Computers & Graphics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Survey Paper

Notions of optimal transport theory and how to implement them on a computer☆

Q1 Bruno Lévy*, Erica L. Schwindt

Inria centre Nancy Grand-Est and LORIA, France

ARTICLE INFO

Article history:
Received 6 October 2017
Revised 12 January 2018
Accepted 16 January 2018
Available online xxx

Keywords:
Mathematics
Physics
Optimal transport
Numerical optimization

ABSTRACT

This article gives an introduction to optimal transport, a mathematical theory that makes it possible to measure distances between functions (or distances between more general objects), to interpolate between objects or to enforce mass/volume conservation in certain computational physics simulations. Optimal transport is a rich scientific domain, with active research communities, both on its theoretical aspects and on more applicative considerations, such as geometry processing and machine learning. This article aims at explaining the main principles behind the theory of optimal transport, introduce the different involved notions, and more importantly, how they relate, to let the reader grasp an intuition of the elegant theory that structures them. Then we will consider a specific setting, called semi-discrete, where a continuous function is transported to a discrete sum of Dirac masses. Studying this specific setting naturally leads to an efficient computational algorithm, that uses classical notions of computational geometry, such as a generalization of Voronoi diagrams called Laguerre diagrams.

1. Introduction

This article presents an introduction to optimal transport, and then focuses on a specific class of numerical methods (semi-discrete). Before diving into the subject, we find it important to mention that besides semi-discrete methods, many other numerical methods exist, the reader is referred to [1] and [2] for a complete overview. We also mention that many alternatives exist for comparing and interpolating between functions, such as Reproducing Kernel Hilbert Space for measures, and diffeomorphism methods for functions and shapes.

This article summarizes and complements a series of conferences given by B. Lévy between 2014 and 2017 on semi-discrete optimal transport. The presentations stays at an elementary level, that corresponds to a computer scientist's vision of the problem. In the article, we stick to using standard notions of analysis (functions, integrals) and linear algebra (vectors, matrices), and give an intuition of the notion of measure. The main objective of the presentation is to understand the overall structure of the reasoning[1], and to follow a continuous path from the theory to an efficient algorithm that can be implemented in a computer.

Optimal transport, initially studied by Monge, [3], is a very general mathematical framework that can be used to model a wide class of application domains. In particular, it is a natural formulation for several fundamental questions in computer graphics [4–6], because it makes it possible to define new ways of *comparing* functions, of measuring *distances* between functions and *interpolating* between two (or more) functions:

Comparing functions. Consider the functions f_1, f_2 and f_3 in Fig. 1. Here we have chosen a function f_1 with a wildly oscillating graph, and a function f_2 obtained by translating the graph of f_1 along the x axis. The function f_3 corresponds to the mean value of f_1 (or f_2). If one measures the relative distances between these functions using the classical L_2 norm, that is $d_{L_2}(f,g) = \int (f(x) - g(x))^2 dx$, one will find that f_1 is nearer to f_3 than f_2. Optimal transport makes it possible to define a distance that will take into account that the graph of f_2 can be obtained from f_1 through a translation (like here), or through a deformation of the graph of f_1. From the point of view of this new distance, the function f_1 will be nearer to f_2 than to f_3.

Interpolating between functions:. Now consider the functions u and v in Fig. 2. Here we suppose that u corresponds to a certain physical quantity measured at an initial time $t = 0$ and that v corresponds to the same phenomenon measured at a final time $t = 1$.

☆ This article was recommended for publication by Dr M Teschner.

* Corresponding author.
E-mail addresses: Bruno.Levy@inria.fr, levy@loria.fr (B. Lévy), schwindt@math.cnrs.fr (E.L. Schwindt).
URL: https://members.loria.fr/BLevy/ (B. Lévy)

[1] Teach principles, not equations. [R. Feynman]

https://doi.org/10.1016/j.cag.2018.01.009

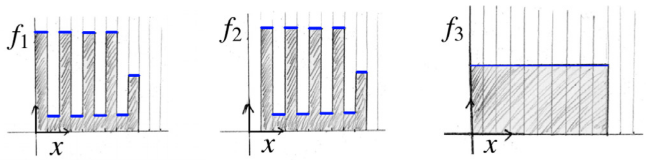

Fig. 1. Comparing functions: one would like to say that f_1 is nearer to f_2 than f_3, but the classical L_2 norm "does not see" that the graph of f_2 corresponds to the graph of f_1 slightly shifted along the x axis.

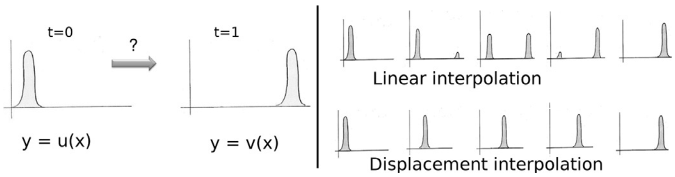

Fig. 2. Interpolating between two functions: linear interpolation makes a hump disappear while the other hump appears; displacement interpolation, stemming from optimal transport, will displace the hump as expected.

The problem that we consider now consists in reconstructing what happened between $t = 0$ and $t = 1$. If we use linear interpolation (Fig. 2, top-right), we will see the left hump progressively disappearing while the right hump progressively appears, which is not very realistic if the functions represent for instance a propagating wave. Optimal transport makes it possible to define another type of interpolation (Mc. Cann's displacement interpolation, Fig. 2, bottom-right), that will progressively displace and morph the graph of u into the graph of v.

Optimal transport makes it possible to define a *geometry* of a space of functions[2], and thus gives a definition of *distance* in this space, as well as means of interpolating between different functions, and in general, defining the barycenter of a weighted family of functions, in a very general context. Thus, optimal transport appears as a fundamental tool in many applied domains. In computer graphics, applications were proposed, to compare and interpolate objects of diverse natures [6], to generate lenses that can concentrate light to form caustics in a prescribed manner [7,8]. Moreover, optimal transport defines new tools that can be used to discretize Partial Differential Equations, and define new numerical solution mechanisms [9]. This type of numerical solution mechanism can be used to simulate for instance fluids [10], with spectacular applications and results in computer graphics [11].

The two sections that follow are partly inspired by [12], [13], [14] and [15], but stay at an elementary level. Here the main goal is to give an intuition of the different concepts, and more importantly an idea of the way the relate together. Finally we will see how they can be directly used to design a computational algorithm with very good performance, that can be used in practice in several application domains.

2. Monge's problem

The initial optimal transport problem was first introduced and studied by Monge, right before the French revolution [3]. We first give an intuitive idea of the problem, then quickly introduce the notion of *measure*, that is necessary to formally state the problem in its most general form and to analyze it.

2.1. Intuition

Monge's initial motivation to study this problem was very practical: supposing you have an army of workmen, how can you transform a terrain with an initial landscape into a given desired target landscape, while minimizing the total amount of work?

[2] or more general objects, called probability measures, more on this later.

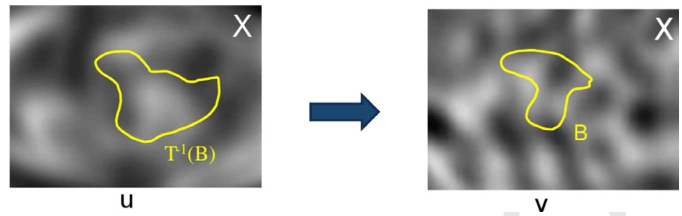

Fig. 3. Given two terrains defined by their height functions u and v, symbolized here as gray levels, Monge's problem consists in transforming one terrain into the other one by moving matter through an application T. This application needs to satisfy a mass conservation constraint.

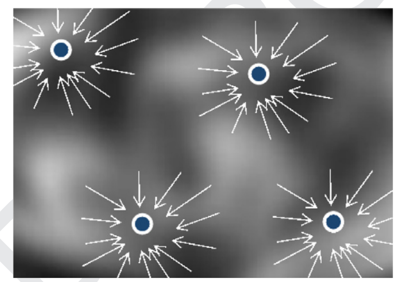

Fig. 4. Transport from a function (gray levels) to a discrete point-set (blue disks).

Monge's initial problem statement was as follows:

$$\inf_{T:X\to X} \int_X c(x, T(x))u(x)dx$$

subject to:

$$\forall B \subset X, \int_{T^{-1}(B)} u(x)dx = \int_B v(x)dx$$

where X is a subset of $\mathbb{R}^2$, u and v are two positive functions defined on X and such that $\int_X u(x)dx = \int_Y v(x)dx$, and $c(\cdot, \cdot)$ is a convex distance (the Euclidean distance in Monge's initial problem statement).

The functions u and v represent the height of the current landscape and the height of the target landscape respectively (symbolized as gray levels in Fig. 3). The problem consists in finding (if it exists) a function T from X to X that transforms the current landscape u into the desired one v, while minimizing the product of the amount of transported earth $u(x)$ with the distance $c(x, T(x))$ to which it was transported. Clearly, the amount of earth is conserved during transport, thus the total quantity of earth should be the same in the source and target landscapes (the integrals of u and v over X should coincide). This *global* matter conservation constraint needs to be completed with a *local* one. The local matter conservation constraint enforces that in the target landscape, the quantity of earth received in any subset B of X corresponds to what was transported here, that is the quantity of earth initially present in the pre-image $T^{-1}(B)$ of B under T. Without this constraint, one could locally create matter in some places and annihilate matter in other places in a counterbalancing way. A map T that satisfies the local mass conservation constraint is called a *transport map*.

2.2. Monge's problem with measures

We now suppose that instead of a "target landscape", we wish to transport earth (or a resource) towards a set of points (that will be denoted by Y for now on), that represent for instance a set of factories that exploit a resource, see Fig. 4. Each factory wishes to

Please cite this article as: B. Lévy, E.L. Schwindt, Notions of optimal transport theory and how to implement them on a computer, Computers & Graphics (2018), https://doi.org/10.1016/j.cag.2018.01.009