STOCHASTIC MASS TRANSFER

Abstract

The theory of optimal transport (OT) has seen a tremendous development in the last 25 years with fascinating applications ranging from geometric and functional inequalities over PDEs and geometry to image analysis and statistics. In recent years, variants of the optimal transport problem with additional stochastic constraints have received increasing attention, e.g. weak optimal transport (WOT), entropic optimal transport (EOT), martingale optimal transport (MOT) and causal/adapted optimal transport (COT).

The aim of this lecture[1] is to serve as an introduction into the stochastic variants of the transport problem. After a quick recall of the classical OT problem we will start investigating the above mentioned probabilistic versions.

Frequently used notation

- X, Y denote Polish spaces
- For a Polish space X we denote the probability measures over X by $\mathcal{P}(\mathsf{X})$, the set of Borel measures by $\mathcal{M}(\mathsf{X})$, and the Borel sets by $\mathcal{B}(\mathsf{X})$.
- For a map $T : \mathsf{X} \to \mathsf{Y}$ and $\lambda \in \mathcal{P}(\mathsf{X})$ we denote the image measure of λ under T by $T(\lambda) = T_\sharp \lambda = \lambda \circ T^{-1}$
- The Lebesgue measure will be denoted by Leb.
- The set of all all couplings between two probability measures μ, ν will be denoted by $\mathsf{Cpl}(\mu, \nu)$.
- $C_b(\mathsf{X})$ continuous and bounded functions on $f : \mathsf{X} \to \mathbb{R}$.
- For integrable $f : \mathsf{X} \to \mathbb{R}$ and $\mu \in \mathcal{M}(\mathsf{X})$ we often write $\mu(f) := \int f d\mu$.

1. The optimal transport problem

In this section we will give a short introduction into the theory of optimal transport. This will serve as a benchmark or guidance for what to expect for the different stochastic variations of the transport problem we will consider in the next sections.

For reference and further reading we refer to the books [San15, AG13, Vil03].

1.1. **On how mass is transported.**

Definition 1.1. *A topological space (X, τ) is called Polish, iff it is separable and there exists a metric d metrizing τ s.t. (X, d) is a complete metric space.*

Let X, Y be Polish spaces and denote the set of probability measures by $\mathcal{P}(\mathsf{X}), \mathcal{P}(\mathsf{Y})$. Given two distributions of mass $\mu \in \mathcal{P}(\mathsf{X}), \nu \in \mathcal{P}(\mathsf{Y})$ we are interested in ways of *transporting mass* distributed according to μ into mass distributed according to ν. In mathematical terms:

Definition 1.2. *For a Borel function $T : \mathsf{X} \to \mathsf{Y}$ we define the push-forward of μ by T or the image measure of μ under T by*

$$T(\mu) := T_\sharp \mu = \mu \circ T^{-1},$$

i.e. $T(\mu)(A) = \mu(T^{-1}(A))$ for all $A \in \mathcal{B}(\mathsf{Y})$. If $T(\mu) = \nu$ we call T a transport map (or Monge transport) from μ to ν.

[1] These notes are based on earlier lectures / lecture notes of Julio Backhoff-Veraguas, Martin Huesmann and Gudmund Pammer.

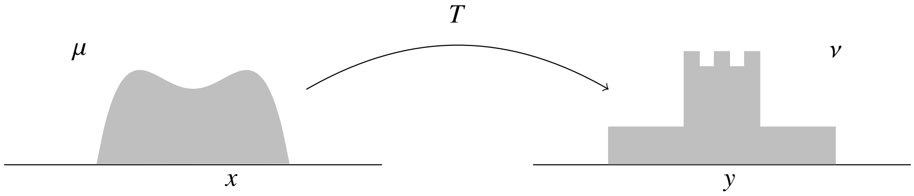

FIGURE 1. A possible transport from a distribution μ to a distribution ν via a map T.

This problem was first formulated by Gaspard Monge in 1781 in the article *"Sur la theorie des déblais et des remblais"* [Mon81] where he was interested in minimizing the transport cost of moving a pile of sand.

Remark 1.3. In $\mathsf{X} = \mathsf{Y} = \mathbb{R}^d$, if μ, ν have densities and T is regular enough, then T is a transport map between μ and ν iff

$$\det(DT)\frac{d\nu}{dx} \circ T = \frac{d\mu}{dx},$$

as follows from the change of variables formula. This is a complicated PDE in the unknown T, called the Monge-Ampère Equation. Finding an optimal map then boils down to finding a solution with further structural properties.

In general, transport maps from μ to ν might not exist:

Example 1.4. Assume $\mu = \delta_0 \in \mathcal{P}(\mathbb{R})$ and $\nu \neq \delta_a$ for all $a \in \mathbb{R}$. Since, $T(\mu) = \delta_{T(0)}$ for any transport map T there cannot be a map T s.t. $T(\mu) = \nu$.

Another problem with the notion of transport maps is that the constraint $T(\mu) = \nu$ is not weakly sequentially closed w.r.t. a reasonable topology.

Definition 1.5. *Let $\mu \in \mathcal{P}(\mathsf{X})$, $\nu \in \mathcal{P}(\mathsf{Y})$. A coupling of μ and ν is a measure $\pi \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$ with marginals μ and ν, i.e.*

$$\pi(A \times \mathsf{Y}) = \mu(A) \text{ for all } A \in \mathcal{B}(\mathsf{X}) \quad \text{and} \quad q(\mathsf{X} \times B) = \nu(B) \text{ for all } B \in \mathcal{B}(\mathsf{Y}).$$

The set of all couplings of μ and ν will be denoted by $\mathsf{Cpl}(\mu, \nu)$.

Stochastically, a coupling π of μ and ν is a joint law of two random variables (X, Y) such that $\mathsf{Law}_\pi(X) = \mu$ and $\mathsf{Law}_\pi(Y) = \nu$. In particular, conditioning on $X = x$ we can interpret the regular conditional probability $\pi(\cdot|X = x)$ as a plan on how to transport the mass at x. Therefore, we will often call coupling transport plans. Analytically, this corresponds to disintegrating π w.r.t. its first marginal μ to obtain a family of probability measures $(\pi_x(dy))_{x\in\mathsf{X}}$ (see Theorem **??**). Writing $\mathrm{proj}_{\mathsf{X}} : \mathsf{X} \times \mathsf{Y} \to \mathsf{X}, (x, y) \mapsto x, \mathrm{proj}_{\mathsf{Y}} : \mathsf{X} \times \mathsf{Y} \to \mathsf{Y}, (x, y) \mapsto y$ a measure $\pi \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$ is an element of $\mathsf{Cpl}(\mu, \nu)$ iff $\mathrm{proj}_{\mathsf{X}}(\pi) = \mu$ and $\mathrm{proj}_{\mathsf{Y}}(\pi) = \nu$.

The set $\mathsf{Cpl}(\mu, \nu)$ is always non-empty. Indeed the product coupling (stochastically, the independent coupling) satisfies $\mu \otimes \nu \in \mathsf{Cpl}(\mu, \nu)$.

Remark 1.6. Observe, that any transport map $T : \mathsf{X} \to \mathsf{Y}$ from μ to ν induces a transport plan $\pi_T := (\mathrm{Id}, T)(\mu) \in \mathsf{Cpl}(\mu, \nu)$. We call π_T a Monge coupling or the coupling induced by the map T.

We give some further examples of transport maps / couplings:

Example 1.7. Let ν be a probability measure on $\mathbb{R}$ and write F^ν for its distribution function. The corresponding quantile function is given by the generalized inverse $q^\nu : (0, 1) \to \mathbb{R}$ defined by

$$q^\nu(u) := \inf\{x : F(x) > 0\}. \tag{1.1}$$

Writing λ for the Lebesgue measure on the unit interval, we have $q^\nu_\#(\lambda) = \nu$, that is q^ν if a Monge-map which takes λ to ν.

If ν has an atom, then so does $F^\nu(\nu)$ and in particular F^ν is *not* a Monge-map from ν to λ.

On the other hand, if μ is a continuous probability on $\mathbb{R}$ (i.e. has no atoms), than $F^\mu_\#(\mu) = \lambda$. In this situation the map $T := q^\nu \circ F^\mu$ is a Monge-transport from μ to ν, the so called *monotone transport* mapping.

Example 1.8. If μ, ν are (non necessarily continuous) measures on the real line, $\pi := (q^\mu, q^\nu)_\#\lambda$ is a coupling of μ, ν, the so called *co-monotone coupling*.

1.2. **The Monge and Kantorovich optimal transport problem.** Fix a Borel measurable function function $c : \mathsf{X} \times \mathsf{Y} \to [0, \infty]$. We will interpret c as the cost of transporting a unit of mass from $x \in \mathsf{X}$ to $y \in \mathsf{Y}$. Therefore, we will call such a function a cost function.

Definition 1.9. *Let $\mu \in \mathcal{P}(\mathsf{X}), \nu \in \mathcal{P}(\mathsf{Y})$ and c a cost function. The Monge problem is to solve*

$$P^M_c := P^M_c(\mu, \nu) := \inf \int c(x, T(x))\, \mu(dx), \tag{MP}$$

where the infimum runs over all transport maps $T : \mathsf{X} \to \mathsf{Y}$ such that $T(\mu) = \nu$. Any map T attaining the infimum in (MP) *is called optimal transport map.*

Definition 1.10. *Let $\mu \in \mathcal{P}(\mathsf{X}), \nu \in \mathcal{P}(\mathsf{Y})$ and c a cost function. The Kantorovich problem is to solve*

$$P^K_c := P^K_c(\mu, \nu) := \inf \int c(x, y)\, \pi(dx, dy), \tag{KP}$$

where the infimum runs over all couplings $q \in \mathsf{Cpl}(\mu, \nu)$. Any coupling π attaining the infimum in (KP) *is called optimal coupling or optimal transport plan.*

As we will see, the Kantorovich problem is much better behaved than the Monge problem. For instance, the following properties are immediate.

Remark 1.11.
- The set $\mathsf{Cpl}(\mu, \nu)$ is convex.
- The map $q \mapsto \int c\, dq$ is linear.

Moreover, $\mathsf{Cpl}(\mu, \nu)$ is compact in a natural topology which will allow us to show existence of optimal couplings under some assumption on the cost function c.

Recall that a sequence of measures $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}(\mathsf{X})$ converges weakly to $\mu \in \mathcal{P}(\mathsf{X})$ iff

$$\int f d\mu_n \to \int f d\mu, \qquad \text{for all } f \in C_b(\mathsf{X}),$$

where $C_b(\mathsf{X})$ denote the continuous and bounded functions on X. We call the induced topology on $\mathcal{P}(\mathsf{X})$ the weak topology.

Theorem 1.12 (Prokhorov)**.** *Let X be a Polish space. A family $A \subseteq \mathcal{P}(\mathsf{X})$ of probability measures on X is relatively compact w.r.t. the weak topology iff it is tight, i.e. for every $\varepsilon > 0$ there exists $K_\varepsilon \subseteq \mathsf{X}$ compact such that*

$$\sup_{\mu\in A} \mu(\mathsf{X} \setminus K_\varepsilon) \le \varepsilon.$$

For a proof we refer to [Bil99].

Lemma 1.13. *If $A_1 \subseteq \mathcal{P}(\mathsf{X}), A_2 \subseteq \mathcal{P}(\mathsf{Y})$ are tight so is $A_3 := \{q \in \mathcal{P}(\mathsf{X} \times \mathsf{Y}) : \mathrm{proj}_{\mathsf{X}}(q) \in A_1$ and $\mathrm{proj}_{\mathsf{Y}}(q) \in A_2\}$.*

Proof. Let $q \in A_3$ and $\varepsilon > 0$ be given. Pick $K_1 \subseteq \mathsf{X}, K_2 \subseteq \mathsf{Y}$ such that $\mu(\mathsf{X} \setminus K_1) \le \varepsilon, \nu(\mathsf{Y} \setminus K_2) \le \varepsilon$ for all $\mu \in A_1, \nu \in A_2$. Since, $K_1 \times K_2 \subseteq \mathsf{X} \times \mathsf{Y}$ is compact the claim follows from

$$q(\mathsf{X} \times \mathsf{Y} \setminus K_1 \times K_2) \le q((\mathsf{X} \setminus K_1) \times \mathsf{Y}) + q(\mathsf{X} \times (\mathsf{Y} \setminus K_2)) = \mu(\mathsf{X} \setminus K_1) + \nu(\mathsf{Y} \setminus K_2) \le 2\varepsilon.$$

□

Corollary 1.14. *The set* $\mathsf{Cpl}(\mu, \nu)$ *is compact.*

Proof. Since $\{\mu\} \subseteq \mathcal{P}(\mathsf{X}), \{\nu\} \subseteq \mathcal{P}(\mathsf{Y})$ are tight, $\mathsf{Cpl}(\mu, \nu)$ is tight by Lemma 1.13. It remains to show that it is closed. Pick $(q_n)_{n\in\mathbb{N}} \subseteq \mathsf{Cpl}(\mu, \nu)$ with limit q. We have to show that q has marginals μ and ν. Pick $\varphi \in C_b(\mathsf{X})$ and define $\bar{\varphi}(x, y) := \varphi(x)$ so that $\bar{\varphi} \in C_b(\mathsf{X} \times \mathsf{Y})$. Then, we know that

$$\int \varphi dq = \int \bar{\varphi} dq = \lim_n \int \bar{\varphi} dq_n = \lim_n \int \varphi dq_n = \int \varphi d\mu$$

so that $\mathrm{proj}_{\mathsf{X}}(q) = \mu$. Similarly, it follows that $\mathrm{proj}_{\mathsf{Y}}(q) = \nu$. □

A function $f : Z \to [0, \infty]$ is lower semi-continuous if for all sequence $z, z_1, z_2, \ldots \in Z$, $\lim_{n\to\infty} z_n = z$ we have $\liminf f(z_n) \geq f(z)$. Equivalently, f is lower semicontinuous if there is a sequence of continuous bounded functions $f_1, f_2, \ldots : Z \to [0, \infty)$ such that $f = \sup_n f_n$.

In this case, also the mapping

$$\pi \to \int f \, d\pi$$

is lower semicontinuous (on $\mathcal{P}(Z)$) since it is a supremum of the continuous bounded functions $\pi \to \int f_n \, d\pi$.

Note also that a lower semicontinuous function attains its infimum on every compact sets. From these observations we obtain:

Theorem 1.15. *Assume that c is lower semi-continous and bounded from below. Then there exists a minimizer π^* to* (KP)*, i.e.* $\pi^* \in \arg\min_{\pi\in\mathsf{Cpl}(\mu,\nu)} \int c d\pi$.

References

[AG13] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[Mon81] G. Monge. Memoire sur la theorie des deblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[San15] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[Vil03] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.