ulm university universität uulm

Fakultät für Mathematik und Wirtschaftswissenschaften

Institut für Angewandte Analysis

Langzeitverhalten von Markov-Ketten mittels diskreter Funktionalungleichungen und entropischer Ricci-Krümmung

Bachelorarbeit

Vorgelegt von:
Lukas Niebel
lukas.niebel@uni-ulm.de

Gutachter:
Prof. Dr. Rico Zacher

2018

Fassung 25. Februar 2018

Inhaltsverzeichnis

1 Einleitung

Die Mathematik dient der Beschreibung der Wirklichkeit. Oft sind die zu beschreibenden Phänomene von kontinuierlicher Struktur, wie zum Beispiel die Modellierung der Wärmeausbreitung in einem Körper. Es gibt jedoch auch Phänomene, bei denen eine diskrete Struktur sinnvoll ist. Ein Beispiel ist das Internet. Wir möchten das Internet als einen Graphen interpretieren. Dazu fassen wir jede Internetseite als eine Ecke auf. Existiert zwischen zwei Internetseiten ein Hyperlink, so sagen wir, dass die jeweiligen Ecken verbunden sind. Zur Vereinfachung nehmen wir an, dass dieser Graph ungerichtet ist. Auf Basis der Ergebnisse, die wir in dieser Arbeit vorstellen, möchten wir eine Methode skizzieren, mit der man die Internetseiten mit den meisten Hyperlinks und den meisten Verlinkungen finden kann.

Wir stellen uns einen sogenannten Zufallssurfer vor. Dieser startet auf einer beliebigen Internetseite, wartet eine zufällige Zeit und klickt dann zufällig auf einen beliebigen Hyperlink. Jedes Mal, wenn er eine neue Internetseite aufruft, zählen wir für die zugehörige Internetseite einen Punkt dazu. Dies führt durch eine geeignete Normierung zu einer Wahrscheinlichkeitsverteilung, die angibt, wie wahrscheinlich es ist, dass der Zufallssurfer sich zu einem gegebenen Zeitpunkt auf einer Internetseite befindet. Wir werden sehen, dass sich diese Verteilung unter gewissen Voraussetzungen an den zugehörigen Graphen für große Zeiten einer stationären Verteilung annähert. Die stationäre Verteilung verschlüsselt die jeweilige Gewichtung der Internetseiten. Diesen Ansatz kann man als die probabilistische Interpretation des PageRank-Algorithmus verstehen. Mit den in dieser Arbeit vorgestellten Methoden kann man die Näherung an die stationäre Verteilung in Zusammenhang mit einer diskreten Entropie setzen. Mithilfe der Entropie kann man dann zum Beispiel Aussagen über die Stationarität der aktuellen Verteilung treffen.

Eine weitere Anwendung finden diskrete Modelle in der numerischen Mathematik. Möchte man ein kontinuierliches Problem und die dazu gehörigen partiellen Differentialgleichungen lösen, so ist man bei den meisten Verfahren auf eine Diskretisierung des Modells angewiesen. Wir möchten dies anhand der sogenannten Fokker-Planck-Gleichung

$$\partial_t u = \operatorname{div}(\nabla u + u\nabla V)$$

erläutern. Diese Gleichung bildet ein zentrales Objekt dieser Arbeit. Die Fokker-Planck-Gleichung kann man als Modellierung der Wärmeleitung unter Präsenz eines Potentials V

interpretieren.

Bei der numerischen Diskretisierung wünscht man sich eine möglichst gute Approximation des eigentlichen Modells. Optimal wäre, dass die Lösungen des diskreten Modells auch für beliebig große Zeiten beliebig genau mit den Lösungen des kontinuierlichen Modells übereinstimmen. Dies ist leider nur selten bis gar nicht möglich. Deshalb scheint es sinnvoll zu fordern, dass das diskrete Modell ähnliche, meist physikalisch motivierte Eigenschaften wie das kontinuierliche Modell aufweist. Eine solche Eigenschaft ist bei der Fokker-Planck-Gleichung zum Beispiel die Positivität von Lösungen zu einem positiven Anfangswert sowie die Massenerhaltung des Anfangswerts entlang der Lösung. Wir stellen ein diskretes Modell der Fokker-Planck-Gleichung vor, das diese gewünschten Eigenschaften erfüllt. Den mathematischen Rahmen für diese Eigenschaften bilden die sogenannten Markov-Halbgruppen. Diese Markov-Halbgruppen untersuchen wir in Kapitel 3 genauer.

Unter gewissen Voraussetzungen an V ist eine weitere wichtige Eigenschaft von Lösungen der Fokker-Planck-Gleichung die exponentiell schnelle Konvergenz der Entropie gegen null für große Zeiten. Es ist wünschenswert, dass dies auch für die Entropie der Diskretisierung gilt. In Kapitel 2 stellen wir zwei verschiedene Wege dar, mit denen man das obige Resultat im kontinuierlichen Fall beweisen kann. Diese Ansätze untersuchen wir dann hinsichtlich ihrer Kompatibilität mit dem diskreten Setting.

Der Ansatz, den wir in den Kapiteln 4 und 5 vorstellen, basiert auf einer Adaption der Theorie des optimalen Transports. Grob formuliert ist die Grundfrage dieser Theorie: Es seien ein Erdloch und ein Sandhaufen mit dem gleichen Volumen gegeben; wie transportiert man den Sand mit möglichst geringem Aufwand in das Erdloch? Diese Theorie liefert einen sehr eleganten Beweis für die Konvergenz der Entropie entlang von Lösungen der Fokker-Planck-Gleichung. Wir versuchen, diese Theorie geeignet auf das diskrete Setting anzuwenden. Ein zentrales Thema des vierten Kapitels ist, wie man diese Theorie in einem diskreten Setting sinnvoll formuliert, sodass man ähnliche Resultate erhält.

Die Ergebnisse, die wir in dieser Arbeit vorstellen, sind meistens in direkter Analogie zu ihrem kontinuierlichen Pendant entwickelt worden. Ein wichtiges Handwerkszeug der Theorie des optimalen Transports ist der sogenannte Otto-Kalkül. Dabei handelt es sich um intuitive Rechenregeln. In Kapitel 4 untersuchen wir die diskrete Version dieser Rechenregeln und werden sehen, dass es dank der diskreten und endlichen Struktur möglich ist, diese Rechenregeln zu formalisieren.

Im Anhang stellen wir eine Einführung zum Begriff der Markov-Ketten, einige Resultate zu riemannschen Mannigfaltigkeiten sowie eine Sammlung technischer Hilfsresultate bereit.

2 Entropie und ihre Anwendungen

Wir möchten in diesem Kapitel den Begriff der Entropie erklären, einige Eigenschaften der Entropie untersuchen und den Zusammenhang zu der Fokker-Planck-Gleichung vorstellen.

2.1 Eine abstrakte Definition

Der Begriff der Entropie wird in den verschiedensten Situationen benutzt. Wir geben hier eine möglichst allgemeine Definition auf Basis der Ausführungen in [Mat07] und [Jü16]. Im Folgenden sei X ein Banachraum. Wir betrachten das abstrakte Cauchy-Problem

$$\begin{cases} \partial_t u + Au = 0 & \text{für } t > 0 \\ u(0) = u_0, \end{cases} \tag{ACP}$$

wobei $A\colon D(A) \to X$ ein linearer Operator mit Definitionsbereich $D(A) \subset X$ ist. Ferner existiere ein eindeutig bestimmtes Equilibrium $u_\infty \in D(A)$, das heißt, es gelte $A(u_\infty) = 0$. Wir nehmen an, dass zu jedem beliebigem Anfangswert $u_0 \in D(A)$ eine glatte Lösung $u\colon [0, \infty) \to D(A)$ des Anfangswertproblems (ACP) existiert.

Definition 2.1.1 (Ljapunovfunktional)**.** Ein Funktional $L\colon D(A) \to \mathbb{R}$ heißt Ljapunovfunktional zum (ACP), falls jede glatte Lösung von (ACP) die Ungleichung

$$\frac{\mathrm{d}}{\mathrm{d}t} L(u_t) \leq 0$$

für alle $t > 0$ erfüllt.

Definition 2.1.2 (Entropie)**.** Sei $H\colon D(A) \to \mathbb{R}$ ein konvexes Ljapunovfunktional. Sei ferner $\varphi\colon \mathbb{R} \to \mathbb{R}$ stetig, monoton wachsend und es gelte $\varphi(0) = 0$. Existiert eine Metrik d auf X, sodass

$$d(u, u_\infty) \leq \varphi(H(u) - H(u_\infty))$$

für alle $u \in D(A)$ gilt, dann nennen wir H Entropie zu (ACP) und die Differenz $H(u|u_\infty) := H(u) - H(u_\infty)$ bezeichnen wir als die relative Entropie.

Bemerkung 2.1.3. (i) Im Vergleich zur physikalischen Entropie ist die oben definierte Entropie fallend entlang der Lösungen des abstrakten Cauchy-Problems und nicht wachsend.

(ii) A priori ist in Definition 2.1.2 nicht klar ob der Term $H(u) - H(u_\infty)$ nichtnegativ ist. Wir wählen deshalb den Definitionsbereich $\mathbb{R}$ für die Funktion φ. Aus der Monotonie, der Bedingung $\varphi(0) = 0$ sowie der Nichtnegativität der Metrik folgt, dass man ohne Beschränkung der Allgemeinheit $\varphi(x) = 0$ für $x \geq 0$ setzen kann.

Definition 2.1.4 (Entropieproduktion)**.** Wir betrachten das (ACP) mit einem gegebenen Entropiefunktional H. Für eine glatte Lösung $u\colon [0,\infty) \to D(A)$ des (ACP), definieren wir die Entropieproduktion $I\colon (0,\infty) \to \mathbb{R}$ durch

$$I(t) = -\frac{\mathrm{d}}{\mathrm{d}t} H(u_t).$$

Die Entropieproduktion wird oft auch Fisher-Information genannt.

2.2 Anwendung auf die Fokker-Planck-Gleichung

In diesem Abschnitt werden wir eine Anwendung von Entropiemethoden auf die Fokker-Planck-Gleichung vorstellen. Die Fokker-Planck-Gleichung im $\mathbb{R}^n$ ist gegeben durch

$$\begin{cases} \partial_t u = \operatorname{div}(\nabla u + u \nabla V) & \text{für } t > 0 \\ u(0) = u_0 \end{cases} \tag{FPG}$$

für ein sogenanntes Potential $V\colon \mathbb{R}^n \to \mathbb{R}$.

Konvention: In den weiteren Ausführungen zur Fokker-Planck-Gleichung werden wir die folgenden Eigenschaften annehmen. Zu jedem positiven Anfangswert existiere eine eindeutig bestimmte, glatte und positive Lösung. Wir betrachten ausschließlich positive Anfangswerte u_0 mit Einheitsmasse. Man kann zeigen, dass dann auch die Lösung von (FPG) Einheitsmasse besitzt, das heißt, es gilt $1 = \int_{\mathbb{R}^n} u_0 \mathrm{d}x = \int_{\mathbb{R}^n} u_t \mathrm{d}x$ für alle $t > 0$. Für ein genügend glattes Potential V mit $e^{-V} \in L^1(\mathbb{R}^n)$ ist das Equilibrium der Fokker-Planck-Gleichung gegeben durch $u_\infty = ce^{-V}$ mit der normalisierenden Konstanten $c := \|e^{-V}\|_{L^1}^{-1}$.

Wir betrachten die Funktion $\varphi\colon [0,\infty) \to \mathbb{R}$, $\varphi(x) := x\log(x) - x + 1$ für $x > 0$ und $\varphi(0) = 0$. Die relative Entropie bezüglich u_∞ ist die durch

$$H(u) := H(u|u_\infty) := \int_{\mathbb{R}^d} \varphi\left(\frac{u}{u_\infty}\right) u_\infty \mathrm{dx}$$

definierte Funktion. Dies stimmt insoweit mit der Definition 2.1.2 überein, dass $H(u_\infty|u_\infty) = 0$ ist und folglich auch $H(u|u_\infty) = H(u|u_\infty) - H(u_\infty|u_\infty) = H(u) - H(u_\infty)$ gilt, wobei $H(u) := H(u|u_\infty)$ bezeichnet. Wir haben jedoch noch nicht gezeigt, dass $H(u)$ die benötigten Eigenschaften besitzt, um eine Entropie im Sinne von Definition 2.1.2 zu sein.

Theorem 2.2.1. *Es sei $H(u_0) < \infty$ und das Potential V erfülle die Bakry-Emery-Bedingung $\nabla^2 V \geq \lambda\,\mathrm{Id}$ für ein $\lambda > 0$. Unter diesen Voraussetzungen konvergiert die glatte Lösung der Fokker-Planck-Gleichung (FPG) zum Anfangswert u_0 exponentiell schnell zum Equilibrium. Insbesondere gilt für alle Zeiten $t > 0$ die Abschätzung*

$$\|u_t - u_\infty\|_{L^1} \leq e^{-\lambda t}\sqrt{2H(u_0)}.$$

Bemerkung 2.2.2. Wir bezeichnen mit $\nabla^2 V$ die Hesse-Matrix von V. Die Bakry-Emery-Bedingung $\nabla^2 V \geq \lambda\,\mathrm{Id}$ ist im Sinne einer gleichmäßigen positiven Semidefinitheit der Matrix $\nabla^2 V - \lambda\,\mathrm{Id}$ zu verstehen. Für ein glattes Potential V folgt aus dieser Bedingung die strikte Konvexität von V.

Beweis. Wir geben nur einen kurzen Überblick über die einzelnen Schritte des Beweises. Für eine genauere Behandlung verweisen wir auf [Mat07] und die dort angegebenen Referenzen.

Teil 1: Im ersten Schritt zeigt man die Nichtnegativität der Entropieproduktion $-\frac{\mathrm{d}}{\mathrm{d}t}H(u_t)$ entlang der Lösung der Fokker-Planck-Gleichung zum Anfangswert u_0. Für alle Zeiten $t > 0$ gilt

$$-\frac{\mathrm{d}}{\mathrm{d}t}H(u_t) \geq 0.$$

Teil 2: Die Idee des Beweises ist die Berechnung und anschließende Abschätzung der zweiten Ableitung der relativen Entropie nach der Zeit. Dies wird oft auch als Bakry-Emery-Methode bezeichnet. Man zeigt, dass

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}H(u_t) \geq -2\lambda\frac{\mathrm{d}}{\mathrm{d}t}H(u_t)$$

für alle Zeiten $t > 0$ gilt.

Teil 3: Die Integration der Ungleichung aus Teil 2 über (s, ∞) liefert dann

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}H(u_t)\right|_{t=s} - \lim_{r\to\infty}\left.\frac{\mathrm{d}}{\mathrm{d}t}H(u_t)\right|_{t=r} \leq -2\lambda\left(H(u_s) - \lim_{r\to\infty}H(u_r)\right).$$

Man kann zeigen, dass die beiden Limiten verschwinden, damit gilt

$$\frac{\mathrm{d}}{\mathrm{d}t} H(u_t)\bigg|_{t=s} \leq -2\lambda H(u_s)$$

und nach Gronwall's Lemma folgt für $s \geq 0$

$$H(u_s) \leq H(u_0)e^{-2\lambda s}.$$

Teil 4: Im letzten Schritt stellt man den Zusammenhang zwischen Entropie und $\|u_t - u_\infty\|_{L^1}$ her. Dazu verwendet man die Csiszár-Kullback-Pinsker-Ungleichung, diese Ungleichung liefert

$$\|u_t - u_\infty\|_{L^1}^2 \leq 2H(u_t) \leq 2H(u_0)e^{-2\lambda t}$$

für alle $t \geq 0$ und somit die Behauptung.

□

Bemerkung 2.2.3. (i) Teil 1 des obigen Beweises zeigt, dass H ein Ljapunovfunktional ist. Wir möchten noch begründen, dass es sich hierbei auch um eine Entropie im Sinne der Defintion 2.1.2 handelt. Wählen wir als Metrik den L^1-Abstand, dann gilt nach Teil 4 die Ungleichung $\|u_t - u_\infty\|_{L^1} \leq \sqrt{2H(u_t)}$ für alle genügend glatten Lösungen u der (FPG). Wählen wir $\psi\colon \mathbb{R} \to \mathbb{R}$, $\psi(x) = \sqrt{2x}$, dann ist $\|u_t - u_\infty\|_{L^1} \leq \psi(H(u_t))$. ψ ist stetig, monoton wachsend und es gilt $\psi(0) = 0$. Da φ konvex ist, folgt die Konvexität der Entropie und insgesamt die Behauptung.

(ii) Wir sehen, dass man mithilfe einer geeigneten Entropie das Langzeitverhalten von Lösungen der Fokker-Planck-Gleichung kontrollieren kann.

2.3 Die Theorie des optimalen Transports

Sei $(\mathcal{X}, d)$ ein vollständiger separabler metrischer Raum. Wir betrachten den Raum der borelschen Wahrscheinlichkeitsmaße auf $\mathcal{X}$ mit endlichem zweiten Moment

$$\mathcal{P}_2(\mathcal{X}) := \left\{\mu\colon \mathcal{B}(\mathcal{X}) \to [0,\infty) \;\middle|\; \mu \text{ Maß}, \mu(\mathcal{X}) = 1 \text{ und } \exists x_0 \in \mathcal{X}\colon \int_{\mathcal{X}} d(x_0,x)^2 \mathrm{d}\mu(x) < \infty\right\}.$$

Für zwei Maße $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$ definieren wir

$$W_2(\mu,\nu) = \inf_{\Gamma \in \Pi(\mu,\nu)} \left(\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y)\right)^{\frac{1}{2}},$$

wobei $\Pi(\mu,\nu)$ die Menge aller Kopplungen von μ und ν bezeichnet, das heißt,

$$\Pi(\mu,\nu) = \left\{\Gamma \in \mathcal{P}_2(\mathcal{X}\times\mathcal{X}) \;\middle|\; \Gamma(A\times\mathcal{X}) = \mu(A) \text{ und } \Gamma(\mathcal{X}\times A) = \nu(A)\ \forall A \in \mathcal{B}(\mathcal{X})\right\}.$$

Diese Abbildung wird meistens als L^2-Wasserstein-Metrik bezeichnet, wir werden sie im Folgenden vereinfacht Wasserstein-Metrik nennen. Den Wert $W_2(\mu,\nu)$ kann man als die minimalen Transportkosten interpretieren, die benötigt werden, um eine Masse von ihrem Anfangszustand μ in den Endzustand ν zu überführen. Die Theorie des optimalen Transports beschäftigt sich mit einer verallgemeinerten Version dieser Fragestellung. Es gilt das folgende Lemma, welches uns wichtige Eigenschaften der Wasserstein-Metrik bereitstellt und unter anderem ihren Namen begründet.

Bemerkung 2.3.1. (i) Wir statten den metrischen Raum $\mathcal{X}\times\mathcal{X}$ mit der Metrik definiert durch $d_\times((x_1,x_2),(y_1,y_2)) := \sqrt{d(x_1,y_1)^2 + d(x_2,y_2)^2}$ für $(x_1,x_2),(y_1,y_2) \in \mathcal{X}\times\mathcal{X}$ aus. Dann ist $(\mathcal{X}\times\mathcal{X}, d_\times)$ ebenso ein vollständiger metrischer Raum.

(ii) Für zwei gegebene Maße $\mu,\nu \in \mathcal{P}_2(\mathcal{X})$ gilt stets $\Pi(\mu,\nu) \neq \emptyset$. Denn zu den Wahrscheinlichkeitsmaßen μ und ν wählen wir das Produktmaß $\mu\otimes\nu$, dann ist dieses Maß ein Wahrscheinlichkeitsmaß und es gilt

$$\int_{\mathcal{X}\times\mathcal{X}} d_\times((x,y),(x_0,y_0))^2 \mathrm{d}\,(\mu\otimes\nu)\,(x,y) = \int_{\mathcal{X}} d(x,x_0)^2 \mathrm{d}\mu(x) + \int_{\mathcal{X}} d(y,y_0)^2 \mathrm{d}\nu(y) < \infty,$$

wobei $x_0, y_0 \in \mathcal{X}$ nach der Definition von $\mathcal{P}_2(\mathcal{X})$ so gewählt werden, dass die Terme auf der rechten Seite in obiger Gleichung endlich sind. Das Produktmaß erfüllt per definitionem die Eigenschaft, dass $(\mu\otimes\nu)(A\times\mathcal{X}) = \mu(A)\nu(\mathcal{X}) = \mu(A)$ und $(\mu\otimes\nu)(\mathcal{X}\times A) = \nu(A)$ für alle $A \in \mathcal{B}(\mathcal{X})$ ist. Dies zeigt $\mu\otimes\nu \in \Pi(\mu,\nu)$.

Lemma 2.3.2. Unter den obigen Voraussetzungen ist das Paar $(\mathcal{P}_2(\mathcal{X}), W_2)$ ein separabler vollständiger metrischer Raum.

Beweis. Einen Beweis dieser Aussage findet man in [LN09, Theorem 2.10] mit der Bemerkung, dass ein separabler vollständiger metrischer Raum ein Beispiel für einen polnischen Raum ist. □

Definition 2.3.3 (Relative Entropie)**.** Für ein gegebenes Referenzwahrscheinlichkeitsmaß $\nu \in \mathcal{P}_2(\mathcal{X})$ definieren wir die relative Entropie als die Abbildung

$$H_\nu\colon \mathrm{AC}_\nu(\mathcal{X}) \to [0,\infty],\ H_\nu(\mu) = \int_{\mathcal{X}} \rho \log \rho \mathrm{d}\nu.$$

Hier bezeichnen wir mit $\mathrm{AC}_\nu(\mathcal{X})$ die Teilmenge aller Maße in $\mathcal{P}_2(\mathcal{X})$, die absolut stetig bezüglich ν sind und bezeichnen dann mit $\rho = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ die Radon-Nikodym-Dichte der beiden Maße.

Die Entropie zusammen mit der Wasserstein-Metrik sind der Ausgangspunkt für viele schöne Resultate. Zwei davon werden wir hier vorstellen. Wir betrachten die Dirichlet-Energie $\mathcal{E}\colon L^2(\mathbb{R}^n) \to \mathbb{R} \cup \{\infty\}$ definiert durch

$$\mathcal{E}(u) = \begin{cases} \frac{1}{2} \int_{\mathbb{R}^n} |\nabla u(x)|^2 \mathrm{d}x & : u \in W^{1,2}(\mathbb{R}^n) \\ \infty & : \text{sonst.} \end{cases}$$

Man kann zeigen, dass die Lösungen der Wärmeleitungsgleichung auch Trajektorien des Gradientenflusses der Dirichlet-Energie sind. Richard Jordan, David Kinderlehrer und Felix Otto haben erstmals gezeigt, dass man die Lösungen der Wärmeleitungsgleichung auch als Gradientenfluss der relativen Entropie auffassen kann. Sie bewiesen in [JKO98] eine Version des folgenden bemerkenswerten Theorems.

Theorem 2.3.4 (JKO-Theorem)**.** *Wir betrachten ein glattes Potential $V\colon \mathbb{R}^n \to [0, \infty)$ und die zugehörige Fokker-Planck-Gleichung $\partial_t u = \operatorname{div}(\nabla u + u\nabla V)$. Dann sind die Lösungen der Fokker-Planck-Gleichung bei gleichem Anfangswert auch die Trajektorien des Gradientenflusses der freien Energie bezüglich der Wasserstein-Metrik. Die freie Energie ist definiert durch*

$$\Phi(u) = \int_{\mathbb{R}^n} u \log(u) \mathrm{d}x + \int_{\mathbb{R}^n} V u \mathrm{d}x$$

für geeignete Funktionen $u\colon \mathbb{R}^n \to \mathbb{R}$.

Bemerkung 2.3.5. (i) Der Gradientenfluss der Dirichlet-Energie ist wohldefiniert, da wir uns in einem Hilbertraum befinden. A priori ist in Theorem 2.3.4 jedoch nicht klar, was man unter einem Gradientenfluss bezüglich einer metrischen Struktur versteht. Darauf möchten wir hier nicht näher eingehen und verweisen für eine genauere Behandlung auf [JKO98] und [Vil09, Kapitel 23].

(ii) Betrachten wir in Theorem 2.3.4 den Spezialfall $V \equiv 0$, dann erhalten wir eine Aussage über die Wärmeleitungsgleichung. Genauer bedeutet das, dass sich die Verteilung der Wärme zu jeder Zeit in gewisser Weise in Richtung des stärksten Abstiegs der Entropie entwickelt. Beachtet man, dass die von uns betrachtete Entropie der negativen physikalischen Entropie entspricht, stimmt dies mit der physikalischen Intuition überein, dass sich der Verteilung der Wärme in Richtung des stärksten Anstiegs der physikalischen Entropie entwickelt.

(iii) Schreiben wir $\mathrm{d}\nu = u_\infty \mathrm{d}x$ und $\mathrm{d}\mu = \frac{u}{u_\infty} u_\infty \mathrm{d}x = \rho\nu$ für $\rho = \frac{u}{u_\infty}$, dann gilt

$$\begin{aligned} H_\nu(\mu) = \int_{\mathbb{R}^n} \rho \log(\rho) \mathrm{d}\nu &= \int_{\mathbb{R}^n} \frac{u}{u_\infty} \log(u) u_\infty \mathrm{d}x - \int_{\mathbb{R}^n} u \log(u_\infty) \mathrm{d}x \\ &= \int_{\mathbb{R}^n} u \log(u) \mathrm{d}x + \int_{\mathbb{R}^n} uV \mathrm{d}x = \Phi(u). \end{aligned}$$

Das heißt, die freie Energie ist gleich der relativen Entropie zum Referenzwahrscheinlichkeitsmaß $\mathrm{d}\nu = u_\infty \mathrm{d}x$.

2.4 Entropie und untere Schranken an die Ricci-Krümmung

Eine weitere bemerkenswerte Eigenschaft der Entropie ist die Verbindung zur Ricci-Krümmung. Zur Erinnerung wiederholen wir die Definition einer unteren Schranke an die Ricci-Krümmung und erläutern den Zusammenhang zur Entropie. Sprechen wir im Folgenden von einer riemannschen Mannigfaltigkeit $\mathcal{X}$, so werden wir stets annehmen, dass diese Mannigfaltigkeit zusammenhängend, vollständig und endlich dimensional ist. Mit $T_p\mathcal{X}$ bezeichnen wir den Tangentialraum an $p \in \mathcal{X}$.

Definition 2.4.1. Sei $(\mathcal{X}, g)$ eine riemannsche Mannigfaltigkeit mit Ricci-Krümmungstensor $\mathrm{Ric}_p \colon T_p\mathcal{X} \times T_p\mathcal{X} \to \mathbb{R}$ für $p \in \mathcal{X}$. Wir sagen, die Ricci-Krümmung auf $(\mathcal{X}, g)$ ist von unten durch $\lambda \in \mathbb{R}$ beschränkt, falls für alle $p \in \mathcal{X}$ die Ungleichung

$$\mathrm{Ric}_p(\xi, \xi) \geq \lambda g_p(\xi, \xi)$$

für alle $\xi \in T_p\mathcal{X}$ erfüllt ist.

Bemerkung 2.4.2. Wir werden an dieser Stelle nicht näher auf die Definition und die Eigenschaften des Ricci-Krümmungstensors eingehen. Für eine ausführliche Einführung verweisen wir auf [Lee97] und [Car13]. Ein tieferes Verständnis des Ricci-Krümmungstensors ist nicht notwendig für die folgenden Ausführungen.

Bei bestimmten Anwendungen, unter anderem bei derjenigen, die wir in diesem Kapitel vorstellen, ist es sinnvoll, nicht das Volumenmaß auf der Mannigfaltigkeit $\mathcal{X}$, sondern ein Maß der Form $e^{-V}\mathrm{Vol}$ zu betrachten. In dieser Situation ist eine allgemeinere Definition des Ricci-Krümmungstensors angemessen. Wir werden hier einen Spezialfall dieses verallgemeinerten Ricci-Krümmungstensors vorstellen und verweisen für eine ausführliche Behandlung auf [Vil09, Kapitel 14]. Der Unterschied zu der hier vorgestellten Definition ist, dass die folgende Definition die Dimension der Mannigfaltigkeit außer Acht lässt. In den allgemeineren Ausführungen in [Vil09, Kapitel 14] betrachtet man einen weiteren verallgemeinerten Ricci-Krümmungstensor, der auch von der Dimension abhängt.

Definition 2.4.3. Auf einer riemannschen Mannigfaltigkeit $(\mathcal{X}, g)$, ausgestattet mit dem Referenzmaß $\nu = e^{-V}\mathrm{Vol}$ für ein glattes Potential $V \colon \mathcal{X} \to \mathbb{R}$, definieren wir den verallgemeinerten Ricci-Krümmungstensor

$$\mathrm{Ric}_{\infty,\nu} := \mathrm{Ric} + \nabla^2 V.$$

Hier fassen wir die Hesse-Matrix $\nabla^2 V$ von V als bilineare reelle Abbildung auf. Wir sagen, die Ricci-Krümmung auf $(\mathcal{X}, g, \nu)$ ist von unten durch $\lambda \in \mathbb{R}$ beschränkt, falls für alle $p \in \mathcal{X}$ die Ungleichung

$$\mathrm{Ric}_p(\xi, \xi) + \langle \nabla^2 V(p)\xi, \xi \rangle \geq \lambda g_p(\xi, \xi)$$

für alle $\xi \in T_p\mathcal{X}$ erfüllt ist. In diesem Fall sagen wir, dass $(\mathcal{X}, g, \nu)$ die Krümmungsbedingung $CD(\lambda, \infty)$ erfüllt.

2.4.1 Funktionalungleichungen

Definition 2.4.4 (Fisher-Information)**.** Zu einem gegebenen Referenzwahrscheinlichkeitsmaß ν definieren wir die Abbildung $I_\nu \colon AC_\nu \to \mathbb{R}$ durch die Vorschrift

$$I_\nu(\mu) = \int_{\mathcal{X}} \frac{|\nabla \rho|^2}{\rho} \mathrm{d}\nu.$$

Dabei bezeichnet ρ wieder die Radon-Nikodym-Dichte von μ bezüglich ν.

Theorem 2.4.5. *Wir betrachten eine riemannsche Mannigfaltigkeit $(\mathcal{X}, g)$. Das Referenzwahrscheinlichkeitsmaß sei von der Form $\nu = e^{-V}\mathrm{Vol}$ für ein $V \in C^2(\mathcal{X})$. Erfüllt ferner $(\mathcal{X}, g, \nu)$ die Krümmungsbedingung $CD(\lambda, \infty)$ für ein $\lambda > 0$, dann erfüllen alle Wahrscheinlichkeitsmaße μ der Form $\mu = \rho\nu$ für eine glatte Funktion $\rho \colon \mathbb{R}^n \to \mathbb{R}$ die Ungleichung*

$$\int_{\mathcal{X}} \rho \log \rho \mathrm{d}\nu = H_\nu(\rho) \leq \frac{1}{2\lambda} I_\nu(\rho) = \frac{1}{2\lambda} \int_{\mathcal{X}} \frac{|\nabla \rho|^2}{\rho} \mathrm{d}\nu.$$

Eine Ungleichung dieser Art bezeichnet man auch als logarithmische Sobolev-Ungleichung mit Konstante λ.

Beweis. Einen Beweis und die zugehörigen Ausführungen findet man in [Vil09, Theorem 21.2]. □

Eine Folgerung aus einer allgemeineren Form der obigen logarithmischen Sobolev-Ungleichung ist das folgende Theorem.

Theorem 2.4.6. *Wir betrachten den $\mathbb{R}^n$ mit der euklidischen Struktur als riemannsche Mannigfaltigkeit ausgestattet mit dem Referenzwahrscheinlichkeitsmaß $\mathrm{d}\nu(x) = u_\infty \mathrm{d}x$ für $u_\infty =$*

ce^{-V} für ein glattes Potential $V\colon \mathbb{R}^n \to \mathbb{R}$ und $c = \|e^{-V}\|_{L^1}^{-1}$. Wir setzen voraus, dass $(\mathbb{R}^n, \langle\cdot,\cdot\rangle, \nu)$ die Krümmungsbedingung $CD(\lambda,\infty)$ für ein $\lambda > 0$ erfüllt. Dann gilt für jede glatte Lösung $(u_t)_{t\geq 0}$ der (FPG) zum Anfangswert u_0 die Abschätzung

$$H_\nu(\mu_t) \leq e^{-2\lambda t} H_\nu(\mu_0)$$

für alle $t > 0$. Hier bezeichnet μ_t das Maß der Form $\mathrm{d}\mu_t = \frac{u_t}{u_\infty} u_\infty \mathrm{d}x = u_t \mathrm{d}x$.

Beweis. Wir bemerken, dass für den $\mathbb{R}^n$ mit der euklidischen Struktur das Volumenmaß mit dem Lebesgue-Maß übereinstimmt. Dann ist dieses Theorem ein Spezialfall der Aussage (i) aus [Vil09, Theorem 24.7], wobei $K = \lambda$, $N = \infty$ und $U(\rho) = \rho\log(\rho)$. □

Wir möchten untersuchen, in welchem Zusammenhang dies mit dem Resultat der exponentiell schnellen Konvergenz von Lösungen der (FPG) aus Theorem 2.2.1 steht.

Beweis von Theorem 2.2.1. Wir möchten Theorem 2.4.6 anwenden, dazu müssen wir eine positive untere Schranke an die verallgemeinerte Ricci-Krümmung finden. Die Bakry-Emery-Bedingung kontrolliert den Term $\nabla^2 V$, ferner ist der Ricci-Krümmungstensor des $\mathbb{R}^n$ mit der euklidischen Struktur schon konstant gleich 0. Zusammen folgt

$$\mathrm{Ric}_{\infty,\nu} = \mathrm{Ric} + \nabla^2 V = \nabla^2 V \geq \lambda\,\mathrm{Id}$$

und damit erfüllt $(\mathbb{R}^n, \langle\cdot,\cdot\rangle, ce^{-V}\mathrm{d}x)$ die Krümmungsbedingung $CD(\lambda,\infty)$ mit $\lambda > 0$. Theorem 2.4.6 zeigt dann die exponentielle Konvergenz der Entropie, das heißt, für alle $t > 0$ gilt

$$H_\nu(\mu_t) \leq e^{-2\lambda t} H_\nu(\mu_0).$$

Um daraus eine Schranke für $\|u_t - u_\infty\|$ zu folgern, werden wir wieder die Csiszár-Kullback-Pinsker-Ungleichung anwenden. Es gilt

$$H_\nu(\mu_t) = \int_{\mathbb{R}^n} \frac{u_t}{u_\infty} \log\left(\frac{u_t}{u_\infty}\right) u_\infty \mathrm{d}x = \int_{\mathbb{R}^n} \frac{u_t}{u_\infty} \log\left(\frac{u_t}{u_\infty}\right) u_\infty \mathrm{d}x + \int_{\mathbb{R}^n} \left(-\frac{u_t}{u_\infty} + 1\right) u_\infty \mathrm{d}x,$$

denn

$$\int_{\mathbb{R}^n} \left(-\frac{u_t}{u_\infty} + 1\right) u_\infty \mathrm{d}x = -\int_{\mathbb{R}^n} u_t \mathrm{d}x + \int_{\mathbb{R}^n} u_\infty \mathrm{d}x = -1 + 1 = 0.$$

Folglich haben wir auch eine Abschätzung an die Entropie aus Abschnitt 2.2 mit der Funktion $\varphi(x) := x\log(x) - x + 1$. Auf diese können wir die Csiszár-Kullback-Pinsker-Ungleichung anwenden und erhalten

$$\|u_t - u_\infty\|_{L^1}^2 \leq 2H(u_t) = H_\nu(\mu_t) \leq 2H(\mu_0)e^{-2\lambda t},$$

die gewünschte Ungleichung. □

Bemerkung 2.4.7. Es stellt sich die Frage, welche der beiden vorgestellten Beweisarten von Theorem 2.2.1 mehr Vorteile mit sich bringt. Interessant ist vor allem, welche dieser Techniken auch im diskreten Fall zu Ergebnissen führt. Der zuerst vorgestellte Beweis lässt sich nicht direkt auf den diskreten Fall anwenden, da der Beweis auf Resultaten der kontinuierlichen Differentialrechnung basiert. Eine Kettenregel gibt es für die diskreten Differenzen in dieser Form nicht. Ein Vorteil des Beweises mithilfe der Theorie des optimalen Transports ist seine Allgemeinheit. So ist Theorem 2.4.6 auch gültig für allgemeinere Mannigfaltigkeiten und die Fokker-Planck-Gleichung auf diesen. Ferner kann man mithilfe dieser Theorie auch Aussagen zur Optimalität der Konvergenzkonstanten treffen. Eine weitere schöne Eigenschaft ist, dass sie viele Ansatzpunkte für eine Verallgemeinerung auf metrische Räume gibt. Wie bereits erwähnt, lässt sich diese Theorie jedoch nicht ohne Weiteres auf den diskreten Fall anwenden. Die Frage, warum dies nicht geht, werden wir im nächsten Abschnitt beantworten.

2.4.2 Ricci-Krümmung auf metrischen Maßräumen

Definition 2.4.8. Ein metrischer Raum $(\mathcal{X}, d)$ heißt geodätisch, falls für alle $x_0, x_1 \in \mathcal{X}$ eine Kurve $\gamma\colon [0,1] \to \mathcal{X}$ mit $\gamma_0 = x_0$ und $\gamma_1 = x_1$ existiert, die für alle $s, t \in [0,1]$ die Gleichung

$$d(\gamma_t, \gamma_s) = |t-s| d(x_0, x_1)$$

erfüllt. Eine solche Kurve nennen wir eine Geodäte mit konstanter Geschwindigkeit.

Beispiel 2.4.9. (i) Jede vollständige und zusammenhängende riemannsche Mannigfaltigkeit ist, ausgestattet mit dem riemannschen Abstand nach dem Hopf-Rinow-Theorem, ein geodätischer metrischer Raum.

(ii) Ist $(\mathcal{X}, d)$ ein geodätischer Raum, dann ist auch der metrische Raum $(\mathcal{P}_2(\mathcal{X}), W_2)$ geodätisch. Für einen Beweis verweisen wir auf [Vil09, Kapitel 7].

Theorem 2.4.10. *Sei $\lambda \in \mathbb{R}$ und $(\mathcal{X}, g, \nu)$ eine riemannsche Mannigfaltigkeit mit Referenzmaß $\nu = e^{-V}\mathrm{Vol}$. Dann sind die folgenden Aussagen äquivalent.*

(i) Es gilt $\mathrm{Ric}_{\infty,\nu} \geq \lambda$ auf $\mathcal{X}$.

(ii) Für alle Wahrscheinlichkeitsmaße $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ existiert eine W_2-Geodäte mit konstanter Geschwindigkeit $\mu\colon [0,1] \to \mathcal{P}_2(\mathcal{X})$, sodass für alle $t \in [0,1]$ die Ungleichung

$$H_\nu(\mu_t) \leq (1-t)H_\nu(\mu_0) + tH_\nu(\mu_1) - \frac{\lambda}{2}t(t-1)W_2(\mu_0,\mu_1)^2$$

erfüllt ist.

Beweis. Dieses Theorem ist ein Spezialfall von [Vil09, Theorem 17.15] mit $N = \infty$ und $U(\rho) = \rho\log(\rho)$. Wir bemerken, dass die Aussage (ii) in obigem Theorem, unter Verwendung von [Vil09, Definition 16.5] und der dort nachfolgenden Bemerkung, der Aussage (ii) in [Vil09, Theorem 17.15] entspricht. □

Bemerkung 2.4.11. Im Spezialfall $V \equiv 0$ liefert Theorem 2.4.10 eine Charakterisierung von unteren Schranken an die Ricci-Krümmung aus Definition 2.4.1. Für die Formulierung der Bedingung (i) in Theorem 2.4.10 benötigt man die riemannsche Struktur der Menge $\mathcal{X}$. Für die äquivalente Bedingung (ii) reicht schon eine Metrik auf $\mathcal{X}$ sowie ein Referenzmaß. Diese Beobachtung ist eine geeignete Möglichkeit, um einen synthetischen Begriff von Schranken an die Ricci-Krümmung auf mit einem Borelmaß ausgestatteten metrischen Räumen, den sogenannten metrischen Maßräumen, zu definieren.

Definition 2.4.12. Sei $(\mathcal{X}, d, \nu)$ ein separabler metrischer Maßraum. Wir sagen, die Ricci-Krümmung von $(\mathcal{X}, d, \nu)$ ist von unten durch $\lambda \in \mathbb{R}$ beschränkt, falls alle Wahrscheinlichkeitsmaße $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ durch eine Geodäte mit konstanter Geschwindigkeit $\mu\colon [0,1] \to \mathcal{P}_2(\mathcal{X})$ verbunden werden können, entlang welcher die Ungleichung

$$H_\nu(\mu_t) \leq (1-t)H_\nu(\mu_0) + tH_\nu(\mu_1) - \frac{\lambda}{2}t(t-1)W_2(\mu_0,\mu_1)^2$$

für alle $t \in [0,1]$ erfüllt ist. In diesem Fall schreiben wir $\mathrm{Ric} = \mathrm{Ric}(\mathcal{X}, d\nu) \geq \lambda$.

Lemma 2.4.13. Sei $(\mathcal{X}, d)$ ein diskreter metrischer Raum. In diesem Fall ist jede Geodäte mit konstanter Geschwindigkeit bezüglich der Wasserstein-Metrik W_2 schon konstant.

Beweis. Sei $\mu\colon [0,1] \to \mathcal{P}_2(\mathcal{X})$ eine Geodäte mit konstanter Geschwindigkeit, das heißt, es gelte $W_2(\mu_t, \mu_s) = |t-s|W_2(\mu_0,\mu_1)$ für alle $s, t \in [0,1]$. Wir wählen ein $a \in \mathcal{X}$ und betrachten für $t \in [0,1]$ die Abbildung $\alpha(t) := [\mu_t](\{a\})$. Da $\mathcal{X}$ diskret ist, existiert ein $\varepsilon > 0$, sodass $d(a,y) \geq \varepsilon$ für alle $y \in \mathcal{X} \setminus \{a\}$ ist. Zu gegebenen $s,t \in [0,1]$, sei $\Gamma \in \Pi(\mu_s,\mu_t)$ eine beliebige Kopplung. Das heißt, es gilt $\Gamma(A \times \mathcal{X}) = \mu_s(A)$ sowie $\Gamma(\mathcal{X} \times A) = \mu_t(A)$. Da Γ eine Kopplung ist, gilt $\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y) < \infty$. Unter diesen Voraussetzungen folgt

$$\begin{aligned}
\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y) &= \sum_{x,y\in\mathcal{X}} d(x,y)^2 \Gamma(\{x\}\times\{y\}) = \sum_{x\in\mathcal{X}}\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} d(x,y)^2\Gamma(\{x\}\times\{y\}) \\
&\geq \varepsilon^2 \sum_{x\in\mathcal{X}}\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} \Gamma(\{x\}\times\{y\}) \geq \varepsilon^2 \sum_{\substack{y\in\mathcal{X}\\ y\neq a}} \Gamma(\{a\}\times\{y\}) \\
&= \varepsilon^2 \left(\Gamma(\{a\}\times\mathcal{X}) - \Gamma(\{a\}\times\{a\})\right) \geq \varepsilon^2 \left(\Gamma(\{a\}\times\mathcal{X}) - \Gamma(\mathcal{X}\times\{a\})\right)
\end{aligned}$$

$$= \varepsilon^2 \left(\alpha(s) - \alpha(t)\right)$$

unter anderem aufgrund der Nichtnegativität der Kopplung Γ. Wenden wir in der zweiten Gleichung den Satz von Fubini an, dann erhalten wir

$$\begin{aligned}
\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{X}} d(x,y)^2 \Gamma(\{x\}\times\{y\}) = \sum_{y\in\mathcal{X}} \sum_{x\in\mathcal{X}} d(x,y)^2 \Gamma(\{x\}\times\{y\}) \\
&= \sum_{y\in\mathcal{X}} \sum_{\substack{x\in\mathcal{X}\\ x\neq y}} d(x,y)^2 \Gamma(\{x\}\times\{y\}) \geq \varepsilon^2 \sum_{y\in\mathcal{X}} \sum_{\substack{x\in\mathcal{X}\\ x\neq y}} \Gamma(\{x\}\times\{y\}) \\
&\geq \varepsilon^2 \sum_{\substack{x\in\mathcal{X}\\ x\neq a}} \Gamma(\{x\}\times\{a\}) = \varepsilon^2 \left(\Gamma(\mathcal{X}\times\{a\}) - \Gamma(\{a\}\times\{a\})\right) \\
&\geq \varepsilon^2 \left(\Gamma(\mathcal{X}\times\{a\}) - \Gamma(\{a\}\times\mathcal{X})\right) = \varepsilon^2 \left(\alpha(t) - \alpha(s)\right).
\end{aligned}$$

Insgesamt folgt für feste $s, t \in [0,1]$ und jede beliebige Kopplung $\Gamma \in \Pi(\mu_s, \mu_t)$

$$\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y) \geq \varepsilon |\alpha(s) - \alpha(t)|$$

und damit auch

$$W_2(\mu(s), \mu(t)) = \inf_{\Gamma\in\Pi(\mu(s),\mu(t))} \left(\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \mathrm{d}\Gamma(x,y)\right)^{\frac{1}{2}} \geq \varepsilon \sqrt{|\alpha(s) - \alpha(t)|}.$$

Unter Verwendung der konstanten Geschwindigkeit der Geodäte μ erhalten wir

$$|\alpha(s) - \alpha(t)| \leq \frac{1}{\varepsilon^2} W_2(\mu(s), \mu(t))^2 = \frac{1}{\varepsilon^2} |s - t|^2,$$

das heißt, die Abbildung $[0,1] \to \mathbb{R}$, $\alpha(t) = \mu_t(x)$ ist Hölder-stetig mit Exponent 2. Jede solche Abbildung ist schon konstant. Da $a \in \mathcal{X}$ beliebig war, ist μ_t konstant und somit ist jede Geodäte schon konstant. □

Bemerkung 2.4.14. Wir sehen, dass wir die Theorie des optimalen Transports nicht direkt auf das diskrete Setting anwenden können, um etwa Aussagen über die Konvergenz der Entropie mittels der verallgemeinerten Ricci-Krümmung treffen zu können. Diesem Defekt werden wir uns in den Kapiteln 4 und 5 widmen und ihn mittels einer neuen diskreten Wasserstein-Metrik beheben.

3 Markov-Halbgruppen

3.1 Eine kurze Einführung

In diesem Abschnitt geben wir eine Einführung in die Theorie der Markov-Halbgruppen auf einem messbaren Zustandsraum $(\mathcal{X}, \Sigma)$. Wir orientieren uns dabei vor allem an [BGL14]. Wir werden sehen, in welchem Zusammenhang Markov-Halbgruppen und die sogenannten Markov-Prozesse stehen. Für eine Definition dieser Markov-Prozesse verweisen wir auf Kapitel A des Anhangs. Genauer werden wir zeigen, dass eine Markov-Kette, das heißt ein Markov-Prozess, der nur abzählbar viele Werte annimmt, unter bestimmten Voraussetzungen mit einer Markov-Halbgruppe identifiziert werden kann. Die Theorie der Markov-Halbgruppen steht in engem Zusammenhang zur Wärmeleitungsgleichung auf Graphen. Diesen Zusammenhang werden wir in Abschnitt 3.3 genauer studieren. In Abschnitt 3.4 stellen wir vor, wie eine diskrete Fokker-Planck-Gleichung aussehen könnte.

Definition 3.1.1 (Markov-Operator)**.** Sei $(\mathcal{X}, \Sigma)$ ein messbarer Raum. Ein linearer Operator $P\colon \mathcal{M}_b(\mathcal{X}, \Sigma) \to \mathcal{M}_b(\mathcal{X}, \Sigma)$ heißt Markov-Operator, falls die folgenden Eigenschaften erfüllt sind.

(i) Die Funktion $\mathbb{1}\colon \mathcal{X} \to \mathbb{R},\ x \mapsto 1$ erfüllt $P\mathbb{1} = \mathbb{1}$. (Markov-Eigenschaft)

(ii) Ist $f \geq 0$, dann folgt $Pf \geq 0$. (Positivität)

Wir möchten an dieser Stelle einige Eigenschaften von Markov-Operatoren vorstellen.

Lemma 3.1.2. Jeder Markov-Operator ist ein stetiger linearer Operator auf dem Raum der messbaren beschränkten Funktionen $\mathcal{M}_b(\mathcal{X}, \Sigma)$, ausgestattet mit der Supremumsnorm.

Beweis. Sei P ein Markov-Operator und $f\colon \mathcal{X} \to \mathbb{R}$ messbar mit $\|f\|_\infty \leq 1$. Teile f in Positiv- und Negativteil auf, das heißt, $f = f_+ - f_-$ für messbare Funktionen $f_+ \geq 0$ und $f_- \geq 0$. Unter dieser Zerlegung gilt sowohl $(\mathbb{1} - f_+) \geq 0$ als auch $(\mathbb{1} - f_-) \geq 0$ und folglich $P(\mathbb{1} - f_+) \geq 0$ sowie $P(\mathbb{1} - f_-) \geq 0$. Aus der Linearität folgt dann

$$-\mathbb{1} = -P(\mathbb{1}) \leq -P(f_-) \leq P(f_+) - P(f_-) = P(f) \leq P(f_+) \leq P(\mathbb{1}) = \mathbb{1},$$

da $P(f_+) \geq 0$ und $P(f_-) \geq 0$ ist. Somit gilt $\|Pf\|_\infty \leq 1$ für alle $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$ mit $\|f\|_\infty \leq 1$, dies zeigt die Stetigkeit des Markov-Operators P. □

Lemma 3.1.3. Sei $P\colon \mathcal{M}_b(\mathcal{X}, \Sigma) \to \mathcal{M}_b(\mathcal{X}, \Sigma)$ ein Markov-Operator und $\varphi\colon \mathbb{R} \to \mathbb{R}$ konvex. Dann gilt für alle $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$ die Ungleichung

$$\varphi(Pf) \leq P(\varphi(f)).$$

Beweis. Wir zeigen die Aussage zunächst für Treppenfunktionen. Seien $A_1, \dots, A_n \in \Sigma$ eine disjunkte Zerlegung von $\mathcal{X}$. Wir betrachten f von der Form $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ für Konstanten $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Für alle $i \in \{1, \dots, n\}$ folgt aufgrund der Positivität aus $\mathbb{1}_{A_i} \geq 0$ auch $P(\mathbb{1}_{A_i}) \geq 0$. Weiter gilt $\sum_{i=1}^n P(\mathbb{1}_{A_i}) = P(\sum_{i=1}^n \mathbb{1}_{A_i}) = P(\mathbb{1}) = \mathbb{1}$ als Konsequenz der Linearität und der Markov-Eigenschaft. Sei $x \in \mathcal{X}$, dann gilt für $\xi_i := [P(\mathbb{1}_{A_i})](x)$

$$[\varphi(Pf)](x) = \varphi\left(\frac{\sum\limits_{i=1}^n \alpha_i [P(\mathbb{1}_{A_i})](x)}{1}\right) = \varphi\left(\frac{\sum\limits_{i=1}^n \alpha_i \xi_i}{\sum\limits_{i=1}^n \xi_i}\right) \leq \frac{\sum\limits_{i=1}^n \xi_i \varphi(\alpha_i)}{1} = [P(\varphi(f))](x)$$

nach der Jensenschen Ungleichung. Sei nun $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$ beliebig, dann existiert nach Satz C.0.1 eine Folge von Treppenfunktionen $(t_n)_{n \in \mathbb{N}}$, die gleichmäßig gegen f konvergiert. Dann folgt die Behauptung aus der Stetigkeit von P und aus der Stetigkeit konvexer Funktionen zusammen mit obiger Ungleichung für Treppenfunktionen. Genauer gilt auf $\mathcal{X}$

$$\varphi(Pf) = \varphi\left(P \lim_{n \to \infty} t_n\right) = \lim_{n \to \infty} \varphi(Pt_n) \leq \lim_{n \to \infty} P(\varphi(t_n)) = P(\varphi(f)).$$

Diese Ungleichung zeigt die Behauptung. □

Beispiel 3.1.4 (Markov-Ketten Teil 1)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Wir betrachten eine zeithomogene Markov-Kette $(X_t)_{t \geq 0}$ auf einem höchstens abzählbar unendlichen Zustandsraum $(\mathcal{X}, \Sigma)$ der Form $\mathcal{X} = \{x_i\colon i \in \mathbb{N}\}$. Für $i, j \in \mathbb{N}$, $s \geq 0$ und $t > 0$ definieren wir die Zahlen

$$p_t(i,j) = \mathbb{P}(X_{s+t} = x_j | X_s = x_i) = \mathbb{P}(X_t = x_j | X_0 = x_i) \geq 0.$$

Die zweite Gleichung gilt nach Definition A.0.7, da die Zahlen $p_t(i,j)$ nicht von s abhängen. Diese (unendliche) Matrix spielt eine wichtige Rolle in der Analyse von Markov-Ketten. Wir definieren für $t \geq 0$ den Operator

$$P(t) : \mathcal{M}_b(\mathcal{X}, \Sigma) \to \mathcal{M}_b(\mathcal{X}, \Sigma),\ [P(t)f](x_i) = \sum_{j=1}^{\infty} p_t(i,j) f(x_j).$$

Dann ist $P(t)$ für jedes $t \geq 0$ ein wohldefinierter Markov-Operator. Ferner gilt $P(0) = \mathrm{Id}$.

Beweis. Wir betrachten zunächst die Funktion $\mathbb{1} \colon \mathcal{X} \to \mathbb{R}$, $x \mapsto 1$, dann gilt für $i \in \mathbb{N}$

$$\begin{aligned}
[P(t)\mathbb{1}](x_i) &= \sum_{j=1}^{\infty} p_t(i,j)\mathbb{1}(x_j) = \sum_{j=1}^{\infty} p_t(i,j) = \mathbb{P}\left(\bigcup_{j=1}^{\infty}\{X_t = x_j\}\middle| X_0 = x_i\right) \\
&= \mathbb{P}(X_t \in \mathcal{X} | X_0 = x_i) = 1.
\end{aligned}$$

Dies zeigt die Markov-Eigenschaft des Operators $P(t)$. Sei $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$, dann gilt für $i \in \mathbb{N}$

$$|[P(t)f](x_i)| \leq \sum_{j=1}^{\infty} p_t(i,j)|f(x_j)| \leq \|f\|_\infty \sum_{j=1}^{\infty} p_t(i,j) = \|f\|_\infty$$

und folglich ist $P(t)f$ eine beschränkte Funktion. Ferner ist $P(t)f$ als punktweiser Grenzwert messbarer Funktionen wieder messbar. Insgesamt ist $P(t)f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$. Die Linearität der $P(t)$ ist die Konsequenz der Linearität der Summe und des Grenzwerts. Wählen wir ein beliebiges $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$ mit $f \geq 0$, dann gilt für $i \in \mathbb{N}$

$$[P(t)f](x_i) = \lim_{n\to\infty} \sum_{j=1}^{n} p_t(i,j)f(x_j) \geq 0,$$

da $p_t(i,j) \geq 0$ für alle $i, j \in \mathbb{N}$ und $t \geq 0$ ist. Insgesamt folgt, dass $P(t)$ ein Markov-Operator ist. Betrachte nun $P(0)$, es gilt für $i, j \in \mathbb{N}$

$$p_0(i,j) = \mathbb{P}(X_0 = x_j | X_0 = x_i) = \delta_{ij}$$

und folglich ist $P(0) = \mathrm{Id}$. $\square$

Definition 3.1.5 (Invariantes Maß)**.** Ein σ-endliches Maß μ auf $(\mathcal{X}, \Sigma)$ nennen wir invariantes Maß zur Operatorenfamilie $(P(t))_{t\geq 0}$, falls für alle $t \geq 0$ und für alle $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$

$$\int_{\mathcal{X}} P(t)f \mathrm{d}\mu = \int_{\mathcal{X}} f \mathrm{d}\mu$$

gilt.

Bemerkung 3.1.6. Sei μ ein invariantes Maß zur Operatorenfamilie $(P(t))_{t\geq 0}$ und $1 \leq p < \infty$. Wählen wir in Lemma 3.1.3 die konvexe Funktion $\varphi(x) = |x|^p$, dann folgt

$$\|P(t)f\|_{L^p}^p = \int_{\mathcal{X}} |P(t)f|^p \mathrm{d}\mu \leq \int_{\mathcal{X}} [P(t)](|f|^p) \mathrm{d}\mu = \int_{\mathcal{X}} |f|^p \mathrm{d}\mu = \|f\|_{L^p}^p$$

für alle $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$, da μ ein invariantes Maß ist. Es liegt $\mathcal{M}_b(\mathcal{X}, \Sigma)$ dicht in $L^p(\mu)$ bezüglich der $\|\cdot\|_{L^p}$ Norm und folglich können wir $P(t)$ zu einer Kontraktion auf L^p für alle $1 \leq p < \infty$ fortsetzen.

Definition 3.1.7 (Markov-Halbgruppe)**.** Sei $(\mathcal{X}, \Sigma)$ ein messbarer Raum und $(P(t))_{t\geq 0}$ eine Familie von linearen Operatoren auf den beschränkten messbaren Funktionen. Sei μ ein invariantes Maß zu $(P(t))_{t\geq 0}$. Die Familie $(P(t))_{t\geq 0}$ heißt Markov-Halbgruppe, falls sie die folgenden Eigenschaften erfüllt:

(i) $P(t)$ ist für jedes $t \geq 0$ ein Markov-Operator. (Markov-Eigenschaft)

(ii) Es gilt $P(0) = \mathrm{Id}$. (Anfangsbedingung)

(iii) Für alle $s, t \geq 0$ gilt $P(s+t) = P(s)P(t)$. (Halbgruppeneigenschaft)

(iv) Es gilt $\lim_{t\to 0} P(t)f = f$ in $L^2(\mu)$ für alle $f \in L^2(\mu)$. (C_0-Eigenschaft)

Bemerkung 3.1.8. Meistens wird im Rahmen der Definition einer Markov-Halbgruppe noch eine Bedingung an den messbaren Raum $(\mathcal{X}, \Sigma)$ gestellt. Wir werden in dieser Ausführung auf diese Bedingung verzichten. Das hat den Grund, dass wir in den folgenden Kapiteln nur endliche Mengen $\mathcal{X}$ untersuchen werden. Bei diesen Mengen müssen wir uns nicht mit der Regularität des messbaren Raums beschäftigen.

Beispiel 3.1.9 (Markov-Ketten Teil 2)**.** Die Operatorenfamilie $(P(t))_{t\geq 0}$ zu einer Markov-Kette aus Beispiel 3.1.4 erfüllt die Halbgruppeneigenschaft. Das heißt, für alle $s, t \geq 0$ gilt $P(s+t) = P(s)P(t)$.

Beweis. Seien $s, t \geq 0$ beliebig. Wir betrachten $i, j \in \mathbb{N}$, sodass $\mathbb{P}(X_0 = i) > 0$ gilt, denn sonst ist nichts zu zeigen. In dieser Situation gilt die Gleichung

$$
\begin{aligned}
&\sum_{k\in\mathbb{N}} p_s(i,k)p_t(k,j) = \sum_{k\in\mathbb{N}} \mathbb{P}(\{X_s = x_k\}|\{X_0 = x_i\})\mathbb{P}(\{X_t = x_j\}|\{X_0 = x_k\}) \\
&= \sum_{k\in\mathbb{N}} \mathbb{P}(\{X_s = x_k\}|\{X_0 = x_i\})\mathbb{P}(\{X_{s+t} = x_j\}|\{X_s = x_k\}) \\
&= \sum_{k\in\mathbb{N}} \mathbb{P}(\{X_s = x_k\}|\{X_0 = x_i\})\mathbb{P}(\{X_{s+t} = x_j\}|\{X_s = x_k\} \cap \{X_0 = x_i\}) \\
&= \sum_{k\in\mathbb{N}} \frac{\mathbb{P}(\{X_s = x_k\} \cap \{X_0 = x_i\})}{\mathbb{P}(\{X_0 = x_i\})} \frac{\mathbb{P}(\{X_{s+t} = x_j\} \cap \{X_s = x_k\} \cap \{X_0 = x_i\})}{\mathbb{P}(\{X_s = x_k\} \cap \{X_0 = x_i\})} \\
&= \sum_{k\in\mathbb{N}} \frac{\mathbb{P}(\{X_{s+t} = x_j\} \cap \{X_0 = x_i\} \cap \{X_s = x_k\})}{\mathbb{P}(\{X_0 = x_i\})} \\
&= \frac{\mathbb{P}(\{X_{s+t} = x_j\} \cap \{X_0 = x_i\})}{\mathbb{P}(\{X_0 = x_i\})} = \mathbb{P}(\{X_{s+t} = x_j\}|\{X_0 = x_i\}) = p_{s+t}(i,j).
\end{aligned}
$$

Dabei betrachten wir in den Summen nur die $k \in \mathbb{N}$, für die $\mathbb{P}(\{X_s = k\} \cap \{X_0 = i\}) > 0$ gilt. In der dritten Zeile verwenden wir die Markov-Eigenschaft der Markov-Kette und in der sechsten Zeile benutzen wir die σ-Additivität von $\mathbb{P}$. Obige Gleichung wird auch Chapman-Kolmogorov-Gleichung genannt. Ist $f \in \mathcal{M}_b(\mathcal{X}, \Sigma)$ und $i \in \mathbb{N}$, dann folgt

$$\begin{aligned}
[P(s)P(t)f](x_i) &= \left(\sum_{k \in \mathbb{N}} p_s(i,k) \left(\sum_{j \in \mathbb{N}} p_t(k,j) f(x_j) \right) \right) = \sum_{k \in \mathbb{N}} \sum_{j \in \mathbb{N}} p_s(i,k) p_t(k,j) f(x_j) \\
&= \sum_{j \in \mathbb{N}} \sum_{k \in \mathbb{N}} p_s(i,k) p_t(k,j) f(x_j) = \sum_{j \in \mathbb{N}} f(x_j) \sum_{k \in \mathbb{N}} p_s(i,k) p_t(k,j) \\
&= \sum_{j \in \mathbb{N}} p_{s+t}(i,j) f(x_j) = [P(s+t)f](x_i)
\end{aligned}$$

und somit die Behauptung. Hierbei rechtfertigt die Endlichkeit der letzten Summe die Vertauschung der Summationsreihenfolge nach dem Satz von Fubini. □

Beispiel 3.1.10 (Markov-Ketten Teil 3)**.** Wir betrachten die Familie $(P(t))_{t \geq 0}$ der Operatoren zu einer Markov-Kette wie in Beispiel 3.1.4. Wir werden zusätzlich annehmen, dass der Zustandsraum $\mathcal{X}$ endlich ist, das heißt, $\mathcal{X} = \{x_1, \dots, x_n\}$. Unter diesen Voraussetzungen gilt $\lim\limits_{t \to 0} p_t(i,j) = \delta_{ij}$ für alle $i, j \in \{1, \dots, n\}$. Insbesondere folgt, dass $(P(t))_{t \geq 0}$ eine C_0-Halbgruppe auf $(\mathcal{M}_b(\mathcal{X}, \mathscr{P}(\mathcal{X})), \|\cdot\|_\infty)$ ist.

Beweis. Wir betrachten für $i \in \{1, \dots, n\}$ und $l \in \mathbb{N}$ die Mengen

$$A(i,l) := \left\{ \omega \in \Omega \colon \exists t_0 \in \left[0, \frac{1}{l}\right], \text{sd. für den in } \xi_i \in \mathcal{X} \text{ startenden Pfad } t \mapsto X_t \text{ gilt } X_{t_0}(\omega) = \xi_i \right\}.$$

Es gilt $A(i,l) \subset A(i,k)$ für alle $l \leq k$ und zusätzlich $\Omega = \bigcup\limits_{l=1}^{\infty} A(i,l)$ aufgrund der rechtsseitigen Stetigkeit der Pfade sowie der Diskretheit von $\mathcal{X}$. Aus der Stetigkeit von Maßen bezüglich Ausschöpfungen folgt damit

$$\lim_{l \to \infty} \mathbb{P}(A(i,l)) = 1.$$

Es gilt $\mathbb{P}(A(i,n)) \leq p_t(i,i)$ für alle $t \leq \frac{1}{n}$ und folglich ist $\lim\limits_{t \to 0^+} p_t(i,i) = 1$. Seien nun $i, j \in \{1, \dots, n\}$ und $i \neq j$, dann gilt $\sum\limits_{k=1}^{n} p_t(i,k) = 1$ und zusammen mit $p_t(i,j) \geq 0$ folgt $0 \leq p_t(i,j) \leq 1 - p_t(i,i)$ für alle $t \geq 0$. Damit ist

$$0 \leq \lim_{t \to 0^+} p_t(i,j) \leq 1 - \lim_{t \to 0^+} p_t(i,i) = 0.$$

Da $\mathcal{X}$ endlich ist, ist die Konvergenz in $\mathcal{M}_b(\mathcal{X}, \mathscr{P}(\mathcal{X}))$ äquivalent zur punktweisen Konvergenz. Dies zeigt die Behauptung. □

Wir haben gesehen, dass die Operatorenfamilie $(P(t))_{t\geq 0}$ zu einer Markov-Kette eine C_0-Halbgruppe definiert. Unser nächstes Ziel ist es zu zeigen, dass diese Operatorenfamilie, unter gewissen Voraussetzungen, auch eine Markov-Halbgruppe im Sinne der Definition 3.1.7 ist. Es fehlen dazu ein passendes invariantes Maß μ und die Stetigkeit in 0 bezüglich der $L_2(\mu)$ Topologie. Dazu werden wir ein neues Objekt, die sogenannten Markov-Kerne, einführen und untersuchen.

3.2 Markov-Tripel

Konvention: Zu einer gegebenen endlichen Menge $\mathcal{X}$ wählen wir die Potenzmenge als σ-Algebra. Wir möchten weiter annehmen, dass $\mathcal{X}$ mit einer nicht näher spezifizierten Metrik d ausgestattet ist. Es gelte $\mathcal{X} = \{x_1, \ldots, x_n\}$ für $x_i \neq x_j$, falls $i \neq j$ ist, das heißt, es ist insbesondere $\#\mathcal{X} = n$. Jede lineare Abbildung Q auf dem Raum $\mathbb{R}^{\mathcal{X}}$ identifizieren wir mit der Darstellung als Matrix zur Basis $(\mathbb{1}_{\{x_1\}}, \ldots, \mathbb{1}_{\{x_n\}})$. Die Einträge dieser Matrix bezeichnen wir mit $Q(i,j)$ für $i, j \in \{i, \ldots, n\}$. Oft schreiben wir auch $Q(x,y)$ für $x, y \in \mathcal{X}$ und definieren $\mathcal{M}_b(\mathcal{X}) := \mathcal{M}_b(\mathcal{X}, \mathscr{P}(\mathcal{X}))$.

Bemerkung 3.2.1. Man könnte anstelle von einer beliebigen endlichen Menge $\mathcal{X}$ auch die Menge der natürlichen Zahlen $\{1, \ldots, n\}$ betrachten. Wir werden meistens jedoch die Notation $\mathcal{X}$ verwenden, um zu unterstreichen, dass $\mathcal{X}$ auch noch mehr Struktur besitzen kann. So werden wir zum Beispiel den Fall untersuchen, in dem die Menge $\mathcal{X}$ die Menge der Ecken eines Graphs ist.

Definition 3.2.2 (Markov-Kern)**.** Wir nennen eine Matrix $Q \in \mathbb{R}^{n\times n}$ Markov-Kern auf $\mathcal{X}$, falls sie die folgenden Eigenschaften erfüllt:

(i) Es gilt $Q(i,j) \geq 0$ für alle $i, j \in \{1, \ldots, n\}$ mit $i \neq j$.

(ii) Die Gleichung $\sum\limits_{j=1}^{n} Q(i,j) = 0$ gilt für alle $i \in \{1, \ldots, n\}$.

Beispiel 3.2.3 (Markov-Ketten Teil 4)**.** Wir betrachten die C_0-Halbgruppe aus Beispiel 3.1.4 auf einem endlichen Zustandsraum $\mathcal{X}$. Dann ist aufgrund der endlichen Dimension von $\mathcal{M}_b(\mathcal{X})$ die Familie $(P(t))_{t\geq 0}$ eine normstetige Halbgruppe. Jede normstetige Halbgruppe $(P(t))_{t\geq 0}$ ist schon von der Form $P(t) = e^{tQ}$ für einen beschränkten linearen Operator $Q\colon \mathcal{M}_b(\mathcal{X}) \to \mathcal{M}_b(\mathcal{X})$. Einen Beweis für diese Argumentation findet man zum Beispiel in [EN06, Kapitel 1]. Der Operator Q ist der Generator der C_0-Halbgruppe. Seien

$i, j \in \{1, \dots, n\}$ beliebig, dann existiert der Grenzwert

$$Q(i,j) = [Q\mathbb{1}_{\{x_j\}}](x_i) = \left[\lim_{h \to 0^+} \frac{P(t)\mathbb{1}_{\{x_j\}} - \mathbb{1}_{\{x_j\}}}{h}\right](x_i).$$

Ist $i = j$, dann gilt

$$Q(i,i) = \lim_{h \to 0^+} \frac{p_h(i,i) - 1}{h}$$

und es ist

$$Q(i,j) = \lim_{h \to 0^+} \frac{p_h(i,j)}{h}$$

für $i \neq j$. Dabei verwenden wir, dass die Punktauswertung aufgrund der endlichen Dimension stetig ist. Die Matrixdarstellung von Q ist ein Markov-Kern.

Beweis.

(i) Seien $i, j \in \{1, \dots, n\}$ sowie $i \neq j$, dann gilt

$$Q(i,j) = \lim_{h \to 0^+} \frac{p_h(i,j)}{h} \geq 0,$$

denn nach Definition ist $p_h(i,j) \geq 0$ für alle $i, j \in \{1, \dots, n\}$ und alle $h \geq 0$.

(ii) Sei $i \in \{1, \dots, n\}$. Wie in Beispiel 3.1.4 gezeigt wurde, gilt für jedes $t \geq 0$ die Gleichung $\sum_{j=1}^{n} p_t(i,j) = 1$. Die rechtsseitige Differentiation beider Seiten in 0 liefert dann

$$\sum_{j=1}^{n} Q(i,j) = 0.$$

□

Das folgende Lemma gibt eine Charakterisierung des invarianten Maßes einer Markov-Halbgruppe. Diese Charakterisierung wird bei der Suche nach einem invarianten Maß zu einer C_0-Halbgruppe hilfreich sein.

Lemma 3.2.4. Sei $(P(t))_{t \geq 0}$ eine Markov-Halbgruppe auf dem Raum $\mathcal{M}_b(\mathcal{X})$ mit Generator $Q \in \mathbb{R}^{n \times n}$ und sei μ ein Maß auf $\mathcal{X}$, dann sind die folgenden Aussagen äquivalent:

(i) μ ist ein invariantes Maß zur Operatorenfamilie $(P(t))_{t \geq 0}$.

(ii) Es gilt $\int_{\mathcal{X}} Qf \,\mathrm{d}\mu = 0$ für alle $f \in \mathcal{M}_b(\mathcal{X})$.

(iii) Fassen wir das Maß μ als Vektor des $\mathbb{R}^n$ auf, dann gilt $\mu^T Q = 0$.

Beweis.

(i) ⇒ (ii) Sei μ ein invariantes Maß, dann gilt für jede Funktion $f \in \mathcal{M}_b(\mathcal{X})$

$$\int_{\mathcal{X}} P(t) f \mathrm{d}\mu = \int_{\mathcal{X}} f \mathrm{d}\mu$$

und folglich für jedes $t > 0$ auch

$$\int_{\mathcal{X}} \frac{P(t)f - f}{t} \mathrm{d}\mu = 0.$$

Es konvergiert $\frac{P(t)f-f}{t}$ gegen Qf für $t \to 0^+$ und folglich gilt auch

$$\int_{\mathcal{X}} Qf \mathrm{d}\mu = \int_{\mathcal{X}} \lim_{t\to 0^+} \frac{P(t)f - f}{t} \mathrm{d}\mu = \lim_{t\to 0^+} \int_{\mathcal{X}} \frac{P(t)f - f}{t} \mathrm{d}\mu = 0.$$

Dabei ist das Vertauschen der Grenzwerte aufgrund der Endlichkeit von $\mathcal{X}$ gerechtfertigt.

(ii) ⇒ (iii): Angenommen, es gilt $\int_{\mathcal{X}} Qf \mathrm{d}\mu = 0$ für alle $f \in \mathcal{M}_b(\mathcal{X})$. Für $f = \mathbb{1}_{\{x_j\}}$ mit $j \in \{1, \dots, n\}$ können wir dies aufgrund der Endlichkeit von $\mathcal{X}$ auch als

$$0 = \sum_{i=1}^{n} [Q\mathbb{1}_{\{x_j\}}](x_i)\mu(\{x_i\}) = \sum_{i=1}^{n}\sum_{k=1}^{n} Q(i,k)\delta_{jk}\mu(\{x_i\}) = \sum_{i=1}^{n} \mu(\{x_i\})Q(i,j)$$

schreiben. Ein Maß μ auf der endlichen Menge $\mathcal{X}$ ist eindeutig durch die Zahlen $\mu_i := \mu(\{x_i\})$ bestimmt. Bezeichnen wir mit $\mu = (\mu_1, \dots, \mu_n)^T$ und identifizieren wir die lineare Abbildung Q mit ihrer Darstellung als Matrix, dann gilt $\mu^T Q = 0$.

(iii) ⇒ (i): Gilt $\mu^T Q = 0$, dann folgt unter Verwendung der Generatoren-Eigenschaft $\frac{\mathrm{d}}{\mathrm{d}t}P(t) = QP(t)$ die Gleichung

$$0 = \mu^T QP(t) = \mu^T \frac{\mathrm{d}}{\mathrm{d}t} P(t) = \frac{\mathrm{d}}{\mathrm{d}t} \mu P(t).$$

Aus diesem Grund ist $\mu^T P(t)$ ein konstanter Vektor. Es ist $P(0) = \mathrm{Id}$ und deshalb folgt $\mu^T P(t) = \mu^T$ für alle $t \geq 0$. Sei $j \in \{1, \dots, n\}$, dann folgt

$$\int_{\mathcal{X}} P(t)\mathbb{1}_{\{x_j\}} \mathrm{d}\mu = \sum_{i=1}^{n} \mu(\{x_i\})[P(t)](i,j) = \mu(\{x_j\}) = \int_{\mathcal{X}} \mathbb{1}_{\{x_j\}} \mathrm{d}\mu$$

für alle $t \geq 0$. Für ein beliebiges $f \in \mathcal{M}_b(\mathcal{X})$ folgt dann aufgrund der Linearität des Integrals und der $P(t)$ sowie obiger Gleichung $\int_{\mathcal{X}} P(t) f \mathrm{d}\mu = \int_{\mathcal{X}} f \mathrm{d}\mu$ für alle $t \geq 0$.

Dies zeigt die Invarianz des Maßes μ.

$\square$

Theorem 3.2.5. *Ein Markov-Kern induziert über die Zuordnung $P(t) = e^{tQ}$ eine Markov-Halbgruppe. Ferner ist jede Markov-Halbgruppe auf einem endlichen Zustandsraum schon von dieser Form. Insbesondere existiert zu jedem Markov-Kern ein invariantes Maß.*

Beweis. Man rechnet leicht nach, dass die Operatorenfamilie $(P(t))_{t\geq 0}$ die Halbgruppeneigenschaft erfüllt. Auch die Anfangsbedingung $P(0) = \mathrm{Id}$ ist erfüllt. Es gilt ferner

$$P(t)\mathbb{1} = \sum_{k=0}^{\infty} \frac{t^k Q^k}{k!}\mathbb{1} = \mathbb{1} + \sum_{k=1}^{\infty} \frac{t^k Q^k}{k!}\mathbb{1} = \mathbb{1},$$

denn $Q^k\mathbb{1} = 0$ für alle $k \geq 1$. Wir widmen uns jetzt der fehlenden Eigenschaft (iv) der Definition einer Markov-Halbgruppe. Dazu benötigen wir ein invariantes Maß μ auf $\mathcal{X}$. Wählen wir $\alpha > \max_{i\in\{1,\ldots,n\}}|Q(i,i)|$, dann folgt, dass die Matrix $A := \mathrm{Id} + \frac{1}{\alpha}Q$ nichtnegativ ist, insbesondere ist auch A^t nichtnegativ. Mit dieser Wahl von α gilt $A(i,j) \in [0,1]$ für alle $i,j \in \{1,\ldots,n\}$. Die Matrix A ist zeilen-stochastisch und somit gilt nach Lemma C.0.3 $\rho(A) = 1$, daraus folgt dann $\rho(A^t) = 1$. Aus dem Satz von Perron-Frobenius C.0.2 folgt, dass ein $\mu \geq 0$ existiert, sodass $A^t\mu = \rho(A^t)\mu = \mu$ gilt. Nach der Definition von A gilt dann schon $Q^t\mu = 0$. Der Vektor μ ist nichtnegativ und definiert folglich ein Maß auf $\mathcal{X}$. In Lemma 3.2.4 haben wir gezeigt, dass die Bedingung $\mu^t Q = 0$ für die Invarianz des Maßes hinreichend ist. Es wurde bereits gezeigt, dass $(P(t))_{t\geq 0}$ eine C_0-Halbgruppe auf $\mathcal{M}_b(\mathcal{X})$ ist. Betrachten wir nun den Raum $L_2(\mathcal{X})$ und sei $f \in L_2(\mathcal{X})$, dann gilt

$$\lim_{t\to 0} \|P(t)f - f\|_2 \leq \lim_{t\to 0} \|P(t)f - f\|_\infty \mu(\mathcal{X}) = 0.$$

Es folgt die Behauptung. Wir können sogar μ als Wahrscheinlichkeitsmaß wählen, dazu normieren wir den zugehörigen Vektor in der $\|\cdot\|_1$ Norm.

Sei nun $(P(t))_{t\geq 0}$ eine Markov-Halbgruppe, dann ist aufgrund der endlichen Dimension und der daraus resultierenden Normstetigkeit die Halbgruppe schon von der Form $P(t) = e^{tQ}$ für einen linearen Operator $Q\colon \mathcal{M}_b(\mathcal{X}) \to \mathcal{M}_b(\mathcal{X})$. Wir müssen noch zeigen, dass der lineare Operator Q ein Markov-Kern ist. Es gilt

$$Q\mathbb{1} = \lim_{t\to 0^+} \frac{P(t)\mathbb{1} - \mathbb{1}}{t} = \lim_{t\to 0^+} \frac{\mathbb{1} - \mathbb{1}}{t} = 0.$$

Für $i \neq j$ folgt

$$Q(i,j) = [Q\mathbb{1}_{x_j}](x_i) = \left[\lim_{t\to 0^+} \frac{P(t)\mathbb{1}_{\xi_j} - \mathbb{1}_{\xi_j}}{t}\right](x_i) = \lim_{t\to 0^+} \frac{[P(t)\mathbb{1}_{\xi_j}](x_i) - [\mathbb{1}_{\xi_j}](x_i)}{t}$$

$$= \lim_{t\to 0^+} \frac{[P(t)\mathbb{1}_{\xi_j}](x_i)}{t} \geq 0$$

als Konsequenz der Positivität von $P(t)$. Wir haben gezeigt, dass Q ein Markov-Kern ist. □

Beispiel 3.2.6 (Markov-Ketten Teil 5)**.** Wir haben bereits gesehen, dass der Generator zu der C_0-Halbgruppe $(P(t))_{t\geq 0}$ einer Markov-Kette ein Markov-Kern ist. Theorem 3.2.5 zeigt nun, dass es sich bei $(P(t))_{t\geq 0}$ aufgrund der Eindeutigkeit des Generators um eine Markov-Halbgruppe handelt.

Definition 3.2.7 (Markov-Tripel)**.** Sei $\mathcal{X}$ eine endliche Menge und Q ein Markov-Kern mit zugehörigem invariantem Maß μ. Das Tripel $(\mathcal{X}, Q, \mu)$ nennen wir Markov-Tripel.

Bemerkung 3.2.8. Die Existenz eines invarianten Maßes zu einem beliebigen Markov-Kern Q folgt aus Theorem 3.2.5.

Definition 3.2.9. Wir nennen einen Markov-Kern Q auf $\mathcal{X}$ irreduzibel, falls für alle verschiedenen $x, y \in \mathcal{X}$ ein Pfad $x = \gamma_0, \ldots, \gamma_m = y$ existiert, für welchen $Q(\gamma_i, \gamma_{i+1}) > 0$ für alle $i = 0, \ldots, m-1$ gilt. Ein Markov-Tripel heißt irreduzibel, falls Q es ist.

Bemerkung 3.2.10. Ist ein Markov-Kern Q irreduzibel, dann ist auch die zugehörige Matrix irreduzibel im Sinne der Definition C.0.4 der Irreduzibilität von Matrizen.

Lemma 3.2.11. Sei Q ein irreduzibler Markov-Kern, dann gilt $e^{tQ} > 0$ für alle $t > 0$.

Beweis. Definiere die Matrix $K = Q + \alpha \operatorname{Id}$, wobei $\alpha = \max_{x\in\mathcal{X}} |Q(x,x)| + 1$ ist, dann gilt $K(x,x) > 0$ für alle $x \in \mathcal{X}$. Ferner ist nach der Definition eines Markov-Kerns auch $K(x,y) = Q(x,y) \geq 0$ für alle $x \neq y$. Die Matrix K ist irreduzibel, wir können also Satz C.0.5 anwenden. Folglich existiert ein $m \in \mathbb{N}$, sodass $K^m > 0$ erfüllt ist. Insbesondere gilt $e^{tK} > 0$ für alle $t > 0$, denn

$$e^{tK} = \sum_{k=0}^{\infty} \frac{(tK)^k}{k!} \geq \frac{t^m K^m}{m!} > 0.$$

Für alle $t > 0$ folgt

$$e^{tQ} = e^{t(K - \alpha \operatorname{Id})} = e^{tK} e^{-\alpha t \operatorname{Id}} = e^{-\alpha t} e^{tK} > 0.$$

□

Lemma 3.2.12. Sei $(\mathcal{X}, Q, \mu)$ ein Markov-Tripel und ρ eine Wahrscheinlichkeitsdichte bezüglich π, das heißt, es gilt $\rho \geq 0$ und $\int_{\mathcal{X}} \rho \mathrm{d}\pi = 1$, dann ist auch $e^{tQ}\rho$ eine Wahrscheinlichkeitsdichte für alle $t \geq 0$.

Beweis. Wir haben gesehen, dass e^{tQ} eine Markov-Halbgruppe ist, dies zeigt $e^{tQ}\rho \geq 0$ für alle $t \geq 0$. Für beliebiges $t \geq 0$ gilt

$$\int_{\mathcal{X}} e^{tQ}\rho \mathrm{d}\pi = \int_{\mathcal{X}} \rho \mathrm{d}\pi = 1$$

aufgrund der Invarianz des Maßes π. Dies zeigt die Behauptung. □

Lemma 3.2.13. Zu einem irreduziblen Markov-Kern auf einem endlichen Zustandsraum $\mathcal{X}$ existiert ein eindeutig bestimmtes, positives und invariantes Maß.

Beweis. Wie bereits gesehen, gilt bei einem irreduziblen Markov-Kern $e^{tQ} > 0$ für alle $t > 0$. Insbesondere gilt für festes $t > 0$ die Ungleichung $0 < m(t) := \min_{x,y \in \mathcal{X}} [e^{tQ}](x,y)$. Theorem 3.2.5 zeigt die Existenz eines nichtnegativen invarianten Maßes μ auf $\mathcal{X}$. Angenommen, es gilt $\mu(\{x\}) = 0$ für ein $x \in \mathcal{X}$, dann folgt

$$0 = \mu(\{x\}) = \sum_{y \in \mathcal{X}} [e^{tQ}](x,y)\mu(\{y\}) \geq m(t) \sum_{y \in \mathcal{X}} \mu(\{y\}) = m(t) > 0,$$

dies ist ein Widerspruch. Wir schließen die Positivität des Maßes μ. Nehmen wir an es existieren zwei verschiedene invariante Maße μ_1 und μ_2. In diesem Fall gilt $\mu_i^t e^{tQ} = \mu_i^t$ für $i = 1, 2$ und es folgt

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} \mu_1(\{y\})[e^{tQ}](y,x) = \sum_{x \in \mathcal{X}} \mu_1(\{x\}) = 1 = \sum_{x \in \mathcal{X}} \mu_2(\{x\}) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} \mu_2(\{y\})[e^{tQ}](y,x).$$

Subtraktion der rechten von der linken Seite obiger Gleichung liefert

$$0 = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} \left(\mu_1(\{y\}) - \mu_2(\{y\})\right) [e^{tQ}](y,x).$$

Es ist $[e^{tQ}](y,x) > 0$ für alle $x, y \in \mathcal{X}$ und somit muss schon $\mu_1(\{y\}) - \mu_2(\{y\}) = 0$ für alle $y \in \mathcal{X}$ gelten. Dies zeigt die Eindeutigkeit des invarianten Maßes. □

Definition 3.2.14 (Reversibilität)**.** Sei Q ein Markov-Kern und π ein Maß auf $\mathcal{X}$. Wir nennen π reversibel, falls für alle $x, y \in \mathcal{X}$ die Gleichung

$$\pi(\{x\})Q(x,y) = \pi(\{y\})Q(y,x)$$

erfüllt ist. Diese Gleichung nennt man auch die Bedingung des detaillierten Gleichgewichts.

Lemma 3.2.15. Sei π ein reversibles Maß zu einem Markov-Kern Q, dann ist π auch ein invariantes Maß zur Operatorenfamilie $(e^{tQ})_{t \geq 0}$. Ein Markov-Tripel $(\mathcal{X}, Q, \pi)$ heißt reversibel, falls π es ist.

Beweis. Sei $x \in \mathcal{X}$. Wir summieren die Reversibilitätsbedingung über $y \in \mathcal{X}$, dann folgt

$$0 = \pi(\{x\}) \sum_{y \in \mathcal{X}} Q(x,y) = \sum_{y \in \mathcal{X}} \pi(\{x\}) Q(x,y) = \sum_{y \in \mathcal{X}} \pi(\{y\}) Q(y,x) = [\pi^t Q](x).$$

Mit Lemma 3.2.4 folgt die Invarianz des Maßes π. $\square$

3.3 Geometrische Interpretation

In diesem Abschnitt werden wir vorstellen, wie man die Markov-Halbgruppen geometrisch interpretieren kann. Wir werden untersuchen, in welchem Zusammenhang die Markov-Halbgruppen und die Wärmeleitungsgleichung auf Graphen stehen.

Definition 3.3.1. Sei $V \neq \emptyset$ eine Menge sowie $E \subset V \times V$. Wir nennen das geordnete Paar (V,E) einen Graphen. Gilt $(x,y) \in E$ genau dann, wenn $(y,x) \in E$, dann nennen wir den Graphen ungerichtet. Der Graph heißt endlich, falls V eine endliche Menge ist. Ist $(x,y) \in E$, so schreiben wir $x \sim y$.

Bemerkung 3.3.2. Man kann die Menge der Kanten eines Graphen auch durch eine Abbildung charakterisieren. Wir definieren dazu

$$\omega(x,y) = \begin{cases} 1 & : (x,y) \in E \\ 0 & : (x,y) \notin E, \end{cases}$$

dann gilt offensichtlich $\omega(x,y) = 1$ genau dann, wenn $(x,y) \in E$ ist. Man könnte die Werte der Funktion ω als Gewichte der Kanten interpretieren. In unserem Fall hätte jede Kante das Gewicht 1. Interessant sind vor allem Graphen, bei denen die Kanten unterschiedlich gewichtet sind. Dies führt zu der folgenden Definition.

Definition 3.3.3. Ein ungerichteter Graph (V,E) heißt gewichtet, falls eine Abbildung $\omega\colon V \times V \to \mathbb{R}$ existiert, sodass für alle verschiedenen $x,y \in V$ das Paar (x,y) genau dann eine Kante ist, falls $\omega(x,y) \neq 0$ gilt. Den Wert $\omega(x,y)$ nennen wir das Gewicht der Kante (x,y). Für einen solchen Graphen schreiben wir abkürzend auch (V,E,ω).

Definition 3.3.4. Ein Graph heißt zusammenhängend, falls es für alle $x,y \in V$ einen Pfad gibt, der diese beiden Punkte verbindet. Genauer existieren Kanten $(\gamma_i, \gamma_{i+1})_{i=0,\dots,m-1} \subset E$ derart, dass $\gamma_0 = x$ und $\gamma_m = y$ gilt.

Konvention: Wir werden in der folgenden Ausführung lediglich nichtnegativ gewichtete ungerichtete endliche Graphen (V,E,ω) betrachten. Ferner setzen wir zusätzlich voraus, dass

$\omega(x,x) = 0$ für alle $x \in V$ ist. Für einen solchen Graphen gilt $\omega(x,y) = \omega(y,x)$ für alle $x,y \in V$.

Definition 3.3.5. Seien ein Graph (V,E) und eine Abbildung $\mu\colon V \to (0,\infty)$ gegeben. Für eine beliebige Funktion $f\colon V \to \mathbb{R}$ definieren wir

$$\Delta f(x) = \frac{1}{\mu(x)} \sum_{y \sim x} \omega(x,y)\,(f(y) - f(x))\,.$$

Die Abbildung $\Delta\colon V^{\mathbb{R}} \to V^{\mathbb{R}}$ nennen wir den (diskreten) Laplace-Operator auf dem Graphen (V,E) zur Funktion μ.

Definition 3.3.6 (Wärmeleitungsgleichung auf Graphen)**.** Sei (V,E) ein Graph. Eine Abbildung $u\colon [0,\infty) \times V \to \mathbb{R}$ heißt Lösung der Wärmeleitungsgleichung zum Anfangswert $u_0 \in V^{\mathbb{R}}$, falls sie

$$\begin{cases} \partial_t u = \Delta u & \text{auf } [0,\infty) \times V \\ u(0,x) = u_0(x) \end{cases}$$

erfüllt.

Theorem 3.3.7. *Sei $(\mathcal{X},Q,\pi)$ ein reversibles Markov-Tripel, dann definieren wir die Gewichtsfunktion*

$$\omega\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R},\ \omega(x,y) = \begin{cases} Q(x,y)\pi(\{x\}) & : x \neq y \\ 0 & : x = y \end{cases}$$

für $x,y \in \mathcal{X}$. Das Paar $(\mathcal{X},\omega)$ ist ein gewichteter ungerichteter endlicher Graph. Der Laplace-Operator auf $(\mathcal{X},\omega)$ zur Abbildung $\mu\colon \mathcal{X} \to \mathbb{R}$, definiert durch $\mu(x) = \pi(\{x\})$ für alle $x \in \mathcal{X}$, erfüllt die Gleichung

$$\Delta f(x) = \sum_{y \in \mathcal{X}} Q(x,y)\,(f(y) - f(x)) = \sum_{y \in \mathcal{X}} Q(x,y) f(y)$$

für alle $f \in \mathcal{X}^{\mathbb{R}}$ und alle $x \in \mathcal{X}$. Dies bezeichnen wir als den Laplace-Operator zur Geometrie von Q. Ist Q irreduzibel, dann ist der zugehörige Graph $(\mathcal{X},\omega)$ zusammenhängend.

Beweis. Die Funktion ω ist nichtnegativ und sie ist symmetrisch als Konsequenz der Reversibilität. Das zeigt, dass ω die Gewichtsfunktion eines ungerichteten Graphs ist. Sei $f \in \mathcal{X}^{\mathbb{R}}$ und $x \in \mathcal{X}$, dann gilt

$$\begin{aligned} \Delta f(x) &= \frac{1}{\mu(x)} \sum_{y \sim x} \omega(x,y)\,(f(y) - f(x)) = \frac{1}{\pi(\{x\})} \sum_{y \in \mathcal{X}} Q(x,y)\pi(\{x\})\,(f(y) - f(x)) \\ &= \sum_{y \in \mathcal{X}} Q(x,y)\,(f(y) - f(x)) = \sum_{y \in \mathcal{X}} Q(x,y) f(y) - f(x) \sum_{y \in \mathcal{X}} Q(x,y) \end{aligned}$$

$$= \sum_{y \in \mathcal{X}} Q(x,y) f(y)$$

□

Theorem 3.3.8. *Seien ein Graph* (V, E, ω) *mit* $V = \{x_1, \ldots, x_n\}$ *und eine Abbildung* $\mu \colon V \to (0, \infty)$ *gegeben. Wir definieren die Matrix* $Q \in \mathbb{R}^{n \times n}$ *mit den Einträgen* $Q(i,j)$ *durch die Vorschrift*

$$Q(i,j) = \begin{cases} -\sum\limits_{j=1}^{n} \frac{\omega(x_i,x_j)}{\mu(x_i)} & : i = j \\ \frac{\omega(x_i,x_j)}{\mu(x_i)} & : i \neq j. \end{cases}$$

In diesem Fall ist Q *ein Markov-Kern auf* V*. Ferner stimmen die Wärmeleitungsgleichung auf dem Graphen sowie das Anfangswertproblem*

$$\begin{cases} \partial_t u = Qu \text{ für } t \geq 0 \\ u(0) = u_0 \end{cases} \tag{ACP}$$

überein. Ist ω *symmetrisch, dann existiert ein reversibles Maß* π*, gegeben durch*

$$\pi(\{x_i\}) = \frac{\mu(x_i)}{\sum\limits_{j=1}^{n} \mu(x_j)}$$

für $i = \{1, \ldots, n\}$*. Ist* (V, E) *zusammenhängend, so ist* Q *irreduzibel.*

Beweis. Wir zeigen zunächst, dass Q ein Markov-Kern auf V ist. Für $i \in \{1, \ldots, n\}$ gilt

$$\sum_{j=1}^{n} Q(i,j) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\omega(x_i, x_j)}{\mu(x_i)} - \sum_{j=1}^{n} \frac{\omega(x_i, x_j)}{\mu(x_i)} = -\omega(x_i, x_i) = 0,$$

denn $\omega(x_i, x_i) = 0$ gilt nach unserer Konvention. Unter der Voraussetzung, dass ω nichtnegativ ist, folgt auch $Q(i,j) \geq 0$ für alle $i \neq j$. Wir haben gezeigt, dass Q ein Markov-Kern ist. Für die Übereinstimmung des (ACP) mit der Wärmeleitungsgleichung reicht es zu zeigen, dass $Qu = \Delta u$ für alle $u \in V^{\mathbb{R}}$ erfüllt ist. Sei $u \in V^{\mathbb{R}}$ und $x_i \in V$, dann gilt

$$\begin{aligned} [Qu](x_i) &= \sum_{j=1}^{n} Q(i,j) u(x_j) - u(x_i) \sum_{j=1}^{n} Q(i,j) = \sum_{j=1}^{n} Q(i,j) u(x_j) - \sum_{j=1}^{n} Q(i,j) u(x_i) \\ &= \sum_{j=1}^{n} Q(i,j) \left(u(x_j) - u(x_i) \right) = \sum_{\substack{j=1 \\ j \neq i}}^{n} Q(i,j) \left(u(x_j) - u(x_i) \right) + Q(i,i)(u(x_i) - u(x_i)) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mu(x_i)} \sum_{\substack{j=1 \\ j \neq i}}^{n} \omega(x_i, x_j)\,(u(x_j) - u(x_i)) = \frac{1}{\mu(x_i)} \sum_{y \sim x_i} \omega(x_i, y)\,(u(y) - u(x_i)) \\
&= [\Delta u](x_i),
\end{aligned}
$$

denn es gilt $y \sim x_i$ genau dann, wenn $\omega(x_i, y) > 0$ ist. Die oben definierte Abbildung π definiert ein Wahrscheinlichkeitsmaß, denn es gilt

$$
\sum_{i=1}^{n} \pi(\{x_i\}) = \frac{\sum_{i=1}^{n} \mu(x_i)}{\sum_{j=1}^{n} \mu(x_j)} = 1.
$$

Für $i \neq j$ gilt ferner

$$
\pi(\{x_i\})Q(i,j) = \frac{1}{\sum_{j=1}^{n} \mu(x_j)} \left(\mu(x_i) \frac{\omega(x_i, x_j)}{\mu(x_i)} \right) = \frac{\omega(x_i, x_j)}{\sum_{j=1}^{n} \mu(x_j)}
$$

und so folgt zusammen mit der Symmetrie von ω die Reversibilität des Maßes π. Sei nun V ein zusammenhängender Graph. Zu $i, j \in 1, \dots, n$ existiert dann eine Folge von aufeinanderfolgenden Kanten, die die Ecken x_i und x_j verbindet. Anders formuliert, existieren $\gamma_0 = x_i, \gamma_1, \dots, \gamma_m = x_j$ derart, dass $\omega(\gamma_i, \gamma_{i+1}) > 0$ für alle $i = 0, \dots, m-1$ erfüllt ist. Da $\mu(x_i) > 0$ für alle $i \in \{1, \dots, n\}$ gilt, folgt die Irreduzibilität des Markov-Kerns. □

3.4 Eine diskrete Fokker-Planck-Gleichung

Wir möchten eine diskrete Fokker-Planck-Gleichung untersuchen. Eine Möglichkeit, eine solche zu erhalten, ist die Diskretisierung der kontinuierlichen Version. Damit wir eine endliche Diskretisierung dieser Gleichung erhalten, werden wir die Fokker-Planck-Gleichung auf dem Intervall $[0,1]$ betrachten. Diese Version der Fokker-Planck-Gleichung und ihr diskretes Pendant sind Gegenstand des nächsten Abschnitts.

Mit der Wahl von sogenannten Robin-Randbedingungen lautet die Fokker-Planck-Gleichung auf dem Intervall $[0,1]$

$$
\begin{cases}
\partial_t u = \partial_x(\partial_x u + uV') & \text{für } t > 0, x \in (0,1) \\
\partial_x u(t,x) + u(t,x)V'(x) = 0 & \text{für } t > 0, x \in \{0,1\} \\
u(0) = u_0 & \text{in } (0,1).
\end{cases}
$$

Hier ist V ein genügend glattes Potential. Wir möchten die Gleichung zunächst in eine andere Form bringen. Wie wir in Kapitel 2 bemerkt haben, ist auch hier $u_\infty = \frac{1}{\|e^{-V}\|_{L^1}} e^{-V}$ eine

Equilibriumslösung. Es gilt $u_\infty > 0$ in $[0,1]$ und damit ist

$$\begin{aligned}\partial_x\left(u_\infty\partial_x\left(\frac{u}{u_\infty}\right)\right) &= \partial_x\left(u_\infty\frac{u_\infty\partial_x u - u\partial_x u_\infty}{u_\infty^2}\right) = \partial_x\left(\partial_x u - \frac{u\partial_x u_\infty}{u_\infty}\right)\\ &= \partial_x\left(\partial_x u - \frac{-u_\infty V' u}{u_\infty}\right) = \partial_x(\partial_x u + V'u) = \partial_t u\end{aligned}$$

in $(0,1)$. Für $x \in \{0,1\}$ folgt ebenso $\partial_x\left(\frac{u}{u_\infty}\right) = 0$. Dies führt uns zu folgender Formulierung der Fokker-Planck-Gleichung:

$$\begin{cases}\partial_t u = \partial_x\left(u_\infty\partial_x\left(\frac{u}{u_\infty}\right)\right) & \text{für } t>0, x\in(0,1)\\ \partial_x\left(\frac{u}{u_\infty}\right) = 0 & \text{für } t>0, x\in\{0,1\}\\ u(0) = u_0 & \text{in } [0,1].\end{cases} \tag{FPG}$$

Beispiel 3.4.1 (Diskrete FPG)**.** Für beliebiges $n \in \mathbb{N}$, $n \geq 2$ unterteilen wir das Intervall $[0,1]$ in äquidistante Teilintervalle $[x_i^n, x_{i+1}^n]$ für $i \in \{0,\dots,n-1\}$, wobei $x_i^n = \frac{i}{n}$ für $0 \leq i \leq n$ ist. Wir definieren $\pi_i = \int_{x_{i-1}^n}^{x_i^n} u_\infty(x)\mathrm{d}x$ für $i = 1,\dots n$. Wir suchen Funktionen $\rho_i(t)$, die die relative Masse der Funktion u auf dem Teilintervall $[x_{i-1}^n, x_i^n]$ approximieren, das heißt,

$$\rho_i(t) \approx \frac{1}{\pi_i}\int_{x_{i-1}^n}^{x_i^n} u(t,x)\mathrm{d}x.$$

Diesen Ansatz nennt man in der numerischen Mathematik die Finite-Volumen-Methode. Wir möchten eine Differentialgleichung für die Funktionen $\rho_i(t)$ herleiten. Sei dazu $i \in \{1,\dots,n\}$, dann ist

$$\begin{aligned}\partial_t\int_{x_{i-1}^n}^{x_i^n} u(t,x)\mathrm{d}x &= \int_{x_{i-1}^n}^{x_i^n}\partial_t u(t,x)\mathrm{d}x = \int_{x_{i-1}^n}^{x_i^n}\partial_x\left(u_\infty\partial_x\left(\frac{u}{u_\infty}\right)\right)\mathrm{d}x\\ &= \left[\left(u_\infty\partial_x\left(\frac{u}{u_\infty}\right)\right)\right]_{x_{i-1}^n}^{x_i^n}\\ &= u_\infty(x_i^n)\partial_x\left(\frac{u(t,x_i^n)}{u_\infty(x_i^n)}\right) - u_\infty(x_{i-1}^n)\partial_x\left(\frac{u(t,x_{i-1}^n)}{u_\infty(x_{i-1}^n)}\right).\end{aligned}$$

Es gilt nun, die Terme auf der rechten Seite obiger Gleichung in Abhängigkeit der $\rho_i(t)$ zu approximieren. Für die partielle Ableitung wählen wir die Näherung

$$\partial_x\left(\frac{u(t,x_i^n)}{u_\infty(x_i^n)}\right) \approx n\left(\frac{u(t,x_{i+1})}{u_\infty(x_{i+1})} - \frac{u(t,x_i)}{u_\infty(x_i)}\right) \approx n\left(\frac{u(t,x_{i+1})}{n\pi_{i+1}} - \frac{u(t,x_i)}{n\pi_i}\right)$$

$$\approx n \left(\frac{\int_{x_i^n}^{x_{i+1}^n} u(t,x)\mathrm{d}x}{\pi_{i+1}} - \frac{\int_{x_{i-1}^n}^{x_i^n} u(t,x)\mathrm{d}x}{\pi_i} \right) = n\left(\rho_{i+1} - \rho_i\right).$$

Anstelle der Approximation von $u_\infty(x_i) \approx n\pi_i$ wählen wir die Näherung $u_\infty(x_i) \approx \sqrt{n\pi_i n\pi_{i+1}}$, um die in der Differenz auftretenden Massen gleichermaßen zu berücksichtigen. Dies führt zu der folgenden Näherung:

$$u_\infty(x_i^n)\partial_x \left(\frac{u(t,x_i^n)}{u_\infty(x_i^n)} \right) \approx u_\infty(x_i^n) n\left(\rho_{i+1} - \rho_i\right) \approx \sqrt{n\pi_i n\pi_{i+1}} n\left(\rho_{i+1} - \rho_i\right) = \kappa_i\left(\rho_{i+1} - \rho_i\right)$$

für $\kappa_i = n^2\sqrt{\pi_i\pi_{i+1}}$ und $i = 1, \ldots, n-1$. Wir setzen weiter $\kappa_0 = \kappa_n = 0$ für eine gute Darstellung der Randbedingungen. Wir fassen unsere Beobachtungen zusammen, es gilt

$$\partial_t\rho_i(t) = \frac{\kappa_i\left(\rho_{i+1} - \rho_i\right) - \kappa_{i-1}\left(\rho_i - \rho_{i-1}\right)}{\pi_i} = \frac{\kappa_i}{\pi_i}\rho_{i+1} - \frac{\kappa_i + \kappa_{i-1}}{\pi_i}\rho_i + \frac{\kappa_{i-1}}{\pi_i}\rho_{i-1}$$

für $i = 2, \ldots, n-1$. In den Endpunkten erhalten wir die Differentialgleichungen

$$\partial_t\rho_1 = \frac{\kappa_1}{\pi_1}\left(\rho_2 - \rho_1\right) \text{ sowie } \partial_t\rho_n = \frac{\kappa_{n-1}}{\pi_n}\left(\rho_{n-1} - \rho_n\right).$$

Wir schreiben die lineare Differentialgleichung für ρ in der zugehörigen Matrixdarstellung $\partial_t\rho = Q^{(n)}\rho$, wobei $Q^{(n)} \in \mathbb{R}^{n\times n}$ die folgende Matrix bezeichnet

$$Q^{(n)} = \begin{pmatrix} -\frac{\kappa_1}{\pi_1} & \frac{\kappa_1}{\pi_1} & 0 & & & & & \\ \frac{\kappa_1}{\pi_2} & -\frac{\kappa_1+\kappa_2}{\pi_2} & \frac{\kappa_2}{\pi_2} & 0 & & & & \\ 0 & \ddots & \ddots & \ddots & & & & \\ & 0 & \frac{\kappa_{i-1}}{\pi_i} & -\frac{\kappa_{i-1}+\kappa_i}{\pi_i} & \frac{\kappa_i}{\pi_i} & 0 & & \\ & & & \ddots & \ddots & \ddots & 0 & \\ & & & 0 & \frac{\kappa_{n-2}}{\pi_{n-1}} & -\frac{\kappa_{n-2}+\kappa_{n-1}}{\pi_{n-1}} & \frac{\kappa_{n-1}}{\pi_{n-1}} & \\ & & & & 0 & \frac{\kappa_{n-1}}{\pi_n} & -\frac{\kappa_{n-1}}{\pi_n} & \end{pmatrix}.$$

Sei u_0 der Anfangswert zu der kontinuierlichen Fokker-Planck-Gleichung auf dem Intervall $[0,1]$. Wir bezeichnen mit $\rho_0 := \rho(0)$ den Anfangswert der diskreten Differentialgleichung.

Das folgende Anfangswertproblem erklären wir als die diskrete Fokker-Planck-Gleichung

$$\begin{cases} \partial_t \rho = Q^{(n)} \rho \\ \rho(0) = \rho_0. \end{cases} \tag{DFPG^n}$$

Hierbei bezeichnet $n \geq 2$ die Abhängigkeit von der Anzahl der Teilintervalle.

Lemma 3.4.2. Für jedes $n \in \mathbb{N}$, $n \geq 2$ ist die Matrix $Q^{(n)}$ der diskreten Fokker-Planck-Gleichung (DFPG^n) ein reversibler und irreduzibler Markov-Kern. Das reversible Maß ist gegeben durch die Zahlen $\pi = (\pi_1, \dots, \pi_n)$. Wir schreiben $(\{1, \dots, n\}, Q^{(n)}, \pi)$ für das zugehörige Markov-Tripel.

Beweis. Seien $n \in \mathbb{N}$ und $n \geq 2$ aufgrund der Positivität von u_∞, ist auch der Vektor π positiv. Daraus folgt auch die Positivität der κ_i und zusammen $Q(i,j) \geq 0$ für alle $i \neq j$. Für alle $i \in \{2, \dots, n-2\}$ gilt $\frac{1}{\pi_i}(\kappa_{i-1} - \kappa_{i-1} - \kappa_i + \kappa_i) = 0$ und für $i = 1, n-1$ ist die Zeilensumme auch gleich null, folglich ist $Q^{(n)}$ ein Markov-Kern. Wir möchten die Reversibilität von π bezüglich $Q^{(n)}$ zeigen, dazu berechnen wir für $i \in \{1, \dots, n\}$

$$\pi_i Q(i,j) = \begin{cases} \pi_i n^2 \frac{\kappa_{i-1}}{\pi_i} & : j = i-1 \text{ und } i > 2 \\ -\pi_i n^2 \left(\frac{\kappa_i + \kappa_{i-1}}{\pi_i} \right) & : j = i \\ \pi_i n^2 \frac{\kappa_i}{\pi_i} & : j = i+1 \text{ und } i < n \\ 0 & : \text{sonst.} \end{cases}$$

Für $i \in \{1, \dots, n\}$ gilt weiter

$$\pi_j Q(j,i) = \begin{cases} \pi_{i-1} n^2 \frac{\kappa_{i-1}}{\pi_{i-1}} & : j = i-1 \text{ und } i > 2 \\ -\pi_i n^2 \left(\frac{\kappa_i + \kappa_{i-1}}{\pi_i} \right) & : j = i \\ \pi_{i+1} n^2 \frac{\kappa_i}{\pi_{i+1}} & : j = i+1 \text{ und } i < n \\ 0 & : \text{sonst.} \end{cases}$$

Anhand dieser Darstellung erkennen wir die Gleichheit $\pi_i Q(i,j) = \pi_j Q(j,i)$, indem wir die $\pi_{(\cdot)}$ kürzen und mit dem gewünschten Index erweitern. Die Irreduzibilität ist eine Konsequenz der Positivität der Einträge auf der Nebendiagonalen. Zu $i, j \in \{1, \dots, n\}$ und $i < j$ wählen wir $i = \gamma_0, \gamma_1 = i+1, \dots, \gamma_{j-i} = j$, dann gilt $Q(\gamma_k, \gamma_{k+1}) > 0$ für alle $k \in \{1, \dots, j-i-1\}$. Dies zeigt, dass Q ein reversibler und irreduzibler Markov-Kern ist. $\square$

Bemerkung 3.4.3. (i) Für $V \equiv 0$ erhalten wir $u_\infty \equiv 1$ und folglich $\pi_i = \frac{1}{n}$. Dann ist die Matrix $Q^{(n)}$ gerade der Markov-Kern zur diskreten Wärmeleitungsgleichung auf dem Graphen (V, E), gegeben durch $V = \{x_i^n | i = 0, \dots, n\}$ und $E = \{x_i x_{i+1} | i =$

$0, \dots, n-1\}$ mit den Gewichten n^2. In diesem Sinne stimmt die Diskretisierung der Fokker-Planck-Gleichung, hier der Wärmeleitungsgleichung, mit der diskreten Wärmeleitungsgleichung überein.

(ii) Gilt für den kontinuierlichen Anfangswert $\int_0^1 u_0(x)\mathrm{d}x = 1$ und $u_0 \geq 0$, dann ist der diskrete Anfangswert $\rho_0 = \rho(0) \geq 0$ und er erfüllt $\sum_{i=1}^n \rho_i(0)\pi_i = 1$, das heißt, ρ_0 ist eine Dichte bezüglich dem Maß π. Die Lösung der diskreten Fokker-Planck-Gleichung ist gegeben durch $\rho_t = e^{tQ}\rho_0$. Folglich gilt nach Lemma 3.2.12, dass auch ρ_t eine Dichte ist. Ist $\rho_0 > 0$, dann folgt aus der Irreduzibilität $e^{tQ}\rho_0 > 0$ für alle $t \geq 0$. Das heißt, die diskrete Fokker-Planck-Gleichung ist ebenso wie die kontinuierliche Fokker-Planck-Gleichung Masse und Positivität erhaltend.

(iii) Man kann eine solche Diskretisierung der Fokker-Planck-Gleichung auch in höheren Dimensionen durchführen. Ein möglicher Ansatz wird in [ARM16] beschrieben. Auch bei dem dort vorgestellten Ansatz erhalten die Autoren einen Markov-Kern.

4 Eine Theorie des optimalen Transports auf diskreten Mengen

Wir haben in Kapitel 2 bereits gesehen, dass wir die Theorie des optimalen Transports nicht direkt auf das diskrete Setting anwenden können. Einen möglichen Ausweg, um dieses Problem zu beheben, möchten wir in diesem Kapitel vorstellen. Dieser Ansatz ist motiviert durch die in Kapitel 2 angesprochene kontinuierliche Theorie. Ein besonders wichtiger Anhaltspunkt ist die Benamou-Brenier-Formel. Diese Formel und die zugehörige Theorie möchten wir an dieser Stelle kurz skizzieren.

Wir betrachten den Spezialfall $\mathcal{X} = \mathbb{R}^n$. Für zwei Wahrscheinlichkeitsmaße $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ untersuchen wir nicht die Menge aller Kopplungen wie in Abschnitt 2.3, sondern die Menge aller Interpolationen in stetiger Zeit. Das heißt, wir betrachten Abbildungen $\mu\colon [0,1] \to \mathcal{P}_2(\mathbb{R}^n)$, die unter geeigneten, nicht näher spezifizierten Regularitätsbedingungen die sogenannte Kontinuitätsgleichung

$$\partial_t \mu_t + \nabla \cdot (\mu_t \psi_t) = 0 \tag{KG}$$

im distributionellen Sinne erfüllen. Hier bezeichnen wir mit $\psi\colon [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$ ein zeitabhängiges Vektorfeld. Die Benamou-Brenier-Formel besagt, dass wir die Wasserstein-Metrik als

$$W_2(\mu_0, \mu_1) = \sqrt{\inf_{(\mu,\psi) \in \mathcal{CE}(\mu_0,\mu_1)} \left\{ \int_0^1 \int_{\mathbb{R}^n} |\psi_t(x)|^2 \mathrm{d}\mu_t(x) \mathrm{d}t \right\}}$$

darstellen können. Dabei bezeichne $\mathcal{CE}(\mu_0, \mu_1)$ die Menge aller hinreichend regulären Lösungen (μ, ψ) der Kontinuitätsgleichung mit Anfangswert μ_0 und Endwert μ_1.

Zu einer gegebenen Kurve $\mu\colon (-\delta, \delta) \to \mathcal{P}_2(\mathbb{R}^n)$ existieren verschiedene Vektorfelder ψ, die die Kontinuitätsgleichung in $t = 0$ erfüllen. Es gilt jedoch, dass nur eines dieser Vektorfelder ein verallgemeinerter Gradient ist. Wir sagen, ψ ist ein verallgemeinerter Gradient, falls ψ Element des $L^2(\mathbb{R}^n, \mu_0; \mathbb{R}^n)$-Abschlusses der Menge $\{\nabla\varphi \mid \varphi \in C_c^\infty(\mathbb{R}^n)\}$ ist. Der verallgemeinerte Gradient ψ erfüllt ferner die nützliche Eigenschaft, dass er das Funktional der kinetischen Energie $\int_{\mathbb{R}^n} |\psi(x)|^2 \mathrm{d}\mu_0(x)$ in der Klasse aller Vektorfelder, die die Kontinuitätsgleichung in $t = 0$ erfüllen, minimiert. Felix Otto interpretiert in [Ott01] erstmals die Benamou-Brenier-Formel als riemannschen Abstand im Sinne einer intuitiven riemannschen

Struktur. Wir möchten diese riemannsche Struktur kurz umreißen.

Zu einer glatten Kurve $(\mu_t)_{t\in(-\delta,\delta)}$ in $\mathcal{P}_2(\mathbb{R}^n)$ mit $\mu_0 = \mu$ bezeichne $\nabla\varphi = \psi$ den eindeutig bestimmten verallgemeinerten Gradienten, welcher die Kontinuitätsgleichung in $t = 0$ erfüllt. Das heißt, ψ ist der tangentiale Vektor an die Kurve μ in $t = 0$. Wir identifizieren den tangentialen Vektorraum an μ mit dem Raum der verallgemeinerten Gradienten. Für $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ wählen wir das $L^2(\mathbb{R}^n, \mu; \mathbb{R}^n)$ Skalarprodukt auf diesem tangentialen Vektorraum

$$\langle \nabla\varphi_1, \nabla\varphi_2\rangle_\mu = \int_{\mathbb{R}^n} \langle \nabla\varphi_1(x), \nabla\varphi_2(x)\rangle \mathrm{d}\mu(x).$$

Diese Wahl induziert eine formale riemannsche Struktur auf $\mathcal{P}(\mathcal{X})$. Wir betrachten den durch diese Struktur induzierten riemannschen Abstand

$$\begin{aligned}
d(\mu_0,\mu_1) &= \inf\left\{ \int_0^1 \sqrt{\langle \nabla\varphi_t, \nabla\varphi_t\rangle_{\mu_t}}\mathrm{d}t \;\middle|\; \mu\colon [0,1] \to \mathcal{P}_2(\mathbb{R}^n) \text{ glatt } \mu(0) = \mu_0, \mu(1) = \mu_1 \right\} \\
&= \inf_{(\mu,\psi)\in\mathcal{CE}(\mu_0,\mu_1)} \int_0^1 \sqrt{\langle \psi_t, \psi_t\rangle_{\mu_t}}\mathrm{d}t \\
&= \sqrt{\inf_{(\mu,\psi)\in\mathcal{CE}(\mu_0,\mu_1)} \left\{ \int_0^1 \langle \psi_t, \psi_t\rangle_{\mu_t}\mathrm{d}t \right\}} \\
&= \sqrt{\inf_{(\mu,\psi)\in\mathcal{CE}(\mu_0,\mu_1)} \left\{ \int_0^1 \|\psi_t\|^2_{L^2(\mu)}\mathrm{d}t \right\}} = W_2(\mu_0,\mu_1).
\end{aligned}$$

Die zweite Gleichung begründet sich dadurch, dass das tangentiale Vektorfeld zu einer glatten Kurve die Kontinuitätsgleichung erfüllt. Dieses tangentiale Vektorfeld liegt in der Menge der verallgemeinerten Gradienten. Man kann jedoch zeigen, dass man beliebige Vektorfelder zulassen kann, ohne diesen Wert zu verbessern. Ein Argument für die Gültigkeit der dritten Gleichung findet man in [AC08, Proposition 1.3]. Insgesamt sehen wir, dass der riemannsche Abstand mit der Benamou-Brenier-Formel übereinstimmt.

Die obigen Beobachtungen bilden die Grundlage des folgenden Kapitels. Wir werden uns vor allem an der Arbeit [Maa11] von Jan Maas orientieren, der auf Basis der Benamou-Brenier-Formel eine diskrete Wasserstein-Metrik definiert. Wir werden zeigen, dass auch diese Metrik von einer riemannschen Struktur induziert ist. Ferner ist der zugehörige metrische Raum geodätisch, das heißt, diese Metrik weist nicht den Defekt der Wasserstein-Metrik auf (vergleiche Lemma 2.4.13). Wir haben in Kapitel 2 bereits erwähnt, dass die Lösungen der Wärmeleitungsgleichung auch die Trajektorien des Gradientenfluss der Entropie sind. Diese Aussage gilt auch für die von Jan Maas vorgeschlagene diskrete Wasserstein-Metrik und die diskrete Wärmeleitungsgleichung. Dies unterstreicht die Tatsache, dass die erstmals von Jan Maas eingeführte diskrete Wasserstein-Metrik eine gute Wahl für eine Theorie des optimalen Transports auf diskreten Mengen ist.

Wir beginnen mit der Festlegung einiger Notationen und Rahmenbedingungen für das folgende Kapitel. Wir werden dann zunächst die Eigenschaften der diskreten Wasserstein-Metrik diskutieren und darauf die zugehörige riemannsche Struktur vorstellen. Auf dieser riemannschen Struktur werden wir Gradientenflüsse studieren und letztlich den Gradientenfluss der Entropie untersuchen.

4.1 Eine diskrete Wasserstein-Metrik

Definition 4.1.1 (Rahmenbedingungen)**.** Das zentrale Objekt dieses Kapitels ist ein irreduzibles und reversibles Markov-Tripel $(\mathcal{X}, Q, \pi)$ auf einer endlichen Menge $\mathcal{X}$. Zur Erinnerung listen wir alle relevanten Eigenschaften eines solchen auf.

(i) Es gilt $Q(x, y) \geq 0$ für alle $x, y \in \mathcal{X}$ mit $x \neq y$.

(ii) Für alle $x \in \mathcal{X}$ gilt $\sum\limits_{y \in \mathcal{X}} Q(x, y) = 0$.

(iii) Die Irreduzibilität bedeutet, dass der zugehörige Graph zusammenhängend ist und impliziert $e^{tQ} > 0$ für alle $t \geq 0$.

(iv) Das invariante Maß π ist eindeutig bestimmt und positiv. Dieses Maß ist außerdem reversibel, das heißt, es gilt

$$\pi(\{x\})Q(x, y) = \pi(\{y\})Q(y, x)$$

für alle $x, y \in \mathcal{X}$.

Wir betrachten die Menge

$$\mathcal{D}(\mathcal{X}) = \left\{ \rho \colon \mathcal{X} \to \mathbb{R} \;\middle|\; \rho(x) \geq 0 \; \forall x \in \mathcal{X} \text{ und } \sum_{x \in \mathcal{X}} \rho(x)\pi(\{x\}) = 1 \right\}$$

aller Wahrscheinlichkeitsdichten auf $\mathcal{X}$ bezüglich des Maßes π. Die Teilmenge aller positiven Wahrscheinlichkeitsdichten bezeichnen wir mit

$$\mathcal{D}_*(\mathcal{X}) = \left\{ \rho \colon \mathcal{X} \to \mathbb{R} \;\middle|\; \rho(x) > 0 \; \forall x \in \mathcal{X} \text{ und } \sum_{x \in \mathcal{X}} \rho(x)\pi(\{x\}) = 1 \right\}.$$

Diese beiden Mengen sind invariant unter der Halbgruppe e^{tQ} nach Lemma 3.2.12. Wir untersuchen die diskrete (relative) Entropie $H\colon \mathcal{D}(\mathcal{X}) \to \mathbb{R}$ definiert durch

$$H(\rho) = \sum_{x\in\mathcal{X}} \rho(x)\log(\rho(x))\pi(\{x\}) = \int_{\mathcal{X}} \rho\log(\rho)\mathrm{d}\pi$$

mit der Konvention, dass $\rho(x)\log(\rho(x)) = 0$ ist, falls $\rho(x) = 0$ gilt. Diese Abbildung entspricht einem diskreten Analogon zu $H_\pi[\mu]$ aus Definition 2.3.3. Wir statten den Raum $\mathbb{R}^{\mathcal{X}}$ mit dem Skalarprodukt

$$\langle\varphi,\psi\rangle = \int_{\mathcal{X}} \varphi\psi\mathrm{d}\mathbb{1} = \sum_{x\in\mathcal{X}} \varphi(x)\psi(x)$$

für $\varphi,\psi \in \mathbb{R}^{\mathcal{X}}$ aus, dann ist dieser ein Hilbertraum.

Für zwei Matrizen $A, B \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ bezeichnen wir mit $A \bullet B$ die Matrix mit den Einträgen $(A \bullet B)(x, y) = A(x, y)B(x, y)$.

Bemerkung 4.1.2. In seiner Arbeit [Maa11] betrachtet Jan Maas lediglich Markov-Kerne, die $0 \leq Q(x, y) \leq 1$ für alle $x, y \in \mathcal{X}$ und $x \neq y$ erfüllen. Wir betrachten hier Markov-Kerne, die auch beliebige Werte größer als eins annehmen können. Wir werden mehrere Resultate aus seinen Arbeiten ohne Beweis zitieren. Den Beweis, dass diese Resultate auch unter den schwächeren Voraussetzungen ihre Gültigkeit behalten, sparen wir uns an dieser Stelle. Weder in den Aussagen dieser Resultate noch in deren Beweisen spielt oben genannte Eigenschaft eine wichtige Rolle.

Lemma 4.1.3. Für alle $\rho \in \mathcal{D}(\mathcal{X})$ gilt $H(\rho) \geq 0$. Ist $H(\rho) = 0$ für $\rho \in \mathcal{D}(\mathcal{X})$, dann gilt $\rho = \mathbb{1}$.

Beweis. Sei $\rho \in \mathcal{D}(\mathcal{X})$. Wir setzen $\mu(\{x\}) := \rho(x)\pi(\{x\})$ für alle $x \in \mathcal{X}$, dann definiert μ ein Wahrscheinlichkeitsmaß. Die Jensensche Ungleichung impliziert

$$\begin{aligned} H(\rho) &= \int_{\mathcal{X}} \rho(x)\log(\rho(x))\mathrm{d}\pi = \sum_{x\in\mathcal{X}} \rho(x)\log(\rho(x))\pi(\{x\}) = -\sum_{x\in\mathcal{X}_*} \log\left(\frac{1}{\rho(x)}\right)\mu(\{x\}) \\ &= -\int_{\mathcal{X}_*} \log\left(\frac{1}{\rho(x)}\right)\mathrm{d}\mu \geq -\log\left(\int_{\mathcal{X}_*} \frac{1}{\rho(x)}\mathrm{d}\mu\right) = -\log(1) = 0. \end{aligned}$$

Dabei bezeichnen wir mit $\mathcal{X}_* = \{x \in \mathcal{X} \mid \rho(x) \neq 0\}$. Die Gleichheit in der Jensenschen Ungleichung gilt nur, falls ρ konstant ist, das heißt, es muss schon $\rho = \mathbb{1}$ gelten. □

Definition 4.1.4 (Logarithmisches Mittel)**.** Das logarithmische Mittel zweier Zahlen $a, b \geq 0$ ist definiert durch die Abbildungsvorschrift

$$\Lambda\colon [0,\infty)\times[0,\infty) \to [0,\infty), \quad \Lambda(a, b) = \int_0^1 a^{1-p}b^p\mathrm{d}p.$$

Sind a und b positiv sowie $a \neq b$, dann gilt

$$\Lambda(a,b) = \frac{b-a}{\log(b) - \log(a)}.$$

Für ein $\rho \in \mathcal{D}(\mathcal{X})$ bezeichnen wir mit $\hat{\rho} \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ die Matrix mit den Einträgen $\hat{\rho}(x,y) = \Lambda(\rho(x), \rho(y))$.

Bemerkung 4.1.5. Wir werden die folgenden Eigenschaften des logarithmischen Mittels verwenden.

(i) Das logarithmische Mittel ist eine stetige Funktion auf $[0,\infty) \times [0,\infty)$.

(ii) Λ ist unendlich oft stetig differenzierbar auf $(0,\infty) \times (0,\infty)$.

(iii) Es gilt $\Lambda(a,b) = \Lambda(b,a)$ für alle $a, b \geq 0$.

(iv) Für alle $a, b \geq 0$ und $c > 0$ gilt $\Lambda(ca, cb) = c\Lambda(a,b)$.

(v) Für alle $0 \leq a \leq b$ und $c \geq 0$ gilt $\Lambda(a,c) \leq \Lambda(b,c)$.

(vi) Für alle $a, b \geq 0$ gilt $\Lambda(a,b) \geq \sqrt{ab}$.

Wir werden an dieser Stelle keinen Beweis dieser Eigenschaften vorführen. Es sei jedoch bemerkt, dass die Eigenschaften (i) und (ii) eine direkte Konsequenz der Integraldarstellung sind. Eigenschaft (iii) folgt durch Substitution der Integrationsvariable. Für Eigenschaft (iv) genügt die Verwendung der Potenzgesetze. Aussage (v) kann man mittels der Symmetrie auf die Monotonie von $f(x) = x^p$ zurückführen. Ein Beweis der letzten Eigenschaft wird zum Beispiel in [Car72, Theorem 1] geführt.

Bemerkung 4.1.6. Das logarithmische Mittel hat auch eine physikalische Bedeutung. Wir betrachten einen sogenannten Wärmetauscher, das ist zum Beispiel das folgende Konstrukt. Wir nehmen zwei lange zylinderförmige Rohre mit unterschiedlichem Durchmesser und stecken diese ineinander. Dann lassen wir durch das innere Rohr kaltes Wasser und durch das äußere Rohr warmes Wasser fließen. In diesem Fall tauschen die beiden Flüssigkeiten über die Trennwand Wärme aus, vermischen sich aber nicht. Wir definieren die mittlere Temperaturdifferenz

$$\Delta T_m = \frac{\Delta T_1 - \Delta T_2}{\log \Delta T_1 - \log \Delta T_2},$$

wobei ΔT_1 die Temperaturdifferenz der beiden Flüssigkeiten am einen Ende des Rohres und ΔT_2 die Temperaturdifferenz der beiden Flüssigkeiten am anderen Ende des Rohres bezeichnet. Der Wärmestrom des Wärmetauschers beträgt dann

$$\frac{\delta \mathcal{Q}}{\delta t} = \kappa A \Delta T_m = \kappa A \frac{\Delta T_1 - \Delta T_2}{\log \Delta T_1 - \log \Delta T_2}$$

mit dem Wärmedurchgangskoeffizient κ und der Wärmeübertragungsfläche A. $\mathcal{Q}$ steht hier für die übertragene Wärmemenge und steht (noch) in keinem Zusammenhang zu dem Markov-Kern Q.

Bemerkung 4.1.7. Jan Maas betrachtet anstelle des logarithmischen Mittels eine größere Klasse von Funktionen. Es stellt sich jedoch heraus, dass für das logarithmische Mittel die Lösungen der Wärmeleitungsgleichung den Trajektorien des Gradientenfluss der Entropie entsprechen. Aus diesem Grund werden wir uns in dieser Arbeit lediglich mit dem logarithmischen Mittel auseinandersetzen.

Definition 4.1.8 (Diskrete Wasserstein-Metrik)**.** Für zwei Wahrscheinlichkeitsdichten $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}(\mathcal{X})$ definieren wir

$$\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) := \sqrt{\inf_{(\rho,\psi)\in\mathcal{CE}_1(\tilde{\rho}_0,\tilde{\rho}_1)} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y\in\mathcal{X}} (\psi_t(x) - \psi_t(y))^2 \, Q(x,y) \Lambda(\rho_t(x), \rho_t(y)) \pi(\{x\}) \mathrm{d}t \right\}}.$$

Mit $\mathcal{CE}_T(\tilde{\rho}_0, \tilde{\rho}_1)$ bezeichnen wir für $T > 0$ die Menge aller $(\rho, \psi) \in PC^1([0,T]; \mathbb{R}^{\mathcal{X}}) \times \mathcal{M}([0,T]; \mathbb{R}^{\mathcal{X}})$, welche die folgenden Bedingungen erfüllen:

$$\begin{cases} (i) & \rho_0 = \tilde{\rho}_0 \text{ und } \rho_T = \tilde{\rho}_1 \\ (ii) & \rho_t \in \mathcal{D}(\mathcal{X}) \text{ für alle } t \in [0,T] \\ (iii) & \psi\colon [0,T] \to \mathbb{R}^{\mathcal{X}} \text{ ist Borel-messbar} \\ (iv) & \forall x \in \mathcal{X} \text{ und bis auf endlich viele } 0 < t < T \text{ gilt} \\ & \dot{\rho}_t(x) + \sum\limits_{y\in\mathcal{X}} \left[\psi_t(y) - \psi_t(x)\right] Q(x,y) \Lambda(\rho_t(x), \rho_t(y)) = 0. \end{cases}$$

Bemerkung 4.1.9. (i) Auch die diskrete Wasserstein-Metrik kann man als die Transportkosten interpretieren, die benötigt werden, um die Masse von ihrer Ausgangskonfiguration $\tilde{\rho}_0$ in ihren Endzustand $\tilde{\rho}_1$ zu überführen. Ein Unterschied zu ihrem kontinuierlichen Vorbild ist der folgende. Die Transportkosten, um eine Masse von x nach y zu transportieren, sind hier auch davon abhängig, wie viel Masse bereits jeweils in x und y vorhanden ist.

(ii) Die Bedingung $\rho \in PC^1([0,T], \mathbb{R}^{\mathcal{X}})$ ist so zu verstehen, dass für alle $x \in \mathcal{X}$ die Abbildung $t \mapsto \rho_t(x)$ stückweise stetig differenzierbar ist. Insbesondere ist dadurch die Gleichung in (iv) wohldefiniert.

(iii) Die Differentialgleichung in Bedingung (iv) ist das diskrete Analogon zur Kontinuitätsgleichung. Dies werden wir später noch genauer untersuchen. In den folgenden Ausführungen bezeichnen wir diese Gleichung stets als (diskrete) Kontinuitätsgleichung.

Beispiel 4.1.10. Das einfachste, aber dennoch ein sehr wichtiges Beispiel ist der Fall einer

Menge der Form $\mathcal{X} = \{a, b\}$. Wir betrachten den Markov-Kern

$$Q_{p,q} = \begin{pmatrix} -p & p \\ q & -q \end{pmatrix}$$

für $p, q \in (0, \infty)$. Das zugehörige reversible Maß ist gegeben durch $\pi(\{a\}) = \frac{q}{p+q}$ und $\pi(\{b\}) = \frac{p}{p+q}$, denn

$$\pi(a)Q(a,b) = \frac{pq}{p+q} = \pi(b)Q(b,a).$$

Bezeichnen wir mit δ_a und δ_b das Diracmaß zu den Mengen $\{a\}$ und $\{b\}$, dann ist jedes Wahrscheinlichkeitsmaß μ auf der Menge $\mathcal{X}$ von der Form

$$\mu = \frac{1}{2}\left((1-\beta)\delta_a + (1+\beta)\delta_b\right)$$

für ein $\beta \in [-1, 1]$. Die zugehörige Dichte $\rho\colon \{a,b\} \to \mathbb{R}$ bezüglich des Maßes π ist dann gegeben durch

$$\rho^\beta(a) := \frac{p+q}{q}\frac{1-\beta}{2}, \quad \rho^\beta(b) = \frac{p+q}{p}\frac{1+\beta}{2}.$$

In [Maa11, Beispiel 2.6] wird die folgende Darstellung der diskreten Wasserstein-Metrik gezeigt. Für $-1 \leq \alpha \leq \beta \leq 1$ gilt

$$\mathcal{W}(\rho^\alpha, \rho^\beta) = \frac{1}{\sqrt{2}} \int_\alpha^\beta \sqrt{\frac{\log(q(1+r)) - \log(p(1-r))}{q(1+r) - p(1-r)}} \mathrm{d}r.$$

Ist $p = q$, dann gilt die Gleichung

$$\mathcal{W}(\rho^\alpha, \rho^\beta) = \frac{1}{\sqrt{2p}} \int_\alpha^\beta \sqrt{\frac{\mathrm{Artanh} r}{r}} \mathrm{d}r.$$

Nach Lemma C.0.6 ist die rechte Seite obiger Gleichung für alle α, β endlich und damit ist auch die diskrete Wasserstein-Metrik auf $\{a, b\}$ endlich, dies werden wir später verwenden, um die Endlichkeit der Metrik auf beliebigen Mengen $\mathcal{X}$ zu zeigen.

Wir möchten eine Darstellung der Abbildung $\mathcal{W}$ vorstellen, an welcher man die Ähnlichkeit zur Benamou-Brenier-Formel erkennt. Dazu führen wir einige Notationen ein. Diese Bezeichnungen werden im Laufe der folgenden Kapitel eine wichtige Rolle spielen.

Definition 4.1.11 (Diskrete Differentialoperatoren)**.** Für $\psi \in \mathbb{R}^{\mathcal{X}}$ definieren wir den diskreten Gradienten $\nabla\psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ durch

$$[\nabla\psi](x,y) = \psi(x) - \psi(y) \text{ für alle } x, y \in \mathcal{X}.$$

Die diskrete Divergenz $\nabla \cdot \Psi \in \mathbb{R}^{\mathcal{X}}$ einer Abbildung $\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ ist definiert durch

$$[\nabla \cdot \Psi](x) := \frac{1}{2}\sum_{y\in\mathcal{X}} Q(x,y)(\Psi(y,x) - \Psi(x,y))$$

für $x \in \mathcal{X}$.

Definition 4.1.12. Wir führen drei Bilinearformen ein. Für $\varphi, \psi \in \mathbb{R}^{\mathcal{X}}$ definieren wir

$$\langle \varphi, \psi \rangle_\pi = \int_{\mathcal{X}} \varphi\psi \mathrm{d}\pi = \sum_{x\in\mathcal{X}} \varphi(x)\psi(x)\pi(\{x\}).$$

Ferner definieren wir für $\Phi, \Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ die Form

$$\langle \Phi, \Psi \rangle_\pi = \frac{1}{2}\sum_{x,y\in\mathcal{X}} \Phi(x,y)\Psi(x,y)Q(x,y)\pi(\{x\})$$

sowie

$$\langle \Phi, \Psi \rangle_\rho = \frac{1}{2}\sum_{x,y\in\mathcal{X}} \Phi(x,y)\Psi(x,y)Q(x,y)\Lambda(\rho(x),\rho(y))\pi(\{x\})$$

für gegebenes $\rho \in \mathcal{D}(\mathcal{X})$.

Lemma 4.1.13 (Partielle Integration)**.** Für alle $\varphi \in \mathbb{R}^{\mathcal{X}}$ und $\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ gilt

$$\langle \nabla\varphi, \Psi \rangle_\pi = -\langle \varphi, \nabla \cdot \Psi \rangle_\pi.$$

Beweis. Sei $\varphi \in \mathbb{R}^{\mathcal{X}}$ und $\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ gegeben, dann gilt

$$\begin{aligned}
\langle \nabla\varphi, \Psi \rangle_\pi &= \frac{1}{2}\sum_{x,y\in\mathcal{X}} [\nabla\varphi](x,y)\Psi(x,y)Q(x,y)\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} [\varphi(x) - \varphi(y)]\Psi(x,y)Q(x,y)\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(x)\Psi(x,y)Q(x,y)\pi(\{x\}) - \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(y)\Psi(x,y)Q(x,y)\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(x)\Psi(x,y)Q(x,y)\pi(\{x\}) - \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(y)\Psi(x,y)Q(y,x)\pi(\{y\}) \\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(x)\Psi(x,y)Q(x,y)\pi(\{x\}) - \frac{1}{2}\sum_{y,x\in\mathcal{X}} \varphi(x)\Psi(y,x)Q(x,y)\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \varphi(x)[\Psi(x,y) - \Psi(y,x)]Q(x,y)\pi(\{x\})
\end{aligned}$$

$$= \frac{1}{2} \sum_{x \in \mathcal{X}} \varphi(x) \pi(\{x\}) \sum_{y \in \mathcal{X}} Q(x,y)[\Psi(x,y) - \Psi(y,x)]$$
$$= - \sum_{x \in \mathcal{X}} \varphi(x)[\nabla \cdot \Psi](x) \pi(\{x\}) = -\langle \varphi, \nabla \cdot \Psi \rangle_\pi .$$

Für die vierte Gleichung verwenden wir die Bedingung der Reversibilität und in der fünften Gleichung benennen wir die Variablen der zweiten Summe um. □

Bemerkung 4.1.14. Ist $\Psi = \nabla \psi$ für ein $\psi \in \mathbb{R}^{\mathcal{X}}$, dann gilt

$$(\nabla \cdot \nabla \psi) = (\nabla \cdot \Psi) = \frac{1}{2} \sum_{y \in \mathcal{X}} Q(x,y)(\psi(y) - \psi(x) - \psi(x) + \psi(y)) = \sum_{y \in \mathcal{X}} Q(x,y)(\psi(y) - \psi(x))$$

und folglich ist $(\nabla \cdot \nabla \psi)$ der diskrete Laplace-Operator bezüglich der durch Q festgelegten Geometrie im Sinne von Theorem 3.3.7.

Lemma 4.1.15. Für $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}(\mathcal{X})$ gilt

$$\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) = \sqrt{\inf_{(\rho,\psi) \in \mathcal{CE}_1(\tilde{\rho}_0, \tilde{\rho}_1)} \left\{ \int_0^1 \langle \nabla \psi_t, \nabla \psi_t \rangle_{\rho_t} \mathrm{d}t \right\}},$$

wobei wir die Kontinuitätsgleichung nun als

$$\dot{\rho}_t + \nabla \cdot (\hat{\rho}_t \bullet \nabla \psi_t) = 0$$

für alle bis auf endlich viele $0 < t < T$ schreiben können.

Beweis. Die Behauptung folgt direkt aus den eingeführten Definitionen der jeweiligen Terme. □

Bemerkung 4.1.16. (i) An obiger Darstellung der diskreten Wasserstein-Metrik erkennt man die Ähnlichkeit zur Benamou-Brenier-Formel sehr gut. Im Gegensatz zum kontinuierlichen Fall betrachten wir die Menge aller Dichtefunktionen zum Maß π und nicht die Menge aller Maße auf $\mathcal{X}$. Hier macht dies jedoch keinen relevanten Unterschied, denn da das Maß π positiv ist, können wir jedes beliebige Maß μ bezüglich π mittels einer Dichtefunktion darstellen.

(ii) Wir möchten eine mögliche physikalische Interpretation dieses Unterschieds im Sinne von Bemerkung 4.1.6 geben. Wir stellen uns die Menge $(\mathcal{X}, Q)$ als einen Graphen vor, dessen Kanten aus Wärmetauschern bestehen. Das heißt, die zu transportierende Masse liegt in Form von Wärme vor. Den Term $Q(x,y)$ interpretieren wir als Produkt des Wärmedurchgangskoeffizienten κ und der Wärmeübertragungsfläche A. Die momentane Dichte $\rho(x)$ entspreche der Temperaturdifferenz ΔT_1 und die Dichte $\rho(y)$ entspreche der Temperaturdifferenz ΔT_2 am anderen Ende des Wärmetauschers. Das heißt, zu zwei Ecken $x, y \in \mathcal{X}$ gilt für die Wärmeleistung

$$\frac{\delta \mathrm{Q}}{\delta t} = Q(x,y)\Delta T_m = Q(x,y)\Lambda(\rho(x),\rho(y))$$

oder anders formuliert, hängt die von der Ecke x zur Ecke y übertragene Wärme zusätzlich von dieser Gewichtung ab. Dies gibt eine mögliche Erklärung für das Auftreten des Terms $Q(x,y)\Lambda(\rho(x),\rho(y))$ im Integranden der diskreten Wasserstein-Metrik.

Bemerkung 4.1.17. Wir möchten noch eine weitere, vor allem für Rechnungen nützliche Darstellung vorstellen. Dazu definieren wir für $\rho \in \mathcal{D}(\mathcal{X})$ die beiden Abbildungen $A(\rho), B(\rho) \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ durch

$$[A(\rho)](x,y) := \begin{cases} \sum\limits_{\substack{z\in\mathcal{X}\\ z\neq x}} Q(x,z)\Lambda(\rho(x),\rho(z))\pi(\{x\}) & : x = y \\ -Q(x,y)\Lambda(\rho(x),\rho(y))\pi(\{x\}) & : x \neq y \end{cases}$$

sowie

$$[B(\rho)](x,y) = \begin{cases} \sum\limits_{\substack{z\in\mathcal{X}\\ z\neq x}} Q(x,z)\Lambda(\rho(x),\rho(z)) & : x = y \\ -Q(x,y)\Lambda(\rho(x),\rho(y)) & : x \neq y. \end{cases}$$

Bezeichnen wir mit $\Pi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ die Diagonalmatrix mit den Einträgen $\Pi(x,x) = \pi(\{x\})$ für $x \in \mathcal{X}$, dann gilt

$$A(\rho) = \Pi B(\rho).$$

Lemma 4.1.18. Für alle $\rho \in \mathcal{D}(\mathcal{X})$ ist die Matrix $A(\rho)$ symmetrisch und positiv semidefinit. Insbesondere gilt für alle $\rho \in \mathcal{D}(\mathcal{X})$ und $\psi \in \mathbb{R}^{\mathcal{X}}$ die Ungleichung $\langle A(\rho)\psi, \psi\rangle \geq 0$.

Beweis. Seien $x, y \in \mathcal{X}$ und $x \neq y$, dann gilt

$$[A(\rho)](x,y) = -\Lambda(\rho(x),\rho(y))Q(x,y)\pi(\{x\}) = -\Lambda(\rho(y),\rho(x))Q(y,x)\pi(\{y\}) = [A(\rho)](y,x)$$

aufgrund der Symmetrie von Λ und der Reversibilität von Q. Dies zeigt die Symmetrie von $A(\rho)$. Weiter gilt für ein beliebiges $x \in \mathcal{X}$

$$0 \leq \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} |[A(\rho)](x,y)| = \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} Q(x,y)\Lambda(\rho(x),\rho(y))\pi(\{x\}) = [A(\rho)](x,x),$$

folglich ist $A(\rho)$ schwach diagonaldominant. Aus der Symmetrie und der Nichtnegativität der

Diagonaleinträge folgt mit der schwachen Diagonaldominanz und Lemma C.0.7 die positive Semidefinitheit von $A(\rho)$. □

Lemma 4.1.19. Seien $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$, dann gilt die Darstellung

$$\mathcal{W}(\rho_0, \rho_1) = \sqrt{\inf_{(\rho,\psi)\in\mathcal{CE}_1(\rho_0,\rho_1)} \left\{ \int_0^1 \langle A(\rho_t)\psi_t, \psi_t\rangle \mathrm{d}t \right\}}$$

und die Kontinuitätsgleichung lässt sich schreiben als $\dot{\rho}_t = B(\rho_t)\psi_t$.

Beweis. Seien $\rho \in \mathcal{D}(\mathcal{X})$ und $\psi \in \mathbb{R}^{\mathcal{X}}$ beliebig. Wir rechnen die Gleichheit des Integranden nach. Es gilt

$$\begin{aligned}
2\langle A(\rho)\psi, \psi\rangle &= 2\sum_{x,y\in\mathcal{X}} [A(\rho)](x,y)\psi(x)\psi(y) = \sum_{x,y\in\mathcal{X}} [A(\rho)](x,y)2\psi(x)\psi(y) \\
&= \sum_{x,y\in\mathcal{X}} [A(\rho)](x,y)(\psi(x)^2 + \psi(y)^2 - (\psi(x) - \psi(y))^2) \\
&= \sum_{x\in\mathcal{X}} \psi(x)^2 \sum_{y\in\mathcal{X}} [A(\rho)](x,y) + \sum_{y\in\mathcal{X}} \psi(y)^2 \sum_{x\in\mathcal{X}} [A(\rho)](x,y) \\
&\quad - \sum_{x,y\in\mathcal{X}} [A(\rho)](x,y) \left(\psi(x) - \psi(y)\right)^2 \\
&= 0 + \sum_{y\in\mathcal{X}} \psi(y)^2 \sum_{x\in\mathcal{X}} [A(\rho)](y,x) - \sum_{x,y\in\mathcal{X}} [A(\rho)](x,y) \left(\psi(x) - \psi(y)\right)^2 \\
&= 0 - \sum_{x,y\in\mathcal{X}} [A(\rho)](x,y) \left(\psi(x) - \psi(y)\right)^2 \\
&= -\sum_{\substack{x,y\in\mathcal{X} \\ x\neq y}} [A(\rho)](x,y) \left(\psi(x) - \psi(y)\right)^2 \\
&= \sum_{\substack{x,y\in\mathcal{X} \\ x\neq y}} \left(\psi(x) - \psi(y)\right)^2 Q(x,y)\Lambda(\rho(x), \rho(y))\pi(\{x\}) \\
&= \sum_{x,y\in\mathcal{X}} \left(\psi(x) - \psi(y)\right)^2 Q(x,y)\Lambda(\rho(x), \rho(y))\pi(\{x\}).
\end{aligned}$$

Für $x \in \mathcal{X}$ ist

$$\begin{aligned}
[B(\rho)\psi](x) &= \sum_{y\in\mathcal{X}} [B(\rho)](x,y)\psi(y) \\
&= -\sum_{\substack{y\in\mathcal{X} \\ y\neq x}} Q(x,y)\Lambda(\rho(x), \rho(y))\psi(y) + \sum_{\substack{z\in\mathcal{X} \\ z\neq x}} Q(x,z)\Lambda(\rho(x), \rho(z))\psi(x)
\end{aligned}$$

$$= -\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} Q(x,y)\Lambda(\rho(x),\rho(y))\left(\psi(y)-\psi(x)\right),$$

dies zeigt die Darstellung der Kontinuitätsgleichung. □

Lemma 4.1.20. Sei $T > 0$, dann gilt die Darstellung

$$\mathcal{W}(\rho_0,\rho_1) = \sqrt{T}\sqrt{\inf_{(\rho,\psi)\in\mathcal{CE}_T(\rho_0,\rho_1)}\left\{\int_0^T \langle A(\rho_t)\psi_t,\psi_t\rangle \mathrm{d}t\right\}}$$

für alle $\rho_0,\rho_1 \in \mathcal{D}(\mathcal{X})$.

Beweis. Sei $T > 0$ und seien $\rho_0,\rho_1 \in \mathcal{D}(\mathcal{X})$. Wir definieren für $(\rho,\psi) \in \mathcal{CE}_1(\rho_0,\rho_1)$ die Abbildungen $\rho'_t := \rho_{\frac{t}{T}}$ und $\psi'_t := \frac{1}{T}\psi_{\frac{t}{T}}$ für $t \in [0,T]$. Für alle $x \in \mathcal{X}$ folgt

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho'_t(x) = \frac{1}{T}\dot{\rho}_{\frac{t}{T}} = \frac{1}{T}B(\rho_{\frac{t}{T}})\psi_{\frac{t}{T}} = B(\rho'_t)\frac{1}{T}\psi_{\frac{t}{T}} = B(\rho'_t)\psi'_t$$

unter Verwendung der Darstellung aus Lemma 4.1.19. Mithilfe dieser Gleichung sieht man, dass $(\rho',\psi') \in \mathcal{CE}_T(\rho_0,\rho_1)$ gilt. Wir berechnen den Integranden der rechten Seite obiger Darstellung für (ρ',ψ'):

$$\int_0^T \langle A(\rho'_t)\psi'_t,\psi'_t\rangle\mathrm{d}t = \frac{1}{T^2}\int_0^T \langle A(\rho_{\frac{t}{T}})\psi_{\frac{t}{T}},\psi_{\frac{t}{T}}\rangle\mathrm{d}t = \frac{1}{T}\int_0^1 \langle A(\rho_s)\psi_s,\psi_s\rangle\mathrm{d}s.$$

Der Term der letzten Gleichung ist der Integrand zur Darstellung der Wasserstein-Metrik aus Lemma 4.1.19. Wir haben gezeigt, für alle $(\rho,\psi) \in \mathcal{CE}_1(\rho_0,\rho_1)$ ist $(\rho',\psi') \in \mathcal{CE}_T(\rho_0,\rho_1)$ und es gilt

$$\int_0^T \langle A(\rho'_t)\psi'_t,\psi'_t\rangle\mathrm{d}t = \frac{1}{T}\int_0^1 \langle A(\rho_s)\psi_s,\psi_s\rangle\mathrm{d}s.$$

Dies zeigt die Ungleichung "$\geq$" der gewünschten Gleichung. Für die Richtung "$\leq$" argumentiert man ähnlich und skaliert dazu die Elemente von $\mathcal{CE}_T(\rho_0,\rho_1)$ zu Elementen der Menge $\mathcal{CE}_1(\rho_0,\rho_1)$. □

Lemma 4.1.21. Für alle $\rho_0,\rho_1 \in \mathcal{D}(\mathcal{X})$ gilt

$$d_{TV}(\rho_0,\rho_1) \leq \sqrt{2\|\Lambda\|_\infty}\mathcal{W}(\rho_0,\rho_1),$$

wobei d_{TV} die Norm der totalen Variation bezeichnet. Aufgrund der Endlichkeit von $\mathcal{X}$ stimmt diese mit dem $L^1(\mathcal{X},\pi)$ Abstand zwischen ρ_0 und ρ_1 überein. Genauer gilt

$$d_{TV}(\rho_0,\rho_1) = \int_{\mathcal{X}} |\rho_0-\rho_1|\mathrm{d}\pi = \sum_{x\in\mathcal{X}} |\rho_0(x)-\rho_1(x)|\pi(\{x\})$$

sowie $\|\Lambda\|_\infty = \sup\{\Lambda(s,t) | 0 \leq s,t \leq (\min_{x\in\mathcal{X}} \pi(\{x\}))^{-1}\}$.

Beweis. Seien $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ beliebig. Wir nehmen an, dass $\mathcal{W}(\rho_0,\rho_1) < \infty$ gilt, denn sonst ist nichts zu zeigen. Zu einem beliebigen $\varepsilon > 0$ wählen wir mit Lemma 4.1.19 ein $(\rho,\psi) \in \mathcal{CE}_1(\rho_0,\rho_1)$, sodass

$$\left(\int_0^1 \langle A(\rho_t)\psi_t, \psi_t\rangle \mathrm{d}t\right)^{\frac{1}{2}} < \mathcal{W}(\rho_0,\rho_1) + \varepsilon$$

gilt. Für eine beliebige Funktion $\varphi\colon \mathcal{X} \to \mathbb{R}$ gilt

$$\begin{aligned}
\left|\sum_{x\in\mathcal{X}} \varphi(x)\left(\rho_0(x) - \rho_1(x)\right)\pi(\{x\})\right| &= \left|\sum_{x\in\mathcal{X}} \varphi(x)\pi(\{x\}) \int_0^1 \dot{\rho}_t(x)\mathrm{d}t\right| = \left|\int_0^1 \langle \Pi\varphi, \dot{\rho}_t\rangle \mathrm{d}t\right| \\
&= \left|\int_0^1 \langle \Pi\varphi, B(\rho_t)\psi_t\rangle \mathrm{d}t\right| = \left|\int_0^1 \langle \varphi, A(\rho_t)\psi_t\rangle \mathrm{d}t\right| \\
&\leq \int_0^1 \left|\langle \varphi, A(\rho_t)\psi_t\rangle\right| \mathrm{d}t,
\end{aligned}$$

denn $\Pi B(\rho) = A(\rho)$. Wir haben bereits gesehen, dass $A(\rho)$ symmetrisch und positiv semidefinit ist, folglich können wir die Cauchy-Schwarz-Ungleichung auf die Bilinearform $\langle A\cdot,\cdot\rangle$ anwenden und erhalten

$$\begin{aligned}
\int_0^1 \left|\langle A(\rho_t)\varphi, \psi_t\rangle\right| \mathrm{d}t &\leq \int_0^1 \sqrt{\langle A(\rho_t)\varphi,\varphi\rangle}\sqrt{\langle A(\rho_t)\psi_t,\psi_t\rangle}\mathrm{d}t \\
&\leq \left(\int_0^1 \langle A(\rho_t)\varphi,\varphi\rangle \mathrm{d}t\right)^{\frac{1}{2}} \left(\int_0^1 \langle A(\rho_t)\psi_t,\psi_t\rangle \mathrm{d}t\right)^{\frac{1}{2}}.
\end{aligned}$$

Es gilt $\rho_t \in \mathcal{D}(\mathcal{X})$ für alle $t \in [0,1]$, deswegen ist $\rho_t(x) \geq 0$ für alle $x \in \mathcal{X}$ und alle $t \in [0,1]$. Bezeichnen wir mit $m = \min_{x\in\mathcal{X}} \pi(\{x\})$, dann ist $\rho_t(x) \leq \frac{1}{m}$ für alle $x \in \mathcal{X}$ und $t \in [0,1]$. Denn für $x \in \mathcal{X}$ gilt

$$1 = \sum_{y\in\mathcal{X}} \rho_t(y)\pi(y) \geq m \sum_{y\in\mathcal{X}} \rho_t(y) \geq m\rho_t(x).$$

Ferner ist $\pi(x) > 0$ für alle $x \in \mathcal{X}$ und folglich ist auch $m > 0$. Aufgrund der Monotonie und der Stetigkeit des logarithmischen Mittels folgt $0 < \Lambda(\rho_t(x),\rho_t(y)) \leq \|\Lambda\|_\infty < \infty$ für alle $x,y \in \mathcal{X}$ und $t \in [0,1]$.

Wir widmen uns nun zuerst dem Integranden des linken Terms in der rechten Seite obiger

Abschätzung. Für $t \in [0,1]$ gilt

$$\begin{aligned}
\langle A(\rho_t)\varphi, \varphi\rangle &= \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} (\varphi(x)-\varphi(y))^2\, Q(x,y)\Lambda(\rho_t(x),\rho_t(y))\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} (\varphi(x)-\varphi(y))^2\, Q(x,y)\Lambda(\rho_t(x),\rho_t(y))\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} (\varphi(x)-\varphi(y))^2\, (Q(x,y)+\mathrm{Id}(x,y))\,\Lambda(\rho_t(x),\rho_t(y))\pi(\{x\}) \\
&= \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} (\varphi(x)-\varphi(y))^2\, (Q(x,y)+\mathrm{Id}(x,y))\,\Lambda(\rho_t(x),\rho_t(y))\pi(\{x\}) \\
&\leq \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} \left(2\varphi(x)^2+2\varphi(y)^2\right) (Q(x,y)+\mathrm{Id}(x,y))\,\Lambda(\rho_t(x),\rho_t(y))\pi(\{x\}) \\
&\leq 2\|\varphi\|_\infty^2\|\Lambda\|_\infty \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} (Q(x,y)\pi(\{x\})+\mathrm{Id}(x,y)\pi(\{x\})) \\
&\leq 2\|\varphi\|_\infty^2\|\Lambda\|_\infty \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} Q(x,y)\pi(\{x\}) + 2\|\varphi\|_\infty^2\|\Lambda\|_\infty \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} \mathrm{Id}(x,y)\pi(\{x\}) \\
&= 0 + 2\|\varphi\|_\infty^2\|\Lambda\|_\infty \sum_{x\in\mathcal{X}} \pi(\{x\}) = 2\|\varphi\|_\infty^2\|\Lambda\|_\infty.
\end{aligned}$$

In der vorletzten Gleichung verwenden wir die Charakterisierung der Invarianz eines Maßes aus Lemma 3.2.4. Wir definieren $\varphi\colon \mathcal{X}\to\mathbb{R}$, $\varphi(x) = \operatorname{sgn}(\rho_0(x)-\rho_1(x))$ für $x\in\mathcal{X}$, dann ist $|\rho_0(x)-\rho_1(x)| = \varphi(x)\,(\rho_0(x)-\rho_1(x))$ für alle $x\in\mathcal{X}$ und $\|\varphi\|_\infty = 1$. Insgesamt folgt

$$\begin{aligned}
d_{TV}(\rho_0,\rho_1) = \sum_{x\in\mathcal{X}} |\rho_0(x)-\rho_1(x)|\pi(\{x\}) &= \left|\sum_{x\in\mathcal{X}} \varphi(x)\,(\rho_0(x)-\rho_1(x))\,\pi(\{x\})\right| \\
&\leq \left(\int_0^1 \langle A(\rho_t)\varphi,\varphi\rangle \mathrm{d}t\right)^{\frac{1}{2}} \left(\int_0^1 \langle A(\rho_t)\psi_t,\psi_t\rangle \mathrm{d}t\right)^{\frac{1}{2}} \\
&\leq \sqrt{2\|\Lambda\|_\infty}\|\varphi\|_\infty \left(\int_0^1 \langle A(\rho_t)\psi_t,\psi_t\rangle \mathrm{d}t\right)^{\frac{1}{2}} \\
&< \sqrt{2\|\Lambda\|_\infty}\,(\mathcal{W}(\rho_0,\rho_1)+\varepsilon)\,.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig ist, folgt die Behauptung. □

Satz 4.1.22. Das Paar $(\mathcal{D}(\mathcal{X}),\mathcal{W})$ ist ein metrischer Raum.

Für den Beweis der Dreiecksungleichung sowie der Endlichkeit der Metrik benötigen wir zwei weitere Lemmata. Wir führen eine weitere Darstellung ein. Diese Beschreibung der diskreten

Wasserstein-Metrik wird auch von Bedeutung sein, wenn wir die riemannsche Struktur der Metrik untersuchen.

Lemma 4.1.23. Für die Wasserstein-Metrik gilt für alle $T > 0$ und $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ die Darstellung

$$\mathcal{W}(\rho_0, \rho_1) = \inf_{(\rho,\psi)\in\mathcal{CE}_T(\rho_0,\rho_1)} \int_0^T \sqrt{\langle A(\rho_t)\psi_t, \psi_t\rangle}\mathrm{d}t.$$

Beweis. Seien $T > 0$ und $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ beliebig. Wir bezeichnen mit $\mathcal{M}(\rho_0, \rho_1)$ die rechte Seite obiger Gleichung. Ist $(\rho, \psi) \in \mathcal{CE}_1(\rho_0, \rho_1)$, dann können wir dies wie im Beweis von Lemma 4.1.20 zu einem Element $(\rho', \psi') \in \mathcal{CE}_T(\rho_0, \rho_1)$ skalieren. Ferner gilt dann unter Verwendung der Cauchy-Schwarz-Ungleichung

$$\int_0^T \sqrt{\langle A(\rho'_t)\psi'_t, \psi'_t\rangle}\mathrm{d}t \le \sqrt{\int_0^T \mathrm{d}t}\sqrt{\int_0^T \langle A(\rho'_t)\psi'_t, \psi'_t\rangle\mathrm{d}t)} = \sqrt{T}\frac{1}{\sqrt{T}}\sqrt{\int_0^1 \langle A(\rho_t)\psi_t, \psi_t\rangle\mathrm{d}t}.$$

Dies impliziert $\mathcal{M}(\rho_0, \rho_1) \le \mathcal{W}(\rho_0, \rho_1)$. Für den Beweis der anderen Richtung sei $(\rho, \psi) \in \mathcal{CE}_T(\rho_0, \rho_1)$. Für $\varepsilon > 0$ und $t \in [0, T]$ definieren wir

$$\Phi_\varepsilon(t) := \int_0^t \sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s.$$

Dann ist die Funktion Φ_ε streng monoton wachsend und nach dem Hauptsatz der Differential- und Integralrechung differenzierbar. Aus diesen Gründen existiert die inverse Funktion, das heißt, $\Phi_\varepsilon^{-1}\colon [0, \Phi_\varepsilon(T)] \to [0, T]$ und für ihre Ableitung $(\Phi_\varepsilon^{-1})'$ gilt

$$(\Phi_\varepsilon^{-1})'(r) = (\Phi_\varepsilon^{-1})'(\Phi_\varepsilon(s)) = \frac{1}{\sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}}$$

für alle $s \in [0, T]$ und $r = \Phi_\varepsilon(s)$. Für $t \in [0, \Phi_\varepsilon(T)]$ definieren wir

$$\rho_t^\varepsilon := \rho_{\Phi_\varepsilon^{-1}(t)} \text{ sowie } \psi_t^\varepsilon := (\Phi_\varepsilon^{-1})'(t)\psi_{\Phi_\varepsilon^{-1}(t)}$$

und behaupten, dass $(\rho^\varepsilon, \psi^\varepsilon) \in \mathcal{CE}_{\Phi_\varepsilon(T)}(\rho_0, \rho_1)$ ist. Die Bedingungen (i) bis (iii) sind eine direkte Konsequenz unserer Defintion von $(\rho^\varepsilon, \psi^\varepsilon)$. Für $x \in \mathcal{X}$ und $t \in [0, \Phi_\varepsilon(T)]$ gilt

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t^\varepsilon(x) = (\Phi_\varepsilon^{-1})'(t)\dot{\rho}_{\Phi_\varepsilon^{-1}(t)}(x) = -(\Phi_\varepsilon^{-1})'(t)[B(\rho_{\Phi_\varepsilon^{-1}(t)})\psi_{\Phi_\varepsilon^{-1}(t)}](x) = -[B(\rho_t^\varepsilon)\psi_t^\varepsilon](x)$$

und damit gilt auch die Kontinuitätsgleichung. Mithilfe von Lemma 4.1.20 folgern wir

$$\mathcal{W}(\rho_0, \rho_1) \le \sqrt{\Phi_\varepsilon(T) \int_0^{\Phi_\varepsilon(T)} \langle A(\rho_t^\varepsilon)\psi_t^\varepsilon, \psi_t^\varepsilon\rangle\mathrm{d}t}$$

$$
\begin{aligned}
&= \sqrt{\Phi_\varepsilon(T)\int_0^{\Phi_\varepsilon(T)} \langle A(\rho_{\Phi_\varepsilon^{-1}(t)})(\Phi_\varepsilon^{-1})'(t)\psi_{\Phi_\varepsilon^{-1}(t)}, (\Phi_\varepsilon^{-1})'(t)\psi_{\Phi_\varepsilon^{-1}(t)}\rangle \mathrm{d}t} \\
&= \sqrt{\Phi_\varepsilon(T)\int_0^{T} \langle A(\rho_s)(\Phi_\varepsilon^{-1})'(\Phi_\varepsilon(s))\psi_s, (\Phi_\varepsilon^{-1})'(\Phi_\varepsilon(s))\psi_s\rangle \frac{1}{(\Phi_\varepsilon^{-1})'(\Phi_\varepsilon(s))}\mathrm{d}s} \\
&= \sqrt{\Phi_\varepsilon(T)\int_0^{T} \langle A(\rho_s)\psi_s, \psi_s\rangle (\Phi_\varepsilon^{-1})'(\Phi_\varepsilon(s))\mathrm{d}s} \\
&= \sqrt{\Phi_\varepsilon(T)\int_0^{T} \frac{\langle A(\rho_s)\psi_s, \psi_s\rangle}{\sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}}\mathrm{d}s} \\
&= \sqrt{\Phi_\varepsilon(T)\int_0^{T} \frac{\langle A(\rho_s)\psi_s, \psi_s\rangle}{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s} \\
&\leq \sqrt{\Phi_\varepsilon(T)\int_0^{T} \sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s} \\
&= \sqrt{\left(\int_0^{T} \sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s\right)^2} \\
&= \int_0^{T} \sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s,
\end{aligned}
$$

wobei wir in der letzten Gleichung die Definition von $\Phi_\varepsilon(T)$ verwenden. Wir haben gezeigt, dass für alle $(\rho, \psi) \in \mathcal{CE}_T(\rho_0, \rho_1)$

$$
\mathcal{W}(\rho_0, \rho_1) \leq \int_0^{T} \sqrt{\varepsilon + \langle A(\rho_s)\psi_s, \psi_s\rangle}\mathrm{d}s
$$

gilt, dies zeigt $\mathcal{W}(\rho_0, \rho_1) \leq \mathcal{M}(\rho_0, \rho_1)$ für $\varepsilon \to 0^+$. □

Das folgende Lemma benötigen wir, um die Endlichkeit der Abbildung $\mathcal{W}$ zu zeigen. Wir haben in Beispiel 4.1.10 gesehen, dass die Wasserstein-Metrik auf der zweipunktigen Menge stets endlich ist. Das Lemma vergleicht die Wasserstein-Metrik auf $\mathcal{X}$ von bestimmten Dichten mit der Wasserstein-Metrik auf einer zweipunktigen Menge.

Lemma 4.1.24. Seien $x, y \in \mathcal{X}$ zwei verschiedene Punkte mit $Q(x, y) > 0$. Wir definieren $p := Q(x, y)\pi(\{x\})$. Seien ρ_0, ρ_1 zwei Wahrscheinlichkeitsdichten, die auf der Menge $\mathcal{X} \setminus \{x, y\}$ übereinstimmen, das heißt, es gelte $\rho_0(z) = \rho_1(z)$ für alle $z \in \mathcal{X} \setminus \{x, y\}$. Wir betrachten die Menge $\{x, y\}$ ausgestattet mit dem Markov-Kern $Q_{p,p}$ aus Beispiel 4.1.10.

Für $i = 0, 1$ seien $\tilde{\rho}_i \colon \{x, y\} \to \mathbb{R}_+$ definiert durch

$$\tilde{\rho}_i(x) := \frac{2\rho_i(x)\pi(\{x\})}{\rho_i(x)\pi(\{x\}) + \rho_i(y)\pi(\{y\})} \text{ und } \tilde{\rho}_i(y) := \frac{2\rho_i(y)\pi(\{y\})}{\rho_i(x)\pi(\{x\}) + \rho_i(y)\pi(\{y\})}.$$

Dann gilt die Ungleichung

$$\mathcal{W}(\rho_0, \rho_1) \leq \mathcal{W}_p(\tilde{\rho}_0, \tilde{\rho}_1) < \infty,$$

wobei $\mathcal{W}_p$ die Wasserstein-Metrik auf $(\{x, y\}, Q_{p,p})$ bezeichnet.

Beweis. Einen Beweis dieser Aussage findet man in [Maa11, Lemma 3.14] unter der Verwendung der Eigenschaft (iv) des logarithmischen Mittels aus Bemerkung 4.1.5. Die Endlichkeit der Wasserstein-Metrik auf $\{x, y\}$ haben wir bereits in Beispiel 4.1.10 angesprochen. □

Beweis von Satz 4.1.22. Lemma 4.1.21 impliziert $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) \geq 0$ für alle $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}(\mathcal{X})$ und dass $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) = 0$ genau dann gilt, wenn schon $\tilde{\rho}_0 = \tilde{\rho}_1$ ist. Dies impliziert die positive Definitheit der Metrik. Wir widmen uns der Symmetrie der Abbildung $\mathcal{W}$. Seien $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}(\mathcal{X})$. Wir möchten zeigen, dass $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) = \mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_0)$ gilt. Sei dazu $(\rho, \psi) \in \mathcal{CE}(\tilde{\rho}_0, \tilde{\rho}_1)$. Wir definieren (ρ', ψ') durch $\rho'_t = \rho_{1-t}$ und $\psi'_t = -\psi_{1-t}$ für $t \in [0, 1]$. Dann gilt $(\rho', \psi') \in \mathcal{CE}(\tilde{\rho}_1, \tilde{\rho}_1)$, denn per definitionem sind die charakterisierenden Eigenschaften (i) bis (iii) aus Definition 4.1.8 erfüllt. Wir rechnen noch die Kontinuitätsgleichung nach. Es gilt

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho'_t = \frac{\mathrm{d}}{\mathrm{d}t}\rho_{1-t} = -\dot{\rho}_{1-t} = \nabla \cdot (\hat{\rho}_{1-t} \bullet \nabla\psi_{1-t}) = -\nabla \cdot (\hat{\rho}_{1-t} \bullet \nabla(-\psi_{1-t})) = -\nabla \cdot ((\hat{\rho'})_t \bullet \nabla\psi'_t)$$

und somit ist $(\rho', \psi') \in \mathcal{CE}(\tilde{\rho}_1, \tilde{\rho}_1)$. Weiter gilt

$$\begin{aligned}
\int_0^1 \langle A(\rho'_t)\psi'_t, \psi'_t\rangle \mathrm{d}t &= \int_0^1 \sum_{x,y\in\mathcal{X}} \left(\psi'_t(x) - \psi'_t(y)\right)^2 Q(x,y)\Lambda(\rho'_t(x), \rho'_t(y))\pi(x)\mathrm{d}t \\
&= \int_0^1 \sum_{x,y\in\mathcal{X}} \left(-\psi_{1-t}(x) + \psi_{1-t}(y)\right)^2 Q(x,y)\Lambda(\rho_{1-t}(x), \rho_{1-t}(y))\pi(x)\mathrm{d}t \\
&= \int_1^0 \sum_{x,y\in\mathcal{X}} \left(-\psi_s(x) + \psi_s(y)\right)^2 Q(x,y)\Lambda(\rho_s(x), \rho_s(y))\pi(x)\mathrm{d}s \\
&= \int_0^1 \sum_{x,y\in\mathcal{X}} \left(\psi_s(x) - \psi_s(y)\right)^2 Q(x,y)\Lambda(\rho_s(x), \rho_s(y))\pi(x)\mathrm{d}s.
\end{aligned}$$

Wir haben gezeigt, dass für alle $(\rho, \psi) \in \mathcal{CE}(\tilde{\rho}_0, \tilde{\rho}_1)$ ein $(\rho', \psi') \in \mathcal{CE}(\tilde{\rho}_1, \tilde{\rho}_0)$ derart existiert, dass die Terme, für deren Infimum wir uns interessieren, denselben Wert annehmen. Wir schließen, $\mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_0) \leq \mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1)$. Mithilfe obiger Argumente kann man durch Vertauschen von $\tilde{\rho}_0$ und $\tilde{\rho}_1$ auch zeigen, dass $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) \leq \mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_0)$ ist und somit folgt die Symmetrie $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) = \mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_0)$.

Wir werden nun die Dreiecksungleichung beweisen. Seien $\tilde{\rho}_0, \tilde{\rho}_1, \tilde{\rho}_2 \in \mathcal{D}(\mathcal{X})$. Für ein $\varepsilon > 0$ beliebig, wählen wir nach Lemma 4.1.23 $(\rho^*, \psi^*) \in \mathcal{CE}_1(\tilde{\rho}_0, \tilde{\rho}_1)$ und $(\rho', \psi') \in \mathcal{CE}_1(\tilde{\rho}_1, \tilde{\rho}_2)$, sodass

$$\int_0^1 \sqrt{\langle A(\rho_t^*)\psi_t^*, \psi_t^* \rangle} \mathrm{d}t < \mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) + \varepsilon$$

und

$$\int_0^1 \sqrt{\langle A(\rho_t')\psi_t', \psi_t' \rangle} \mathrm{d}t < \mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_2) + \varepsilon$$

gilt. Wir definieren $(\rho, \psi) \in PC^1([0,1]; \mathcal{D}(\mathcal{X})) \times \mathcal{M}([0,1]; \mathbb{R}^{\mathcal{X}})$ durch

$$\rho_t = \begin{cases} \rho_{2t}^* & t \in \left[0, \frac{1}{2}\right] \\ \rho_{2t-1}' & t \in \left[\frac{1}{2}, 1\right] \end{cases}$$

sowie

$$\psi_t = \begin{cases} 2\psi_{2t}^* & t \in \left[0, \frac{1}{2}\right] \\ 2\psi_{2t-1}' & t \in \left[\frac{1}{2}, 1\right] . \end{cases}$$

Dann ist $(\rho, \psi) \in \mathcal{CE}(\tilde{\rho}_0, \tilde{\rho}_2)$, denn die benötigten Eigenschaften (i) bis (iii) sind nach Konstruktion erfüllt. Für die Gültigkeit der Kontinuitätsgleichung verweisen wir auf den Beweis von Lemma 4.1.23, dort haben wir eine ähnliche Rechnung bereits durchgeführt. Wir schließen

$$\begin{aligned} \int_0^1 \sqrt{\langle A(\rho_t)\psi_t, \psi_t \rangle} \mathrm{d}t &= \int_0^{\frac{1}{2}} \sqrt{\langle A(\rho_{2t}^*) 2\psi_{2t}^*, 2\psi_{2t}^* \rangle} \mathrm{d}t + \int_{\frac{1}{2}}^1 \sqrt{\langle A(\rho_{2t-1}') 2\psi_{2t-1}', 2\psi_{2t-1}' \rangle} \mathrm{d} \\ &= \int_0^1 2\sqrt{\langle A(\rho_s^*)\psi_s^*, \psi_s^* \rangle} \frac{1}{2} \mathrm{d}s + \int_0^1 2\sqrt{\langle A(\rho_s')\psi_s', \psi_s' \rangle} \frac{1}{2} \mathrm{d}s \\ &< \mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1) + \mathcal{W}(\tilde{\rho}_1, \tilde{\rho}_2) + 2\varepsilon \end{aligned}$$

und da $\varepsilon > 0$ beliebig war, folgt die Dreiecksungleichung.

Wir zeigen zuletzt noch die Endlichkeit der Abbildung $\mathcal{W}(\tilde{\rho}_0, \tilde{\rho}_1)$ für beliebige $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}(\mathcal{X})$. Sei $\mathcal{X} = \{x_1, \ldots, x_n\}$ für paarweise verschiedene $x_1, \ldots, x_n$. Für $i \in \{0, \ldots, n\}$ definieren wir

$$\rho^{(i)}(x) = \begin{cases} \sum\limits_{k=1}^{i} \rho_0(x_k) \frac{\pi(\{x_k\})}{\pi(\{x_1\})} & : x = x_1 \\ 0 & : x = \{x_2, \ldots, x_i\} \\ \rho_0(x) & : x \in \{x_{i+1}, \ldots, x_n\}, \end{cases}$$

dann gilt insbesondere $\tilde{\rho}_0 = \rho^{(0)}$, $\rho^{(n)} = \frac{1}{\pi(\{x_1\})}\mathbb{1}_{\{x_1\}}$ und $\rho^{(i)}(x) = \rho^{(i+1)}(x)$ für alle $x \in \mathcal{X} \setminus \{x_1, x_{i+1}\}$. Wir wenden die Dreiecksungleichung an und erhalten

$$\mathcal{W}(\tilde{\rho}_0, \rho^{(n)}) \leq \sum_{i=0}^{n-1} \mathcal{W}(\rho^{(i)}, \rho^{(i+1)}).$$

Wir würden gerne Lemma 4.1.24 auf die Dichten $\rho^{(i)}$ und $\rho^{(i+1)}$ und die Menge $\{x_1, x_{i+1}\}$ für $i \in \{1, \dots, n-1\}$ anwenden, um eine obere Schranke für $\mathcal{W}(\rho^{(i)}, \rho^{(i+1)})$ zu erhalten. Das funktioniert in den Fällen, in denen $Q(x_1, x_{i+1}) > 0$ ist. In einem solchen Fall folgt mit Lemma 4.1.24, dass $\mathcal{W}(\rho^{(i)}, \rho^{(i+1)}) < \infty$ gilt. Im Fall $Q(x_1, x_{i+1}) = 0$ existiert aufgrund der Irreduzibilität ein Pfad $x_1 = \gamma_0, \gamma_1, \dots, \gamma_{n_i} = x_{i+1}$ mit $Q(\gamma_k, \gamma_{k+1}) > 0$ für alle $k \in \{0, \dots, n_i - 1\}$. Wir können ohne Beschränkung der Allgemeinheit annehmen, dass $\left(\rho^{(i)}(\gamma_0) - \rho^{(i+1)}(\gamma_0)\right) \geq 0$ gilt, denn sonst vertauschen wir mithilfe der Symmetrie von $\mathcal{W}$ die Position der beiden Dichten. Wir definieren die Dichte $\rho^{(i,1)}$ durch

$$\rho^{(i,1)}(x) := \begin{cases} \rho^{(i)}(x) & : x \in \mathcal{X} \setminus \{\gamma_0, \gamma_1\} \\ \rho^{(i+1)}(x) & : x = \gamma_0 \\ \rho^{(i)}(\gamma_1) + \left(\rho^{(i)}(\gamma_0) - \rho^{(i+1)}(\gamma_0)\right) \frac{\pi(\{\gamma_0\})}{\pi(\{\gamma_1\})} & : x = \gamma_1 \end{cases}$$

und für $l \in \{2, \dots, n_i - 1\}$ die Dichten

$$\rho^{(i,l)}(x) := \begin{cases} \rho^{(i,l-1)}(x) & : x \in \mathcal{X} \setminus \{\gamma_{l-1}, \gamma_l\} \\ \rho^{(i,l-1)}(x) - \left(\rho^{(i)}(\gamma_0) - \rho^{(i+1)}(\gamma_0)\right) \frac{\pi(\{\gamma_0\})}{\pi(\{\gamma_{l-1}\})} & : x = \gamma_{l-1} \\ \rho^{(i,l-1)}(\gamma_{l+1}) + \left(\rho^{(i)}(\gamma_0) - \rho^{(i+1)}(\gamma_0)\right) \frac{\pi(\{\gamma_0\})}{\pi(\{\gamma_l\})} & : x = \gamma_l. \end{cases}$$

Wir setzen $\rho^{(i,0)} = \rho^{(i)}$ und $\rho^{(i,n_i)} = \rho^{(i+1)}$. Mit jedem Schritt in l transportieren wir die Masse $\left(\rho^{(i)}(\gamma_0) - \rho^{(i+1)}(\gamma_0)\right)$ entlang des Pfades ein Stück weiter. Die Definition der $\rho^{(i,l)}$ impliziert, dass $\rho^{(i,1)} \in \mathcal{D}(\mathcal{X})$ für alle $l \in \{0, \dots, n_i\}$ gilt. Ferner ist $\rho^{(i,l)}(x) = \rho^{(i,l+1)}(x)$ für alle $x \in \mathcal{X} \setminus \{\gamma_l, \gamma_{l+1}\}$ sowie $Q(\gamma_l, \gamma_{l+1}) > 0$ nach obiger Wahl des Pfades. Das heißt, wir können Lemma 4.1.24 auf die Mengen $\{\gamma_l, \gamma_{l+1}\}$ und die zugehörigen Dichten anwenden und erhalten $\mathcal{W}(\rho^{(i,l)}, \rho^{(i,l+1)}) < \infty$. Insgesamt schließen wir mithilfe der Dreiecksungleichung

$$\mathcal{W}(\rho^{(i)}, \rho^{(i+1)}) \leq \sum_{l=0}^{n_i - 1} \mathcal{W}(\rho^{(i,l)}, \rho^{(i,l+1)}) < \infty.$$

Dies zeigt die Endlichkeit der $\mathcal{W}(\rho^{(i)}, \rho^{(i+1)})$. Obige Anwendung der Dreiecksungleichung zeigt dann $\mathcal{W}(\rho_0, \rho^{(n)}) < \infty$. Aufgrund der Symmetrie der Metrik zusammen mit obigem

Argument erhält man auch

$$\mathcal{W}(\rho^{(n)}, \rho_1) = \mathcal{W}(\rho_1, \rho^{(n)}) < \infty$$

und damit die Endlichkeit von $\mathcal{W}(\rho_0, \rho_1)$ nach erneuter Anwendung der Dreiecksungleichung. Insgesamt folgt die Behauptung, dass $(\mathcal{D}(\mathcal{X}), \mathcal{W})$ ein metrischer Raum ist. □

Bemerkung 4.1.25. Die Endlichkeit der diskreten Wasserstein-Metrik verschlüsselt eine wichtige Information. Da $\mathcal{W}(\rho_0, \rho_1)$ für alle $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ endlich ist, folgt auch, dass $\mathcal{CE}_1(\rho_0, \rho_1)$ nicht leer ist. Insbesondere existiert also für alle $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ eine mögliche Lösung der Kontinuitätsgleichung beziehungsweise eine Möglichkeit, die Masse mit der Dichte ρ_0 in diesem Sinne in den Zustand der Dichte ρ_1 zu überführen.

Satz 4.1.26. Seien $\rho_n, \rho \in \mathcal{D}_*(\mathcal{X})$, dann sind die folgenden Aussagen äquivalent:

(i) $\lim\limits_{n\to\infty} d_{TV}(\rho, \rho_n) = 0$

(ii) $\lim\limits_{n\to\infty} \mathcal{W}(\rho, \rho_n) = 0$.

Insbesondere stimmen die von $\mathcal{W}$ erzeugte Topologie und die durch d_{TV} erzeugte Topologie überein. Wir bemerken, da $\mathcal{D}(\mathcal{X})$ endlich ist, ist die durch d_{TV} erzeugte Topologie die euklidische Topologie.

Beweis. Die Richtung (ii) $\Rightarrow$ (i) ist eine Konsequenz von Lemma 4.1.21. Für die Rückrichtung verweisen wir auf [Maa11, Theorem 3.21]. □

4.2 Riemannsche Struktur der diskreten Wasserstein-Metrik

In diesem Abschnitt möchten wir die riemannsche Struktur der Menge $\mathcal{D}_*(\mathcal{X})$ untersuchen. Es stellt sich heraus, dass $\mathcal{D}_*(\mathcal{X})$ mit der durch $\mathcal{W}$ induzierten Topologie eine riemannsche Mannigfaltigkeit ist. Die zugehörige riemannsche Metrik induziert dann auf natürliche Weise einen Abstand auf $\mathcal{D}_*(\mathcal{X})$. Wir werden sehen, dass dieser Abstand mit der diskreten Wasserstein-Metrik übereinstimmt.

Satz 4.2.1. Der metrische Raum $(\mathcal{D}_*(\mathcal{X}), \mathcal{W})$ ist eine $(n-1)$-dimensionale glatte Mannigfaltigkeit.

Beweis. Wir führen die Menge $N(\pi) = \left\{\psi \in \mathbb{R}^{\mathcal{X}} \;\middle|\; \sum_{x\in\mathcal{X}} \psi(x)\pi(\{x\}) = 0\right\}$ ein. Es gilt

$$\mathcal{D}_*(\mathcal{X}) = \left\{\rho \in \mathbb{R}^{\mathcal{X}} \;\middle|\; \rho(x) > 0 \;\forall x \in \mathcal{X}, \sum_{x\in\mathcal{X}} \pi(\{x\})\rho(x) = 1\right\} \subset \mathbb{1} + N(\pi).$$

Die rechte Seite ist ein $(n-1)$-dimensionaler affiner Unterraum von $\mathbb{R}^{\mathcal{X}}$ und folglich nach Lemma B.0.1 eine glatte Mannigfaltigkeit bezüglich der euklidischen Topologie. Ferner ist $\mathcal{D}_*(\mathcal{X})$ eine relativ offene Teilmenge dieses affinen Unterraums und somit eine glatte Mannigfaltigkeit bezüglich der euklidischen Topologie. Wir haben in Satz 4.1.26 gesehen, dass die euklidische Topologie mit der von $\mathcal{W}$ erzeugten Topologie übereinstimmt. Folglich ist $(\mathcal{D}_*(\mathcal{X}), \mathcal{W})$ eine glatte Mannigfaltigkeit. □

Wir möchten den Tangentialraum an $\rho \in \mathcal{D}_*(\mathcal{X})$ untersuchen. Dazu stellen wir zunächst eine Eigenschaft der Matrix $B(\rho)$ vor.

Lemma 4.2.2. Die Einschränkung $B(\rho)\colon N(\mathbb{1}) \to N(\pi)$ ist ein Isomorphismus. Dabei bezeichnet $N(\mathbb{1}) = \{\psi \in \mathbb{R}^{\mathcal{X}} \mid \sum\limits_{s\in\mathcal{X}} \psi(s) = 0\}$.

Beweis. Es ist $B(\rho) = \Pi^{-1}A(\rho)$. Es reicht zu zeigen, dass $\ker(A(\rho)) = \{\alpha\mathbb{1}_{\mathcal{X}} \mid \alpha \in \mathbb{R}\}$ ist. In diesem Fall folgt $\ker(B(\rho)) = \ker(A(\rho))$. Die Verwendung der Reversibilitätsbedingung zeigt $\int_{\mathcal{X}} B(\rho) f \,\mathrm{d}\pi = 0$ für alle $f \in \mathbb{R}^{\mathcal{X}}$, wir schließen $\operatorname{Ran}(B(\rho)) \subset N(\pi)$ und aus Dimensionsgründen auch $\operatorname{Ran}(B(\rho)) = N(\pi)$. Es gilt

$$[A(\rho)\mathbb{1}_{\mathcal{X}}](x) = \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} -Q(x,y)\Lambda(\rho(x),\rho(y))\pi(\{x\}) + \sum_{\substack{z\in\mathcal{X}\\ z\neq x}} Q(x,z)\Lambda(\rho(x),\rho(z))\pi(\{x\}) = 0$$

und damit auch $\{\alpha\mathbb{1}_{\mathcal{X}} \mid \alpha \in \mathbb{R}\} \subset \ker(A(\rho))$. Für die andere Inklusion sei $f \in \ker(A(\rho))$. Dann gilt nach der Rechnung im Beweis von Lemma 4.1.19

$$0 = 2\langle A(\rho)f, f\rangle = -\sum_{\substack{x,y\in\mathcal{X}\\ x\neq y}} [A(\rho)](x,y)\left(f(x) - f(y)\right)^2.$$

Folglich ist $f(x) = f(y)$ für alle $x, y \in \mathcal{X}$. Denn angenommen, es existieren $x, y \in \mathcal{X}$, sodass $f(x) \neq f(y)$ ist, dann existiert aufgrund der Irreduzibilität auch ein Pfad $x = \gamma_0, \ldots, \gamma_n = y$ derart, dass $Q(\gamma_i, \gamma_{i+1}) > 0$ für alle $i \in \{0, \ldots, n-1\}$ ist. Dann gilt auch $\Lambda(\rho(\gamma_i), \rho(\gamma_{i+1})) > 0$ und folglich $[A(\rho)](\gamma_i, \gamma_{i+1}) < 0$ für alle $i \in \{0, \ldots, n-1\}$. Mit obiger Rechnung folgt dann

$$0 = -\sum_{\substack{x,y\in\mathcal{X}\\ x\neq y}} [A(\rho)](x,y)\left(f(x) - f(y)\right)^2 \geq -\sum_{i=0}^{n} [A(\rho)](\gamma_i, \gamma_{i+1})\left(f(\gamma_i) - f(\gamma_{i+1})\right)^2,$$

denn $[A(\rho)](x, y) \leq 0$ für alle $x \neq y$. Unter dieser Bedingung muss $f(\gamma_i) = f(\gamma_{i+1})$ für alle $i \in \{0, \dots, n-1\}$ gelten und somit ist auch $f(x) = f(\gamma_1) = \dots = f(\gamma_{n-1}) = f(y)$. Dies zeigt die Behauptung. □

Auf natürliche Weise können wir den Tangentialraum an ρ mit der Menge $N(\pi)$ aus dem Beweis von Satz 4.2.1 identifizieren. Das hat den Grund, dass man den Tangentialraum jedes Punktes einer affinen Mannigfaltigkeit auf natürliche Weise mit dem affinen Unterraum selbst identifizieren kann. Für einen Beweis dieses Arguments verweisen wir auf [Lee13, Proposition 3.8]. Ferner kann man den tangentialen Vektor an eine glatte Kurve ρ_t mit dem Vektor der Ableitungen der Punktauswertungen $\dot{\rho}_t$ identifizieren. Wir möchten eine weitere Darstellung des Tangentialraums untersuchen. Der folgende Satz beschreibt den Tangentialraum als Menge von Gradienten $\nabla\psi$ für Funktionen $\psi \in \mathbb{R}^{\mathcal{X}}$. Im darauffolgenden Satz klären wir den Zusammenhang des tangentialen Vektors $\dot{\rho}_t$ und der Identifikation mit dem Gradienten. Die folgenden zwei Sätze kann man als die diskrete Analogie zum eingangs vorgestellten Otto-Kalkül interpretieren.

Satz 4.2.3. Sei $\rho \in \mathcal{D}_*(\mathcal{X})$, wir definieren die Menge $T_\rho = \{\nabla\psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}} \mid \psi \in \mathbb{R}^{\mathcal{X}}\}$, dann ist die Abbildung

$$\mathcal{I}_\rho \colon N(\pi) \to T_\rho,\ \psi \mapsto \nabla(B(\rho)^{-1}\psi)$$

ein Isomorphismus.

Beweis. Wir bemerken, dass $N(\pi)$ sowie T_ρ endlich-dimensionale Vektorräume über $\mathbb{R}$ sind. Die Abbildungsvorschrift $\mathcal{I}_\rho$ ist nach Lemma 4.2.2 wohldefiniert und linear. Wir wissen bereits, dass $B(\rho)$ ein Isomorphismus ist, folglich reicht es zu zeigen, dass die Abbildung

$$N(\mathbb{1}) \to T_\rho,\ \psi \mapsto \nabla\psi$$

ein Isomorphismus ist. Wir zeigen die Injektivität, sei dazu $\psi \in N(\mathbb{1})$, sodass $\nabla\psi(x, y) = 0$ für alle $x, y \in \mathcal{X}$ gilt. Dann folgt $\psi(x) = \psi(y)$ für alle $x, y \in \mathcal{X}$ und folglich ist ψ von der Form $\psi = \alpha\mathbb{1}_{\mathcal{X}}$ für ein $\alpha \in \mathbb{R}$. Da $\psi \in N(\mathbb{1})$ ist, muss $\alpha = 0$ und somit $\psi = 0$ gelten. Sei nun $\Phi \in T_\rho$, dann existiert ein $\psi \in \mathbb{R}^{\mathcal{X}}$, sodass $\Phi = \nabla\psi$ gilt. Definieren wir $\psi'(x) := \psi(x) - \frac{1}{|\mathcal{X}|}\sum_{y\in\mathcal{X}}\psi(y)$ für $x \in \mathcal{X}$, dann ist per definitionem $\psi' \in N(\mathbb{1})$ und es folgt $\nabla(\psi')(x, y) = \psi'(x) - \psi'(y) = \psi(x) - \psi(y) = \nabla\psi(x, y) = \Phi(x, y)$ für alle $x, y \in \mathcal{X}$. Wir haben gezeigt, dass $\mathcal{I}_\rho$ eine lineare und bijektive Abbildung ist, das heißt, $\mathcal{I}_\rho$ ist ein Isomorphismus. □

Satz 4.2.4. Sei $\rho \colon [0, 1] \to \mathcal{D}_*(\mathcal{X})$ eine glatte Kurve. Dann gilt für alle $t \in (0, 1)$, dass $\mathcal{I}_{\rho_t}\dot{\rho}_t = \nabla\psi_t \in T_{\rho_t}$ der eindeutige Vektor ist, welcher die Gleichung

$$\dot{\rho}_t(x) + [\nabla \cdot (\hat{\rho}_t \bullet \nabla\psi_t)](x) = 0$$

für alle $x \in \mathcal{X}$ erfüllt. Insbesondere zeigt dies, dass $(\rho, \psi) \in \mathcal{CE}(\rho_0, \rho_1)$ ist. Hierbei bezeichnet $\mathcal{I}_\rho$ die gleichnamige Abbildung aus Satz 4.2.3.

Beweis. Die Kombination der Beweise von Lemma 4.1.15 und Lemma 4.1.19 zeigt $B(\rho)\psi = -\nabla \cdot (\hat{\rho} \bullet \nabla\psi)$ für alle $\rho \in \mathcal{D}(\mathcal{X})$ und $\psi \in \mathbb{R}^{\mathcal{X}}$. Da $\mathcal{I}_\rho$ auf $N(\pi)$ ein Isomorphismus ist, folgt

$$\dot{\rho}_t = \mathcal{I}_{\rho_t}^{-1}(\nabla\psi_t) = \mathcal{I}_{\rho_t}^{-1}(\nabla(B(\rho)^{-1}B(\rho)\psi_t)) = B(\rho_t)\psi_t = -\nabla \cdot (\hat{\rho}_t \bullet \nabla\psi_t)$$

für alle $t \in (0,1)$. Bezeichnen wir mit $R^{-1}\colon T_\rho \to N(\mathbb{1})$ die Inverse der Abbildung

$$N(\mathbb{1}) \to T_\rho,\ \psi \mapsto \nabla\psi$$

aus dem Beweis von Lemma 4.2.3, dann gilt $\psi_t = R^{-1}(\mathcal{I}_{\rho_t}\dot{\rho}_t)$. Insbesondere ist damit die Abbildung $(0,1) \to N(\mathbb{1}),\ t \mapsto \psi_t$ stetig und somit Borel-messbar. Dies zeigt, dass $(\rho, \psi) \in \mathcal{CE}_1(\rho_0, \rho_1)$ ist. $\square$

Wir möchten zeigen, dass $\mathcal{D}_*(\mathcal{X})$ eine riemannsche Mannigfaltigkeit ist. Wir werden den Tangentialraum an $\rho \in \mathcal{D}_*(\mathcal{X})$ stets mit T_ρ wie in Satz 4.2.3 identifizieren. Wir führen ein Skalarpodukt auf T_ρ ein und werden zeigen, dass die Familie dieser Skalarprodukte eine riemannsche Metrik auf $\mathcal{D}_*(\mathcal{X})$ definiert.

Satz 4.2.5. Für $\rho \in \mathcal{D}_*(\mathcal{X})$ definieren wir auf dem Tangentialraum T_ρ die Form $\langle\cdot,\cdot\rangle_\rho\colon T_\rho \times T_\rho \to \mathbb{R}$ durch

$$\langle\nabla\varphi, \nabla\psi\rangle_\rho = \frac{1}{2}\sum_{x,y\in\mathcal{X}} (\varphi(x) - \varphi(y))\,(\psi(x) - \psi(y))\,Q(x,y)\Lambda(\rho(x), \rho(y))\pi(\{x\})$$

für beliebige $\nabla\varphi, \nabla\psi \in T_\rho$. Diese bilineare Form ist für jedes $\rho \in \mathcal{D}_*(\mathcal{X})$ ein Skalarprodukt auf T_ρ.

Beweis. Die Linearität und Symmetrie sind klar. Es bleibt die positive Definitheit zu zeigen. Es gilt $\langle\nabla\varphi, \nabla\varphi\rangle_\rho \geq 0$ für alle $\varphi \in \mathbb{R}^{\mathcal{X}}$, denn $Q(x,y) \geq 0$ für alle $x \neq y$ und für $x = y$ verschwindet der gesamte Term. Ferner sind $\rho(x), \rho(y) > 0$ und somit auch $\Lambda(\rho(x), \rho(y)) > 0$. Ist $\varphi = 0$, dann gilt auch $\langle\nabla\varphi, \nabla\varphi\rangle_\rho = 0$. Sei $\langle\nabla\varphi, \nabla\varphi\rangle_\rho = 0$ für ein $\varphi \in \mathbb{R}^{\mathcal{X}}$. Es gilt

$$0 = \langle\nabla\varphi, \nabla\varphi\rangle_\rho = \langle A(\rho)\varphi, \varphi\rangle$$

und folglich ist nach der Argumentation im Beweis von Lemma 4.2.2 notwendigerweise schon $\varphi(x) = \varphi(y)$ für alle $x, y \in \mathcal{X}$. Damit verschwindet $\nabla\varphi = 0$ und insgesamt gilt, dass $\langle\cdot,\cdot\rangle_\rho$ ein Skalarprodukt definiert. $\square$

Lemma 4.2.6. Sei $x \in \mathcal{X}$. Dann ist die Abbildung $\eta\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$, $\rho \mapsto \rho(x)$ differenzierbar.

Beweis. Wir wählen eine Basis $v_1, \ldots, v_{n-1}$ von $N(\pi)$. Für jedes $j \in \{1, \ldots, n-1\}$ existieren Konstanten $\alpha_i^{(j)} \in \mathbb{R}$ derart, dass $v_j = \sum_{i=1}^n \alpha_i^{(j)} e_i$ gilt. Hierbei bezeichnet $e_1, \ldots, e_n$ die Standardbasis von $\mathbb{R}^{\mathcal{X}}$, das heißt, insbesondere ist $e_m = \mathbb{1}_{\{x\}}$ für ein $m \in \{1, \ldots, n-1\}$. Nach dem Beweis von Lemma B.0.1 ist die inverse Abbildung von

$$K\colon \mathbb{R}^{n-1} \to \mathbb{1} + N(\pi),\ K(\xi) = \mathbb{1} + \sum_{i=1}^{n-1} \xi_i v_i$$

eine Karte von $\mathbb{1} + N(\pi)$ und damit ist die inverse Abbildung der Einschränkung auf $\mathcal{D}_*(\mathcal{X})$

$$K\colon K^{-1}(\mathcal{D}_*(\mathcal{X})) \to D_*,\ K(\xi) = \mathbb{1} + \sum_{i=1}^{n-1} \xi_i v_i$$

eine Karte von $\mathcal{D}_*(\mathcal{X})$. Das Paar $(\mathcal{D}_*(\mathcal{X}), K^{-1})$ ist ein glatter Atlas von $\mathcal{D}_*(\mathcal{X})$. Für die Differenzierbarkeit der Abbildung η betrachten wir

$$\begin{aligned}\eta(K(\xi_1, \ldots, \xi_{n-1})) &= \eta\left(\mathbb{1} + \sum_{k=1}^{n-1} \xi_k v_k\right) = \eta\left(\mathbb{1} + \sum_{k=1}^{n-1} \xi_k \left(\sum_{i=1}^{n} \alpha_i^{(k)} e_i\right)\right) \\ &= \eta\left(\sum_{i=1}^{n}\left(\sum_{k=1}^{n-1} \xi_k \alpha_i^{(k)} + 1\right) e_i\right) = \left(\sum_{k=1}^{n-1} \xi_k \alpha_m^{(k)} + 1\right).\end{aligned}$$

Folglich ist die Verknüpfung $\eta \circ K$ im klassischen Sinne differenzierbar und damit ist auch die Abbildung $\eta\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$ differenzierbar. □

Theorem 4.2.7. *Das Paar $(\mathcal{D}_*(\mathcal{X}), \langle\cdot,\cdot\rangle_\rho)$ ist eine riemannsche Mannigfaltigkeit.*

Beweis. In Satz 4.2.5 haben wir gezeigt, dass für jedes $\rho \in \mathcal{D}_*(\mathcal{X})$ die Bilinearform $\langle\cdot,\cdot\rangle_\rho$ ein Skalarpodukt auf dem Tangentialraum T_ρ ist. Wir müssen noch die glatte Abhängigkeit von $\rho \in \mathcal{D}_*(\mathcal{X})$ nachweisen. Wie bereits bemerkt, können wir in unserem Fall den Tangentialraum an $\rho \in \mathcal{D}_*(\mathcal{X})$ jeweils mit T_ρ identifizieren. Wir wählen eine Basis $z_1, \ldots, z_{n-1}$ von T_ρ. Zwei glatte Vektorfelder V, W auf $\mathcal{D}_*(\mathcal{X})$ können wir dann als

$$V(\rho) = \sum_{j=1}^{n-1} v_j(\rho) z_j \text{ und } W(\rho) = \sum_{j=1}^{n-1} w_j(\rho) z_j$$

für glatte Komponentenfunktionen $v_j\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$ und $w_j\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$ schreiben. Wählen wir die Standardbasis der Matrizen $e_1, \ldots, e_{n^2}$, dann existiert für jeden der z_j eine Darstellung der Form $z_j = \sum_{i=1}^{n^2} \alpha_i^{(j)} e_i$ für $\alpha_1^{(j)}, \dots, \alpha_{n^2}^{(j)} \in \mathbb{R}$. Mit diesen Bezeichnungen gilt

$$V(\rho) = \sum_{j=1}^{n-1} \sum_{i=1}^{n^2} \alpha_i^{(j)} v_j(\rho) e_i = \sum_{i=1}^{n^2} \left(\sum_{j=1}^{n-1} \alpha_i^{(j)} v_j(\rho) \right) e_i$$

sowie

$$W(\rho) = \sum_{j=1}^{n-1} \sum_{i=1}^{n^2} \alpha_i^{(j)} v_j(\rho) e_i = \sum_{i=1}^{n^2} \left(\sum_{j=1}^{n-1} \alpha_i^{(j)} v_j(\rho) \right) e_i.$$

Insbesondere sind auch die Koeffizienten der e_i glatte Funktionen. Dies erklärt, analog zu Lemma 4.2.6, warum die Punktauswertung $\rho \mapsto [V(\rho)](x, y)$ der Vektorfelder für beliebige $x, y \in \mathcal{X}$ glatt ist. Betrachten wir nun

$$\rho \mapsto \langle V(\rho), W(\rho) \rangle_\rho = \frac{1}{2} \sum_{x,y \in \mathcal{X}} ([V(\rho)](x, y)) \, ([W(\rho)](x, y)) \, Q(x, y) \Lambda(\rho(x), \rho(y)) \pi(\{x\}),$$

dann müssen wir uns nur noch mit der Differenzierbarkeit der Abbildung $\rho \mapsto \Lambda(\rho(x), \rho(y))$ für $x, y \in \mathcal{X}$ beschäftigen. Die Punktauswertung $\rho \mapsto \rho(x)$ ist nach Lemma 4.2.6 glatt. Ferner ist $\Lambda \in C^\infty((0, \infty)^2)$, dies zeigt die Differenzierbarkeit der Abbildung $\rho \mapsto \langle V(\rho), W(\rho) \rangle_\rho$ als Summe, Produkt und Komposition von glatten Funktionen. Insgesamt folgt, dass $\langle \cdot, \cdot \rangle_\rho$ eine riemannsche Metrik ist und dies zeigt die Behauptung. $\square$

Bemerkung 4.2.8. Betrachten wir die riemannsche Mannigfaltigkeit $(\mathcal{D}_*(\mathcal{X}), \langle \cdot, \cdot \rangle_\rho)$. Dann stimmt in diesem Fall die durch die riemannsche Struktur induzierte Metrik d aus Definition B.0.8 mit der diskreten Wasserstein-Metrik $\mathcal{W}$ eingeschränkt auf $\mathcal{D}_*(\mathcal{X})$ überein. Dies möchten wir an dieser Stelle kurz kommentieren. Sei $(\rho_t)_{t \in [0,1]}$ eine glatte Kurve in $\mathcal{D}_*(\mathcal{X})$. Dann ist der tangentiale Vektor ψ_t an ρ_t eindeutig bestimmt durch $\dot{\rho}_t + [\nabla \cdot (\hat{\rho}_t \bullet \nabla \psi_t)] = 0$. Die Länge dieser Kurve ist dann gegeben durch $\int_0^1 \sqrt{\langle \nabla \psi_t, \nabla \psi_t \rangle_{\rho_t}} \mathrm{d}t$. Vergleichen wir dies mit der Darstellung der diskreten Wasserstein-Metrik aus Lemma 4.1.15, so erkennen wir unter Verwendung von Lemma 4.1.23 sofort die Gleichheit mit dem riemannschen Abstand.

4.3 Gradientenfluss der diskreten Entropie

In diesem Abschnitt werden wir zeigen, dass die Lösungen des Wärmeleitungsgleichung zu Q auch die Trajektorien des Gradientenflusses der diskreten Entropie bezüglich der diskreten Wasserstein-Metrik sind. Dazu werden wir den Gradienten der diskreten Entropie untersuchen und anschließend das tangentiale Vektorfeld entlang von Lösungen der Wärmeleitungsgleichung bestimmen. Wir haben bereits gesehen, dass $(\mathcal{D}_*(\mathcal{X}), \mathcal{W})$ von riemannscher Struktur ist. Wir interessieren uns aber auch für Wahrscheinlichkeitsdichten, die nur nichtnegativ sind. Ist der Anfangswert $\rho_0 \in \mathcal{D}(\mathcal{X})$, dann ist die zugehörige Lösung der Wärmeleitungsgleichung durch $\rho_t = e^{tQ}\rho \in \mathcal{D}_*(\mathcal{X})$ für alle positiven Zeiten $t > 0$ gegeben. Dies ermöglicht es uns, eine größere Menge von Startwerten zuzulassen. Diese Eigenschaft folgt aus der Irreduzibilität und Lemma 3.2.11.

Wir erinnern an die Notationen des vorherigen Abschnitts. Die Menge $\mathcal{D}_*(\mathcal{X})$ sei in diesem Abschnitt stets mit ihrer riemannschen Struktur versehen. Für $\rho \in \mathcal{D}_*(\mathcal{X})$ identifizieren wir den Tangentialraum an ρ mit $T_\rho = \{\nabla\psi \mid \psi \in \mathbb{R}^{\mathcal{X}}\}$. Sei $t \mapsto \rho_t$ eine glatte Kurve, dann bezeichnen wir mit $t \mapsto \nabla\psi_t \in T_{\rho_t}$ das tangentiale Vektorfeld entlang dieser Kurve im Sinne der Identifikation in Satz 4.2.4. Den Gradienten eines glatten Funktionals $F\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$, ausgewertet in $\rho \in \mathcal{D}_*(\mathcal{X})$, bezeichnen wir mit $\mathrm{grad}F(\rho) \in T_\rho$.

Lemma 4.3.1 (Gradient der diskreten Entropie)**.** Wir betrachten die diskrete Entropie eingeschränkt auf $\mathcal{D}_*(\mathcal{X})$, das heißt, $H\colon \mathcal{D}_*(\mathcal{X}) \to \mathbb{R}$. Diese Abbildung ist differenzierbar und für $\rho \in \mathcal{D}_*(\mathcal{X})$ gilt

$$\mathrm{grad}H(\rho) = \nabla \log(\rho).$$

Beweis. Es gilt $H(\rho) = \sum_{x\in\mathcal{X}} \rho(x)\log(\rho(x))\pi(\{x\})$ und für alle $x \in \mathcal{X}$ ist die Abbildung $\rho \mapsto \rho(x)$ differenzierbar. Folglich ist $\rho \mapsto H(\rho)$ als Summe, Produkt und Komposition differenzierbarer Funktionen wieder differenzierbar. Wir berechnen den Gradienten von H in $\rho \in \mathcal{D}_*(\mathcal{X})$. Der Gradient ist eindeutig bestimmt durch die Auswertung aller Derivationen an $\rho \in \mathcal{D}_*(\mathcal{X})$ in H. Sei also v eine beliebige Derivation, dann können wir diese nach Lemma B.0.4 mithilfe einer glatten Kurve $(\rho_t)_{t\in(-\varepsilon,\varepsilon)}$ als $v(H) = \frac{\mathrm{d}}{\mathrm{d}t}(H \circ \rho)\Big|_{t=0}$ für $\rho_0 = \rho$ darstellen. Es bezeichne $\nabla\psi_t \in T_{\rho_t}$ das tangentiale Vektorfeld dieser Kurve, dann gilt

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}H(\rho_t) &= \sum_{x\in\mathcal{X}} \log(\rho_t(x))\dot\rho_t(x)\pi(\{x\}) + \sum_{x\in\mathcal{X}} \dot\rho_t(x)\pi(\{x\}) \\
&= -\sum_{x\in\mathcal{X}} \log(\rho_t)[\nabla\cdot(\hat\rho_t \bullet \nabla\psi_t)](x)\pi(\{x\}) - \sum_{x\in\mathcal{X}}[B(\rho_t)\psi_t](x)\pi(\{x\}) \\
&= -\langle \log(\rho_t), \nabla\cdot(\hat\rho_t \bullet \nabla\psi_t)\rangle_\pi + 0 \\
&= \langle \nabla\log(\rho_t), \hat\rho_t \bullet \nabla\psi_t\rangle_\pi \\
&= \langle \nabla\log(\rho_t), \nabla\psi_t\rangle_{\rho_t}.
\end{aligned}$$

Dabei verwenden wir die Formel der partiellen Integration 4.1.13 und dass nach Lemma 4.2.2 die Aussage $[(B(\rho_t)\psi_t](s) \in N(\pi)$ gilt. Dies zeigt

$$v(H) = \frac{\mathrm{d}}{\mathrm{d}t}H(\rho_t)\Big|_{t=0} = \langle \nabla\log(\rho_0), \nabla\psi_0\rangle_{\rho_0} = \langle \nabla\log(\rho), \nabla\psi_0\rangle_\rho.$$

Da v beliebig war und ψ_0 der zugehörige tangentiale Vektor ist, folgt die Behauptung. $\square$

Lemma 4.3.2 (Tangentiales Vektorfeld von Lösungen der Wärmeleitungsgleichung)**.** Sei $\rho \in \mathcal{D}(\mathcal{X})$. Es bezeichne $\rho_t = e^{tQ}\rho$ für $t \geq 0$. Dann ist die Abbildung $t \mapsto \rho_t$ glatt auf $(0, \infty)$ und das tangentiale Vektorfeld ist gegeben durch $-\nabla \log(\rho_t) \in T_{\rho_t}$ für alle $t > 0$. Insbesondere erfüllt das Paar $(\rho_r, -\log(\rho_r))_{r\in[0,t]}$ für alle $t > 0$ die Kontinuitätsgleichung, das heißt, es ist $(\rho, -\log(\rho)) \in \mathcal{CE}_t(\rho_0, \rho_t)$.

Beweis. Wir bemerken zunächst, dass für $t > 0$ nach Lemma 3.2.11 $e^{tQ} > 0$ gilt und deshalb auch $\rho_t \in \mathcal{D}_*(\mathcal{X})$ ist. Fassen wir e^{tQ} als Matrix und ρ als Vektor des $\mathbb{R}^n$ auf, dann ist die Differenzierbarkeit der Abbildung $t \mapsto \rho_t$ ein Resultat der gewöhnlichen Differentialrechnung. Die Differenzierbarkeit als Abbildung $(0, \infty) \to \mathcal{D}_*(\mathcal{X})$ folgt dann durch die Wahl einer geeigneten Basis von $N(\pi)$, ähnlich wie in den Beweisen von Lemma 4.2.6 und Theorem 4.2.7. Wir behaupten, dass für alle $\rho \in \mathcal{D}_*(\mathcal{X})$ die Gleichung $\Delta\rho = \nabla \cdot (\nabla\rho) = \nabla \cdot (\hat{\rho} \bullet \nabla(\log \circ \rho))$ erfüllt ist. Sei dazu $x \in \mathcal{X}$, dann ist

$$\begin{aligned}
[\nabla \cdot (\nabla\rho)](x) &= \frac{1}{2}\sum_{y\in\mathcal{X}} Q(x,y)([\nabla\rho](y,x) - [\nabla\rho](x,y)) \\
&= \sum_{y\in\mathcal{X}} Q(x,y)(\rho(y) - \rho(x))\frac{\log(\rho(y)) - \log(\rho(x))}{\log(\rho(y)) - \log(\rho(x))} \\
&= \sum_{y\in\mathcal{X}} Q(x,y)(\log(\rho(y)) - \log(\rho(x)))\Lambda(\rho(y), \rho(x)) \\
&= \frac{1}{2}\sum_{y\in\mathcal{X}} Q(x,y)(\log(\rho(y)) - \log(\rho(x)) + \log(\rho(y)) - \log(\rho(x)))\Lambda(\rho(y), \rho(x)) \\
&= \frac{1}{2}\sum_{y\in\mathcal{X}} Q(x,y)([\nabla\log(\rho)](y,x) - [\nabla\log(\rho)](x,y))\Lambda(\rho(y), \rho(x)) \\
&= \frac{1}{2}\sum_{y\in\mathcal{X}} Q(x,y)(\Lambda(\rho(y), \rho(x))[\nabla\log(\rho)](y,x) - \Lambda(\rho(x), \rho(y))[\nabla\log(\rho)](x,y)) \\
&= [\nabla \cdot (\hat{\rho} \bullet \nabla(\log \circ \rho))](x).
\end{aligned}$$

Folglich gilt für alle $t > 0$ auch

$$\dot{\rho}_t = \Delta\rho_t = \nabla \cdot (\hat{\rho}_t \bullet \nabla(\log \circ \rho_t)).$$

Sei $t > 0$, dann folgt unter der Verwendung von Lemma 4.2.4, dass der zu ρ_t gehörende tangentiale Vektor durch $-\nabla \log(\rho_t)$ gegeben ist. □

Wir möchten zeigen, dass die Lösungen der Wärmeleitungsgleichung die Trajektorien des Gradientenflusses der Entropie sind. Dies ist eine Folgerung der obigen Beobachtungen.

Theorem 4.3.3. *Sei $\tilde{\rho} \in \mathcal{D}(\mathcal{X})$ und bezeichne $\rho_t = e^{tQ}\tilde{\rho}$ für $t \geq 0$. Dann ist die Abbildung $t \mapsto \rho_t$ differenzierbar auf $(0, \infty)$. Das tangentiale Vektorfeld von ρ_t erfüllt*

$$\nabla\psi_t = -\operatorname{grad} H(\rho_t)$$

für alle $t > 0$. Ferner ist $t \mapsto \rho_t$ stetig in $t = 0$ mit $\rho_0 = \tilde{\rho}$ bezüglich der d_{TV} Metrik.

Beweis. Für beliebiges $t_0 > 0$ ist $\rho_{t_0} \in \mathcal{D}_*(\mathcal{X})$. Ferner ist $(0, \infty) \to D_*(\mathcal{X})$, $s \mapsto e^{sQ}\rho_{t_0}$ nach Lemma 4.3.2 differenzierbar. Da $t_0 > 0$ beliebig ist und aufgrund der Halbgruppeneigenschaft der ρ_t, folgt die Differenzierbarkeit der Abbildung $(0, \infty) \to \mathcal{D}_*(\mathcal{X})$, $t \mapsto \rho_t$. Ferner folgt unter Benutzung von Lemma 4.3.2 und Lemma 4.3.1 die Gleichung

$$\nabla\psi_t = -\nabla\log(\rho_t) = -\operatorname{grad} H(\rho_t)$$

für alle $t > 0$. Da $(e^{tQ})_{t\geq 0}$ eine Markov-Halbgruppe ist, folgt zusammen mit der Normäquivalenz die Stetigkeit in 0 bezüglich der d_{TV} Metrik. □

Bemerkung 4.3.4. In Kapitel 2, genauer in Theorem 2.3.4, haben wir gesehen, dass die Lösungen der Fokker-Planck-Gleichung auch die Trajektorien des Gradientenflusses der freien Energie bezüglich der Wasserstein-Metrik sind. In Bemerkung 2.3.5 haben wir nachgerechnet, dass die freie Energie mit der relativen Entropie zum Referenzwahrscheinlichkeitsmaß $u_\infty \mathrm{d}x$ übereinstimmt. Wir haben gezeigt, dass die von uns gewählte Diskretisierung der Fokker-Planck-Gleichung einen Markov-Kern definiert. Folglich zeigt Theorem 4.3.3, dass die diskrete Fokker-Planck-Gleichung der Gradientenfluss der diskreten relativen Entropie ist.

Definition 4.3.5 (Freie Energie)**.** Sei $V\colon \mathcal{X} \to \mathbb{R}$ eine beliebige Abbildung, dann definieren wir für $\rho \in \mathcal{D}(\mathcal{X})$ die potentielle Energie

$$\mathcal{V}(\rho) = \int_{\mathcal{X}} V\rho \mathrm{d}\pi.$$

Die freie Energie $\mathcal{F}$ von ρ ist definiert durch $\mathcal{F}(\rho) = H(\rho) + \mathcal{V}(\rho)$.

Lemma 4.3.6. Die potentielle Energie, eingeschränkt auf $\mathcal{D}_*(\mathcal{X})$, ist ein differenzierbares Funktional und es gilt

$$\operatorname{grad} \mathcal{V}(\rho) = \nabla V.$$

Beweis. Wie im Beweis von Lemma 4.3.1 zeigt man mithilfe von Lemma 4.2.6 die Differenzierbarkeit von $\mathcal{V}$. Sei nun $t \mapsto \rho_t$ eine glatte Kurve in $\mathcal{D}_*(\mathcal{X})$ mit tangentialem Vektorfeld $\nabla\psi_t \in T_{\rho_t}$. In diesem Fall folgern wir mithilfe von Lemma 4.1.15 sowie der Formel für die partielle Integration 4.1.13 die Gleichung

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(\rho_t) &= \sum_{x\in\mathcal{X}} V(x)\dot{\rho}_t(x)\pi(\{x\}) = -\sum_{x\in\mathcal{X}} V(x)(\nabla\cdot(\hat{\rho}\bullet\nabla\psi_t))(x)\pi(\{x\})\\ &= -\langle V, (\nabla\cdot(\hat{\rho}\bullet\nabla\psi_t))\rangle_\pi = \langle\nabla V, \hat{\rho}_t\bullet\nabla\psi_t\rangle_\pi = \langle\nabla V, \nabla\psi_t\rangle_{\rho_t}.\end{aligned}$$

Diese Gleichung impliziert $\operatorname{grad}\mathcal{V}(\rho) = \nabla V$ für alle $\rho\in\mathcal{D}_*(\mathcal{X})$. □

Bemerkung 4.3.7. Es stellt sich die Frage, wie man eine diskrete Fokker-Planck-Gleichung auf Graphen definieren könnte. Dazu möchten wir hier noch einen Ansatz skizzieren. Im kontinuierlichen Fall betrachten wir den $\mathbb{R}^n$ mit der euklidischen Struktur und statten diese riemannsche Mannigfaltigkeit dann mit einem Referenzwahrscheinlichkeitsmaß der Form $e^{-V}\mathrm{d}x$ aus. Dies entspricht in Theorem 2.3.4 der Addition des Term $\int_{\mathbb{R}^n} Vu\mathrm{d}x$ zur Entropie. Diese Veränderung führt dann zur Fokker-Planck-Gleichung. Im diskreten Setting stellt sich die Frage, welches Maß π dem Lebesgue-Maß $\mathrm{d}x$ entspricht. Einen möglichen Ansatz möchten wir im folgenden Beispiel erklären.

Beispiel 4.3.8. Zu einem gegebenen Graphen betrachten wir das Markov-Tripel aus Theorem 3.3.8. Dies führt zu der Wärmeleitungsgleichung $\dot{\rho} = Q\rho$ als Gradientenfluss der Entropie $H(\rho) = \int_{\mathcal{X}}\rho\log(\rho)\mathrm{d}\pi$. In Analogie zu Theorem 2.3.4 untersuchen wir den Gradientenfluss der freien Energie $H(\rho)+\mathcal{V}(\rho)$. Zu einem beliebigem Startwert $\rho_0\in\mathcal{D}_*(\mathcal{X})$ erfüllt das tangentiale Vektorfeld die Gleichung

$$\nabla\psi_t = -\operatorname{grad}\mathcal{F}(\rho_t) = -\nabla\log(\rho_t) - \nabla V.$$

Dies impliziert

$$\dot{\rho}_t = -\nabla\cdot(\hat{\rho}_t\bullet(-\nabla\log(\rho_t)-\nabla V)) = \Delta\rho_t + \nabla\cdot(\hat{\rho}_t\bullet\nabla V) = Q\rho_t + \nabla\cdot(\hat{\rho}_t\bullet\nabla V)$$

nach Theorem 4.3.3. Wie bereits bemerkt, können wir diese Gleichung auch als $\dot{\rho}_t = Q\rho_t - B(\rho_t)V$ schreiben.

Bemerkung 4.3.9. (i) Wir sehen, dass man eine diskrete Fokker-Planck-Gleichung auf mindestens zwei verschiedene Arten interpretieren kann. Bei der Diskretisierung der Fokker-Planck-Gleichung in Beispiel 3.4.1 speichern wir die Wirkung des Potentials V in den Kantengewichten. Bei der oben vorgestellten diskreten Fokker-Planck-Gleichung betrachten wir zu einer gegebenen Geometrie zusätzlich die Wirkung des Potentials V.

(ii) Wir möchten an dieser Stelle daran erinnern, dass die Diskretisierung der kontinuierlichen Fokker-Planck-Gleichung zu einer linearen Differentialgleichung der Form

$\dot{\rho}_t = Q\rho_t$ geführt hat. Die zuletzt vorgestellte Version der diskreten Fokker-Planck-Gleichung ist eine nichtlineare Differentialgleichung der Form $\dot{\rho}_t = Q\rho_t + f(\rho_t)$ mit der nichtlinearen Funktion $f(\rho) = -B(\rho)V$. Dieser Unterschied macht es unmöglich, diese beiden verschiedenen Modelle der Fokker-Planck-Gleichung zu vereinen. Das heißt, man findet kein nichttriviales Potential V, sodass die Fokker-Planck-Gleichung auf dem Graphen, gegeben durch die Ecken $\{x_i^n | i = 0, \ldots, n\}$ und den mit n^2 Gewichteten Kanten $\{x_i x_{i+1} | i = 0, \ldots, n-1\}$, mit der Diskretisierung der Fokker-Planck-Gleichung aus Beispiel 3.4.1 übereinstimmt.

(iii) Anhand der obigen Darstellung können wir die diskrete Fokker-Planck-Gleichung auch als Differentialgleichung auf der größeren Menge $\mathcal{D}(\mathcal{X})$ betrachten. Problematisch ist, dass das logarithmische Mittel am Rand von $(0,\infty)^2$ nicht lokal Lipschitz-stetig ist. Somit können wir für Anfangswerte in $\mathcal{D}(\mathcal{X}) \backslash \mathcal{D}_*(\mathcal{X})$ nicht den Satz von Picard-Lindelöf anwenden.

(iv) In [CHLZ12] werden ebenfalls zwei verschiedene diskrete Fokker-Planck-Gleichungen vorgestellt. Beide der dort vorgestellten Gleichungen sind Gradientenflüsse der freien Energie, jedoch zu verschiedenen riemannschen Strukturen. Die hier eingeführte Fokker-Planck-Gleichung unterscheidet sich ebenfalls von den beiden in [CHLZ12] vorgestellten Gleichungen. Eine mögliche Begründung für den Unterschied zur Fokker-Planck-Gleichung I aus [CHLZ12] ist, dass die dort verwendete riemannsche Struktur auch von der Potentialfunktion V abhängt.

Eine interessante Ähnlichkeit zur kontinuierlichen Version der Fokker-Planck-Gleichung möchten wir in dem folgenden Lemma untersuchen.

Lemma 4.3.10. Ein Equilibrium der diskreten Fokker-Planck-Gleichung $\dot{\rho}_t = Q\rho_t - B(\rho_t)V$ ist gegeben durch $\rho_\infty \colon \mathcal{X} \to \mathbb{R}$, $\rho_\infty(x) = c^{-1}e^{-V(x)}$ für die Konstante $c = \int_{\mathcal{X}} e^{-V} \mathrm{d}\pi$.

Beweis. Wir müssen lediglich zeigen, dass ρ_∞ die Gleichung $Q\rho_\infty - B(\rho_\infty)V = 0$ erfüllt. Es ist klar, dass $\rho_\infty \in \mathcal{D}_*(\mathcal{X})$ erfüllt ist. Es gilt

$$\begin{aligned}
[B(\rho_\infty)V](x) &= -\sum_{\substack{y\in\mathcal{X}\\ y\neq x}} Q(x,y)\Lambda(\rho_\infty(x),\rho_\infty(y))V(y) + \sum_{\substack{z\in\mathcal{X}\\ z\neq x}} Q(x,z)\Lambda(\rho_\infty(x),\rho_\infty(z))V(x) \\
&= \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} Q(x,y)\frac{\rho_\infty(y)-\rho_\infty(x)}{\log\rho_\infty(y)-\log\rho_\infty(x)}\,(V(x)-V(y)) \\
&= \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} Q(x,y)\frac{\rho_\infty(y)-\rho_\infty(x)}{-V(y)+V(x)}\,(V(x)-V(y))
\end{aligned}$$

$$= \sum_{\substack{y \in \mathcal{X} \\ y \neq x}} Q(x,y) \left(\rho_\infty(y) - \rho_\infty(x)\right)$$

und

$$\left[Q\rho_\infty\right](x) = \sum_{y \in \mathcal{X}} Q(x,y)\rho_\infty(y) = \sum_{y \in \mathcal{X}} Q(x,y) \left(\rho_\infty(y) - \rho_\infty(x)\right)$$

für alle $x \in \mathcal{X}$. Die Kombination dieser Gleichungen zeigt dann die Behauptung. □

Wie bereits bemerkt, kann man die zuletzt vorgestellte diskrete Fokker-Planck-Gleichung auch als Differentialgleichung auf der Menge $\mathcal{D}(\mathcal{X})$ betrachten. Weitere wichtige qualitative Eigenschaften dieser diskreten Fokker-Planck-Gleichung behandelt der folgende Satz.

Satz 4.3.11. Sei $\rho_0 \in \mathcal{D}(\mathcal{X})$. Dann existiert die Lösung $t \mapsto \rho_t$ der diskreten Fokker-Planck-Gleichung zum Anfangswert ρ_0 global nach rechts und es gilt $\rho_t \in \mathcal{D}_*(\mathcal{X})$ für alle $t > 0$.

Beweis. Wir setzen das logarithmische Mittel durch 0 stetig auf den $\mathbb{R}^2$ fort. Damit können wir die Differentialgleichung als Gleichung für Funktionen mit Werten im $\mathbb{R}^n$ auffassen. Für $\rho \in \mathbb{R}^n$ bezeichnen wir mit $\rho(k)$ die k-te Komponente des Vektors. Die Abbildung $\rho \mapsto Q\rho - B(\rho)V$ ist stetig auf dem $\mathbb{R}^n$. Ferner gilt für alle $\rho \in \mathbb{R}^n$ mit $\rho \geq 0$ und $\rho(k) = 0$ auch

$$\begin{aligned} [Q\rho](k) - [B(\rho)V](k) &= \sum_{j=1}^{n} Q(k,j)\rho(j) + \sum_{\substack{j=1 \\ j \neq k}}^{n} Q(k,j)\Lambda(\rho(k),\rho(j))(V(k) - V(j)) \\ &= \sum_{\substack{j=1 \\ j \neq k}}^{n} Q(k,j)\rho(j) + 0 \geq 0. \end{aligned}$$

Dies zeigt die Quasipositivität der rechten Seite dieser Differentialgleichung. Sei $\rho_0 \in \mathcal{D}(\mathcal{X})$, das heißt, aufgefasst als Element des $\mathbb{R}^n$ ist ρ_0 ein nichtnegativer Vektor mit $\langle \rho_0, \pi \rangle = 1$. Aufgrund der Stetigkeit der rechten Seite können wir mit dem Satz von Peano die Existenz einer (nicht notwendigerweise eindeutigen) Lösung $\rho\colon [0, t^+(\rho_0)) \to \mathbb{R}^n$ folgern. Diese Lösung erfüllt nach Lemma 3.2.4 die Gleichung

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho_t, \pi \rangle = \langle Q\rho_t - B(\rho_t)V, \pi \rangle = 0,$$

da $\operatorname{Ran} B(\rho_t) \in N(\pi)$ gilt. Als Konsequenz dieser Gleichung ist somit auch $\langle \rho_t, \pi \rangle = \langle \rho_0, \pi \rangle = 1$ für alle $t \in [0, t^+(\rho_0))$. Insbesondere impliziert die Quasipositivität der Differentialgleichung auch die Nichtnegativität der Lösung. Zusammen schließen wir $\rho_t \in \mathcal{D}(\mathcal{X})$ für alle $t \in [0, t^+(\rho_0))$ und aufgrund der Kompaktheit von $D(\mathcal{X})$ auch $t^+(\rho_0) = \infty$.

Wir möchten zeigen, dass ein $\delta > 0$ existiert, sodass $\rho_t > 0$ für alle $t \in (0, \delta)$ gilt. Für den Beweis dieser Behauptung definieren wir die Indexmenge $G \subseteq \{1, \ldots, n\}$ als

$$G = \left\{k \in \{1, \ldots, n\} \middle| \exists t_k > 0 \colon \rho_t(k) > 0 \ \forall t \in (0, t_k)\right\}.$$

Es gilt

$$\emptyset \neq \{k \in \{1, \ldots, n\} \big| \rho_0(k) > 0\} \subseteq G$$

aufgrund der Stetigkeit der Lösung ρ_t sowie aufgrund der Anfangsbedingung. Da G endlich ist existiert ein $\delta > 0$, sodass $\rho_t(k) > 0$ für alle $k \in G$ und alle $t \in (0, \delta)$ ist. Angenommen es gilt $G \subsetneq \{1, \ldots, n\}$, dann existieren aufgrund der Irreduzibilität von Q Indizes $k \in G^c$ und $j \in G$ derart, dass $Q(k, j) > 0$ ist. Insbesondere existiert ein $t_k \in (0, \delta)$, sodass $\rho_{t_k}(k) = 0$ ist. Dies impliziert $\rho_t(k) = 0$ für alle $t \in [0, t_k]$, denn sonst wäre $k \in G$. Sei nun $\varepsilon \in (0, t_k)$ beliebig, dann folgt der Widerspruch

$$\dot{\rho}_\varepsilon(k) = \sum_{\substack{j=1 \\ j \neq k}}^{n} Q(k, j)\rho_\varepsilon(j) \geq Q(k, j)\rho_\varepsilon(j) > 0.$$

Dies zeigt, dass $G^c = \emptyset$ gelten muss. Dies zeigt, dass für alle $t \in (0, \delta)$ auch $\rho_t > 0$ gilt.

Bezeichnen wir mit $t_0^{(k)} := \sup\{t > 0 \mid \rho_s(k) > 0 \ \forall s \in (0, t]\}$ für $k \in \{1, \ldots, n\}$. Dann gilt $t_0^{(k)} > 0$ aufgrund der obigen Eigenschaft für alle $k \in \{1, \ldots, n\}$. Wir bezeichnen mit $t_0 = \min\{t_0^{(k)} \mid k \in \{1, \ldots, n\}\}$. Angenommen es gilt $t_0 < \infty$, dann ist $\arg\min\{t_0^{(k)} \mid k \in \{1, \ldots, n\}\} \neq \emptyset$. Da $\langle \rho_{t_0^{(k)}}, \pi \rangle = 1$ ist, gibt es mindestens ein $j \in \{1, \ldots, n\}$, sodass $\rho_{t_0^{(k)}}(j) > 0$ ist. Aufgrund der Irreduzibilität existiert wieder ein $k \in \arg\min\{t_0^{(k)} \mid k \in \{1, \ldots, n\}\}$ und ein $j \in \{1, \ldots, n\}$, sodass $\rho_{t^{(k)}}(j) > 0$ und $Q(k, j) > 0$ ist. Nun gilt $\rho_{t_0^{(k)}} = 0$ und $\rho_t > 0$ für alle $t \in (0, t_0^{(k)})$. In diesem Fall folgt aber mit der selben Argumentation wie oben $\rho_t(k) > 0$ für $t \in (t_0^{(k)}, t_0^{(k)} + \varepsilon)$ für ein hinreichend kleines $\varepsilon > 0$. Das heißt, $\rho_t(k)$ hat in $t_0^{(k)}$ ein Minimum, damit ist jedoch auch $\dot{\rho}_{t_0^{(k)}}(k) = 0$. Dies steht im Widerspruch zu

$$\dot{\rho}_{t_0^{(k)}}(k) \geq \sum_{\substack{j=1 \\ j \neq k}}^{n} Q(k, j)\rho_{t_0^{(k)}}(j) > 0.$$

Wir schließen $\rho_t \in \mathcal{D}_*(\mathcal{X})$ für alle $t \in [0, \infty)$. Ist $\rho_0 > 0$, dann zeigt obige Argumentation auch $\rho_t > 0$ für alle $t \geq 0$. □

Bemerkung 4.3.12. (i) Wir möchten an dieser Stelle noch einmal die Eindeutigkeit der Lösungen kommentieren. Ist der Anfangswert $\rho_0 \in \mathcal{D}_*(\mathcal{X})$, dann zeigt Satz 4.3.11, dass eine globale Lösung mit Werten in $D_*(\mathcal{X})$ existiert. Ferner ist die rechte Seite der diskreten Fokker-Planck-Gleichung in $D_*(\mathcal{X})$ lokal Lipschitz-stetig. Dies zeigt, dass für

Anfangswerte in $\mathcal{D}_*(\mathcal{X})$ die zugehörige Lösung auch eindeutig ist.

(ii) Eine genauere Untersuchung der Funktion $\mathcal{F}$ zeigt, dass $\mathcal{F}$ eine strikte Ljapunov-Funktion für diese diskrete Fokker-Planck-Gleichung ist. Mithilfe dieser Aussage kann man dann folgern, dass ρ_∞ aus Lemma 4.3.10 das einzige Equilibrium in $\mathcal{D}(\mathcal{X})$ ist. Mit denselben Argumenten wie im Beweis von Satz 5.3.11 kann man dann die Konvergenz von Lösungen mit Anfangswerten in $\mathcal{D}(\mathcal{X})$ gegen das Equilibrium zeigen.

5 Diskrete Ricci-Krümmung

5.1 Definition und Charakterisierung

Wir haben gesehen, dass die Theorie zur diskreten Wasserstein-Metrik ähnliche Resultate wie die kontinuierliche Theorie des optimalen Transports liefert. So ist auch die diskrete Wasserstein-Metrik von riemannscher Struktur und der Gradientenfluss der Entropie stimmt mit der Wärmeleitungsgleichung überein. Motiviert durch das Theorem 2.4.10, führen wir deshalb den Begriff der diskreten Ricci-Krümmung ein. Dieser Abschnitt basiert größtenteils auf der Arbeit [EM12].

Definition 5.1.1. Sei $(\mathcal{X}, Q, \pi)$ ein irreduzibles reversibles Markov-Tripel. Wir sagen, dass dieses Markov-Tripel eine von unten durch $\lambda \in \mathbb{R}$ beschränkte (diskrete entropische) Ricci-Krümmung hat, falls für alle Wahrscheinlichkeitsdichten $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ eine $\mathcal{W}$-Geodäte $\rho \colon [0,1] \to \mathcal{D}(\mathcal{X})$ mit konstanter Geschwindigkeit existiert, derart, dass

$$H(\rho_t) \leq (1-t)H(\rho_0) + tH(\rho_1) - \frac{\lambda}{2} t(1-t)\mathcal{W}(\rho_0, \rho_1)^2$$

für alle $t \in [0,1]$ erfüllt ist. In diesem Fall sagen wir auch, dass $(\mathcal{X}, Q, \pi)$ die Krümmungsbedingung $CD(\lambda, \infty)$ erfüllt.

Bemerkung 5.1.2. Es stellt sich die Frage, ob für alle $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ eine Geodäte mit konstanter Geschwindigkeit existiert. Im Falle $\rho_0, \rho_1 \in \mathcal{D}_*(\mathcal{X})$ können wir dies bereits beantworten. Nach Theorem 4.2.7 ist $\mathcal{D}_*(\mathcal{X})$ eine vollständige riemannsche Mannigfaltigkeit. Ferner stimmt der durch die riemannsche Struktur induzierte Abstand mit $\mathcal{W}$ überein. In Bemerkung 2.4.9 haben wir bereits erwähnt, dass dies bedeutet, dass zu beliebigen Dichten $\rho_0, \rho_1 \in \mathcal{D}_*(\mathcal{X})$ eine Geodäte mit konstanter Geschwindigkeit existiert. Die Frage nach der Existenz von Geodäten in der größeren Menge $\mathcal{D}(\mathcal{X})$ beantwortet das folgende Theorem.

Theorem 5.1.3. *Für alle $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{X})$ existiert eine Geodäte mit konstanter Geschwindigkeit. Insbesondere ist damit $\mathcal{D}(\mathcal{X})$ ein geodätischer metrischer Raum.*

Beweis. Für einen Beweis verweisen wir an dieser Stelle auf [EM12, Theorem 3.2]. $\square$

Für eine Funktion $f\colon (0,\infty) \to \mathbb{R}$ bezeichnen wir die rechte obere Dini-Ableitung mit

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f(t) = \limsup_{h \to 0^+} \frac{f(t+h) - f(t)}{h}.$$

Theorem 5.1.4. *Sei $(\mathcal{X}, Q, \pi)$ ein reversibles und irreduzibles Markov-Tripel. Dann sind für $\lambda \in \mathbb{R}$ die folgenden Aussagen äquivalent:*

(i) Das Markov-Tripel erfüllt die Krümmungsbedingung $CD(\lambda, \infty)$.

(ii) Für alle $\rho_0, \sigma \in \mathcal{D}(\mathcal{X})$ gilt

$$\frac{1}{2} \frac{\mathrm{d}^+}{\mathrm{d}t} \mathcal{W}(e^{tQ} \rho_0, \sigma)^2 + \frac{\lambda}{2} \mathcal{W}(e^{tQ} \rho_0, \sigma)^2 \leq H(\sigma) - H(e^{tQ} \rho_0)$$

für alle $t \geq 0$.

(iii) Für alle $\tilde{\rho}_0, \tilde{\rho}_1 \in \mathcal{D}_*(\mathcal{X})$ existiert eine Geodäte mit konstanter Geschwindigkeit $(\rho_t)_{t \in [0,1]}$, sodass $\rho_0 = \tilde{\rho}_0$, $\rho_1 = \tilde{\rho}_1$ gilt und für alle $t \in [0,1]$ die Ungleichung

$$H(\rho_t) \leq (1-t) H(\rho_0) + t H(\rho_1) - \frac{\lambda}{2} t(1-t) \mathcal{W}(\rho_0, \rho_1)^2$$

erfüllt ist.

Beweis. Dies ist ein Spezialfall von [EM12, Theorem 4.5]. □

5.2 Berechnung von Schranken an die Ricci-Krümmung

Weder die Definition 5.1.1 der Krümmungsbedingung $CD(\lambda, \infty)$ noch die Charakterisierung in Theorem 5.1.4 liefert genaue Informationen darüber, wie man zu gegebenen $(\mathcal{X}, Q, \pi)$ ein Krümmungsbedingung nachweist. Ferner wissen wir auch nicht, ob jedes Markov-Tripel $(\mathcal{X}, Q, \pi)$ überhaupt eine Krümmungsbedingung erfüllt. Diesen Problemen werden wir in diesem Abschnitt nachgehen.

Bemerkung 5.2.1. Man kann zeigen, dass jedes Markov-Tripel eine Krümmungsbedingung erfüllt. Diese Aussage entspricht Theorem 4.1 in [Mie13]. Alexander Mielke betrachtet in seiner Arbeit [Mie13] dieselben Markov-Kerne wie die in Definition 3.2.2. Er führt den Begriff der diskreten Ricci-Krümmung anhand der Bedingung (iii) aus Theorem 5.1.4 ein. Ein Unterschied ist, dass er die Menge der positiven Wahrscheinlichkeitsmaße $\mathcal{P}_*(\mathcal{X})$ untersucht und nicht die Menge der positiven Dichten $\mathcal{D}_*(\mathcal{X})$. Wie bereits bemerkt, ist dies jedoch nicht wirklich relevant. Man kann nachrechnen, dass die von ihm eingeführte riemannsche

Struktur dieselbe ist wie die unsere. Damit sind nach Theorem 5.1.4 die beiden Begriffe von diskreter Ricci-Krümmung kompatibel.

Wir möchten nun Markov-Kerne von tridiagonaler Form betrachten. Wir werden sehen, dass man bei den zugehörigen Markov-Tripeln Schranken an die Ricci-Krümmung angeben kann. Seien $\alpha_1, \ldots, \alpha_{n-1}$ und $\beta_2, \ldots, \beta_n$ positive reelle Zahlen. Wir betrachten die Matrix $Q \in \mathbb{R}^{n \times n}$ definiert durch

$$Q = \begin{pmatrix} -\alpha_1 & \alpha_1 & 0 & & & \\ \beta_2 & -\beta_2 - \alpha_2 & \alpha_2 & 0 & & \\ 0 & \beta_3 & -\beta_3 - \alpha_3 & \alpha_3 & 0 & \\ & & \ddots & \ddots & \ddots & \\ & & 0 & \beta_{n-1} & -\beta_{n-1} - \alpha_{n-1} & \alpha_{n-1} \\ & & & 0 & \beta_n & -\beta_n \end{pmatrix}$$

und setzen $\alpha_n = 0$ sowie $\beta_1 = 0$.

Lemma 5.2.2. Ist Q eine tridiagonale Matrix der obigen Form, dann ist $(\{1, \ldots, n\}, Q, \pi)$ ein reversibles und irreduzibles Markov-Tripel. π ist definiert durch $\pi(\{i+1\}) = \frac{\alpha_i}{\beta_{i+1}} \pi(\{i\})$ für $i \in \{1, \ldots, n-1\}$ und $\pi(\{1\})$ ist so gewählt, dass π ein Wahrscheinlichkeitsmaß ist.

Beweis. Die Behauptung, dass Q ein Markov-Kern ist, folgt sofort aus der Definition von Q. Die Irreduzibilität ist die Konsequenz der Positivität der Einträge auf der Nebendiagonalen, aufgrund dieser Einträge kann man zwischen zwei beliebigen Nachbarn springen und damit auch zwischen beliebigen $i, j \in \{1, \ldots, n\}$. Einen Beweis in einer ähnlichen Situation haben wir bereits in 3.4.2 geführt. Für die Reversibilität genügt die Gültigkeit von

$$\pi(\{i\})Q(i, i+1) = \frac{\beta_{i+1}}{\alpha_i} \pi(\{i+1\})Q(i, i+1) = \beta_{i+1}\pi(\{i+1\}) = \pi(\{i+1\})Q(i+1, i)$$

für alle $i \in \{1, \ldots, n-1\}$, denn in allen anderen Einträgen sind beide Seiten gleich null. □

Theorem 5.2.3. *Wir betrachten das Markov-Tripel* $(\{1, \ldots, n\}, Q, \pi)$ *zu einem tridiagonalen Markov-Kern obiger Form. Es gelte* $\alpha_i \geq \alpha_{i+1}$ *für alle* $i \in \{1, \ldots, n-1\}$ *sowie* $\beta_i \leq \beta_{i+1}$ *für alle* $i \in \{1, \ldots, n-1\}$*. Wir definieren die Abbildung*

$$G \colon [0, \infty) \times [0, \infty) \to \mathbb{R}, \ G(a, b) = \frac{a+b}{2} + \frac{1}{2} \inf_{r,s>0} \Lambda(r, s) \left(\frac{a}{r} + \frac{b}{s} \right).$$

Gilt für ein $\lambda > 0$ die Ungleichung

$$G(\alpha_i - \alpha_{i+1}, \beta_{i+1} - \beta_i) \geq \lambda$$

für alle $i \in \{1, \ldots, n-1\}$, dann erfüllt das Markov-Tripel $(\{1, \ldots, n\}, Q, \pi)$ die Krümmungsbedingung $CD(\lambda, \infty)$.

Beweis. Für einen Beweis dieses Theorems verweisen wir auf [FM16, Theorem 4.1]. In diesem Artikel werden auch Rechenregeln für Schranken an die Ricci-Krümmung weiterer Markov-Kerne bereitgestellt. □

Beispiel 5.2.4. Wir betrachten die diskrete Fokker-Planck-Gleichung aus Beispiel 3.4.1 und untersuchen, ob $(\{1, \ldots, n\}, Q^{(n)}, \pi)$ eine Krümmungsbedingung erfüllt. Dazu bemerken wir zunächst, dass $Q^{(n)}$ von tridiagonaler Form ist, dazu wählen wir $\alpha_i = n^2 \frac{\kappa_i}{\pi_i}$ für $i = 1, \ldots, n-1$ und $\beta_i = n^2 \frac{\kappa_{i-1}}{\pi_i}$ für $i \in \{2, \ldots, n\}$. Gilt $\inf_{x \in [0,1]} V''(x) \geq \hat{\lambda} \geq 0$, dann erfüllt $(Q^{(n)}, \{1, \ldots, n\}, \pi)$ die Krümmungsbedingung $C(\lambda_n, \infty)$ mit $\lambda_n = 2n^2 \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right)$. Insbesondere gilt $\lambda_n \to \hat{\lambda}$ für $n \to \infty$.

Beweis. Wir möchten Theorem 5.2.3 anwenden und weisen dazu zuerst die Monotonieeigenschaften der α_i und β_i nach. Es gilt

$$\alpha_i - \alpha_{i+1} = \frac{\kappa_i}{\pi_i} - \frac{\kappa_{i+1}}{\pi_{i+1}} = n^2 \left(\frac{\sqrt{\pi_{i+1}}}{\sqrt{\pi_i}} - \frac{\sqrt{\pi_{i+2}}}{\sqrt{\pi_{i+1}}} \right) \geq n^2 \frac{\sqrt{\pi_{i+1}}}{\sqrt{\pi_i}} \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) \geq 0$$

für alle $i \in \{1, \ldots, n-2\}$. Dabei verwenden wir eine Abschätzung, die die π_i zueinander in Beziehung setzt, deren Beweis wir in Lemma C.0.9 führen. Für $i \in \{2, \ldots, n-1\}$ gilt

$$\beta_{i+1} - \beta_i = \frac{\kappa_i}{\pi_{i+1}} - \frac{\kappa_{i-1}}{\pi_i} = n^2 \left(\frac{\sqrt{\pi_i}}{\sqrt{\pi_{i+1}}} - \frac{\sqrt{\pi_{i-1}}}{\sqrt{\pi_i}} \right) \geq n^2 \frac{\sqrt{\pi_i}}{\sqrt{\pi_{i+1}}} \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) \geq 0,$$

dies zeigt die Monotonie der α_i und der β_i. Wir möchten nun noch eine untere Schranke an $G(\alpha_i - \alpha_{i+1}, \beta_{i+1} - \beta_i)$ finden. Es gilt

$$\begin{aligned} G(\alpha_i - \alpha_{i+1}, \beta_{i+1} - \beta_i) &\geq G\left(n^2 \frac{\sqrt{\pi_{i+1}}}{\sqrt{\pi_i}} \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right), n^2 \frac{\sqrt{\pi_i}}{\sqrt{\pi_{i+1}}} \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) \right) \\ &= n^2 \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) G\left(\frac{\sqrt{\pi_{i+1}}}{\sqrt{\pi_i}}, \frac{\sqrt{\pi_i}}{\sqrt{\pi_{i+1}}} \right) \\ &\geq 2n^2 \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) \sqrt{\frac{\sqrt{\pi_{i+1}}}{\sqrt{\pi_i}} \frac{\sqrt{\pi_i}}{\sqrt{\pi_{i+1}}}} = 2n^2 \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) \end{aligned}$$

für alle $i \in \{2, \ldots, n-2\}$. Dabei verwenden wir die folgenden, in Lemma C.0.8 bewiesenen Eigenschaften der Abbildung G. Zuerst benutzen die Monotonie von G, dann die Homogenität von G und zuletzt die Abschätzung $G(a,b) \geq 2\sqrt{ab}$ für alle $a, b \geq 0$. Dies zeigt die gewünschte Krümmungsbedingung. Es gilt

$$\lim_{n\to\infty} 2n^2 \left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right) = \lim_{n\to\infty} \frac{\left(1 - e^{-\frac{\hat{\lambda}}{2n^2}}\right)}{\frac{1}{2n^2}} = -\frac{\mathrm{d}}{\mathrm{d}t} e^{-\hat{\lambda} t}\bigg|_{t=0} = \hat{\lambda}$$

un somit die Behauptung. □

5.3 Funktionalungleichungen

In den folgenden Ausführungen betrachten wir stets das irreduzible und reversible Markov-Tripel $(\mathcal{X}, Q, \pi)$.

Theorem 5.3.1 (Diskrete Csiszár-Kullback-Pinsker-Ungleichung)**.** *Sei* $\rho \in \mathcal{D}(\mathcal{X})$*, dann gilt*

$$d_{TV}(\rho, \mathbb{1}) \leq \sqrt{2H(\rho)}.$$

Insbesondere gilt für alle $\rho \in \mathcal{D}(\mathcal{X})$ *die Ungleichung*

$$d_{TV}(e^{tQ}\rho, \mathbb{1}) \leq \sqrt{2H(e^{tQ}\rho)}$$

für alle Zeiten $t > 0$*.*

Beweis. Sei $\rho \in D(\mathcal{X})$ beliebig. Wir bemerken zunächst, dass

$$H(\rho) = \int_{\mathcal{X}} \rho \log(\rho) \mathrm{d}\pi = \int_{\mathcal{X}} \rho \log(\rho) \mathrm{d}\pi - \int_{\mathcal{X}} \rho \mathrm{d}\pi + \int_{\mathcal{X}} \mathbb{1} \mathrm{d}\pi = \int_{\mathcal{X}} \left(\rho \log(\rho) - \rho + \mathbb{1}\right) \mathrm{d}\pi$$

gilt. Es bezeichne $\varphi(x)\colon (0,\infty) \to [0,\infty)$ die Abbildung, definiert durch $\varphi(x) = x(\log(x) - 1) + 1$. Wir betrachten die Taylorentwicklung erster Ordnung in dem Punkt $x_0 = 1$, das heißt,

$$\begin{aligned}
\varphi(s+1) &= \varphi(1) + \varphi'(1)(s+1-1) + \int_1^{s+1} (s+1-t)\varphi''(t)\mathrm{d}t = \int_1^{s+1} (s+1-t)\varphi''(t)\mathrm{d}t \\
&= \int_0^s (s-r)\varphi''(1+r)\mathrm{d}r = \int_0^1 (s-s\tau)\varphi''(1+s\tau)s\mathrm{d}\tau \\
&= s^2 \int_0^1 (1-\tau)\varphi''(1+s\tau)\mathrm{d}\tau.
\end{aligned}$$

Damit erhalten wir die folgende Darstellung der Entropie

$$
\begin{aligned}
H(\rho) &= \int_{\mathcal{X}} \varphi(\rho(x))\mathrm{d}\pi = \int_{\mathcal{X}} \varphi(\rho(x) - \mathbb{1}(x) + \mathbb{1}(x))\mathrm{d}\pi \\
&= \int_{\mathcal{X}} (\rho(x)-1)^2 \int_0^1 (1-\tau)\varphi''(\tau(\rho(x)-1)+1)\mathrm{d}\tau\mathrm{d}\pi.
\end{aligned}
$$

Wir schließen

$$
\begin{aligned}
&\frac{1}{4} d_{TV}(\rho,\mathbb{1})^2 = \left(\int_{\mathcal{X}} \frac{1}{2}|\rho(x)-1|\right)^2 = \left(\int_{\mathcal{X}}\int_0^1 (1-\tau)\mathrm{d}\tau|\rho(x)-1|\mathrm{d}\pi\right)^2 \\
&= \left(\int_{\mathcal{X}}\int_0^1 (1-\tau)|\rho(x)-1|\mathrm{d}\tau\mathrm{d}\pi\right)^2 \\
&= \left(\int_{\mathcal{X}}\int_0^1 \sqrt{1-\tau}\sqrt{\varphi''(\tau(\rho(x)-1)+1)}|\rho(x)-1|\frac{\sqrt{1-\tau}}{\sqrt{\varphi''(\tau(\rho(x)-1)+1)}}\mathrm{d}\tau\mathrm{d}\pi\right)^2 \\
&\leq \int_{\mathcal{X}}\int_0^1 (1-\tau)\varphi''(\tau(\rho(x)-1)+1)|\rho(x)-1|^2\mathrm{d}\tau\mathrm{d}\pi \int_{\mathcal{X}}\int_0^1 \frac{1-\tau}{\varphi''(\tau(\rho(x)-1)+1)}\mathrm{d}\tau\mathrm{d}\pi \\
&= H(\rho)\int_{\mathcal{X}}\int_0^1 \frac{1-\tau}{\varphi''(\tau(\rho(x)-1)+1)}\mathrm{d}\tau\mathrm{d}\pi.
\end{aligned}
$$

Es gilt $\varphi''(t) = \frac{1}{t}$ für alle $t > 0$, dies zeigt die Integrabilität von $\frac{1}{\varphi''}$. Ferner zeigt dies auch

$$
\begin{aligned}
\int_{\mathcal{X}}\int_0^1 \frac{1-\tau}{\varphi''(\tau(\rho(x)-1)+1)}\mathrm{d}\tau\mathrm{d}\pi &= \int_{\mathcal{X}}\int_0^1 \left(\tau(1-\tau)(\rho(x)-1)+(1-\tau)\right)\mathrm{d}\tau\mathrm{d}\pi \\
&= \frac{1}{2} + \int_0^1 \tau(1-\tau)\int_{\mathcal{X}} (\rho(x)-\mathbb{1})\,\mathrm{d}\pi\mathrm{d}\tau = \frac{1}{2}.
\end{aligned}
$$

Insgesamt schlussfolgern wir

$$
d_{TV}(\rho,\mathbb{1}) \leq \sqrt{2H(\rho)}
$$

für alle $\rho \in \mathcal{D}(\mathcal{X})$. □

Bemerkung 5.3.2. (i) Theorem 5.3.1 ist das in der Einleitung erwähnte Resultat, das die Entropie in Zusammenhang mit der Näherung an die stationäre Verteilung setzt. Wir bemerken, dass für die Berechnung der Entropie jedoch bereits das Wissen über die stationäre Verteilung bekannt sein muss. Das heißt, man kann dies nicht direkt in einem praktischen Algorithmus realisieren. Ein möglicher Ausweg wäre zum Beispiel die Berechnung des Terms

$$
\int_{\mathcal{X}} \varphi(\rho)\mathrm{d}\mathbb{1},
$$

mit $\varphi(x) = x(\log(x)-1)+1$, der in jedem Fall größer ist als die Entropie. Ist obiger Wert klein, dann ist folglich auch die Entropie klein und somit der Abstand zur stationären Verteilung.

(ii) Dieses Theorem zeigt, dass die diskrete Entropie eine Entropie im Sinne der eingangs vorgestellten Definition 2.1.2 ist. Dazu argumentieren wir wie in Bemerkung 2.2.3 (i). Für den Beweis, dass die Entropie eine Ljapunovfunktion ist, verweisen wir auf den Beweis von Satz 5.3.11.

Lemma 5.3.3. Sei $(\mathcal{X}, Q, \pi)$ ein irreduzibles reversibles Markov-Tripel und $\rho_0 \in \mathcal{D}(\mathcal{X})$. Wir bezeichnen mit $\rho_t = e^{tQ}\rho_0$ die zugehörige Lösung der Wärmeleitungsgleichung. Dann gilt

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\rho_t) = -\frac{1}{2} \sum_{x,y \in \mathcal{X}} \left(\log(\rho_t(x)) - \log(\rho_t(y))\right) \left(\rho_t(x) - \rho_t(y)\right) Q(x,y)\pi(\{x\})$$

für alle $t > 0$.

Beweis. Dies ist eine Folgerung von Lemma 4.3.2 und der Rechnung im Beweis von Lemma 4.3.1. Das tangentiale Vektorfeld entlang der Lösungen der Wärmeleitungsgleichung ist gegeben durch $-\nabla \log(\rho_t)$ für alle $t > 0$. Dann gilt

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} H(\rho_t) &= -\langle \nabla \log(\rho_t), \nabla \log(\rho_t) \rangle_{\rho_t} \\
&= -\frac{1}{2} \sum_{x,y \in \mathcal{X}} \left(\log(\rho_t(x)) - \log(\rho_t(y))\right)^2 Q(x,y) \Lambda(\rho(x), \rho(\xi)) \pi(\{x\}) \\
&= -\frac{1}{2} \sum_{x,y \in \mathcal{X}} \left(\log(\rho_t(x)) - \log(\rho_t(y))\right) \left(\rho_t(x) - \rho_t(y)\right) Q(x,y)\pi(\{x\})
\end{aligned}$$

nach der Rechnung im Beweis von Lemma 4.3.1 zusammen mit der Eigenschaft $(\log(a) - \log(b))\Lambda(a,b) = a - b$. □

Bemerkung 5.3.4. Wir definieren die Abbildung $\mathcal{I} \colon \mathcal{D}(\mathcal{X}) \to \mathbb{R}$ durch die Abbildungsvorschrift

$$\begin{aligned}
\mathcal{I}(\rho) &= \frac{1}{2} \sum_{x,y \in \mathcal{X}} \left(\log(\rho(x)) - \log(\rho(y))\right) \left(\rho(x) - \rho(y)\right) Q(x,y)\pi(\{x\}) \\
&= \langle \nabla \log(\rho), \nabla \log(\rho) \rangle_\rho
\end{aligned}$$

für $\rho \in \mathcal{D}_*(\mathcal{X})$ und $\mathcal{I}(\rho) = \infty$, falls $\rho \in \mathcal{D}(\mathcal{X}) \setminus \mathcal{D}_*(\mathcal{X})$ ist. $\mathcal{I}$ ist das diskrete Analogon der Fisher-Information aus Definition 2.4.4. Mit diesen Bezeichnungen ist die negative Entropieproduktion entlang von Lösungen der Wärmeleitungsgleichung gleich der Fisher-Information.

Lemma 5.3.5. Seien $\rho_0, \sigma \in \mathcal{D}(\mathcal{X})$. Bezeichnen wir mit $\rho_t = e^{tQ}\rho_0$, dann gilt für alle $t > 0$ die Ungleichung

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t, \rho_1) \leq \sqrt{\mathcal{I}(\rho_t)}.$$

Beweis. Wie bereits gesehen, gilt $\rho_t \in \mathcal{D}_*(\mathcal{X})$ für alle $t > 0$. Ferner ist die Abbildung $t \mapsto \rho_t$ nach Theorem 4.3.3 differenzierbar. Es gilt

$$\begin{aligned}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t, \sigma) &= \limsup_{s\to 0^+} \frac{\mathcal{W}(\rho_{t+s}, \sigma) - \mathcal{W}(\rho_t, \sigma)}{s} \\ &\leq \limsup_{s\to 0^+} \frac{\mathcal{W}(\rho_{t+s}, \rho_t) + \mathcal{W}(\rho_t, \sigma) - \mathcal{W}(\rho_t, \sigma)}{s} \leq \limsup_{s\to 0^+} \frac{\mathcal{W}(\rho_t, \rho_{t+s})}{s}\end{aligned}$$

aufgrund der Dreiecksungleichung für $\mathcal{W}$. Wir möchten diesen Ausdruck abschätzen. Nach Lemma 4.3.2 gilt $(\rho_r, -\log(\rho_r))_{r\in[0,s]} \in \mathcal{CE}_s(\rho_t, \rho_{t+s})$. Mit Lemma 4.1.23 sowie der Gleichheit $\langle A(\rho_t)\log(\rho_t), \log(\rho_t)\rangle = \langle \nabla \log(\rho_t), \nabla \log(\rho_t)\rangle_{\rho_t}$ für alle $t > 0$ folgt die Abschätzung

$$\mathcal{W}(\rho_t, \rho_{t+s}) \leq \int_0^s \sqrt{\langle \nabla \log(\rho_t), \nabla \log(\rho_t)\rangle_{\rho_t}}\mathrm{d}t.$$

Diese Abschätzung impliziert

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t, \sigma) \leq \limsup_{s\to 0^+} \frac{\mathcal{W}(\rho_t, \rho_{t+s})}{s} \leq \limsup_{s\to 0^+} \frac{1}{s}\int_0^s \sqrt{\langle \nabla \log(\rho_t), \nabla \log(\rho_t)\rangle_{\rho_t}}\mathrm{d}t.$$

Verwenden wir, dass $\mathcal{I}(\rho_t) = \langle \nabla \log(\rho_t), \nabla \log(\rho_t)\rangle_{\rho_t}$ gilt und $t \mapsto \mathcal{I}(\rho_t)$ stetig ist, dann folgt

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t, \sigma) \leq \sqrt{\mathcal{I}(\rho_t)}$$

für alle $t > 0$. □

Theorem 5.3.6. *Erfüllt $(\mathcal{X}, Q, \pi)$ die Krümmungsbedingung $CD(\lambda, \infty)$ für ein $\lambda \in \mathbb{R}$, dann gilt die Ungleichung*

$$H(\rho) \leq \mathcal{W}(\rho, \mathbb{1})\sqrt{\mathcal{I}(\rho)} - \frac{\lambda}{2}\mathcal{W}(\rho, \mathbb{1})^2 \tag{H$\mathcal{WI}$}$$

für alle $\rho \in \mathcal{D}(\mathcal{X})$.

Beweis. Sei $\rho \in \mathcal{D}(\mathcal{X})$ beliebig. Ist $\rho(x) = 0$ für ein $x \in \mathcal{X}$, dann folgt $\mathcal{I}(\rho) = \infty$ und die gewünschte Ungleichung ist trivialerweise erfüllt. Sei also $\rho \in \mathcal{D}_*(\mathcal{X})$. Da $(\mathcal{X}, Q, \pi)$ die Krümmungsbedingung $CD(\lambda, \infty)$ für $\lambda \in \mathbb{R}$ erfüllt, können wir Theorem 5.1.4 auf $(\rho, \mathbb{1})$

anwenden und erhalten für $t \geq 0$ die Ungleichung

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(e^{tQ}\rho,\mathbb{1})^2 + \frac{\lambda}{2}\mathcal{W}(e^{tQ}\rho,\mathbb{1})^2 \leq H(\mathbb{1}) - H(e^{tQ}\rho).$$

Diese impliziert für $t = 0$

$$H(\rho) \leq -\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho,\mathbb{1})^2\bigg|_{t=0} - \frac{\lambda}{2}\mathcal{W}(\rho,\mathbb{1})^2,$$

denn $H(\mathbb{1}) = 0$. Sei $t > 0$, dann gilt die Ungleichung

$$\begin{aligned}
0 &\leq 2\mathcal{W}(\rho,\rho_t)^2 = \mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}) - \mathcal{W}(\rho,\mathbb{1})^2 \\
&\quad + \mathcal{W}(\rho,\rho_t)^2 - 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}) + \mathcal{W}(\rho,\mathbb{1})^2 \\
&= \mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}) - \mathcal{W}(\rho,\mathbb{1})^2 + \big(\mathcal{W}(\rho,\rho_t)^2 - \mathcal{W}(\rho,\mathbb{1})\big)^2 \\
&\leq \mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}) - \mathcal{W}(\rho,\mathbb{1})^2 + (\mathcal{W}(\rho,\mathbb{1}) + \mathcal{W}(\mathbb{1},\rho_t) - \mathcal{W}(\rho,\mathbb{1}))^2 \\
&= \mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}) - \mathcal{W}(\rho,\mathbb{1})^2 + \mathcal{W}(\mathbb{1},\rho_t)^2
\end{aligned}$$

und aus dieser Ungleichung folgt aufgrund der Symmetrie von $\mathcal{W}$ die Ungleichung

$$\mathcal{W}(\rho,\mathbb{1})^2 - \mathcal{W}(\rho_t,\mathbb{1})^2 \leq \mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1}).$$

Dies zeigt,

$$\begin{aligned}
-\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho,\mathbb{1})^2\bigg|_{t=0} &= -\frac{1}{2}\limsup_{t\to 0^+}\frac{1}{t}\left(\mathcal{W}(\rho_t,\mathbb{1})^2 - \mathcal{W}(\rho,\mathbb{1})^2\right) \\
&= \frac{1}{2}\liminf_{t\to 0^+}\frac{1}{t}\left(\mathcal{W}(\rho,\mathbb{1})^2 - \mathcal{W}(\rho_t,\mathbb{1})^2\right) \\
&\leq \frac{1}{2}\liminf_{t\to 0^+}\frac{1}{t}\left(\mathcal{W}(\rho,\rho_t)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1})\right) \\
&\leq \frac{1}{2}\limsup_{t\to 0^+}\frac{1}{t}\left(\mathcal{W}(\rho_t,\rho)^2 + 2\mathcal{W}(\rho,\rho_t)\mathcal{W}(\rho,\mathbb{1})\right).
\end{aligned}$$

Wir untersuchen den ersten Term der rechten Seite und behaupten, dass

$$\limsup_{t\to 0^+}\frac{\mathcal{W}(\rho,\rho_t)^2}{t} = 0$$

gilt. Wie im Beweis von Lemma 5.3.5 gesehen, gilt für alle $t > 0$, dass

$$(\rho_r, -\log(\rho_r))_{r\in[0,t]} \in \mathcal{CE}(\rho,\rho_t)$$

ist. Mit Lemma 4.1.20 folgt dann

$$\frac{\mathcal{W}(\rho_t, \rho)^2}{t} \leq \frac{1}{t} t \int_0^t \langle \log(\rho_r), \log(\rho_r) \rangle_{\rho_r} \mathrm{d}r,$$

für $t \to 0^+$ verschwindet der rechte Term, dies zeigt die Behauptung. Wir schließen

$$\begin{aligned}
-\frac{1}{2} \frac{\mathrm{d}^+}{\mathrm{d}t} \mathcal{W}(\rho, \mathbb{1})^2 \bigg|_{t=0} &\leq \frac{1}{2} \limsup_{t \to 0^+} \frac{1}{t} \left(\mathcal{W}(\rho_t, \rho)^2 + 2\mathcal{W}(\rho, \rho_t) \mathcal{W}(\rho, \mathbb{1}) \right) \\
&= \limsup_{t \to 0^+} \frac{\mathcal{W}(\rho_t, \rho)^2}{t} + \frac{2}{2} \mathcal{W}(\rho, \mathbb{1}) \frac{\mathrm{d}^+}{\mathrm{d}t} \mathcal{W}(\rho, \rho_t) \bigg|_{t=0} \\
&\leq 0 + \mathcal{W}(\rho, \mathbb{1}) \sqrt{\mathcal{I}(\rho)}
\end{aligned}$$

nach einer Anwendung von Lemma 5.3.5 mit $\rho_0 = \rho_1 = \rho$ und $t = 0$. Dies zeigt die Ungleichung (H$\mathcal{WI}$) für alle $\mathcal{D}(\mathcal{X})$. □

Theorem 5.3.7. *Sei $(\mathcal{X}, Q, \pi)$ ein irreduzibles reversibles Markov-Tripel. Gilt die Krümmungsbedingung $CD(\lambda, \infty)$ mit $\lambda > 0$ für $(\mathcal{X}, Q, \pi)$, dann gilt für alle $\rho \in \mathcal{D}(\mathcal{X})$ die sogenannte modifizierte logarithmische Sobolev-Ungleichung*

$$H(\rho) \leq \frac{1}{2\lambda} \mathcal{I}(\rho). \tag{MLSI}$$

Beweis. $(\mathcal{X}, Q, \pi)$ erfüllt die Krümmungsbedingung $CD(\lambda, \infty)$, folglich gilt auch die (H$\mathcal{WI}$)-Ungleichung mit Konstante $\lambda > 0$. Seien $x, y \in \mathbb{R}$ und $c > 0$. Wir wenden die Peter-Paul-Ungleichung auf $\varepsilon = \frac{1}{2c}$ an und erhalten

$$ab \leq cx^2 + \frac{1}{4c} y^2.$$

Wir setzen $x = \mathcal{W}(\rho, \mathbb{1})$, $y = \sqrt{I(\rho)}$ und $c = \frac{\lambda}{2}$ und wenden obige Ungleichung auf die (H$\mathcal{WI}$)-Ungleichung an. Es folgt

$$H(\rho) \leq \mathcal{W}(\rho, \mathbb{1}) \sqrt{I(\rho)} - \frac{\lambda}{2} \mathcal{W}(\rho, \mathbb{1})^2 \leq \frac{\lambda}{2} \mathcal{W}(\rho, \mathbb{1})^2 + \frac{1}{2\lambda} \mathcal{I}(\rho) - \frac{\lambda}{2} \mathcal{W}(\rho, \mathbb{1})^2 = \frac{1}{2\lambda} \mathcal{I}(\rho)$$

für alle $\rho \in \mathcal{D}(\mathcal{X})$. □

Korollar 5.3.8. Sei $(\mathcal{X}, Q, \pi)$ ein irreduzibles und reversibles Markov-Tripel, welches die Krümmungsbedingung $CD(\lambda, \infty)$ für $\lambda > 0$ erfüllt. In diesem Fall gilt für alle $\rho_0 \in \mathcal{D}(\mathcal{X})$ und alle $t > 0$ die Abschätzung

$$H(e^{tQ} \rho_0) \leq -\frac{1}{2\lambda} \frac{\mathrm{d}}{\mathrm{d}t} H(e^{tQ} \rho_0).$$

Insbesondere ist $H(\rho_t) \leq e^{-2\lambda t} H(\rho_0)$ für alle $t \geq 0$.

Beweis. $(\mathcal{X}, Q, \pi)$ erfüllt nach Theorem 5.3.7 die MLSI-Ungleichung. Sei $\rho_0 \in \mathcal{D}(\mathcal{X})$, dann gilt

$$H(e^{tQ}\rho_0) \leq \frac{1}{2\lambda}\mathcal{I}(e^{tQ}) = -\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2\lambda}H(e^{tQ}\rho_0)$$

nach Lemma 5.3.3. Da die Abbildung $(0, \infty) \to \mathbb{R}$, $t \mapsto H(e^{tQ}\rho_0)$ differenzierbar ist, folgt mit dem Lemma von Gronwall

$$H(\rho_t) \leq e^{-2\lambda t} H(\rho_0)$$

für alle $t \geq 0$. □

Korollar 5.3.9. Für ein irreduzibles und reversibles Markov-Tripel $(\mathcal{X}, Q, \pi)$, welches die Krümmungsbedingung $CD(\lambda, \infty)$ für ein $\lambda > 0$ erfüllt, gilt für jeden Anfangswert $\rho_0 \in \mathcal{D}(\mathcal{X})$, dass die zugehörige Lösung exponentiell schnell gegen die Equilibriumslösung $\mathbb{1}$ konvergiert. Für alle $t \geq 0$ gilt die Ungleichung

$$d(\rho_t, \mathbb{1}) \leq \sqrt{2H(\rho_0)}e^{-\lambda t}.$$

Beweis. Die Behauptung folgt aus einer Kombination von Theorem 5.3.1 und Korollar 5.3.8. □

Bemerkung 5.3.10. Wir haben in Korollar 5.3.9 gezeigt, dass die Markov-Halbgruppe zu einem Markov-Kern, der die Krümmungsbedingung $CD(\lambda, \infty)$ für ein $\lambda > 0$ erfüllt, für $t \to \infty$ konvergiert. In folgendem Satz werden wir zeigen, dass dies für alle irreduziblen und reversiblen Markov-Kerne gilt. Wir bemerken, dass dabei jedoch nicht die exponentiell schnelle Konvergenz jedes irreduziblen und reversiblen Markov-Kerns gezeigt wird, sondern lediglich die Konvergenz der Markov-Halbgruppe ohne eine Aussage über die Konvergenzgeschwindigkeit.

Satz 5.3.11. Sei $(\mathcal{X}, Q, \pi)$ ein irreduzibles und reversibles Markov-Tripel. Für alle $\rho \in \mathcal{D}(\mathcal{X})$ konvergiert $e^{tQ}\rho \to \mathbb{1}$ für $t \to \infty$.

Beweis. Wir fassen $\dot{\rho} = Q\rho$ als Differentialgleichung im $\mathbb{R}^n$ auf. Die Lösung der Differentialgleichung ist gegeben durch $\rho_t = e^{tQ}\rho$ für einen beliebigen Startwert $\rho \in \mathcal{D}(\mathcal{X})$. Wir haben bereits gesehen, dass die Menge $\mathcal{D}(\mathcal{X})$ eine invariante Menge der Differentialgleichung $\dot{\rho} = Q\rho$ ist. Wir möchten zeigen, dass $H \colon \mathcal{D}(\mathcal{X}) \to \mathbb{R}$ eine strikte Ljapunovfunktion ist. In Lemma 5.3.3 haben wir gezeigt, dass die Ungleichung

$$\frac{\mathrm{d}}{\mathrm{d}t}H(\rho_t) = -\langle \nabla \log(\rho_t), \nabla \log(\rho_t)\rangle_{\rho_t} \leq 0$$

entlang der Lösung für alle $t > 0$ erfüllt ist. Ferner gilt wegen der positiven Definitheit dieses Skalarprodukts und wegen $e^{tQ} > 0$ für alle $t \geq 0$ die Gleichheit in obiger Ungleichung genau dann, wenn $\rho_t = \mathbb{1}$ ist. Dies zeigt, dass H eine strikte Ljapunovfunktion ist. Die Menge $\mathcal{D}(\mathcal{X})$ ist nach dem Satz von Heine-Borel kompakt. Die Verwendung von [PW10, Satz 5.5.6] zeigt die Konvergenz der Lösung gegen ein Equilibrium. Da Q ein Markov-Kern ist, können wir mit [Cha75, Theorem 1] folgern, dass Q den Rang $n-1$ hat. Da $\rho_t > 0$ für alle $t \geq 0$ gilt, ist das einzige mögliche Equilibrium gegeben durch $\mathbb{1}$. Dies zeigt die Konvergenz der Lösung gegen die Funktion $\mathbb{1}$. $\square$

A Markov-Prozesse

Wir geben eine kurze Einführung in die Theorie der Markov-Prozesse. Dieses Kapitel stellt lediglich die für die in Kapitel 3 benötigten Begriffe bereit. Ziel ist es, den Begriff der Markov-Kette einzuführen. Markov-Ketten bilden die Grundlage für ein wichtiges Beispiel in Kapitel 3. Für weitergehende Informationen verweisen wir auf die Bücher [Nor98] und [Beh13].

Definition A.0.1 (Stochastischer Prozess)**.** Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{X}, \mathcal{A})$ ein messbarer Raum. Eine Familie von Zufallsvariablen $(X_t)_{t\geq 0}$, wobei $X_t \colon \Omega \to \mathcal{X}$ für alle $t \geq 0$ ist, nennen wir einen stochastischen Prozess. Den messbaren Raum $(\mathcal{X}, \mathcal{A})$ bezeichnen wir dann als Zustandsraum des stochastischen Prozesses.

Ein Markov-Prozess ist ein spezieller stochastischer Prozess. Man könnte sagen, bei einem Markov-Prozess hängt die Zukunft nur von der Gegenwart, nicht aber von der Vergangenheit ab. Für eine rigorose Definition müssen wir ein wenig ausholen.

Definition A.0.2 (Filtration)**.** Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{F}_t)_{t\geq 0}$ eine Familie von Teil-σ-Algebren von Σ. Falls für alle $s < t$ die Bedingung $\mathcal{F}_s \subset \mathcal{F}_t$ gilt, nennen wir $(\mathcal{F}_t)_{t\geq 0}$ eine Filtration. Sei $(X_t)_{t\geq 0}$ ein stochastischer Prozess, dann heißt $(X_t)_{t\geq 0}$ an $(\mathcal{F}_t)_{t\geq 0}$ adaptiert, falls X_t für alle $t \geq 0$ schon $(\mathcal{F}_t, \mathcal{A})$-messbar ist. Ein wichtiges Beispiel für eine solche Filtration ist die sogenannte natürliche Filtration $\mathcal{F}_t := \sigma(\{X_s \mid 0 \leq s \leq t\})$. Dies ist die kleinste σ-Algebra, bezüglich welcher alle X_s für $0 \leq s \leq t$ messbar sind.

Definition A.0.3 (Bedingte Erwartung)**.** Sei $Y \colon \Omega \to \mathcal{X}$ eine integrierbare Zufallsvariable und $\Sigma_0 \subset \Sigma$ eine Unter-σ-Algebra. Wir nennen eine Funktion $\varphi \colon \Omega \to \mathcal{X}$ bedingte Erwartung von Y unter der Bedingung Σ_0, falls sie $(\Sigma_0, \mathcal{A})$-messbar ist und

$$\int_{E_0} Y \mathrm{d}\mathbb{P} = \int_{E_0} \varphi \mathrm{d}\mathbb{P}$$

für alle $E_0 \in \Sigma_0$ erfüllt.

Satz A.0.4. Die bedingte Erwartung aus Definition A.0.3 ist eindeutig bestimmt. Wir bezeichnen sie mit $\mathbb{E}(Y|\Sigma_0)$. Falls $\Sigma_0 = \sigma(Z_1, \ldots, Z_n)$ für die Zufallsvariablen $Z_1, \ldots, Z_n$ ist, dann schreiben wir $\mathbb{E}(Y|Z_1, \ldots, Z_n)$ beziehungsweise $\mathbb{E}(Y|Z)$ für den Fall, dass $\Sigma_0 = \sigma(Z)$.

Beweis. Einen Beweis dieser Aussage findet man zum Beispiel in [Beh13, Satz 1.4.2]. □

Definition A.0.5 (Bedingte Wahrscheinlichkeit)**.** Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\Sigma_0 \subset \Sigma$ eine Unter-σ-Algebra und $E \in \Sigma_0$ ein Ereignis. Die bedingte Erwartung der Zufallsvariable $\mathbb{1}_E$ bezeichnen wir mit $\mathbb{P}(E|\Sigma_0) := \mathbb{E}(\mathbb{1}_E|\Sigma_0)$ und nennen sie die bedingte Wahrscheinlichkeit von E.

Definition A.0.6 (Markov-Prozess)**.** Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_t)_{t\geq 0}$ ein stochastischer Prozess auf dem Zustandsraum $(\mathcal{X}, \mathcal{A})$ und $(\mathcal{F}_t)_{t\geq 0}$ die natürliche Filtration. $(X_t)_{t\geq 0}$ heißt Markov-Prozess, falls für alle $s > 0$ und $0 \leq t < s$ sowie $E \in \Sigma$

$$\mathbb{P}(X_s \in E|\mathcal{F}_t) = \mathbb{P}(X_s \in E|X_t)$$

gilt. Zusätzlich werden wir voraussetzen, dass die Pfade $t \mapsto X_t(\omega)$ für jedes $\omega \in \Omega$ rechtsseitig stetig sind.

Wir möchten nun eine spezielle Klasse von Markov-Prozessen vorstellen, die sogenannten Markov-Ketten. Sei $\mathcal{X}$ im Folgenden eine höchstens abzählbare Menge. Wählen wir als σ-Algebra die Potenzmenge $\mathscr{P}(\mathcal{X})$, dann ist $(\mathcal{X}, \mathscr{P}(\mathcal{X}))$ ein messbarer Raum. Zu einem gegebenen Markov-Prozess $(X_t)_{t\geq 0}$ seien $0 \leq t < s$ und $y \in \mathcal{X}$ beliebig, dann gilt

$$\mathbb{P}(X_s = y|\mathcal{F}_t) = \mathbb{P}(X_s \in \{y\}|\mathcal{F}_t) = \mathbb{P}(X_s \in \{y\}|X_t) = \mathbb{P}(X_s = y|X_t).$$

Die rechte Seite obiger Gleichung können wir explizit berechnen. Sei $x \in \mathcal{X}$, dann hat sie auf den Mengen $\{X_t = x\}$ den konstanten Wert $\mathbb{P}(X_s = y|X_t = x) \in [0, 1]$. Dies führt uns zu der folgenden Definition.

Definition A.0.7 (Markov-Kette)**.** Einen Markov-Prozess auf einem höchstens abzählbaren Zustandsraum nennen wir Markov-Kette. Hängen für alle $0 \leq t < s$ und $x, y \in \mathcal{X}$ die Zahlen

$$\mathbb{P}(X_s = y|X_t = x)$$

nur von x, y und $s - t$ ab, so sprechen wir von einer Markov-Kette in homogener Zeit.

B Riemannsche Mannigfaltigkeiten

In diesem Kapitel sind einige Resultate zusammengetragen, die bei unserer Behandlung der riemannschen Mannigfaltigkeit $(\mathcal{D}_*(\mathcal{X}), \langle\cdot,\cdot\rangle_\rho)$ hilfreich sind. Für weitere Informationen empfehlen wir [Lee13] und [Gri09], die auch eine Grundlage dieser Ausführungen bilden.

Lemma B.0.1. Sei V ein endlich dimensionaler normierter $\mathbb{R}$-Vektorraum, dann ist V eine n-dimensionale glatte Mannigfaltigkeit. Ist ferner $U \subset V$ ein m-dimensionaler affiner Unterraum, dann ist auch U eine glatte Mannigfaltigkeit.

Beweis. Es ist V normiert und folglich ist die Topologie hausdorffsch. Da V von endlicher Dimension ist, wählen wir Vektoren $v_1, \ldots, v_n$, die eine Basis von V bilden. Die Abbildung

$$K_1 \colon \mathbb{R}^n \to V,\ K_1(\xi) = \sum_{i \in \mathbb{N}} \xi_i v_i$$

definiert einen linearen stetigen Isomorphismus. Folglich ist K_1^{-1} eine Karte von ganz V und es genügt, den Atlas (V, K_1^{-1}) zu betrachten. Dieser Atlas ist schon ein differenzierbarer Atlas, denn wir brauchen keine Koordinatenwechsel zu betrachten. Sei U von der Form $U = x + \operatorname{span}\{v_1, \ldots, v_m\}$ für linear unabhängige $v_1, \ldots, v_m$. Wir modifizieren die Abbildung K_2 zu

$$K_2 \colon \mathbb{R}^m \to U,\ K(\xi) = x + \sum_{i=1}^{m} \xi_i v_i.$$

Dann ist auch diese Abbildung ein Isomorphismus und folglich ist nach obigem Argument auch U eine glatte Mannigfaltigkeit. Es gilt sogar, dass die resultierende differentielle Struktur unabhängig von der Wahl der Basis ist. Einen Beweis findet man in [Lee13, Beispiel 1.6]. $\square$

Definition B.0.2. Sei M eine glatte Mannigfaltigkeit. Eine Abbildung $X \colon \mathcal{C}^\infty(M) \to \mathbb{R}$ heißt Derivation in $x \in M$, falls

(i) sie linear ist, und

(ii) für alle $f, g \in \mathcal{C}^\infty(M)$

$$X(fg) = X(f)g(x) + f(x)X(g)$$

gilt.

Die Menge aller Derivationen in x nennen wir den Tangentialraum von M an x und schreiben

$$T_x M = \{X : \mathcal{C}^\infty(M) \to \mathbb{R} \mid X \text{ ist eine } \mathbb{R}\text{-Derivation}\}.$$

Satz B.0.3. Sei M eine n-dimensionale glatte Mannigfaltigkeit, dann ist für jedes $x \in M$ der Tangentialraum $T_x M$ ein Vektorraum der Dimension n.

Beweis. [Gri09, Theorem 3.8] □

Lemma B.0.4. Sei M eine n-dimensionale glatte Mannigfaltigkeit. Für ein Intervall $I \subset \mathbb{R}$ und eine glatte Kurve $\gamma\colon I \to M$ mit $\gamma(0) = x$ und $0 \in \mathring{I}$, definieren wir für $f \in C^\infty(M)$

$$v(f) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)\Big|_{t=0}.$$

Die Abbildung $v\colon C^\infty(M) \to \mathbb{R},\ f \mapsto v(f)$ ist eine Derivation in x. Es gilt sogar, dass jede Derivation schon von obiger Form ist.

Beweis. [Lee13, Lemma 3.10] □

Bemerkung B.0.5. In der Situation von Lemma B.0.1 können wir den Tangentialraum an jeden Punkt mit ein und derselben Menge $TU \simeq V$ identifizieren. Für weitere Informationen dazu verweisen wir auf [Lee13, Proposition 3.8].

Definition B.0.6. Sei M eine n-dimensionale glatte Mannigfaltigkeit. Wir definieren das Tangentenbündel TM als die Menge $TM := \{(x, v) \mid x \in M \text{ und } v \in T_x M\}$. Ein Vektorfeld auf M ist eine Abbildung $V\colon M \to TM$, $x \mapsto (x, V(x))$, sodass $V(x) \in T_x M$ für alle $x \in M$ gilt.

Definition B.0.7. Sei M eine n-dimensionale glatte Mannigfaltigkeit. Eine riemannsche Metrik ist eine Familie $g = (g_x(\cdot,\cdot))_{x \in M}$, sodass für alle $x \in M$ die bilineare Form $g_x(\cdot,\cdot)$ ein Skalarprodukt auf $T_x M$ ist, welches glatt von x abhängt. Das heißt, für alle glatten Vektorfelder $V, W\colon M \to TM$ ist die Abbildung $x \mapsto g_x(V(x), W(x))$ differenzierbar. Das Paar (M, g) nennen wir riemannsche Mannigfaltigkeit.

Auf jeder riemannschen Mannigfaltigkeit (M, g) können wir mithilfe der riemannschen Metrik einen Abstandsbegriff einführen.

Definition B.0.8. Sei $\gamma\colon [0, 1] \to M$ eine stückweise glatte Kurve. Es bezeichne $\gamma'(t)$ den tangentialen Vektor an diese Kurve. Die Länge dieser Kurve ist definiert durch

$$\mathcal{L}(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))}\mathrm{d}t.$$

Für zwei Punkte $x, y \in M$ definieren wir den Abstand zwischen p und q als

$$d(p,q) = \inf \left\{ \mathcal{L}(\gamma) \;\middle|\; \gamma\colon [0,1] \to M \text{ ist stückweise glatt und } \gamma(0) = p, \text{ sowie } \gamma(1) = q \right\}.$$

Lemma B.0.9. Sei (M, g) eine riemannsche Mannigfaltigkeit, dann ist (M, d), wobei d die Abbildung aus Definition B.0.8 bezeichnet, ein metrischer Raum. Ferner stimmen die beiden zugehörigen Topologien überein.

Beweis. [Gri09, Abschnitt 3.11] □

Definition B.0.10 (Gradient)**.** Sei (M, g) eine riemannsche Mannigfaltigkeit. Für ein Funktional $f \in C^\infty(M)$ definieren wir den Gradienten als das eindeutig bestimmte Vektorfeld $\operatorname{grad} f\colon M \to TM$, welches

$$g_x(\operatorname{grad} f(x), X) = X(f)$$

für jeden tangentialen Vektor $X \in T_xM$ und jeden Punkt $x \in M$ erfüllt. Dieses Vektorfeld nennen wir den Gradienten von f.

Bemerkung B.0.11. Es stellt sich die Frage, ob der Gradient wohldefiniert ist. Wir betrachten für $f \in C^\infty(M)$ und für beliebiges $x \in M$ die Abbildung $T_xM \to \mathbb{R}$, $X \mapsto X(f)$. Diese Abbildung ist ein stetiges lineares Funktional auf dem Hilbertraum (T_xM, g_x), somit existiert nach dem Satz von Riesz-Fréchet ein eindeutig bestimmtes $v_x \in T_xM$, sodass

$$g_x(v_x, X) = X(f)$$

für alle $X \in T_xM$ gilt. Setzen wir $\operatorname{grad} f(x) := v_x$, dann ist $\operatorname{grad} f$ eindeutig bestimmt und erfüllt die gewünschte Gleichung.

C Hilfsresultate

Satz C.0.1. Sei (Ω, Σ) ein messbarer Raum. Für alle $f \in \mathcal{M}_b(\Omega, \Sigma)$ existiert eine Folge von Treppenfunktionen, die gleichmäßig gegen f konvergiert.

Beweis. [Sch05, Kapitel 8] $\square$

Satz C.0.2 (Ein Perron-Frobenius-Theorem)**.** Sei $A \in \mathbb{R}^{n\times n}$ mit nichtnegativen Einträgen, dann ist der Spektralradius $\rho(A)$ ein Eigenwert zu einem nichtnegativen Eigenvektor.

Beweis. [HJ13, Theorem 8.3.1] $\square$

Lemma C.0.3 (Spektralradius stochastischer Matrizen)**.** Sei $A \in \mathbb{R}^{n\times n}$ mit den Einträgen a_{ij}, eine zeilen-stochastische Matrix, das heißt, für alle $i \in \{1, \ldots, n\}$ gilt $\sum_{j=1}^{n} a_{ij} = 1$ und $a_{ij} \in [0, 1]$ für alle $i, j \in \{1, \ldots, n\}$. Dann gilt $\rho(A) = 1$ für den Spektralradius $\rho(A)$ von A.

Beweis. Es gilt $A\mathbb{1} = \mathbb{1}$, folglich ist $1 \in \sigma(A)$ und somit $1 \leq \rho(A)$. Weiter ist $\|A\|_\infty = 1$ und deshalb gilt $|\lambda| \leq 1$ für alle $\lambda \in \sigma(A)$. Es folgt $\rho(A) = 1$. $\square$

Definition C.0.4. Eine Matrix $A \in \mathbb{R}^{n\times n}$ mit den Einträgen $(A(i, j))$ heißt irreduzibel, falls für alle $i, j \in \{1, \ldots, n\}$ Indizes $i = i_1, \ldots, i_k = j$ existieren, sodass $A(i_l, i_{l+1}) \neq 0$ für alle $l = 1, \ldots, k-1$ gilt.

Satz C.0.5. Sei $A \in \mathbb{R}^{n\times n}$ eine Matrix mit nichtnegativen Einträgen sowie positiven Einträgen auf der Diagonalen, dann sind die folgenden Aussagen äquivalent:

1. Es existiert ein $k \in \mathbb{N}$, sodass $A^k > 0$.
2. A ist irreduzibel.

Beweis. [KB13, Theorem 8.46] $\square$

Lemma C.0.6. Seien $-1 \leq \alpha \leq \beta \leq 1$, dann gilt

$$\int_\alpha^\beta \sqrt{\frac{\operatorname{Artanh}(r)}{r}}\,\mathrm{d}r < \infty.$$

Beweis. Seien $-1 \leq \alpha \leq \beta \leq 1$ beliebig Wir wenden die Cauchy-Schwarz-Ungleichung auf das Integral

$$\int_\alpha^\beta \sqrt{\frac{\operatorname{Artanh}(r)}{r}}\,\mathrm{d}r \leq \int_{-1}^1 \sqrt{\frac{\operatorname{Artanh}(r)}{r}}\,\mathrm{d}r \leq \int_{-1}^1 \mathrm{d}r \int_{-1}^1 \frac{\operatorname{Artanh}(r)}{r}\mathrm{d}r$$

an. Für alle $x \in (-1,1)$ gilt die Gleichung $\operatorname{Artanh}(x) = \frac{1}{2}(\ln(1+x) - \ln(1-x))$, diese Gleichung findet man zum Beispiel in [BS01, Formel 2.213]. Es gilt

$$\int_{-1}^1 \frac{\ln(1+x)}{x}\mathrm{d}x \leq \int_{-1}^1 \frac{x}{x} \leq 2,$$

denn es ist $\ln(1+x) \leq x$ für alle $x > -1$ und es gilt

$$2 \geq \int_{-1}^1 \frac{\ln(1+r)}{r}\mathrm{d}r = \int_{-1}^1 \frac{-\ln(1-x)}{x}\mathrm{d}x$$

nach der Substitution $x = -r$. Dies zeigt

$$\int_\alpha^\beta \sqrt{\frac{\operatorname{Artanh}(r)}{r}}\,\mathrm{d}r \leq \int_{-1}^1 \mathrm{d}r \int_{-1}^1 \frac{\operatorname{Artanh}(r)}{r}\mathrm{d}r \leq 4 < \infty.$$

□

Lemma C.0.7. Sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix mit nichtnegativen Einträgen auf der Diagonalen. Ist A schwach diagonaldominant, das heißt, es gilt

$$\sum_{i=1}^n |A(i,j)| \leq A(i,i)$$

für alle $i \in \{1,\dots,n\}$, dann ist A positiv semidefinit.

Beweis. Da A symmetrisch ist, hat A nur reelle Eigenwerte. Mithilfe des Satzes über die Gerschgorin-Kreise [HJ13, Theorem 6.1.1] folgt dann aus der schwachen Diagonaldominanz und den nichtnegativen Diagonaleinträgen schon die Nichtnegativität der Eigenwerte. Dies zeigt die positive Semidefinitheit. □

Lemma C.0.8. Die Abbildung

$$G\colon [0,\infty)\times[0,\infty) \to \mathbb{R},\ G(a,b) = \frac{a+b}{2} + \frac{1}{2}\inf_{r,s>0} \Lambda(r,s)\left(\frac{a}{r} + \frac{b}{s}\right)$$

ist monoton, das heißt, für alle $0 \leq c \leq a$ und $0 \leq d \leq b$ gilt $G(c,d) \leq G(a,b)$. Für alle

$\alpha > 0$ gilt $G(\alpha a, \alpha b) = \alpha G(a,b)$ und ferner erfüllt sie für alle $a, b \geq 0$ die Ungleichung $G(a,b) \geq 2\sqrt{ab}$.

Beweis. Seien $0 \leq c \leq a$ und $0 \leq d \leq b$ beliebig, dann existieren zu $\varepsilon > 0$ beliebig zwei Zahlen $r, s > 0$ derart, dass $\frac{a+b}{2} + \frac{1}{2}\Lambda(r,s)\left(\frac{a}{r} + \frac{b}{s}\right) < G(a,b) + \varepsilon$ ist. Wir schließen

$$G(c,d) \leq \frac{c+d}{2} + \frac{1}{2}\Lambda(r,s)\left(\frac{c}{r} + \frac{d}{s}\right) \geq \frac{a+b}{2} + \frac{1}{2}\Lambda(r,s)\left(\frac{a}{r} + \frac{b}{s}\right) < G(a,b) + \varepsilon$$

und da ε beliebig ist, folgt $G(c,d) \leq G(a,b)$. Ferner gilt

$$G(a,b)+\varepsilon > \frac{a+b}{2}+\frac{1}{2}\Lambda(r,s)\left(\frac{a}{r}+\frac{b}{s}\right) \geq \sqrt{ab}+\Lambda(r,s)\left(\frac{as+br}{2sr}\right) \geq \sqrt{ab}+\sqrt{rs}\frac{\sqrt{ab}}{\sqrt{rs}} \geq 2\sqrt{ab}$$

nach der Ungleichung vom arithmetischen und geometrischen Mittel sowie der Eigenschaft (vi) aus Bemerkung 4.1.5 über das logarithmische Mittel. Dies zeigt die gewünschte Ungleichung. Die fehlende Eigenschaft folgt direkt aus der Definition von G. □

Lemma C.0.9. Wir betrachten die in Beispiel 3.4.1 vorgestellte Diskretisierung der Fokker-Planck-Gleichung. Sei $V \in C^2[0,1]$ und erfülle $\inf_{x\in[0,1]} V''(x) \geq \hat{\lambda} \geq 0$. Unter diesen Voraussetzungen gilt für jedes $n \in \mathbb{N}$ und jedes $i \in \{2, \ldots, n-1\}$ die Ungleichung

$$\frac{\pi_i^2}{\pi_{i-1}\pi_{i+1}} \geq e^{\frac{\hat{\lambda}}{n^2}}.$$

Diese impliziert dann

$$-\frac{\sqrt{\pi_{i-1}}\sqrt{\pi_{i+1}}}{\pi_i} \geq -e^{-\frac{\hat{\lambda}}{2n^2}}$$

für alle $i \in \{2, \ldots, n-1\}$.

Beweis. Sei $n \in \mathbb{N}$ beliebig. Wir erinnern an die Definition der π_i, es gilt

$$\pi_i = \int_{x_{i-1}^n}^{x_i^n} u_\infty(x)\mathrm{d}x$$

für $x_i^n = \frac{i}{n}$ und $u_\infty(x) = ce^{-V(x)}$. Die Voraussetzungen implizieren, dass V auf dem Intervall $[0,1]$ konvex ist. Sei $i \in \{2, \ldots, n-1\}$, dann existiert eine eindeutig bestimmte Parabel der Form $\mathrm{par}(x) = \frac{\hat{\lambda}}{2}x^2 + px + q$, die V an den Stellen x_{i-1}^n und x_i^n schneidet. Aufgrund der unteren Schranke an die zweite Ableitung von V gilt $\mathrm{par} \geq V$ beziehungsweise $e^{-V} \geq e^{-par}$ in $[x_{i-1}^n, x_i^n]$ und $\mathrm{par} \leq V$ beziehungsweise $e^{-V} \leq e^{-par}$ in $[0,1] \setminus [x_{i-1}^n, x_i^n]$. Diese

Abschätzungen an e^{-V} implizieren Abschätzungen an die Konstanten π_i. Es gilt

$$\pi_{i+1} = \int_{x_i^n}^{x_{i+1}^n} ce^{-V(x)}\mathrm{d}x \leq c\int_{\frac{i+1}{n}}^{\frac{i}{n}} e^{-\frac{\hat{\lambda}}{2}x^2-px-q}\mathrm{d}x = c\int_{\frac{1}{2n}}^{\frac{3}{2n}} e^{-\frac{\hat{\lambda}}{2}\left(r+\frac{i-\frac{1}{2}}{n}\right)^2-p\left(r+\frac{i-\frac{1}{2}}{n}\right)-q}\mathrm{d}r$$

nach der Substitution $x = r + \frac{i-\frac{1}{2}}{n}$. Mutatis mutandis folgen auch die Abschätzungen

$$\pi_i \geq c\int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-\frac{\hat{\lambda}}{2}\left(r+\frac{i-\frac{1}{2}}{n}\right)^2-p\left(r+\frac{i-\frac{1}{2}}{n}\right)-q}\mathrm{d}r$$

sowie

$$\pi_{i-1} \leq c\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} e^{-\frac{\hat{\lambda}}{2}\left(r+\frac{i-\frac{1}{2}}{n}\right)^2-p\left(r+\frac{i-\frac{1}{2}}{n}\right)-q}\mathrm{d}r.$$

Bezeichne g das Polynom, definiert durch $g(r) = \frac{\hat{\lambda}}{2}\left(r+\frac{i-\frac{1}{2}}{n}\right)^2 + p\left(r+\frac{i-\frac{1}{2}}{n}\right) + q$. Dieses Polynom hat den führenden Koeffizienten $\frac{\hat{\lambda}}{2}$. Hinreichend für die Gültigkeit der Behauptung des Lemmas ist dann die Ungleichung

$$\frac{\left(\int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-g(r)}\mathrm{d}r\right)^2}{\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} e^{-g(r)}\mathrm{d}r \int_{\frac{1}{2n}}^{\frac{3}{2n}} e^{-g(r)}\mathrm{d}r} \geq e^{\frac{\hat{\lambda}}{n^2}}.$$

Für den Beweis dieser Ungleichung wenden wir die Jensensche Unleichung auf die drei verschiedenen Integrale an. Es gilt

$$\int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-g(r)}\mathrm{d}r \geq \frac{1}{n}\exp\left(n\int_{-\frac{1}{2n}}^{\frac{1}{2n}} -g(r)\mathrm{d}r\right)$$

sowie

$$-\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} e^{-g(r)}\mathrm{d}r \leq -\frac{1}{n}\exp\left(n\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} -g(r)\mathrm{d}r\right)$$

und

$$-\int_{\frac{1}{2n}}^{\frac{3}{2n}} e^{-g(r)}\mathrm{d}r \leq -\frac{1}{n}\exp\left(n\int_{\frac{1}{2n}}^{\frac{3}{2n}} -g(r)\mathrm{d}r\right).$$

Insgesamt schließen wir

$$
\begin{aligned}
\frac{\pi^2}{\pi_{i-1}\pi_{i+1}} &\geq \frac{\left(\int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-g(r)}\mathrm{d}r\right)^2}{\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} e^{-g(r)}\mathrm{d}r \int_{\frac{1}{2n}}^{\frac{3}{2n}} e^{-g(r)}\mathrm{d}r} \\
&= \frac{\left(\int_{-\frac{1}{2n}}^{\frac{1}{2n}} e^{-g(r)}\mathrm{d}r\right)^2}{\left(-\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} e^{-g(r)}\mathrm{d}r\right)\left(-\int_{\frac{1}{2n}}^{\frac{3}{2n}} e^{-g(r)}\mathrm{d}r\right)} \\
&\geq \frac{\left(\frac{1}{n}\exp\left(-n\int_{-\frac{1}{2n}}^{\frac{1}{2n}} g(r)\mathrm{d}r\right)\right)^2}{\left(-\frac{1}{n}\exp\left(-n\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} g(r)\mathrm{d}r\right)\right)\left(-\frac{1}{n}\exp\left(-n\int_{\frac{1}{2n}}^{\frac{3}{2n}} g(r)\mathrm{d}r\right)\right)} \\
&= \exp\left(n\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} g(r)\mathrm{d}r + n\int_{\frac{1}{2n}}^{\frac{3}{2n}} g(r)\mathrm{d}r - 2n\int_{-\frac{1}{2n}}^{\frac{1}{2n}} g(r)\mathrm{d}r\right).
\end{aligned}
$$

In den Integralen des letzten Terms spielen nur die Terme der Ordnung r^2 eine Rolle. Die linearen Terme summieren sich aufgrund der Symmetrie zu null und die konstanten Terme heben sich ebenso auf. Eine weitere Rechnung liefert dann

$$
\frac{\pi_i^2}{\pi_{i-1}\pi_{i+1}} \geq \exp\left(n\int_{-\frac{3}{2n}}^{-\frac{1}{2n}} g(r)\mathrm{d}r + n\int_{\frac{1}{2n}}^{\frac{3}{2n}} g(r)\mathrm{d}r - 2n\int_{-\frac{1}{2n}}^{\frac{1}{2n}} g(r)\mathrm{d}r\right) = \exp\left(\frac{\lambda}{n^2}\right).
$$

Dies zeigt die gewünschte Behauptung. $\square$

Literaturverzeichnis

[AC08] ALESSIO FIGALLI ; CÉDRIC VILLANI: *Optimal Transport and Cuvrature*. http://cedricvillani.org/wp-content/uploads/2012/08/P12.CIME_.pdf#page=2&zoom=auto,-265,306. Version: 2008

[AGS08] AMBROSIO, Luigi ; GIGLI, Nicola ; SAVARE, Giuseppe: *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. 2. Birkhäuser Basel, 2008 (Lectures in Mathematics. ETH Zürich). //www.springer.com/de/book/9783764387211. – ISBN 978–3–7643–8721–1

[Ale09] ALEXANDER GRIGORYAN: *Analysis on Graphs*. https://www.math.uni-bielefeld.de/~grigor/aglect.pdf. Version: Oktober 2009

[ARM16] AL REDA, F ; MAURY, B: *Interpretation of Finite Volume discretization schemes for the Fokker Planck equation as gradient flows for the discrete Wasserstein distance*. https://hal.archives-ouvertes.fr/hal-01288983. Version: März 2016

[Beh13] BEHRENDS, Ehrhard: *Markovprozesse und stochastische Differentialgleichungen: vom Zufallsspaziergang zur Black-Scholes-Formel*. Wiesbaden : Springer Spektrum, 2013 (Lehrbuch). – ISBN 978–3–658–00988–5 978–3–658–00987–8. – OCLC: 826589521

[BGL14] BAKRY, Dominique ; GENTIL, Ivan ; LEDOUX, Michel: *Analysis and geometry of Markov diffusion operators*. Cham : Springer, 2014 (Grundlehren der mathematischen Wissenschaften 348). – ISBN 978–3–319–00227–9 978–3–319–00226–2. – OCLC: 867136831

[Bre99] BREMAUD, Pierre: *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. New York, NY : Springer, 1999 (Texts in applied mathematics 31). – ISBN 978–1–4419–3131–3 978–0–387–98509–1. – OCLC: 245810628

[BS01] BRONSTEJN, Ilja N. (Hrsg.) ; SEMENDJAEV, Konstantin A. (Hrsg.): *Taschenbuch der Mathematik*. 5., überarb. und erw. Aufl., unveränd. Nachdr. Thun : Deutsch, 2001. – ISBN 978–3–8171–2005–5 978–3–8171–2015–4. – OCLC: 64650632

[Car72] CARLSON, B. C.: The Logarithmic Mean. In: *The American Mathematical Monthly* 79 (1972), Nr. 6, 615–618. http://dx.doi.org/10.2307/2317088. – DOI 10.2307/2317088. – ISSN 0002–9890

[Car13] CARMO, Manfredo P. d.: *Riemannian Geometry*. Birkhäuser Boston, 2013. – ISBN 978–0–8176–3490–2. – Google-Books-ID: ct91XCWkWEUC

[Cha75] CHAKRAVARTI, I. M.: On a characterization of irreducibility of a non-negative matrix. In: *Linear Algebra and its Applications* 10 (1975), April, Nr. 2, 103–109. http://dx.doi.org/10.1016/0024-3795(75)90002-6. – DOI 10.1016/0024–3795(75)90002–6. – ISSN 0024–3795

[CHLZ12] CHOW, Shui-Nee ; HUANG, Wen ; LI, Yao ; ZHOU, Haomin: Fokker–Planck Equations for a Free Energy Functional or Markov Process on a Graph. In: *Archive for Rational Mechanics and Analysis* 203 (2012), März, Nr. 3, 969–1008. http://dx.doi.org/10.1007/s00205-011-0471-6. – DOI 10.1007/s00205–011–0471–6. – ISSN 0003–9527, 1432–0673

[EF16] ERBAR, Matthias ; FATHI, Max: Poincar\'e, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. In: *arXiv:1612.00514 [math]* (2016), Dezember. http://arxiv.org/abs/1612.00514. – arXiv: 1612.00514

[EM12] ERBAR, Matthias ; MAAS, Jan: Ricci Curvature of Finite Markov Chains via Convexity of the Entropy. In: *Archive for Rational Mechanics and Analysis* 206 (2012), Dezember, Nr. 3, 997–1038. http://dx.doi.org/10.1007/s00205-012-0554-z. – DOI 10.1007/s00205–012–0554–z. – ISSN 0003–9527, 1432–0673

[EMT15] ERBAR, Matthias ; MAAS, Jan ; TETALI, Prasad: Discrete Ricci Curvature bounds for Bernoulli-Laplace and Random Transposition models. In: *Annales de la faculté des sciences de Toulouse Mathématiques* 24 (2015), Nr. 4, 781–800. http://dx.doi.org/10.5802/afst.1464. – DOI 10.5802/afst.1464. – ISSN 0240–2963, 2258–7519

[EN06] ENGEL, Klaus-Jochen ; NAGEL, R.: *A short course on operator semigroups*. New York, N.Y : Springer, 2006 (Universitext). – ISBN 978–0–387–31341–2

[Erb14] ERBAR, Matthias: Gradient flows of the entropy for jump processes. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 50 (2014), August, Nr. 3, 920–945. http://dx.doi.org/10.1214/12-AIHP537. – DOI 10.1214/12–AIHP537. – ISSN 0246–0203

[FM16] FATHI, Max ; MAAS, Jan: Entropic Ricci curvature bounds for discrete interacting systems. In: *The Annals of Applied Probability* 26 (2016), Juni, Nr. 3, 1774–1806. http://dx.doi.org/10.1214/15-AAP1133. – DOI 10.1214/15–AAP1133. – ISSN 1050–5164, 2168–8737

[Gri09] GRIGORYAN, Alexander: *Heat kernel and analysis on manifolds*. Providence, R.I. : American mathematical Society, 2009. – ISBN 978–0–8218–9393–7. – OCLC: 898709479

[HJ13] HORN, Roger A. ; JOHNSON, Charles R.: *Matrix analysis*. Second edition. New York, NY : Cambridge University Press, 2013. – ISBN 978–0–521–54823–6 978–0–521–83940–2. – OCLC: 802400841

[Jü16] JÜNGEL, Ansgar: *Entropy methods for diffusive partial differential equations*. 2016 http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1251402. – ISBN 978–3–319–34219–1 978–3–319–34218–4. – OCLC: 951975239

[JKO98] JORDAN RICHARD ; KINDERLEHRER DAVID ; OTTO FELIX: *The Variational Formulation Of The Fokker–Planck Equation*. http://www-dimat.unipv.it/savare/Ravello2010/JKO.pdf. Version: Januar 1998

[KB13] KNABNER, Peter ; BARTH, Wolf: *Lineare Algebra: Grundlagen und Anwendungen*. Berlin : Springer Spektrum, 2013 (Springer-Lehrbuch). – ISBN 978–3–642–32185–6 978–3–642–32186–3. – OCLC: 823231045

[Lee97] LEE, John M.: *Riemannian manifolds: an introduction to curvature*. New York : Springer, 1997 (Graduate texts in mathematics 176). – ISBN 978–0–387–98271–7 978–0–387–98322–6

[Lee13] LEE, John M.: *Introduction to smooth manifolds*. 2nd ed. New York ; London : Springer, 2013 (Graduate texts in mathematics 218). – ISBN 978–1–4419–9981–8 978–1–4419–9982–5. – OCLC: ocn800646950

[LN09] LUIGI AMBROSIO ; NICOLA GIGLI: *A user's guide to optimal transport*. https://www.math.umd.edu/~yanir/OT/AmbrosioGigliDec2011.pdf. Version: 2009

[Maa11] MAAS, Jan: Gradient flows of the entropy for finite Markov chains. In: *Journal of Functional Analysis* 261 (2011), Oktober, Nr. 8, 2250–2292. http://dx.doi.org/10.1016/j.jfa.2011.06.009. – DOI 10.1016/j.jfa.2011.06.009. – ISSN 0022–1236

[Mat07] MATTHES, Daniel: *Lecture Notes on the Course "Entropy Methods and Related Functional Inequalities"*. http://www-m8.ma.tum.de/personen/matthes/papers/lecpavia.pdf. Version: 2007

[Mie13] MIELKE, Alexander: Geodesic convexity of the relative entropy in reversible Markov chains. In: *Calculus of Variations and Partial Differential Equations* 48 (2013), September, Nr. 1-2, 1–31. http://dx.doi.org/10.1007/s00526-012-0538-8. – DOI 10.1007/s00526–012–0538–8. – ISSN 0944–2669, 1432–0835

[Nor98] NORRIS, J. R.: *Markov chains*. 1st pbk. ed. Cambridge, UK ; New York : Cambridge University Press, 1998 (Cambridge series on statistical and probabilistic mathematics). – ISBN 978–0–521–48181–6 978–0–521–63396–3

[NR17] NAJMAN, Laurent (Hrsg.) ; ROMON, Pascal (Hrsg.): *Modern approaches to discrete curvature*. Cham : Springer, 2017 (Lecture notes in mathematics 2184). – ISBN 978–3–319–58002–9 978–3–319–58001–2. – OCLC: 1008921329

[Ott01] OTTO, Felix: The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. In: *Communications in Partial Differential Equations* 26 (2001), Januar, Nr. 1-2, 101–174. http://dx.doi.org/10.1081/PDE-100002243. – DOI 10.1081/PDE–100002243. – ISSN 0360–5302

[PW10] PRÜSS, Jan ; WILKE, Mathias: *Gewöhnliche Differentialgleichungen und dynamische Systeme*. Basel : Birkhäuser [u.a.], 2010 (Grundstudium Mathematik). – ISBN 978–3–0348–0001–3 978–3–0348–0002–0. – OCLC: 699714497

[Ris96] RISKEN, H.: *The Fokker-Planck equation: methods of solution and applications*. 2nd ed. New York : Springer-Verlag, 1996 (Springer series in synergetics v. 18). – ISBN 978–3–540–61530–9

[Sch05] SCHILLING, René L.: *Measures, integrals and martingales*. Cambridge ; New York : Cambridge University Press, 2005. – ISBN 978–0–521–85015–5 978–0–521–61525–9. – OCLC: ocm60611040

[Vil09] VILLANI, Cédric: *Optimal transport: old and new*. Berlin : Springer, 2009 (Grundlehren der mathematischen Wissenschaften 338). – ISBN 978–3–540–71049–3. – OCLC: ocn244421231

Name: Lukas Niebel Matrikelnummer: 886819

Erklärung

Ich erkläre, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ulm, den 25. Februar 2018 . . . .

Lukas Niebel