1 Anwendungen der Theorie des optimalen Transport

Aus der Übungsgruppe wissen wir, dass einige Teile der Theorie des optimalen Transport durch diskrete Probleme motiviert sind, insbesondere die Kantorovich Dualität. Diese ist motiviert durch Probleme der linearen Optimierung. Die lineare Optimierung hat vor allem wirtschaftswissenschaftliche Anwendungen in Spieltheorie oder bei der Produktionsplanung. Es gibt aber auch andere vielseitige Anwendungen wie zum Beispiel "Mischprobleme". Dabei soll aus verschiedenen Stoffen eine Mischung erzeugt, die gewisse Mindestwerte erfüllt (Konzentrationen etc..). Natürlich gibt es auch eine vielzahl an Anwendungen des allgemeinen Problems, also des stetigen Falls, nicht nur innerhalb der Mathematik. Zunächst wollen wir eine klassische Anwendung vorstellen, die auf Robert McCann zurück geht. Er beschreibt ein Modell für interagierende Gase in einem Equilibriumszustand.

Definition 1.1.
Wir beschreiben alle möglichen Zustände, die ein n-dimensionales Gas annehmen kann, mit einer Menge von Wahrscheinlichkeitsdichten

$$\mathcal{D} := \left\{ f : \mathbb{R}^n \to \mathbb{R} : f \in L^1(\mathbb{R}^n),\, f \geq 0 \text{ a.e. },\, \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 1 \right\}. \tag{1}$$

Zu jeder solchen Verteilung assoziieren wir eine Energie via

$$E[f] := \int_{\mathbb{R}^n} f^\gamma(x)\,\mathrm{d}x + \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x)V(x-y)f(y)\,\mathrm{d}x\,\mathrm{d}y \tag{2}$$

für eine Konstante $\gamma > 1$ und eine nicht negative strikt konvexe Funktion $V : \mathbb{R}^n \to \mathbb{R}$. Das 3. Gesetz von Newton besagt, dass wenn ein Körper auf einen anderen Körper wirkt, so wirkt dieser in umgekehrter Richtung auf den ersten Körper ein. Also ist es sinnvoll anzunehmen, dass $V(x) = V(-x)$.

Der erste Term in dieser Definition stellt die innere Energie der Gasmenge dar:

$$E_1(f) := \int_{\mathbb{R}^n} f^\gamma(x)\,\mathrm{d}x,$$

wobei man sich $U(f) = f^\gamma$ als Energiedichte des Zustands, also als ein "Integral über den Druck" vorstellen kann. Der zweite Teil korrespondiert mit der Energie, die durch die Interaktion des Gases mit sich selber entsteht, also

$$E_2(f) := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x)V(x-y)f(y)\,\mathrm{d}x\,\mathrm{d}y.$$

In einem Equilibrium-Zustand sollte die Gesamtenergie E minimal sein. Unser Ziel ist es also einen Zustand, bzw. eine Verteilung, des Gases $f \in \mathcal{D}$ zu finden, welche E minimiert. Ein wichtiges Werkzeug, um Minimierer eines Funktionals zu finden, ist die Konvexität. Wir schauen uns also zwei Zustände $f_0, f_1 \in \mathcal{D}$ des Gases an. Dabei stellt sich raus, dass E im allgemeinen nicht im klassischen Sinne konvex ist, also im Hinblick auf die Interpolation

$$t \mapsto tf_0 + (1-t)f_1$$

für $t \in [0,1]$. McCann betrachtete also die folgende Interpolation

Definition 1.2 (Verschiebungsinterpolation)**.**
Es seien $f_0, f_1 \in \mathcal{D}$ und $\mu_0 = f_0(x)\,\mathrm{d}x$ und $\mu_1 = f_1(y)\,\mathrm{d}y$. Dann gibt es nach Korollar 3.22 eine konvexe Funktion $\varphi : \mathbb{R}^n \to \mathbb{R}$, die $(\nabla\varphi)_\#(\mu) = \nu$ erfüllt. Wir definieren, dann die Verschiebungsinterpolierte $f_t \in \mathcal{D}$ als die Dichte des Maßes

$$[\mu_0, \mu_1]_t := ((1-t)\mathrm{I}_n + t\nabla\varphi)_\#\,\mu_0$$

für jedes $t \in [0,1]$.

Diese Interpolation ist nichtlinear, erzeugt jedoch eine Konvexitätseigenschaft für E. Wir wollen zunächst einige Eigenschaften der Verschiebungsinterpolierten zeigen, insbesondere, dass f_t tatsächlich existiert.

Proposition 1.3.
Seien $t, t', s \in [0,1]$ und μ_1, μ_0 zwei Wahrscheinlichkeitsmaße, sodass μ_0 absolut stetig ist. Dann gilt

1. $[\mu_0, \mu_1]_t = [\mu_1, \mu_0]_{1-t}$
2. $\Big[[\mu_0, \mu_1]_t, [\mu_0, \mu_1]_{t'}\Big]_s = [\mu_0, \mu_1]_{(1-s)t+st'}$
3. Es ist auch $[\mu_0, \mu_1]_t$ absolut stetig für alle $t \in (0,1)$ mit Dichte $f_t \in \mathcal{D}$.
4. $w_2(\mu_0, \mu_t) = t w_2(\mu_0, \mu_1)$ für jedes t

Beweis. 1./2. sind Übungsaufgaben, wobei man bemerkt, dass wegen $\mu_1 = (\nabla\varphi)_\#\mu_0$ auch $\mu_0 = (\nabla\varphi^*)_\#(\mu_1)$ gilt. Außerdem ist die Inverse zu $\nabla\varphi$ gegeben

durch $\nabla\varphi^*$. Dies folgt aus $y \in \partial\varphi(x) \leftrightarrow x \in \partial\varphi^*(y)$.
Für 3. definieren wir $\varphi_t(x) := t\varphi(x) + (1-t)|x|^2/2$. Dann gilt

$$\langle \nabla\varphi_t(x) - \nabla\varphi_t(y), x - y\rangle \geq (1-t)|x-y|^2$$

was

$$|\nabla\varphi_t(x) - \nabla\varphi_t(y)| \geq (1-t)|x-y|$$

impliziert. Außerdem ist φ_t uniform konvex, also strikt konvex, also φ_t^* fast überall differenzierbar und $\nabla\varphi_t$ ist Lipschitz stetig mit konstante kleiner gleich $(1-t)^{-1}$. Insbesondere gilt also für jede Nullmenge $A \subset \mathbb{R}^n$, dass auch $\nabla\varphi_t^*(A)$ eine Nullmenge ist. Daraus schließen wir

$$\mu_t[A] = \mu_0[\partial\varphi_t^*(A)] = \mu_0[\nabla\varphi_t^*(A)] = 0,$$

da μ_0 absolut stetig war. Hier haben wir Aufgabe 16 verwendet in der ersten Gleichheit.
Für 4. berechnen wir

$$\int_{\mathbb{R}^n} |x - ((1-t)x + t\nabla\varphi(x))|^2 \,\mathrm{d}\mu_0(x) = t^2 \int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2 \,\mathrm{d}\mu_0(x).$$

Da $T = \nabla\varphi$, $T_\#\mu_0 = \mu_1$ erfüllt und optimal ist, folgt die Aussage durch Wurzelziehen. $\square$

Nun zeigen wir, dass E tatsächlich konvex ist.

Theorem 1.4.
Die Abbildung $t \mapsto E[f_t]$ ist konvex auf $[0,1]$ und strikt konvex außer μ_1 ist eine Translation von μ_0.

Beweis. Zunächst gilt

$$\begin{aligned} E_2[f_t] &= \int_{\mathbb{R}^n\times\mathbb{R}^n} f_t(x)V(x-y)f_t(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} f_0(y)V((1-t)(x-y) + t(\nabla\varphi(x) - \nabla\varphi(y))f_0(x)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Also gilt aufgrund der Konvexität von V die Konvexität von E_2. Da V strikt konvex ist, gilt auch die strikte Konvexität, außer

$$x - y = \nabla\varphi(x) - \nabla\varphi(y)$$

für fast alle $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. Also ist $x - \varphi(x)$ unabhängig von x. Das heißt $\nabla\varphi(x) = x + C$ für eine Konstante C und demnach muss μ_1 eine Translation von μ_0 sein. Nun benutzen wir die Monge-Ampére Gleichung für f_t

$$f_0(x) = f_t((1-t)x + \nabla\varphi(x))\det\Big((1-t)\mathrm{I}_n + tD^2\varphi(x)\Big)$$

für fast alle $x \in \mathbb{R}^n$. Dies ist ein Theorem von McCann das wir in der Übung besprochen haben. Dieses besagt auch dass für jede nichtnegative messbare Funktion U mit $U(0) = 0$

$$\int_{\mathbb{R}^n} U(f_t(x))\,\mathrm{d}x = \int_{\mathbb{R}^n} U\left(\frac{f_0(x)}{\det\left((1-t)\mathrm{I}_n + tD^2\varphi(x)\right)}\right)\det\Big((1-t)\mathrm{I}_n + tD^2\varphi(x)\Big)\,\mathrm{d}x \tag{3}$$

gilt. Wir berechnen nun für $U(f) = f^\gamma$

$$\begin{aligned} E_1(f_t) &= \int_{\mathbb{R}^n} f_t^\gamma(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^n} \left[\frac{f_0(x)}{\det\left((1-t)\mathrm{I}_n + tD^2\varphi(x)\right)}\right]^\gamma \det\Big((1-t)\mathrm{I}_n + tD^2\varphi(x)\Big)\,\mathrm{d}x. \end{aligned} \tag{4}$$

Wie betrachten also die positiv definite Matrix $A_t := \Big((1-t)\mathrm{I}_n + tD^2\varphi(x)\Big)$ für $0 \le t < 1$ und fast jedes x. Für positiv definite Matrizen A ist

$$A \mapsto \det(A)^{\frac{1}{n}}.$$

eine konkave Funktion. Also ist die Funktion $\alpha_x(t) := \det\Big((1-t)\mathrm{I}_n + tD^2\varphi(x)\Big)$ für fast alle x konkav. Wir definieren nun eine neue Funktion $\beta_x(t) := \alpha_x(t)^{n(1-\gamma)}$ und berechnen

$$\begin{aligned} \beta_x''(t) &= n(1-\gamma)(n - n\gamma - 1)\alpha_x(t)^{n(1-\gamma)-2}\alpha_x'(t)^2 + (n(1-\gamma)\alpha_x(t)^{n(1-\gamma)-1}\alpha_x''(t) \\ &\ge 0, \end{aligned}$$

da $\gamma > 1$. Also ist β_x konvex für fast alle x. Nun können wir (3) schreiben als

$$E_1(f_t) = \int_{\mathbb{R}^n} f_0(x)\beta_x(t)\,\mathrm{d}x,$$

woraus die Konvexität folgt. □

BEMERKUNG: Allgemeiner können wir die innere Energie auch mit Hilfe einer messbaren Funktion $U : \mathbb{R}_+ \to \mathbb{R} \cup \{\infty\}$ mit $U(0) = 0$ definieren, nämlich

$$E_1(f) := \int_{\mathbb{R}^n} U(f(x))\,\mathrm{d}x.$$

Um dann auch noch Konvexität zu erhalten, lässt sich zeigen, dass U gewissen Eigenschaften erfüllen muss, z.B. $r \mapsto r^n U(r^{-n})$ muss monoton

nicht wachsend sein auf $(0,+\infty)$. Typisch sind $U(r) = r^\gamma$, $\gamma > 1$ oder $U(r) = -r^\gamma$, $\gamma \in [1-1/n, 1]$.
Wir bezeichnen die Menge der absolut stetigen Maße mit $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n) \subset \mathcal{P}(\mathbb{R}^n)$.
Es besteht eine 1 zu 1 Korrespondenz zwischen $\mathcal{P}_{\mathrm{ac}}$ und $\mathcal{D}$. Die Verschiebungskonvexität lässt sich auch auf ganze Teilmengen von $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n)$ ausweiten.

Definition 1.5.

1. Eine Teilmenge $\mathcal{F} \subset \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n)$ heißt *verschiebungskonvex*, wenn für alle $\mu, \nu \in \mathcal{F}$ und alle $t \in [0,1]$ das verschiebungsinterpolierte Maß $[\mu,\nu]_t$ wieder in $\mathcal{F}$ liegt.

2. Ein Funktional $F : \mathcal{F} \to \mathbb{R}$ auf einer verschiebungskonvexen Menge $\mathcal{F}$, heißt *verschiebungskonvex*, falls für $\mu_0, \mu_1 \in \mathcal{F}$ und für die zugehörige Verschiebungsinterpolierte $(\mu_t := [\mu_0,\mu_1]_t)_{t\in[0,1]}$ die Abbildung

$$t \mapsto F(\mu_t)$$

   konvex ist auf $[0,1]$.

3. Für nicht absolut stetige Maße lässt sich auch eine Verschiebungskonvexität definieren. Dabei nehmen wir $\sigma_t : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ als $\sigma_t(x,y) = \Phi_{x,t}(y)$ aus Aufgabe 23. Eine Teilmenge $\mathcal{F} \subset \mathcal{P}_2(\mathbb{R}^n)$ heißt verschiebungskonvex, falls für alle $\mu_0, \mu_1 \in \mathcal{F}$ und alle optimalen $\pi \in \Pi(\mu_0,\mu_1)$ bezüglich quadratischer Kosten, das Maß $\mu_t := \sigma_t \# \pi$ in $\mathcal{F}$ liegt für alle $t \in [0,1]$.

Mit Definition 1.5.3 lässt sich diese Definition auch auf allgemeine strikt konvexe Kostenfunktionen $c(x-y)$ verallgemeinern, indem wir einfach die optimalen π bezüglich der neuen Kosten wählen. Dabei muss aber zusätzlich angenommen werden, dass stets die totalen Kosten endlich sind.