach rechts sind neben dem Originalbild (ganz links) die drei verschiedenen Einträge der Hessematrix zu sehen:  $\frac{\delta}{\delta^2 x}, \frac{\delta}{\delta^2 y}$  und  $\frac{\delta}{\delta x \delta y}$ .

## 3.2 Ergebnisse

Im Rahmen dieser Arbeit wurde eine Applikation in Python erstellt, die die beschriebenen Verfahren implementiert. Um die Kostenmatrix aufstellen zu können verwendet die Applikation neben der Zuordnungsmatrix  $X$  in der gelifteten Version die zuvor beschriebenen Hessematrizen. Nun bleibt nur noch die Frage, in wie weit die zusätzliche Betrachtung der lokalen Strukturen gegenüber der einfachen Massenbetrachtung gerechtfertigt ist. Wir wollen auf zwei verschiedene Weisen zeigen, dass der beschriebene Mehraufwand gerechtfertigt ist und die Theorie des Datenliftings zu Klassifizierung von Daten berücksichtigt werden sollte.

### 3.2.1. Multidimensionales Skalieren

Die Multidimensionale Skalierung (kurz „MDS“) umfasst Verfahren, mittels derer sich die Ähnlichkeit von Objekten untereinander in einem mehrdimensionalen Raum visualisieren lässt (vgl. [BA16]). Je dichter zwei Objekte nach der MDS beieinander liegen, desto ähnlicher sind diese zueinander. Neben der Wahl der Anzahl der Dimensionen, die wir im weiteren aus Visualisierungsgründen auf zwei beschränken, bleibt die Wahl einer geeigneten Metrik. So benötigt die MDS zu jedem Paar von Objekten deren Abstand. Aus diesem Grund ist die Methode ideal zum visualisieren unserer Thematik. Natürlich können wir hier auch den Hellinger-Kantorovich Abstand visualisieren, obgleich dieser keine Metrik im strengen Sinne darstellt.

Als erstes richten wir unsere Aufmerksamkeit auf den Parameter  $\lambda$  des Sinkhorn Algorithmus, der den Regularisierungsterm  $h(P)$  mit  $\frac{1}{\lambda}$  gewichtet (siehe Kapitel 1).

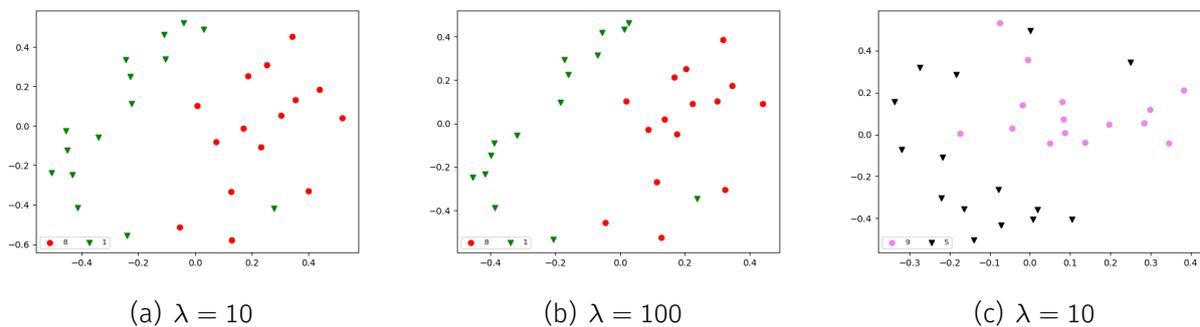


Abbildung 3.3.: MDS der klaren Wasserstein Distanz

Bild (a) und (b) verdeutlichen die Wasserstein-Distanz der Klassen 8 und 1 für zwei verschiedene  $\lambda$ . Bild (c) zeigt die Wasserstein Distanz für die Klassen 9 und 5, die sich als Ziffern ähnlicher sind.

Es wird deutlich, dass sich die Verteilung der Ziffern für verschiedene  $\lambda > 1$  kaum ändern. Zudem zeigt sich, dass der optimale Transport zur Klassifizierung des MNIST Datensatzes geeignet ist, aber wie erwartet für sehr ähnliche Ziffern in Bezug auf die Masse wie die 9 und 5 keine idealen Ergebnisse liefert.

Halten wir das Ergebnis des Bildes (c) der Abbildung 3.3 im Hinterkopf und untersuchen dessen Bilddaten nun im unbalancierten Transport. Für verschiedene  $\kappa$  des Hellinger-Kantorovich Abstands erhalten wir Abbildung 3.4.

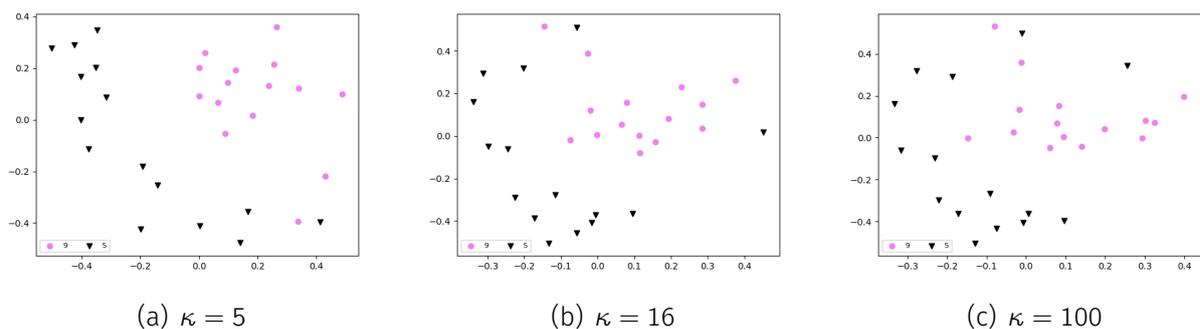


Abbildung 3.4.: MDS des klaren Hellinger-Kantorovich Abstands

Die Bildreihe verdeutlicht den Unterschied des Hellinger-Kantorovich Abstands für verschiedene  $\kappa$ . Wie erwartet erhalten wir für einen großen Wert von  $\kappa$  die Wasserstein Distanz. Die Masseparikel dürfen ihre Masse ändern und wir sehen für kleine  $\kappa$  bessere Ergebnisse als zuvor.

Betrachten wir nun den gewichtenden Parameter  $\lambda$  des Liftings für die Wasserstein Distanz um wie erhofft weitere Verbesserungen zu erreichen.

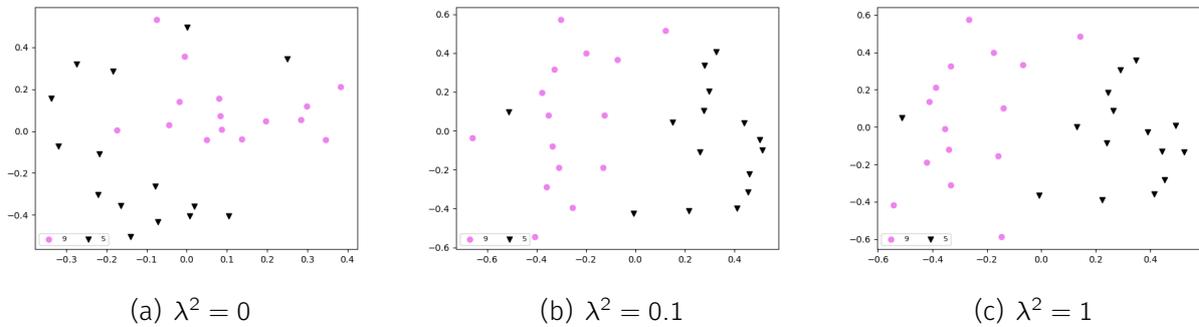


Abbildung 3.5.: MDS der gelifteten Wasserstein Distanz

Wir sehen jetzt an dem Beispiel der Klasse 9 und 5, die zwar ähnliche Masseverteilungen aber stark unterschiedliche Strukturen besitzen, dass man die Elemente deutlich besser voneinander unterscheiden kann. Bringen wir nun die den Abbildungen 3.4 und 3.5 zugrundeliegenden Verfahren zusammen, so erhalten wir den gelifteten Hellinger-Kantorovich Abstand.

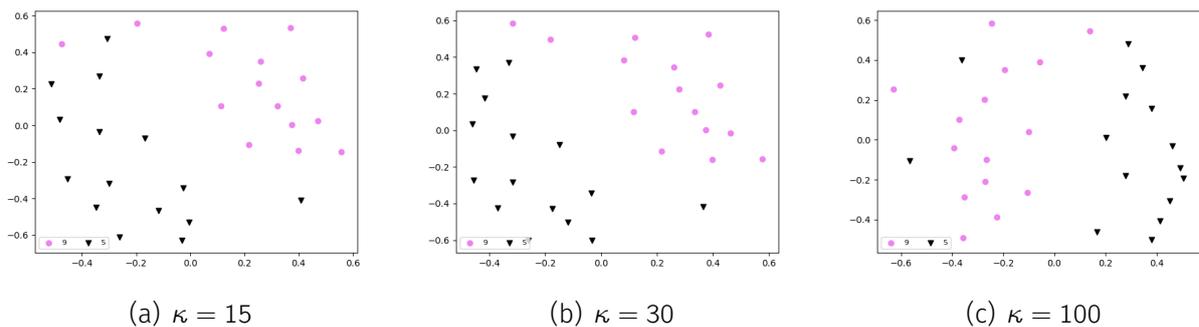


Abbildung 3.6.: MDS des gelifteten Hellinger-Kantorovich Abstands

Hier erhalten wir auf der einen Seite wie zuvor für ein großes  $\kappa$  die geliftete Wasserstein Distanz und für ein kleines  $\kappa$  die skalare Version des Hellinger-Kantorovich Abstands, da

bei diesem Fall die Massepartikel ihre Masse und damit auch ihren zugehörigen Signalwert stark ändern dürfen.

### 3.2.2. Konfusionsmatrix eines Klassifizierers

Nach der MDS testen wir nun die Klassifizierung mit den vier verschiedenen Ansätzen direkt. Hierfür wählen wir fünf Ziffern der Klassen 5 und 9 als Trainingsmenge aus. Unter diesen gehen wir vor wie in 3.2.1 und berechnen den Abstand jedes Paares. Anhand dieser Daten trainieren wir eine einfach Klassifizierungsmethode, die „nearest-neighbor“ Klassifizierung (vgl. [GE17]). Die trainierte Methode wenden wir auf eine Menge von Testdaten an und betrachten das Ergebnis mittels einer Konfusionsmatrix. Eine Konfusionsmatrix stellt nun die vier Werte „richtig positiv“, „falsch negativ“, „falsch positiv“ und „richtig negativ“ in Matrixform dar. Wir erhalten für unsere vier Ansätze auf 5 Trainingsdaten und 500 Testdaten pro Klasse folgendes Ergebnis:

	$d_{W,\Omega}$	$d_{HK,\Omega}$	$d_{W,Y}$	$d_{HK,Y}$
Konfusionsmatrix	$\begin{bmatrix} 472 & 28 \\ 211 & 289 \end{bmatrix}$	$\begin{bmatrix} 487 & 13 \\ 215 & 285 \end{bmatrix}$	$\begin{bmatrix} 420 & 80 \\ 73 & 427 \end{bmatrix}$	$\begin{bmatrix} 494 & 6 \\ 185 & 315 \end{bmatrix}$
Fehlerrate in %	23,8	22,8	15,3	19,1

Tabelle 3.1.: Ergebnisse einer „nearest-neighbor“ Klassifizierung

## 4 | Zusammenfassung und Ausblick

Die Ergebnisse des Kapitels 3 bilden den Abschluss dieser Arbeit gebildet. Sie zeigen uns rückblickend die Motivation hinter der Thematik.

Wir haben begonnen die wichtigsten Grundlagen des optimalen Transports einzuführen. Danach waren wir im Stande mittels Wasserstein Distanz den balancierten Transport und den unbalancierten Transport in Form des Hellinger-Kantorvich Abstands beschreiben zu können. Des Weiteren haben wir uns mit der Numerik hinter den analytischen Formulierungen auseinandergesetzt und am Schluss des ersten Kapitels mit dem Sinkhorn Algorithmus ein Werkzeug zur konkreten Implementation beschrieben.

In Kapitel 2 haben wir die gesamte Thematik auf höherdimensionale Daten erweitert und Anwendungsbeispiele betrachtet in denen das Datenlifting eine wichtige Rolle spielt wie der Farbinterpolation zwischen Bildern oder der Klassifizierung.

Die Klassifizierung spielte in Kapitel 3 eine große Rolle, indem wir die Theorie auf den MNIST Datensatz anwenden konnten und versucht haben die Klassifizierung verschiedener Ziffern zu optimieren. Wir haben gesehen, dass die skalare Wasserstein Distanz bei ähnlichen Klassen an ihre Grenzen stoßen kann. Mit Hilfe der anderen Formulierungen konnten wir diesen Umstand jedoch verbessern.

Im Anschluss an diese Arbeit wäre zunächst eine Vertiefung durch nähere Analysen der AUSLAN Daten in Form von Kapitel 3 denkbar. Zudem ist ein Vergleich zwischen herkömmlicher Klassifizierung und der Klassifizierung basierend auf dem optimalen Transport eine spannende Thematik, die nicht Einzug in diese Arbeit hielt.

Ein besonderer Dank gilt Herrn Dr. Bernhard Schmitzer für die Betreuung dieser Arbeit.

# A | Anhang

## A.1 Kullback-Leibler Divergenz

In diesem Teil des Anhangs widmen wir uns der Kullback-Leibler Divergenz und ihrer Bedeutung. Die Kullback-Leibler Divergenz oder auch relative Entropie bezeichnet ein Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen. Wie bereits in Kapitel 1 gezeigt, formuliert sich die kontinuierliche Kullback-Leibler Divergenz für zwei Verteilungen  $\mu_0$  und  $\mu_1$  mit ihren Dichtefunktionen  $\rho_{\mu_0}(x)$  und  $\rho_{\mu_1}(x)$  durch:

$$\text{KL}(\mu_0||\mu_1) = \int_{-\infty}^{\infty} \rho_{\mu_0}(x) \cdot \log\left(\frac{\rho_{\mu_0}(x)}{\rho_{\mu_1}(x)}\right) dx$$

In diskreter Gestalt und mit der Notation aus Kapitel 1 erhalten wir für diskrete  $r, c \in \Sigma_d$ :

$$\text{KL}(c||r) = \begin{cases} \sum_{i=1}^d c_i \cdot \log\left(\frac{c_i}{r_i}\right) & \text{wenn } [r_i = 0] \Rightarrow [c_i = 0] \\ +\infty & \text{sonst} \end{cases}$$

Die Kullback-Leibler Divergenz ist keine Distanz, da weder Symmetrie noch Dreiecksungleichung im Allgemeinen erfüllt wird. Für unsere Anwendungen in Kapitel 1 soll uns der Umstand ausreichen, dass „je ähnlicher“ sich beide Verteilungen sind, desto kleiner die Kullback-Leibler Divergenz wird (vgl. [CO16]). Um darüber hinaus noch eine kleine direkte Anwendung der Divergenz zu sehen, richten wir unsere Aufmerksamkeit kurz auf das Gebiet der Informationstheorie: Die Huffman-Kodierung, die 1952 von David Huffman entwickelt wurde, beschäftigt sich mit dem Kodieren durch binären Code. Ein Ziel ist damit beispielsweise einen Text zu speichern, bei dem die Buchstaben mit einer Verteilung  $\mu$  auftreten (vgl. [MA13]). Haben wir nun zwei verschiedene Huffman-Trees für eine jeweilige codierte Verteilung vorbereitet, stellt sich die Frage, welcher Tree weniger Speicher-

platz benötigt. Diese Frage kann über die Kullback-Leibler Divergenz geklärt werden (vgl. [MA13]).

**Bemerkung.** Für die diskreten Verteilungen  $r, c$ , die diskrete Menge der Transportpläne  $U(r, c)$  und  $\alpha > 0$  gilt:

$$\{P \in U(r, c) \mid KL(P \parallel rc^T) \leq \alpha\} = \{P \in U(r, c) \mid h(P) \geq h(r) + h(c) - \alpha\}$$

*Beweis.* Sei  $P \in U(r, c)$ , dann gilt:

$$\begin{aligned} KL(P \parallel rc^T) &= \sum_{ij} p_{ij} \cdot \log \left( \frac{p_{ij}}{(rc^T)_{ij}} \right) = \sum_{ij} p_{ij} \cdot (\log(p_{ij}) - \log((rc^T)_{ij})) \\ &= h(P) - \sum_{ij} p_{ij} \cdot \log(\log(r_i) \cdot \log(c_j)) = -h(P) - \sum_i r_i \log(r_i) - \sum_j c_j \log(c_j) \\ &= -h(P) + h(r) + h(c) \end{aligned}$$

Für die vorletzte Gleichheit betrachte man die Matrixform und berücksichtige die Nebenbedingungen von  $P$ . □

# Literaturverzeichnis

- [FI17] FITSCHEN, Jan Hendrik; LAUS, Fiederike; SCHMITZER, Bernhard: *Optimal Transport for Manifold-Valued Images*
- [SA15] SANTAMBROGIO, Filippo: *Optimal Transport for Applied Mathematicians*, 2015
- [TH16] THORPE, Matthew; PARK, Serim; KOLOURI, Soheil; ROHDE, Gustavo K.; DEJAN, Slepcev: *A Transportation  $L^p$  Distance for Signal Analysis*, 2016
- [CH17] CHIZAT Lenaic, PEYPRE Gabriel, SCHMITZER Bernhard, VIALARD Francois-Xavier: *SCALING ALGORITHMS FOR UNBALANCED OPTIMAL TRANSPORT PROBLEMS*, 2017
- [CU15] CUTURI, Marco: *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, 2015
- [FO12] FORSTER, Otto: *Analysis 3 - Maß- und Integrationstheorie, Integralsätze im  $\mathbb{R}^n$  und Anwendungen*, 2012
- [BH00] Weltbild: *Der Brockhaus*, 2000
- [GE17] GERON Aurelien: *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, 2017
- [SC17] SCHMITZER, Bernhard: Persönliche Absprache
- [GO09] GONZALEZ, Rafael: *Digital Image Processing Using MATLAB*, second edition
- [LI16] LIERO, Matthias, MIELKE, Alexander, SAVARÉ, Giuseppe: *Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures*, 2016

[BA16] BACKHAUS, Klaus, ERICHSON, Bernd, PLINKE, Wulff, WEIBER, Rolf *Multivariate Analysemethoden*

[CO06] COVER, Thomas, THOMAS, Joy: ELEMENTS OF INFORMATION THEORY, 2006

[MA13] MAZUMDAR, Arya, DATA COMPRESSION:

[HTTP://PEOPLE.ECE.UMN.EDU/~ARYA/EE5585/LECTURE4.PDF](http://people.ece.umn.edu/~arya/EE5585/LECTURE4.PDF), 2013

# Eidesstattliche Erklärung

Hiermit versichere ich, Lucas Plagwitz, dass die vorliegende Arbeit mit dem Titel „Höherdimensionaler optimaler Transport mit Anwendung für die Klassifizierung von Daten“ 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.

---

(04.09.2017, Lucas Plagwitz)

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.

---

(04.09.2017, Lucas Plagwitz)