Optimaler Transport

Marzena Franek

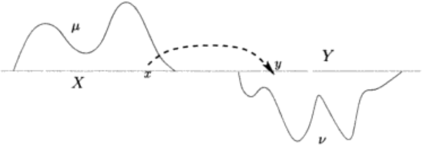
Institut für Numerische und Angewandte Mathematik

Skiseminar Februar 2009

Monges Transportproblem

Das **Transportproblem** wurde als erstes von dem französischen Mathematiker Gespard Monge im Jahre 1781 veröffentlicht.

Idee: Wir wollen einen Haufen Sand (Remblais) in ein Loch (Dèblais) der gleichen Größe transportieren. Dies wollen wir mit minimalem Aufwand tun.



Mathematische Formulierung

Seien X, Y zwei metrische Räume.
$P(X)$ und $P(Y)$ seien die Mengen aller Wahrscheinlichkeitsmaße auf X bzw. Y.
μ sei die gegebene Dichte auf dem Remblais (modelliert den Sandhaufen) und ν sei die Dichte auf dem Dèblais (modelliert das Loch). $\mu \in P(X)$ und $\nu \in P(Y)$. Zusätzlich gelte die Bedingung

$$\mu(X) = \nu(Y) < \infty,$$

d.h. μ und ν sind endlich mit gleicher Gesamtmasse. $\mu(X)$ gibt an, wieviel Sand in X liegt und $\nu(Y)$ gibt an, wieviel Masse nach Y bewegt werden kann.

Mathematische Formulierung

push forward

$r : X \to Y$ sei eine injektive Abbildung. Das **push-forward** von μ durch die Abbildung r ist dann definiert durch:

$$r_{\#}\mu(B) := \mu(r^{-1}(B)) = \nu(B) \quad \forall B \subset Y \text{messbar}$$

Allgemeiner können wir sagen: r **drückt** μ **in das Maß** ν. Es ergibt sich dann

$$\int_X f(r(x))d\mu(x) = \int_Y f(y)dr_{\#}\mu(y)$$

für jede beschränkte (oder $r_{\#}\mu$- integrierbare) Borel-Funktion $f : X \to \mathbb{R}$.

Mathematische Formulierung

Wir wollen wissen, welche Kosten beim Transport, d.h. beim „drücken" der Masse von der Stelle x nach y, entstehen. Als Mess- oder Kostenfunktion betrachten wir die Funktion

$$c : X \times Y \to [0, +\infty].$$

Ziel ist es, die Kosten zu minimieren:

Monge-Problem

$$\inf_r I(r) = \inf_r \left\{ \int_X c(x, r(x)) d\mu(x) | \; r_{\#}\mu = \nu \right\}$$

Kostenfuntkion

Monge beschäftigte sich mit der Kostenfunktion

linear

$$c(x,y) = |x-y|,$$

d.h. mit der Arbeit, die proportional ist zur euklidischen Distanz.
Eine weitere mögliche Kostenfunktion wäre

quadratisch

$$c(x,y) = |x-y|^2,$$

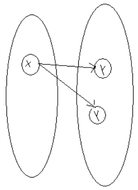
die den quadratischen Abstand beschreibt.

Existenz und Eindeutigkeit einer Lösung

Es kann weder die **Existenz** einer Lösung noch die **Eindeutigkeit** für Monges Transportproblem erwartet werden.

Existenz

Sei $\mu = \delta_0$ ein Dirac-delta-Maß in 0, $\nu = \left(\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}}\right)$. Dann existiert keine Lösung des Monge-Problems, da es keine Abbildung r gibt mit $r_{\#}\mu = \nu$. Monges Problem erlaubt keine Trennung der Massen, d.h an jedem Ort darf nur eine Einheit angeboten bzw. nachgefragt werden.



Existenz und Eindeutigkeit einer Lösung

Eindeutigkeit

Aufgrund der Nichtlinearität des Problems ist das Monge Problem eine schwer zu lösende Aufgabe.

$\Longrightarrow$ Kantorovich Formulierung, eine schwächere Formulierung

Kantorovich-Problem

1942 schlug Kantorovich das **relaxierte Problem** vor:

Anstatt einer Transport-Abbildung zu betrachten, schauen wir uns Transport-Pläne π an. Transport-Pläne sind Wahrscheinlichkeitsmaße auf dem Produktraum $X \times Y$. X und Y sind zwei kompakte Räume mit $\mu \in X$ und $\nu \in Y$. Wir betrachten nur Transport-Pläne $\pi \in P(X \times Y)$ mit Marginalen μ und ν, d.h.

$$\pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B)$$

mit kompakten Teilmengen $A \subseteq X$ und $B \subseteq Y$. Mit $\Pi(\mu, \nu)$ bezeichnen wir die Menge aller Transport-Pläne:

$$\Pi(\mu, \nu) = \{\pi \in P(X \times Y) | \ \pi(A \times Y) = \mu(A), \ \pi(X \times B) = \nu(B)\}$$

Kantorovich-Problem

Die Kostenfunktion c sei eine stetige Funktion auf dem Produktraum $X \times Y$. Wir können dann schreiben

$$\min_{\pi \in \Pi} J[\pi] = \min_{\pi \in \Pi} \left\{ \int_{X \times Y} c(x,y) d\pi(x,y) | \pi \in \Pi(\mu, \nu) \right\}$$

mit den Nebenbedingungen

$$\int_Y d\pi(x,y) = d\mu(x)$$

$$\int_X d\pi(x,y) = d\nu(y)$$

$d\pi(x,y)$ gibt den Betrag der Masse an, die von x nach y transportiert wird. All die Masse die vom Punkt x genommen wird, stimmt mit $d\mu(x)$ überein. All die Masse die nach y transportiert wird, stimmt mit $d\nu(y)$ überein.

Kantorovich-Problem

- Kantorovich hat die Existenz einer Lösung für dieses Problem gezeigt.
- Gangbo und McCann haben gezeigt, dass bei strikt konvexem c das relaxierte Problem und das ursprüngliche Problem die gleiche Lösung haben.

Das optimale Transport-Problem findet Anwendung in der Ökonomie

- Verarbeitung von Materialien mit möglichst wenig Restabfällen.
- Verteilung von Gütern auf mehrere Nachfrageorte.

Aber auch in der Bildverarbeitung

- Image warping
- Medizinische Bildverarbeitung

Duale Formulierung des Kantorovich-Problems

Jedes lineare Problem mit einer konvexen Bedingung hat eine duale Formulierung. Zur Kantorovich-Formulierung können wir also schreiben:

Sei c schwach unterhalbstetig, dann ist das Minimum der Kantorovich Formulierung äquivalent zu

$$\sup_{(\varphi,\psi)} D(\varphi,\psi) = \sup_{(\varphi,\psi)} \left(\int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y) \right)$$

mit $(\varphi,\psi) \in C_b^0(X) \times C_b^0(Y)$.
Folgende Nebenbedingung muss noch erfüllt sein

$$\varphi(x) + \psi(y) \leq c(x,y)$$

Duale Formulierung des Kantorovich-Problems

Wir bezeichnen mit Φ_c die Menge aller solcher Paare (φ, ψ), die die oberen Bedingungen erfüllt. Ohne Beweis können wir sagen

$$\inf_{\Pi(\mu,\nu)} J[\pi] = \sup_{\Phi_c} D(\varphi, \psi)$$

Anstatt nach einer Abbildung r zu suchen, suchen wir jetzt ein Paar (φ, ψ). Dieses Problem ist leichter zu lösen, aufgrund der Nichtlinearität des Monge-Problems.

Definition

Kantorovich-Wasserstein-Distanz

Sei (X, d) ein metrischer Raum.
Die $p-$**Kantorovich-Wasserstein-Distanz** zwischen zwei Maßen im Raum $P(X)$ ist definiert durch

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_X d(x,y)^p d\pi(x,y) \right)^{1/p}$$

Beispiel:

$$W_2(\mu, \nu) = \left(\inf_{\pi in \Pi(\mu,\nu)} \int_X |x-y|^2 d\pi(x,y) \right)^{\frac{1}{2}}$$

regularisierter optimaler Transport

Wir suchen die Lösung des folgenden Variationsproblems:

$$\mathcal{J}(u) = \frac{1}{2} W_2(\nu, u\mathcal{L}^d)^2 + \epsilon E(u) \to \min_{u\mathcal{L}^d \in P(\Omega)}$$

mit ν Wahrscheinlichkeitsmaß auf $\Omega \subset \mathbb{R}^d$ und u Wahrscheinlichkeitsdichte. $\mathcal{L}^d$ ist das Lebesque-Maß auf $\mathbb{R}^d$. Als typische Regularisierungsenergie betrachten wir

TV Regularisierung

$$E(u) = \sup_{g \in C_0^\infty(\Omega;\mathbb{R}^d), \|u\|_\infty \leq 1} \int_\Omega u \; \nabla \cdot g \; dx$$

Regularisierungsterme

Alternativ kann man auch folgende Regularisierungsterme verwenden

quadratische Regularisierung

$$E(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 \, dx$$

log Entropie Regularisierung

$$E(u) = \int_{\Omega} u \; \ln u \; dx$$

Formulierung von Benamou Brenier

Benamou und Brenier schlugen folgende zur L^2-Kantorovich-Wasserstein-Distanz alternative Formulierung vor:

$[0, T]$ sei ein festes Zeitintervall. Wir transferrieren das Problem in ein continuum mechanical framework.
$\rho(x, t) >= 1$ zeitabhängige Dichten.
$v(x, t) \in \mathbb{R}^N, t \in [0, T]$ Geschwindigkeitsfeld.
$x \in \mathbb{R}^N$.

Formulierung von Benamou Brenier

Proposition: Das Quadrat der L^2-Kantorovich-Distanz ist äquivalent zum Infimum von

$$T \int_{\mathbb{R}^N} \int_0^T \rho(x,t)|v(x,t)|^2 dxdt$$

unter den Nebenbedingungen

$$\partial_t \rho + \nabla \cdot (\rho v) = 0$$
$$\rho(0,\cdot) = \rho_0$$
$$\rho(T,\cdot) = \rho_T$$

Numerische Lösung

Wir wollen folgendes Problem minimieren:

$$\begin{aligned} \min_{\rho} \quad & \frac{1}{2}\int_0^1 \int_D \rho v^2 dxdt + \epsilon E(\rho(1)), \\ \text{unter} \quad & \partial_t \rho + \nabla \cdot (\rho v) = 0, \\ \text{mit} \quad & \rho(t=0) = \rho_0. \end{aligned}$$

Die Lagrange-Funktion ergibt sich dann zu

$$L = \epsilon E(\rho(1)) + \frac{1}{2}\int_0^1 \int_D \rho v^2 dxdt + \int_0^1 \int_D (\partial_t \rho + \nabla \cdot (\rho v))\lambda(x,t) dxdt.$$

Numerische Lösung

Die Optimalitätsbedingungen lauten

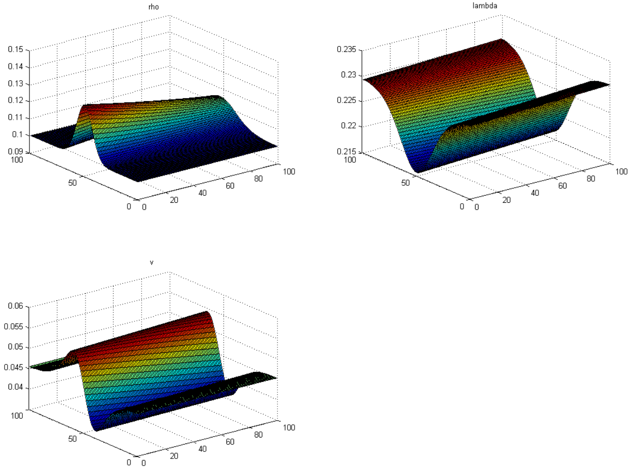
$$
\begin{aligned}
L_v =& \rho v - \nabla \lambda \rho = 0, \\
L_\rho =& \frac{1}{2} v^2 - \partial_t \lambda - \nabla \lambda v = 0, \\
L_\lambda =& \partial_t \rho + \nabla \cdot (\rho v) = 0.
\end{aligned}
$$

Eine weitere Nebenbedingung lautet: $\lambda(1) + \epsilon E'(\rho(1)) = 0$.

Numerische Lösung

Gradienten-Abstiegs-Verfahren:

1. Löse die Erhaltungsgleichung $\partial_t \rho^{k+1} + \nabla \cdot (\rho^{k+1} v^k) = 0$ mit v und ρ_0 gegeben $\Longrightarrow \rho_1$
2. Löse $\epsilon E'(u^{k+1}) + \lambda^{k+1}(1) = 0$ mit $u^{k+1} = \rho^{k+1}(t=1)$ nach $\lambda(1)$
3. Wir haben nun $\lambda(1)$ berechnet, d.h. wir haben den Zustand zum Endzeitpunkt gegeben. Löse rückwärts in der Zeit die adjungierte Gleichung:
$\frac{1}{2} \left| v^k \right|^2 - \partial_t \lambda^{k+1} - \nabla \lambda^{k+1} v^k = 0$
4. Mit dem neuen λ können wir nun ein neues v berechnen $v^{k+1}(1+\tau) = \tau v^k + \nabla \lambda^{k+1}$.


rho
0.15
0.14
0.13
0.12
0.11
0.1
0.09
100
50
0
0
20
40
60
80
100
lambda
0.235
0.23
0.225
0.22
0.215
100
50
0
0
20
40
60
80
100
v
0.06
0.055
0.05
0.045
0.04
100
50
0
0
20
40
60
80
100

Formulierung des Minimierungs-Problems

Kantorovich-Formulierung mit Regularisierungsterm

$$\frac{1}{2}\int_{\Omega\times\Omega} |x-y|^2\, d\Pi(x,y) + \epsilon E(u) \to \min_{\Pi,u}$$

unter

$$\int_{A\times\Omega} d\Pi(x,y) = \int_A d\nu(y)$$

$$\int_{\Omega\times A} d\Pi(x,y) = \int_A u(x)dx$$

für alle $A \subset \Omega$ meßbar, wo u eine Wahrscheinlichkeitsdichte im Raum E ist und Π ist eine Wahrscheinlichkeitsmaß auf $\Omega \times \Omega$

- Für einen konvexen Regularisierungsterm erhalten wir ein konvexes Problem
- Man erhält ein lineares Optimierungsproblem
- Die Dimension des Problems wird verdoppelt
- Formulierung ist nur effizient in einer Raumdimension

Vielen Dank für die Aufmerksamkeit!