Snapshots of modern mathematics
from Oberwolfach

№ 13/2018

A short story on optimal transport and its many applications

Filippo Santambrogio

We present some examples of optimal transport problems and of applications to different sciences (logistics, economics, image processing, and a little bit of evolution equations) through the crazy story of an industrial dynasty regularly asking advice from an exotic mathematician.

1 Mines and factories, and bistochastic matrices

Mr. Hardwork was the owner of 26 iron mines and 26 factories; from each mine exactly one ton of iron per day was extracted, and each factory used precisely the same amount of iron for its industrial production. His father had given names to the factories, and each name started with a different letter from the latin alphabet, and luckily enough the names of the mines also started with the 26 different letters of the alphabet. That's why, when his engineers wondered how to organize the shipping of the metal from mines to factories in the best possible way, with no hesitation he chose the unique reasonable way of proceeding: the whole shipping had to be done in alphabetical order, with the iron production of each mine sent to the factory corresponding to the same letter… But now he was hesitating: after several years, unexpected news risked to force him to change his mind. Some factories were increasing their iron demand, while some others were reducing it, and the same happened to some mines… not to mention the fact that the employees of one of the factories wanted to change its name, and that a plan for buying new mines in Greece and Russia risked to create alphabetical issues. And accountants and engineers insisted for finding a solution which reduced shipping costs...

Mr. Hardwork had been told about an eccentric wise man with an exotic name, called Grasped Mango, who was spending his time writing formulas and strange symbols on a blackboard and solving problems. He wrote him a letter, and he was pleased to receive the following answer:

> Dear Mr. Hardwork,
>
> if you don't mind, I would like to give you an advice. Forget about the alphabetical order...

"he really is a crazy guy", reacted Mr. Hardwork, "the alphabetical naming is the most important point of my whole industrial empire", and continued reading

> ...and just use mathematics. Actually, your problem reminds me of an old problem first studied by a French mathematician, Gaspard Monge (1746–1818).[1] The problem is more general, and it is the same on every time you have a distribution of mass (in your case it's iron, in Monge's case it was a pile of sand) to be moved, you know its current position and the target you want to reach (mines and factories, respectively), you know the shipping cost $c_{x,y}$ to move one unit of mass from source x to target y, and you must choose where to send each particle of material. This is called an *optimal transport problem*, and it is quite easy to attack: we just need to write a 26×26 *cost-matrix* (that is, a table of 26×26 numbers) indicating, at the entry labeled by (x, y), the cost $c_{x,y}$ to ship from the source mine x to the target factory y. Then, we have to choose another matrix with numbers $\gamma_{x,y}$ representing the fraction of mass that goes from x to y. There are some constraints on this matrix: the numbers $\gamma_{x,y}$ must clearly not be negative, and the sum of the elements of each column of the matrix – which we can write as $\sum_y \gamma_{x,y}$ – and of the elements of each row – which we write as $\sum_x \gamma_{x,y}$ – must all be equal to one (that is, the total amount of mass sent from a mine x to all the factories y should be equal to the production of such a mine, and vice versa – I am not yet considering the case where productions and demand have changed).
>
> These matrices $\boldsymbol{\gamma}$ are called *bistochastic matrices*.[2] We then "overlay" such a $\gamma_{x,y}$ matrix on top of the $c_{x,y}$ matrix to get a matrix

[1] You can find the original in [14]

[2] A bistochastic matrix is a matrix whose rows and columns each sum up to one, see [7].

whose entries are $c_{x,y}\,\gamma_{x,y}$ and sum over all entries of this new matrix. This represents the costs given the distribution choice γ. Finding the optimal $\boldsymbol{\gamma} = (\gamma_{x,y})$ is then just a problem of minimizing a linear function under linear constraints. Don't worry: you won't need to organize 26×26 different shipping. In practice, you will see that most of the numbers in the optimal $\boldsymbol{\gamma}$ will be zero, and for each mine x you will still have one unique target factory y... However, it will be in alphabetic order no more. Indeed, you can prove (and I leave it as an exercise for your engineers, or your kids) that the set of bistochastic matrices is the *convex hull*[3] of the set of *permutation matrices*. The permutation matrices are those that only have entries equal to 0 and 1, and the positions of the 1 describe a permutation of the indices. What I mean is that in the high-dimensional space of 26×26 matrices, the bistochastic ones are a polyhedron whose vertices are represented by the permutation matrices, and you will agree that a linear function is always optimized on a vertex.

Moreover, you don't have to be scared if your mines or factories do not produce or need all the same quantity of iron, or even if they are no more in equal numbers. All you have to do is to solve a *linear program*, which is a method to obtain the best outcome given constraints that are linear equations, of the form

$$I_{\min} := \min \left\{ \sum_{x,y} c_{x,y}\gamma_{x,y} \ : \ \gamma_{x,y} \geq 0, \sum_{y} \gamma_{x,y} = \mu_x, \sum_{x} \gamma_{x,y} = \nu_y \right\}. \tag{1}$$

You won't believe how fast such a solution can be obtained.[4]

2 Prices, and Kantorovich duality

Time passed by, and the Hardwork company passed on to Mr. Hardwork's son. The industrial world had evolved and the new owner of the company was more interested in management and in particular in the pricing of the company's production. He decided as well to contact Grasped Mango, who was still alive, and still in front of his blackboard. This time he did it by phone, and was much more precise than his father.

- **Mr. H**: "Let me be very general, Dr. Mango. Suppose that there is a set X of types of goods which are produced, a set Y of types of consumers, and a quantity $v_{x,y}$ which is the value given by the

[3] The convex hull of a set X is the smallest convex set that contains X.

[4] See, for instance, [5].

consumer y to the good x. Suppose also that for each $x \in X$, we know how many items of type x are produced, and we call this quantity μ_x, and we also know the quantity ν_y of consumers of type y. My question is: *what should be the price p_x for each good x given the goods-values $v_{x,y}$ and the data on the production and on the consumers*?"

- **Dr. M**: "This reminds me of the approach that Leonid Kantorovich (1912–1986) used to attack Monge's problem [11]. If I have to be honest, the answer that I gave to your father was already much based on what this Russian researcher, who even got a Nobel prize in Economics in 1975, developed in the 40s. You will soon understand that his ideas have a clear economic interpretation. Indeed, Kantorovich understood that solving the same minimization problem that I considered for your father is intimately related to another optimization problem, that we usually call a *dual problem*. To obtain a dual problem from the original one, which we therefore call *primal one*, you can act in this way (let me give more details, since you seem to be more open to mathematics than your father): Replace the constraints on the sums of rows and columns of the matrix γ_{xy} with a penalization, that you can obtain using the following computation.

Consider the equality

$$\sup_{\phi,\psi} \left\{ \sum_x \phi_x \, \Delta_x + \sum_y \psi_y \, \Delta_y \right\} = \begin{cases} 0 & \text{if } \Delta_x = \Delta_y = 0 \\ +\infty & \text{otherwise,} \end{cases} \tag{2}$$

where we have defined $\Delta_x := \mu_x - \sum_y \gamma_{x,y}$ and $\Delta_y := \nu_y - \sum_x \gamma_{x,y}$ for convenience of notation. Instead of imposing the constraints, you can add the value of the sup in (2) into the minimization problem . In this way, you are adding nothing if the constraints are satisfied, and you are adding $+\infty$ if they are not, which is exactly the same as imposing the constraints, in the end. Then, if you switch the minimization and the maximation procedures, you "magically" end up with a new optimization problem

$$I_{\max} := \max \left\{ \sum_x \phi_x \mu_x + \sum_y \psi_y \nu_y \ : \ \phi_x + \psi_y \le c_{x,y} \right\}. \tag{3}$$

You can prove that the minimum $I_{\min}$ in (2) and the maximum $I_{\max}$ in (3) are the same value. This result depends on advanced convex analysis theorems, but you can imagine a situation similar to that of a mountain pass separating two valleys. Among all possible paths from one valley to the other, find the one which minimizes the highest point it passes through: its altitude is the same as the one you would obtain if you wanted to find the path from one top on the side of the pass to the other, which maximize the elevation of the lowest point it passes through. Moreover, whenever you take the three objects $(\boldsymbol{\gamma}, \phi, \psi)$ you can say that $\boldsymbol{\gamma}$ is optimal in (2) and (ϕ, ψ) in (3) if and only if they are *admissible* (that is, they satisfy the constraints in (2) and (3)) and

$$\phi_x + \psi_y = c_{x,y} \text{ for all pairs } (x, y) \text{ with } \gamma_{x,y} > 0.'' \tag{4}$$

- **Mr. H**: "I understand, Prof. Mango, but what about my pricing question?"

- **Dr. M**: "You see, what you look for is indeed a triple $(\boldsymbol{\gamma}, p, q)$ with the following properties: $\gamma_{x,y}$ stands for how many consumers of type y buy the good x, p_x for the price of x, and q_y for the net utility of the consumer y, in other words, the value of the good he buys, after subtracting the price he paid. The numbers that make up the matrix $\boldsymbol{\gamma}$ must be nonnegative and satisfy the usual constraints on the sums of the rows and columns, as happens in (2). Moreover, if $\gamma_{x,y} > 0$, then this means that some people of type y actually buy a good of type x: this would not happen if x was not the maximizer of $x \mapsto v_{x,y} - p_x$, and in this case we have $q_y = \max_x\{v_{x,y} - p_x\}$. To be precise, this means that you want $p_x + q_y \geq v_{x,y}$, with equality whenever $\gamma_{x,y} > 0$. Now, you see that it is enough to solve (2) for $c_{x,y} = -v_{x,y}$ and take $p = -\phi$ and $q = -\psi$"...

- **Mr. H**: "That's interesting, but in this way I could just add a constant to all the prices p_x, and subtract it from all the utilities q_y, satisfying the equations while getting higher and higher prices. I like adding constant to prices, but that's not how markets work."

- **Dr. M**: "It's because you previously assumed that your consumers are obliged to buy something. Just add an empty good x to your set, and impose that its price must be 0 (I hope you will agree that you cannot charge more than this for buying nothing), and you will see that all the other prices will be fixed. This is a consequence of the fact that, if consumers are not forced to buy but can choose not to buy, then you cannot artificially inflate prices just for the sake of your personal gain. Consumers would, in that case, simply not buy anything!

Whatever choice of model you pick, that is, whether you decide to handle the freedom of consumers to choose or not, I hope I have convinced you that your problem is equivalent to that of your father."

3 Monotone maps, and color rearrangements

Time went on, and technologies went on changing. The Hardwork company was now property of Mr. Hardwork's grand-daughter, who had turned it into a high-tech company mainly specialized in digital pictures. But the tradition to ask Mango's advice had not disappeared and, exactly as her father and her grandfather, she also contacted the old wise man with a question. She did it by email and attached many images and graphs.

From: Ms. Hardwork <director@hardwork.co>
To: Dr. Grasped <grasponge@maths.edu>
Subject: Information request.

My dear Grasped,

here is a natural question from image processing: suppose you need to combine, or compare, two pictures which represent the same object, or two parts of it, but have been taken in different conditions, so that one is "globally" darker than the other; we would like to change the colors of one of the two in order to have the comparable colors, before attempting to compare pictures. If we have B&W pictures, this is not so difficult:

- We draw the histograms of the gray-scale distributions of both pictures (that is, we note how many pixels have each possible level of gray, numbered from 0 to 255).
- We then find the monotonically increasing function $T : \mathbb{R} \to \mathbb{R}$ that allows us to obtain the following: the number of pixels of the first image with gray level smaller than x equals the number of pixels in the second image with level smaller than $T(x)$.
- We transform the colors of the first image, by replacing everywhere the gray level x with the gray level $T(x)$.

I guess that, at this point, you have already understood my question: *what should I do if instead of gray levels I have actual colors? How can I choose the function T in 3D?*

From: Dr. Grasped <grasponge@maths.edu>
To: Ms. Hardwork <director@hardwork.co>
Subject: Re: Information request.

Dear Ms. Hardwork,

your question is once again a question about optimal transport. Let me come back to your grandfather's question; after my general answer suggesting to solve a linear programming problem, he asked me about some specific examples. In particular, he wanted to know what happened if all the mines and the factories were on the same, long, road, and the shipping cost only depended on the distance along the road. That's easy, I told him, if the cost is a power of the distance, and this power is strictly convex (say, $c_{x,y} = |x - y|^p$ for $p > 1$ and x, y lying on a line), the optimal way is to order mines and factories from left to right and to send the iron from the first mine to the first factory, from the second to the second... essentially, to find the monotonically increasing map T which does exactly the job that you were mentioning with pixels. On the other hand, if $p < 1$, this is not optimal, as one single long jump is preferable to many small ones. For instance, if the mines are at points $\{1, 2, \ldots, n\}$ and the factories at points $\{2, 3, \ldots, n, n + 1\}$, for $p > 1$ you will send each point x to $T(x) = x + 1$, and for $p < 1$ you will just send 1 to $T(1) = n + 1$ and not move all the others (that is, $T(x) = x$)... For $p = 1$ you have the choice, with at least two different optimal solutions, which are the two I just described (this is also known as the *book shifting example*). An example of what I just said can be found in Figure 1 attached to this email. You can have a look at [9] if you want, where you can see the important role played by the increasing transport map. Moreover, you will see that there is no need of having mass distributions represented by a finite number of locations, and you can consider any distribution (that we call *measure* in mathematics), in particular those distributions which are "diffuse everywhere", or *continuous*, like water or gas in a container, and therefore can define a density, that is, the quantity defined at every point of the total space available which gives the amount of matter enclosed in a certain volume, once integrated over that volume.

Figure 1: The transport maps in the book shifting example

But your question is multidimensional, and hence more difficult.

This reminds me of an old result[5] by Yann Brenier, see [4]. He was particularly interested in the case $c_{x,y} = |x-y|^2$, which is the most physically relevant, and considered transport maps T which transform one given distribution of mass onto another given one. He found out that the optimal map T exists, and is the gradient of a convex function. A convex function in dimension one is exactly characterized by the fact that its derivative is monotone increasing, while in higher dimensions, if $T = \nabla u$ and u is convex we have

$$\langle T(x) - T(x')\,,\, x - x' \rangle \geq 0 \quad \text{for all } x, x',$$

which can be considered as a notion of higher-dimensional monotonicity.[6] You could try using this transport map T between the distribution of mass on the 3D space of colours given by one image, and the distribution of the other image, and the results should be good![7]

4 Middle points, barycenters, distances, steepest descent

Ms. Hardwork was quite satisfied by the suggestions of her by-now family friend Grasped Mango, which inspired her engineers in developing many useful algorithms, as you can see in Figure 2.

Hence, she decided to ask new questions.

From: Ms. Hardwork <director@hardwork.co>
To: Dr. Grasped <grasponge@maths.edu>
Subject: Changing colors.

Dear Grasped,

how about modifying the colors of both pictures into an intermediate color set? Beware that, for me, the midpoint between a red picture and a yellow one is not half red and half yellow… I would expect something orange.

From: Dr. Grasped <grasponge@maths.edu>

---

[5] To appear more mysterious than what he really was, Grasped Mango always referred to mathematical results calling them "old", independently of how old they really were.

[6] The gradient operator ∇ is the generalisation of derivative in more than one dimension. The inner product $\langle\cdot,\cdot\rangle$ measures the angle between vectors.

[7] Similar color transfers were used, for instance, in [15].

Figure 2: The last image is the first one, with the colors of the second, using optimal transport methods.

To: Ms. Hardwork <director@hardwork.co>
Subject: Re: Changing colors.

Dear. Ms. Hardwork,

That's easy. This reminds me of an old idea by Robert McCann, see [13], who wanted to find a way to move continuously (in the space of distributions) from one mass distribution to another, along a "shortest path" between them, called *geodesic*.[8]. You can interpolate between a distribution μ and another distribution ν by taking the optimal transport map T sending μ onto ν, for instance for the quadratic cost $c_{x,y} = |x-y|^2$ (called the Brenier map). Now, if you take the particles of μ and move each of them from x to $T(x)$, you obtain ν; if you leave them at x, you obtain μ; if you take a number $t \in [0,1]$ and you move them from their position x to $T_t(x) := (1-t)x + tT(x)$ you will obtain a new distribution that you can call μ_t. Letting t vary from $t=0$ to $t=1$ you will obtain a continuous curve of distributions connecting μ to ν. In the application to images that you have in mind, you will gradually transform red into yellow by passing through all shades of orange.

From: Ms. Hardwork <director@hardwork.co>
To: Dr. Grasped <grasponge@maths.edu>
Subject: Re: Re: Changing colors.

[8] This is in analogy with the theory of special relativity, where geodesics represent the natural path followed by test (very small) particles that move through space and time.

Dear Grasped,

this seems to be a nice idea, but what would happen if I had more than two distributions?

From: Dr. Grasped <grasponge@maths.edu>
To: Ms. Hardwork <director@hardwork.co>
Subject: Re: Re: Re: Changing colors.

Dear. Ms. Hardwork,

this is also easy. What you need is to define a notion of *barycenter* (center of mass), and it reminds me of an old paper by Martial Agueh and Guillaume Carlier (which can be found at [1]). What is the barycenter between some pointlike objects with weights λ_i placed at points x_i? If the λ_i are nonnegative numbers that sum up to 1, that is, the weight is given as a percentage of the total mass, then you just set $x := \sum_i \lambda_i \, x_i$ and this defines the position barycenter. I know that you do not want to sum and average mass distributions, neither on the space of colors nor on the space of mine locations, but, please, wait a little bit. The barycenter that I just defined also happens to be the solution of the minimisation problem $\min_x \sum_i \lambda_i |x - x_i|^2$. All you need is to define a suitable distance on the space of mass distributions based on optimal transport ideas. That is, a distance between distributions, so to speak. This distance exists, and it is called *Wasserstein distance* $W_2(\mu, \nu)$ (don't ask me why this name, this is a long story, and not everybody agrees). Since I know that I can use advanced mathematics with you, I will write down the definition for you using integrals instead of sums, it in the following way:

$$W_2(\mu, \nu) := \inf \left\{ \int |T(x) - x|^2 d\mu(x) \ : \ T \text{ transports } \mu \text{ onto } \nu \right\}^{1/2} .$$

If you want to be convinced that the distance defined above does the job we are looking for, think that it generalizes the length of a vector connecting two points to the case where the vector has a "continuous amount of elements". Furthermore, each component in the addition (that is, integral) is "weighted" by the value of the distribution at that point. In this sense, each infinitesimal summand in the integral contributes more or less depending on the value of the measure. If you know a little bit about L^p norms[9], you can guess that the exponent 2 could be replaced by other exponents p, thus leading to W_p distances, and that the power $1/p$ outside the infimum is needed to make it a distance (that is, satisfy the triangle inequality)... Anyway, you can now take a certain number of distributions μ_i on a space X and solve

$$\min \left\{ \sum_i \lambda_i W_2^2(\rho, \mu_i) \ : \ \rho \text{ is a distribution on } X \right\}.$$

The distribution ρ which attains this minimum will be a barycenter between the μ_i's, with weights λ_i. You might say that this distribution is a mixture of the μ_i's with proportions λ_i. You can get an idea from Figure 3, and have a look at [16] for other applications of barycenters in image processing.

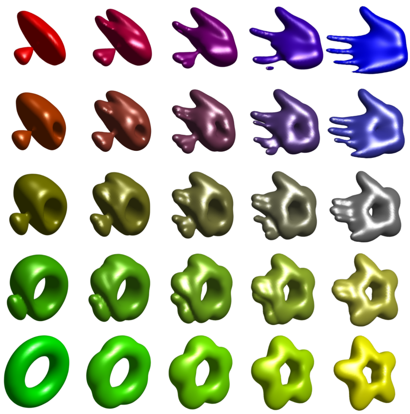

Figure 3: Barycenters between four different shapes, with different weights. Thresholding of the densities has been performed in order to visualize shapes: the barycenters should a priori be mass distributions, not necessarily uniform.

Once he started to tell a story on optimal transport, it was difficult to stop Grasped Mango[10], and he went on without Miss Hardwork having to ask.

[9] The distances known as L^p norms help defining spaces of functions where integrals of powers of these functions, such as $\int dx|f(x)|^p$, have finite values. These norms are paramount in many areas of modern sciences, such as quantum physics.

[10] More or less the same as the author of this snapshot.

You know, Wasserstein distances can be used for many other purposes. Essentially, every time that you have a distance on a set you obtain a distance on the probability measures that are defined on it, or on the histograms on it, you would say. You can use them to compare objects, find middle points, find outliers and clusters... they are widely used now in data processing. But I would like to finish this story with a very different application: after logistics, economics, and image processing, let me speak about evolution equations. Indeed, I mentioned some particular curves in the space of probabilites (geodesics) in my previous email. We can study other curves, and if you imagine that these probabilities are defined through their density ρ, then a curve of probability will be a function $\rho(t, x)$ of time and space, and it will be interesting to study the equations this function solves. Indeed, I would like to tell you something about an old idea by Richard Jordan, David Kinderlehrer, and Felix Otto (printed in [10]). First, give me ten seconds to tell you what is a gradient flow, in case you don't know.

Take a point x_0 in $\mathbb{R}^n$ and a function $F : \mathbb{R}^n \to \mathbb{R}$, and follow the equation $x'(t) = -\nabla F(x(t))$ starting from $x(0) = x_0$, which is the *steepest descent curve* for F stemming from such a point. There is an easy way to approximate such a curve by time-discretization: you fix a time step $\tau > 0$, and you iteratively define x_{k+1} to be a minimizer of $x \mapsto F(x) + \frac{1}{2\tau}|x - x_k|^2$. I guess you know that the gradient of a function vanishes at minima points, which provides the condition, $\frac{1}{\tau}(x_{k+1} - x_k) = -\nabla F(x_{k+1})$, which is a discretization of the above equation. The same iterated minimization can also be performed in the framework of mass distributions: you can iteratively solve

$$\min_{\rho} \left\{ \mathcal{F}(\rho) + \frac{W_2^2(\rho, \rho_k)}{2\tau} \right\},$$

for different choices of the function $\mathcal{F}$, defined on the set of probability measures. For instance, if you take $\mathcal{F}(\rho) := \int F \, d\rho$, the curve you get is just a superposition of trajectories of $x' = -\nabla F(x)$, that is, you have a solution to the partial differential equation $\partial_t \rho - \nabla \cdot (\rho \nabla F) = 0$. If instead you use the entropy functional $\mathcal{F}(\rho) := \int \rho(x) \log \rho(x) dx$ (to be properly defined), you can see that you have a solution to the heat equation $\partial_t \rho - \Delta \rho = 0$ and, summing the two, with $\mathcal{F}(\rho) := \int \rho(x) \log \rho(x) dx + \int F \, \frac{d\rho}{dx} \, dx$ you have the

Fokker-Planck equation

$$\partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla F) = 0.$$

I am sorry if these equations are not easy to obtain from the minimization problem above, but you can have a look at [10] if you want to know more about the details.

Many other equations, in particular those coming from *population dynamics* (because of the conservation of the mass which is implicit in the idea that in optimal transport, the mass is fixed to 1, and this is well adapted to describe the motion of a population), can be studied via these tools: just to note some examples, you can find biological equations such as the Keller-Segel models for chemotaxis (see [3]), or models for crowd motion where the density is constrained not to go beyond a given threshold (see [12]).

Further reading

The Hardwork family fell in love with the stories by Grasped Mango, with optimal transport, and with its many applications. If this also happened to you and you want to know more, there are books which could be of interest (but they are written as research monographs, and require advanced mathematics). The main reference is [19], a book that, later on, the author has expanded into a huge treaty, [20], detailing all the connections with differential geometry that have been omitted in this short story. From the point of view of applications, [17] is an accessible introduction, and [8] focuses on economics; a text specifically devoted to numerical methods is [6]. When talking about "gradient flows", the bible for the general theory is [2], but a recent and shorter survey is available at [18]. Finally, another simple introduction on the topic of optimal transport can be also found in Snapshot 8/2017 *Computational Optimal Transport* by Justin Solomon

Image credits

Figure 1 was designed by the author.

Figure 2 and Figure 3 have been kindly provided by Gabriel Peyré.

References

[1] M. Agueh and G. Carlier, *Barycenters in the Wasserstein Space*, SIAM Journal on Mathematical Analysis **43** (2011), no. 2, 904–924.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Lectures in Mathematics. ETH Zürich, Springer, 2005.

[3] A. Blanchet, V. Calvez, and J. A. Carrillo, *Convergence of the Mass-Transport Steepest Descent Scheme for the Subcritical Patlak-Keller-Segel model*, SIAM Journal on Numerical Analysis **46** (2008), no. 2, 691–721.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on Pure and Applied Mathematics **44** (1991), no. 4, 375–417.

[5] M. Cuturi, *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, Advances in Neural Information Processing Systems 26 (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds.), Curran Associates, Inc., 2013, pp. 2292–2300.

[6] M. Cuturi and G. Peyré, *Computational Optimal Transport*, *preprint*, stat.ML/1803.00567, available at https://arxiv.org.

[7] P. A. Gagniuc, *Markov Chains: From Theory to Implementation and Experimentation*, John Wiley & Sons, 2017.

[8] A. Galichon, *Optimal transport methods in economics*, Princeton University Press, 2016.

[9] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Mathematica **177** (1996), no. 2, 113–161.

[10] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck Equation*, SIAM Journal on Mathematical Analysis **29** (1998), no. 1, 1–17.

[11] L. Kantorovich, *On the transfer of masses*, Doklady Akademii Nauk USSR (1942), no. 37, 7–8.

[12] B. Maury, A. Roudneff-Chupin, and F. Santambrogio, *Macroscopic crowd motion model of gradient flow type*, Mathematical Models and Methods in Applied Sciences **20** (2010), no. 10, 1787–1821.

[13] R. J. McCann, *A Convexity Principle for Interacting Gases*, Advances in Mathematics **128** (1997), no. 1, 153–179.

[14] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris (1781), 666–704.

[15] F. Pitié, A.C. Kokaram, and R. Dahyot., *Automated colour grading using colour distribution transfer*, Computer Vision and Image Understanding **107** (2007), no. 1–2, 123–137.

[16] J. Rabin, G. Peyré, J. Delon, and M. Bernot, *Wasserstein Barycenter and its Application to Texture Mixing*, Springer, 2012.

[17] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications, Springer International Publishing, 2015.

[18] ———, *{Euclidean, metric, and Wasserstein} gradient flows: an overview*, Bulletin of Mathematical Sciences **7** (2017), no. 1, 87–154.

[19] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003.

[20] ———, *Optimal transport: Old and new*, Grundlehren der mathematischen Wissenschaften, Springer, 2008.

Filippo Santambrogio *is a professor of applied mathematics at Université Paris-Sud*
filippo.santambrogio@math.u-psud.fr.

Mathematical subjects
Analysis, Numerics and Scientific Computing

Connections to other fields
Engineering and Technology, Finance, Humanities and Social Sciences

DOI
10.14760/SNAP-2018-013-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Association

IMAGINARY open mathematics