

Effective Brenier Theorem

Applications to Computable Analysis and Algorithmic Randomness

Alex Galicki

University of Auckland
agal629@auckland.ac.nz

Abstract

Brenier's theorem is a landmark result in Optimal Transport. It postulates existence, monotonicity and uniqueness of an optimal map, with respect to the quadratic cost function, between two given probability measures (under some weak regularity conditions). We prove an effective version of Brenier's theorem: we show that for any two computable absolutely continuous measures on \mathbb{R}^n , μ and ν , with some restrictions on their support, there exists a computable convex function ϕ , whose gradient $\nabla\phi$ is the optimal transport map between μ and ν .

The main insight of the paper is the idea that an effective Brenier's theorem can be used to construct effective monotone maps on \mathbb{R}^n with desired (non-)differentiability properties. We use it to solve several problems at the interface of algorithmic randomness and computable analysis. In particular, we show that $z \in \mathbb{R}^n$ is computably random if and only if every computable monotone function on \mathbb{R}^n is differentiable at z . Furthermore, we prove the converse of the effective Aleksandrov theorem (Galicki 2015): we show that if $z \in \mathbb{R}^n$ is not computably random, there exists a computable convex function that is not twice differentiable at z . Finally, we prove several new characterisations of computable randomness in \mathbb{R}^n : in terms of differentiability of computable measures, in terms of a particular Monge-Ampère equation and in terms of critical values of computable Lipschitz functions.

Categories and Subject Descriptors F.1.1 [Models of Computation]: Computability theory

General Terms Algorithmic Randomness, Computable Analysis, Optimal Transport, Differentiability

1. Introduction

This paper studies connections between three established areas of mathematics: algorithmic randomness, classical analysis and optimal transport.

Optimal transport is an influential area of research, with history going back to 1781, related to probability theory, economics and optimization. One of the most important results in that area is known as Brenier's theorem. Very roughly, this theorem asserts existence, uniqueness and monotonicity of a solution to the so

called Monge-Kantorovich problem (with respect to the quadratic cost and under some mild regularity assumptions).

In the Section 4, we will present the following effective version of Brenier's theorem:

Theorem 1.1 (The effective Brenier theorem). *Let μ, ν be absolutely continuous computable probability measures on \mathbb{R}^n such that $\text{supp}(\mu) = [0, 1]^n$ and the support of ν is bounded. There exists a computable convex function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\nabla\phi$ is the optimal transport map from μ to ν .*

A convex function ϕ from the above theorem is sometimes called *Kantorovich potential*. Our result says that under some specific conditions on μ and ν , a Kantorovich potential is, in principle, computable. Since the theory of optimal transport has numerous practical applications, ranging from economics to image processing and machine learning, the above result has important practical implications. This paper however, concentrates on applications of the theorem to problems in computable analysis and algorithmic randomness. A type of questions often arising in those areas is following:

(*) "How much algorithmic randomness is necessary and sufficient for a particular analytical property to hold for effective functions?"

A comprehensive answer to such a question comes in two parts: an upper bound asserting that a particular type of algorithmic randomness R is sufficient and a sharpness result asserting that R is necessary as no weaker notion would suffice. In higher dimensions, that is in \mathbb{R}^n for $n > 1$, sharpness results often are considerably more difficult to demonstrate, as they usually require explicit constructions of objects with prescribed computability and analytical/geometric properties. In this paper we answer a number of questions of the form (*), and in every case the more difficult part of the answer (the sharpness part) comes from using the effective form of Brenier's theorem described in this paper.

From the perspective of analysis, sets of (points of) non-differentiability of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ have been fully characterized by Zahorski in (Zahorski 1946). One crucial step of Zahorski's argument was a construction of a monotone function not differentiable on a given G_δ null-set. This particular construction allowed to characterize sets of non-differentiability of monotone and Lipschitz real-valued functions of one variable. For functions of several variables, the situation is more complicated. It took more than 30 years and a lot of effort to characterize sets of non-differentiability of Lipschitz functions on \mathbb{R}^n . An analogous problem for monotone functions has not been studied. One of the major difficulties related to these questions is the fact that until now it was unclear how to generalize Zahorski's construction to higher dimensions. In this paper we use Brenier's theorem to generalize

Zahorski’s construction to higher dimensions and we argue that ours is the natural generalization.

The aforementioned problems of characterising sets of non-differentiability have natural counterparts at the interface of computable analysis and algorithmic randomness. From (Brattka et al. 2016), (Freer et al. 2014) and (Galicki 2015), we know that on the real line, differentiability properties of computable monotone, computable Lipschitz and computable convex functions are related to the notion of computable randomness. More precisely, the following has been proven:

Theorem 1.0.1 ((Brattka et al. 2016), (Freer et al. 2014) and (Galicki 2015)). *Let $z \in \mathbb{R}$. The following are equivalent:*

1. z is computably random,
2. every computable Lipschitz function $f : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at z ,
3. every computable monotone function $f : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at z , and
4. every computable convex function $f : \mathbb{R} \rightarrow \mathbb{R}$ is twice-differentiable at z .

The connection between computable randomness and those three classes of functions persists in higher dimensions too. An effective version of Rademacher’s theorem in (Galicki and Turetsky 2014) asserts that computable randomness implies differentiability of computable Lipschitz functions on \mathbb{R}^n , while an effective version of the Aleksandrov theorem in (Galicki 2015) states that computable randomness implies twice-differentiability of computable convex functions of several variables.

With the help of those two results and using an effective version of Brenier’s theorem, in this paper we will generalize three of the implications of Theorem 1.0.1 to functions on \mathbb{R}^n to prove the following theorem:

Theorem 1.2. *Let $z \in \mathbb{R}^n$. The following are equivalent:*

1. z is computably random,
2. every computable monotone function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at z , and
3. every computable convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice-differentiable at z .

Moreover, we will show several new characterisations of computable randomness in \mathbb{R}^n in terms of other analytical notions.

2. Background, preliminaries and notation

2.1 Optimal transport and Brenier’s theorem

The history of optimal transport history started in 1781 when a French mathematician Gaspard Monge formulated the problem of *déblais* (a hole in the ground) and *remblais* (a heap of soil). Given the shapes of *déblais* and *remblais*, there are many ways of matching the elements (of soil) from their initial positions in the *remblais* to their final positions in the *déblais*. The cost (for example, in terms of energy) of moving an element from *remblais* to *déblais* depends on both its initial position and on its final position. The Monge’s problem is: to find an optimal (with respect to some cost function) transportation plan of elements from *remblais* to *déblais*.

In the 1940s, the Soviet mathematician Leonid Kantorovich reformulated and extended the original Monge problem. In the more modern formulation, both *déblais* and *remblais* are modeled as probability measures μ and ν on some topological spaces X and Y , while the cost of moving is modeled as some function $c : X \times Y \rightarrow \mathbb{R}$. The *Monge-Kantorovich problem* is to find a probability measure π on $X \times Y$ with marginals μ and ν minimizing the total cost $\int_{X \times Y} c(x, y) d\pi(x, y)$. This is a relaxation of

the original Monge problem, as transportation plan is modeled by a probability measure, not a map from X to Y . For several decades the theory of optimal transport has been used in many problems arising in probability theory, economics and statistical mechanics.

Relatively recently the field of optimal transport gained a new and enormous popularity, mainly due to numerous (new) discoveries of connections to other areas of mathematics. According to Villani (Villani 2003), this new popularity can be traced to a single short note by Yann Brenier (Brenier 1987). What is now known as Brenier’s theorem (a result to which several other persons made significant contributions) states that for a quadratic cost function and under some mild assumptions on μ and ν , an optimal transport map exists, it is, in some specific way, unique and it is monotone.

Since this development, the theory of optimal transport has found a multitude of applications in many fields, both theoretical and practical. Our paper contributes to this trend, as we present an effective version of Brenier’s theorem and show how this theorem can be used to solve non-trivial problems at the interface of computable analysis and algorithmic randomness. While the main focus of the paper is on the mentioned applications, it is worth noting that an effective version of Brenier’s theorem is interesting on its own. To our knowledge, ours is the first effectivization, in the sense of computability theory, of Brenier’s theorem. In the case of discrete measures, finding an optimal transport is an integer programming problem, for which efficient algorithms exist. But in non-discrete cases, the closest to effectivization of Brenier’s theorem are (relatively) few papers on numerical solutions to the optimal transport problem. This will be discussed briefly in the next subsection.

For more information on optimal transport, see two recent books by Villani (Villani 2003, 2009).

2.2 Effective vs. numerical

In the context of this paper we use the term “effective” in the usual sense of computability theory. It means that various notions of computability that we use are rigorously defined and every time we use the word “effective”, it refers to one (or several) of those formally defined notions.

While ours is the first paper studying optimal transport from the perspective of computability theory, there is a growing number of recent papers which study optimal transport from the point of view of numerical methods. These are related areas, where computability theory emphasises rigorous approach to effectiveness while numerical methods’ focus is on efficiency and applicability. It is often possible to derive a computability theorem from a numerical algorithm and vice versa. When both measures are discrete, if there is an optimal solution to the Monge problem, it can be found using one of the known optimisation algorithms, for example, using the auction algorithm. In this paper we are more interested in the case when at least one of the measures is not discrete. This case is not as well understood as the “discrete” one.

In our opinion, two recent companion papers by Benamou, Froese and Oberman, (Benamou et al. 2014) and (Benamou et al. 2013), describe results in numerical methods which are quite close to this paper. From our point of view, there are two interesting results in those papers. Firstly, they describe a numerical method for approximating solutions to a particular Monge problem. That is, for a given ϵ , they describe a numerical algorithm for finding an approximation u^ϵ of the optimal map u . Secondly, they prove that the convergence $u^\epsilon \rightarrow u$ is uniform as $\epsilon \rightarrow 0$. To get a computability theorem out of these results, it would be necessary to show that $\epsilon \mapsto u^\epsilon$ is effective uniformly in ϵ and, more importantly, to assess effectiveness of the $u^\epsilon \rightarrow u$ convergence. The authors remark that “[their result] does not give a rate of convergence, which is typical of this kind of convergence result, since viscosity solutions can be singular.”

2.3 “Almost everywhere” theorems in analysis and in algorithmic randomness

An *almost everywhere theorem* in classical analysis is a result of the following form:

(*) given an object o of class C , the property $P(o, x)$ holds for almost all $x \in X$ with respect to some measure μ on X .

It can be interpreted as saying that objects of class C exhibit a high degree of regularity with respect to property P : given $o \in C$, its *set of irregularity*, $\{x : \neg P(o, x)\}$, is negligible. For a result in such a form, it is natural and often important to study *the converse problem*: given a negligible set N , is there an object of class C , whose set of irregularity contains N ? If the answer is positive, we say, *the converse holds*. For purposes of the paper, an almost everywhere theorem together with the corresponding converse result will be referred to as a *two-directional theorem*.

Algorithmic randomness is an area of mathematics that formalizes the intuitive notion of randomness via a combination of computability theory and measure theory. Intuitively, a real number is random if it does not have any exceptional properties. This approach can be formalized by identifying exceptional properties with effective null sets. Different types of effective null sets correspond to different notions of algorithmic randomness. For an overview of the field, see (Nies 2009) and (Downey and Hirschfeldt 2010).

Computable analysis (Weihrauch 2000) is an area of mathematics dedicated to studying effective versions of notions and results from classical analysis. Effective versions of almost everywhere theorems and their respective converses often provide analytical characterisations of known algorithmic randomness notions and come in the following (two-directional) form:

$x \in R(X) \iff$ for all effective $o \in C$, $P(o, x)$ holds (where $R(X)$ is some algorithmic randomness notion on space X).

Differentiability of functions on \mathbb{R}^n is an area with numerous a.e. theorems that has been extensively studied by both classical analysts and researchers of algorithmic randomness. However, the situation is markedly different for functions on the real line and for functions of several variables: there are many, both effective and classical, two-directional results about functions of one variable and very few about functions of several variables. The main reason for this discrepancy is that in higher dimensions it is very difficult to construct a function with given geometric properties that is not differentiable on points of a given null-set and such a construction is a necessary ingredient for any converse result. On the real line, the construction of Zahorski (Zahorski 1946) provides a highly flexible template. Most constructions on the real line, both effective and classical, are variations of that kind. Until now there was no known generalisation of Zahorski’s construction for functions in higher dimensions. In this paper we argue that Brenier’s theorem provides the natural generalisation of Zahorski’s construction.

2.4 Classical results on almost everywhere differentiability and respective converse problems

It has long been known that certain important classes of functions on \mathbb{R}^n , that are well-behaved in terms of geometric properties, are also well-behaved in terms of differentiability. The notable examples relevant to this paper are:

1. Lipschitz functions on \mathbb{R}^n are almost everywhere differentiable via Rademacher’s theorem, see (Rademacher 1919),
2. monotone functions on \mathbb{R}^n are almost everywhere differentiable too by the result of Mignot from 1976, see (Mignot 1976); and
3. convex functions are almost everywhere twice differentiable by the Aleksandrov theorem, see (Aleksandrov 1939).

All of the mentioned above classical results are formulated as a.e. theorems without converses. In the case of Rademacher’s theorem,

the converse asks whether for a given null-set $A \subseteq \mathbb{R}^n$, there exists a Lipschitz function f whose set of non-differentiability contains A . On the real line, the (positive) answer to this question is relatively simple and follows from the result of Zahorski (Zahorski 1946). Since the early work of D. Preiss (Preiss 1990), it has been known that the straightforward generalization of Zahorski’s result to higher dimensions does not hold: Preiss showed that for $n > 1$ there exists a null-set $A \subseteq \mathbb{R}^n$ that contains a point of differentiability of every real valued Lipschitz function. This surprising discovery started a 30-years long quest to characterise sets of non-differentiability of Lipschitz functions in higher dimensions and to provide a more nuanced converse to Rademacher’s theorem. As of now, the problem is considered solved, but not all of the papers contributing to the solution have been published so far. For further reference see (Preiss and Csörnyei 2011) and (Preiss and Speight 2015). It turns out the converse holds for functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ if and only if $m \geq n$. A major technical problem with showing this result is to construct, for a given null-set $A \subseteq \mathbb{R}^n$, a Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which is not differentiable at every point of A . On the real line, this construction is very simple, but it does not generalise to higher dimensions. The construction for higher dimensions is quite complicated (Giovanni Alberti, personal communication, March 23, 2015) and has still not been published after more than 10 years since announcement of the result. The converse problems for the other two theorems (Aleksandrov’s and Mignot’s ones) have not been studied in analysis.

In the effective setting, the question of characterising sets of non-differentiability of well-behaved functions can easily be seen as related to algorithmic randomness. The case of real valued functions of one variable is fairly well understood ((Brattka et al. 2016), (Freer et al. 2014), (Galicki 2015)). Many natural randomness notions have characterisations in terms of differentiability of effective functions of one variable. In the case of functions of several variables, analogous results are much rarer: an effective version of Rademacher’s theorem (without a converse) has been proven in (Galicki and Turetsky 2014) and an effective version of the Aleksandrov theorem (without a converse) has been described in (Galicki 2015). The main problem with proving converse results in the effective setting is the same as in classical analysis: no known generalization of Zahorski’s construction to higher dimensions.

2.5 Monotone, convex and Lipschitz functions

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a function. We say f is *monotone* if

$$\langle f(x) - f(y), x - y \rangle \geq 0 \text{ for all } x, y \in \mathbb{R}^n.$$

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *convex* if the following condition holds for all $x_0, x_1 \in \mathbb{R}^n$ and all $t \in [0, 1]$:

$$f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1).$$

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is *Lipschitz* if there exists $L \in \mathbb{R}^+$ such that

$$\|f(x) - f(y)\| \leq L \cdot \|x - y\| \text{ for all } x, y \in \mathbb{R}^n.$$

The least such L is called *the Lipschitz constant* for f . We denote it by $\text{Lip}(f)$.

2.6 Differentiability of functions and measures

Definition 2.6.1. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a function and let $x \in \mathbb{R}^n$. We say f is *differentiable* at x if for some linear operator T on \mathbb{R}^n the following holds:

$$\lim_{h \rightarrow 0} \frac{\|f(x+h) - f(x) - T \cdot h\|}{\|h\|} = 0. \quad (1)$$

Then, by definition, the derivative of f at x , $Df(x)$, is equal to T :

$$Df(x) = T.$$

Notation-wise, we use both $Df(x)$ and $f'(x)$ for the derivative of f at x .

A separate clarification is needed for what twice-differentiability means in the statement of the Aleksandrov theorem. Note that the Definition 2.6.1 is formulated for total functions. Convex functions on \mathbb{R}^n are a.e. differentiable in the usual sense. Hence their derivatives are well defined a.e. and are not, in general, total functions. To talk about twice-differentiability of convex functions rigorously, first we need to define differentiability for partial, a.e. defined functions. Differentiability of f at x in the *sense of Aleksandrov* is defined in the same way as in the Definition 2.6.1 with two differences:

1. $f(x)$ must be defined, and
2. in the limit (1), only those values of h are considered for which $f(x + h)$ is defined.

Subtle differences between the two notions of differentiability are not relevant for this paper and from now on we will conflate them and will write about differentiability as if there was only one notion.

For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $x \in \mathbb{R}^n$, its *gradient at x* , that is the matrix of partial derivatives of f at x , is denoted by $\nabla f(x)$. To deal with a.e. differentiability of convex functions, we need the following notion.

Definition 2.6.2. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function and let $x \in \mathbb{R}^n$. The *subdifferential of f at x* , $\partial f(x)$, is defined by

$$\partial f(x) = \{y : \forall z \in \mathbb{R}^n \ f(z) \geq f(x) + \langle y, z - x \rangle\}.$$

∂f is a set-valued function. Moreover, for a convex f , $\partial f(x)$ is a singleton if and only if f is differentiable at x and then $\partial f(x) = \{\nabla f(x)\}$.

Somewhat abusing notation we denote Lebesgue measures on \mathbb{R}^n and 2^ω by λ . We let λ_n be the probability measure on \mathbb{R}^n whose density is $1_{[0,1]^n}$.

Definition 2.6.3. Let μ be a measure on \mathbb{R}^n . The (*symmetric*) *derivative of μ with respect to λ at $z \in \mathbb{R}^n$* , $D_\lambda \mu(z)$, is defined by

$$D_\lambda \mu(z) = \lim_{r \rightarrow 0} \frac{\mu(B_r(z))}{\lambda(B_r(z))},$$

where $B_r(z)$ denotes the open ball with radius r and center z .

2.7 Computable functions on \mathbb{R}^n

There are multiple ways of formalizing computability of real functions, most of which turned out to be equivalent. For more details, see (Pour-El and Richards 1988) and (Weihrach 2000).

In this paper we rely on the Grzegorczyk-Lacombe notion of computability. The formal details are following.

A sequence $(q_i)_{i \in \mathbb{N}}$ of elements of \mathbb{R}^n is a *Cauchy name* if all coordinates of all q_i are rational and if $\|q_k - q_n\| \leq 2^{-n}$ for all n, k with $k \geq n$. If $\lim_{n \rightarrow \infty} q_n = x$, then we say that $(q_i)_{i \in \mathbb{N}}$ is a Cauchy name for x .

We say $x \in \mathbb{R}^n$ is *computable* if there is a computable Cauchy name for x .

Definition 2.7.1. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is *computable* if:

1. $f(q)$ is computable (uniformly in q) where q has all dyadic rational coordinates; and
2. f is *effectively uniformly computable*, that is if there is a computable function $h : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ such that for all $i, j \in \mathbb{N}$ all

$$\begin{aligned} x, y \in \mathbb{R}^n, \\ \|x\| \leq j \wedge \|y\| \leq j \wedge \|x - y\| \leq 2^{-h(i,j)} \text{ implies} \\ \|f(x) - f(y)\| \leq 2^{-i}. \end{aligned}$$

The above definition requires computable functions to be continuous and for this reason it is not well-suited for studying discontinuous functions in the effective setting. Monotone functions can be discontinuous, but their sets of discontinuity are Lebesgue null-sets. This suggests the following definition (first used in (Galicki 2015)).

Definition 2.7.2. We say a monotone function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is *effectively monotone* if $f(x)$ is computable uniformly in x when f is continuous at x .

It can easily be seen that Definition 2.7.2 is strictly weaker than Definition 2.7.1. Since monotone functions are almost everywhere continuous, effectively monotone functions are a.e. computable (see subsections 7.1 and 7.2 in (Rute)).

The following fact is a simple consequence of results on the real line from (Galicki 2015).

Fact 2.7.3. Let $u : \mathbb{R}^n \rightarrow \mathbb{R}$ be a computable convex function. Its gradient ∇u is an effectively monotone function.

2.8 Computable randomness

Computable randomness is one of the more natural notions of algorithmic randomness. It is usually defined in terms of success sets of effective betting strategies: a sequence is said to be computably random if no computable betting strategy can make an unbounded profit while betting on this sequence.

Betting strategies are usually formalized as *martingales* (see (Nies 2009), Chapter 7).

Definition 2.8.1. We say a function $B : 2^{<\omega} \rightarrow \mathbb{Q}_+$ is a *martingale* if the following condition holds for all $\sigma \in 2^{<\omega}$:

$$2B(\sigma) = B(\sigma 0) + B(\sigma 1).$$

$B(\sigma)$ can be interpreted as the value of capital after betting on bits of σ . We say B *succeeds* on $Z \in 2^\omega$ if $\limsup_n B(Z \upharpoonright_n) = \infty$.

We say B *diverges* on Z if

$$\liminf_n B(Z \upharpoonright_n) < \limsup_n B(Z \upharpoonright_n).$$

The *divergence set* of B is defined by

$$D(B) = \{Z : \liminf_n B(Z \upharpoonright_n) < \limsup_n B(Z \upharpoonright_n)\}.$$

Definition 2.8.2. We say $Z \in 2^\omega$ is *computably random* if no computable martingale succeeds on Z .

We say $z = (0.Z_1, \dots, 0.Z_n) \in [0, 1]^n$ is computably random if its binary expansion, that is $Z = Z_1 \oplus \dots \oplus Z_n$, is computably random. Here $0.A$ denotes the real number whose binary expansion is $A \in 2^\omega$.

It is known that divergence sets of computable martingales provide an alternative characterization of computable randomness: $Z \in 2^\omega$ is computably random $\iff Z$ does not belong to the divergence set of a computable martingale.

2.9 Computable measures

In order to write rigorously about computability of measures, we rely on the approach from (Gács 2005) and (Hoyrup and Rojas 2009). Informally, an atomless probability measure μ on \mathbb{R}^n is computable if there is an algorithm computing the value $\mu(Q)$ for every cube with dyadic endpoints Q .

2.10 Transfer maps

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a map. We say T is a *transport map* from μ to ν , or that T *transports* μ onto ν (in symbols, $\nu = T\#\mu$), if for all measurable A , $\nu(A) = \mu(T^{-1}(A))$.

3. Computable randomness implies differentiability of computable monotone functions from \mathbb{R}^n to \mathbb{R}^n

In non-effective setting, a.e. differentiability of monotone function from \mathbb{R}^n to \mathbb{R}^n has been proven by Mignot (Mignot 1976), who used Rademacher's theorem and a fact about monotone functions discovered by Minty (Minty 1962).

In this section we will prove the (1) \implies (2) implication of Theorem 1.2: that computable randomness implies differentiability of computable monotone real functions of several variables. Our proof follows the path similar to the one taken by Mignot - we use the effective form of Rademacher's Theorem from (Galicki and Turetsky 2014) and the following correspondence observed by Minty.

3.1 Minty parameterization

Minty showed that the so called Cayley transformation

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n \text{ defined by } \Phi(x, y) = \frac{1}{\sqrt{2}}(y+x, y-x)$$

transforms the graph of a monotone function into a graph of a 1-Lipschitz function. Note that when $n = 1$ this is a clockwise rotation of $\pi/4$. In general the graph of a monotone function is a Lipschitz manifold. We will rely on the following consequences of Minty's discovery.

Notation 3.1.1. In the rest of the paper we denote the identity function on \mathbb{R}^n by I .

Proposition 3.1.2 (cf. Proposition 1.2 in (Alberti and Ambrosio 1999)). *Let $u : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be monotone. Then $(u + I)$ and $(u + I)^{-1}$ are monotone and $(u + I)^{-1}$ is 1-Lipschitz.*

Proposition 3.1.3 (cf. Theorem 12.65 in (Rockafellar and Wets 1997)). *Let $u : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a continuous monotone function. Let $z \in \mathbb{R}^n$ and define $f = (u + I)^{-1}$ and $\hat{z} = u(z) + z$. The following two are equivalent:*

1. u is differentiable at z , and
2. f is differentiable at \hat{z} and $f'(\hat{z})$ is invertible.

This allows to “translate” questions about differentiability of monotone functions into questions about differentiability of Lipschitz functions.

Now we are ready to explain our proof of the main result of this section.

Overview of the proof

Let $u : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a monotone computable function and let $z \in \mathbb{R}^n$ be computably random. Then $g = u + I$ is monotone and computable and $f = g^{-1}$ is 1-Lipschitz and computable. If we can show that $g(z)$ is computably random, then f is differentiable at $g(z)$. By Proposition 3.1.3, if the derivative of f at $g(z)$ is invertible, then g is differentiable at z .

From the above description, it is clear that we require the following two ingredients to complete the proof:

(preservation property) we need to show that $g(z)$ is computably random when z is, and that

(singularity property) computable randomness of $g(z)$ implies that $f'(g(z))$ is invertible.

In the following two subsections we will prove both of the above.

3.2 Another preservation property

To prove the preservation property mentioned in the previous subsection, we require some terminology and notation from (Rute).

Firstly, we need to extend the notion of computable randomness to \mathbb{R}^n .

Definition 3.2.1. we say $z \in \mathbb{R}^n$ is computably random if its binary expansion (or, equivalently, its fractional part) is computably random.

When $z \in [0, 1]^n$, this characterisation is equivalent to the Definition 2.8.2. Otherwise, when $z \notin [0, 1]^n$, this characterisation is equivalent to computable randomness on some computable translation of the unit cube equipped with the usual Lebesgue measure.

Notation 3.2.2. For every $n \geq 1$, let \mathcal{A}_n be some fixed a.e. decidable cell decomposition of $[0, 1]^n$. For the sake of simplifying the notation, in the rest of this paper, for all $n \geq 1$ and all $\sigma \in 2^{<\omega}$, we denote the cell $[\sigma]_{\mathcal{A}_n}$ by $[\sigma]$.

Definition 3.2.3. A *Martin-Löf test* is a uniformly computable sequence $(U_i)_{i \in \mathbb{N}}$ of Σ_1^0 subsets of $[0, 1]^n$ such that $\lambda(U_i) \leq 2^{-i}$ for all i . We say $(U_i)_{i \in \mathbb{N}}$ *covers* $z \in [0, 1]^n$ if $z \in \bigcap_i U_i$. We say a Martin-Löf test $(U_i)_{i \in \mathbb{N}}$ is *bounded* if there is a computable measure $\nu : 2^{<\omega} \rightarrow [0, \infty)$ satisfying

$$\lambda(U_i \cap [\sigma]) \leq 2^{-i} \nu(\sigma)$$

for all $i \in \mathbb{N}$ and $\sigma \in 2^{<\omega}$.

We require the following characterisation of computable randomness in the unit cube due to Rute:

Proposition 3.2.4 (cf. Theorem 5.3 in (Rute)). *Let $z \in [0, 1]^n$. The following two are equivalent:*

1. z is not computably random, and
2. either z is an unrepresented point, or there is a bounded Martin-Löf test $(U_i)_{i \in \mathbb{N}}$ that covers z .

Remark 3.2.5. For our considerations it is sufficient to know that if z is an unrepresented point, then it is not weakly random.

We are now in position to state and to prove the required preservation property for computable randomness.

Lemma 3.2.6. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a computable injective Lipschitz function and suppose $z \in \mathbb{R}^n$ is not computably random. Then $f(z)$ is not computably random either.*

Proof. Without loss of generality we assume $z \in [0, 1]^n$, $f(z) \in [0, 1]^n$ and $\text{Lip}(f) \leq 1$ (otherwise we may consider $\hat{f}(x) = A \cdot f(x) + B$ for some suitable computable A and B).

Firstly, suppose z is an unrepresented point. Let $P \subset [0, 1]^n$ be a Π_1^0 null set with $z \in P$. Then $f(z) \in f(P) \cap [0, 1]^n$ and since $f(P) \cap [0, 1]^n$ is also a Π_1^0 null set, $f(z)$ is not weakly random.

Let $(V_i)_{i \in \mathbb{N}}$ be a bounded Martin-Löf test with $z \in \bigcap_i V_i$ and let ν be a computable measure such that $\lambda(V_i \cap [\sigma]) \leq 2^{-i} \nu(\sigma)$ for all i, σ .

Define $U_i = f(V_i) \cap [0, 1]^n$ for all i . Since $\lambda(U_i) \leq \lambda(V_i)$ (see Lemma 3.10.12 in (Bogachev 2007)) and f is injective, $(U_i)_{i \in \mathbb{N}}$ is a Martin-Löf test.

Define $\nu_f = \nu \circ f^{-1}$. It is a computable measure and for all i, σ we have

$$\begin{aligned} \lambda(U_i \cap [\sigma]) &= \lambda(f(V_i) \cap [\sigma]) = \\ &= \lambda(f(V_i \cap f^{-1}([\sigma]))) \leq 2^{-i} \nu(f^{-1}([\sigma])) = 2^{-i} \nu_f(\sigma). \end{aligned}$$

It follows that $(U_i)_{i \in \mathbb{N}}$ is a bounded Martin-Löf test that covers $f(z)$ and thus $f(z)$ is not computably random. \square

3.3 An effective version of Sard's theorem for Lipschitz functions

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Recall that $z \in \mathbb{R}^n$ is said to be a *critical point* of f if either f is not differentiable at z , or $\det f'(z) = 0$. If z is a critical point of f , then $f(z)$ is said to be a *critical value* of f .

The main result in this subsection, Theorem 3.3.2, can be seen as an effective version of Sard's theorem for Lipschitz function. Its classical version, proven by Mignot ((Mignot 1976), also see Theorem 9.65 in (Rockafellar and Wets 1997)), states that for a Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the set of its critical values is a null-set.

Lemma 3.3.1. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a Lipschitz function. Suppose $z \in \mathbb{R}^n$ is such that $f'(z)$ is singular. Then for every $\epsilon > 0$, there exists an open neighbourhood of z , O_ϵ such that $\lambda(f(O_\epsilon)) \leq \epsilon \lambda(O_\epsilon)$.*

Proof. Fix $\epsilon > 0$ and let $k = \text{Lip}(f)$.

Define $\epsilon' = \frac{\epsilon}{k^{n-1}2^n(\sqrt{n})^n}$. Since f is differentiable at z , there exists $\delta > 0$ such that

$$\|f(x) - f(z) - f'(z)(x - z)\| \leq \epsilon' \|x - z\| \quad (2)$$

for all $x \in \mathbb{R}^n$ with $|x - z| \leq \delta$. There is an open n -cube C with side length equal to $s = \frac{\delta}{\sqrt{n}}$ such that $z \in C$ and (2) holds for all $x \in C$.

Let L be the mapping defined by $L(x) = f(z) + f'(z)(x - z)$. Since $f'(z)$ is singular, L is not onto and its range is contained in some hyperplane H .

As a consequence of (2) we have $\|f(x) - L(x)\| \leq \epsilon' \delta$ for all $x \in C$. Thus, $f(C) \subseteq L(C) + [-\epsilon' \delta, \epsilon' \delta]^n$. Since L is a k -Lipschitz mapping, the image of C under L lies in the intersection of H with a closed ball with radius $k\delta$ centered at $f(z)$. Then $L(C)$ is contained in a rotated $(n-1)$ -dimensional cube of side $2k\delta$. This shows that $f(C)$ lies in a rotated box \hat{C} with

$$\lambda(\hat{C}) = (2k\delta)^{n-1} 2\epsilon' \delta = 2(2k)^{n-1} \epsilon' (\sqrt{n})^n \left(\frac{\delta}{\sqrt{n}}\right)^n = \epsilon \lambda(C).$$

\square

Theorem 3.3.2 (Effective Sard's theorem for Lipschitz functions). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a computable Lipschitz function and let $z \in \mathbb{R}^n$. If $f(z)$ is computably random, then it is not a critical value of f .*

Proof. The proof is by contraposition. If f is not differentiable at z , then, by the Effective Rademacher Theorem from (Galicki and Turetsky 2014), z is not computably random.

Suppose $f'(z)$ is singular. Let $K_z \subset \mathbb{R}^n$ be the basic dyadic unit cube containing z and let $K_{f(z)} \subset \mathbb{R}^n$ be the basic dyadic unit cube containing $f(z)$.

Let $\nu = \lambda \circ f^{-1}$ and for every $i \in \mathbb{N}$, define $V_i \subseteq K_{f(z)}$ as the union of all $[\sigma]$ such that $\lambda(\sigma) \leq 2^{-i} \nu(\sigma)$ and $f^{-1}([\sigma]) \subseteq K_z$.

Fix i and τ . Let $(\eta_i)_{i \in \mathbb{N}}$ be such that $\bigcup_j [\eta_j] = [\tau] \cap V_i$ and $[\eta_m] \cap [\eta_k] = \emptyset$ for all $k \neq m$. Note that $\lambda(V_i) \leq 2^{-i}$ and

$$\lambda(V_i \cap [\tau]) = \sum_k \lambda(\eta_k) \leq 2^{-i} \sum_k \nu(\eta_k) \leq 2^{-i} \nu(\tau).$$

Thus $(V_i)_{i \in \mathbb{N}}$ is a bounded Martin-Löf test and, by Lemma 3.3.1, it covers $f(z)$. Hence $f(z)$ is not computably random. \square

Later in the paper, using an effective version of Brenier's theorem, we will prove the converse for this result.

3.4 Main result

We are now ready to formulate and prove the main result of this section.

Theorem 3.4.1. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an computable monotone function and let $z \in [0, 1]^n$ be computably random. Then f is differentiable at z .*

Proof. Define $g = (f + I)^{-1}$, so that g is a computable Lipschitz function with $\text{Lip}(g) \leq 1$.

Let $y = f(z) + z$ so that $g(y) = z$. By Lemma 3.2.6, y is computably random and hence g is differentiable at y and by Theorem 3.3.2 $g'(y)$ is invertible. Hence, by Proposition 3.1.3, f is differentiable at z . \square

4. Zahorski's construction and the effective Brenier theorem

4.1 Zahorski's construction

Generalizing (a part of) Zahorski's construction was the main application we had in mind for Brenier's theorem. Hence, before discussing optimal transport and the theorem itself, we briefly review the said construction and its effectivization.

A modern version of Zahorski's argument can be seen in (Fowler and Preiss 2009). Zahorski characterised sets of non-differentiability of continuous real-valued functions on the real line. As was mentioned in the Introduction, a crucial part of his argument was constructing a monotone Lipschitz function not differentiable at a given G_δ null set. For the purposes of this paper we will call this part *the Zahorski construction*, even if Zahorski's argument was more complicated and included other important constructions. Note that generalising this part to higher dimensions would provide a natural converse both for Rademacher's theorem and for Mignot's result on differentiability of monotone functions.

Classical argument Let $A \subset \mathbb{R}$ be a G_δ null-set. The goal is to construct a monotone Lipschitz function not differentiable at any point of A . The very basic idea is to define a bounded measurable function $G : \mathbb{R} \rightarrow \mathbb{R}$ that is not continuous at any point of A and let $F(x) = \int_a^x G(y) dy$ for some fixed a . Then discontinuity of G at any point $x \in \mathbb{R}$ corresponds to non-differentiability of F at x . F is obviously monotone and boundedness of G makes F Lipschitz.

Remark 4.1.1. It is worth mentioning that the original result was formulated for monotone functions. The construction is easily extended to monotone Lipschitz functions by requiring G to be bounded. This paper, however, is focused on monotone functions. As will be discussed later, the same approach (bounding the density) in higher dimensions guarantees Hölder continuity rather than Lipschitz continuity.

Now let us overview an effectivization of this idea from (Brattka et al. 2016). The effective construction is virtually the same, but, as we will see, effectivization forces a very revealing point of view.

Effective Zahorski construction on the real line The starting point of the effective construction is slightly different. Instead of an arbitrary G_δ null-set, we have an effective null-set corresponding to a test of computable randomness. An educated guess, that it has to be computable randomness, rather than some other randomness notion, comes from a careful analysis of an effective version of the "other direction" (that is, the statement "monotone functions $f : \mathbb{R} \rightarrow \mathbb{R}$ are a.e. differentiable").

Let M be a martingale and define a computable probability measure on the unit interval by $\mu_M([\sigma]) = M(\sigma) \cdot 2^{-|\sigma|}$ and let

$$f(x) = \mathbf{cdf}_{\mu_M} = \int_0^x D_{\lambda} \mu_M d\lambda.$$

Then f is not differentiable at any $z \in [0, 1]$ whose binary expansion Z belongs to the divergence set of M . In this construction, the density of μ_M plays the role of G and f corresponds to F . It is important to analyze how divergence of M at Z corresponds to non-differentiability of f at z . It can be seen as a two step process:

- (S1) divergence of M forces non-differentiability of μ_M (and discontinuity of $D_{\lambda} \mu_M$),
- (S2) which, in turn, forces non-differentiability of f (at z).

Two properties of f that are crucial for S2 to work are:

1. $\lambda_1 = f\#\mu_M$ (f is a transfer map from μ_M to λ_1) and
2. monotonicity of f (hence f maps intervals to intervals).

For any function f with such properties, “oscillations” of measure μ_M around z become oscillations of the slope of f in a neighbourhood of z . More formally, we have

$$\frac{\mu_M(B_r(z))}{\lambda(B_r(z))} = \frac{\lambda(f(B_r(z)))}{\lambda(B_r(z))} = \frac{f(z+r) - f(z-r)}{2r}.$$

Taking the limit with $r \rightarrow 0$ makes the left side $D_{\lambda} \mu_M(z)$ and the right side $f'(z)$ (when they exist).

Now we can reformulate the idea behind Zahorski’s construction that will allow us to generalize it to higher dimensions. The idea is: *given a null-set A , find a probability measure μ_A that is not differentiable at points of A and a monotone transfer map f_A from μ_A to λ_1 . Then f_A is not differentiable at any point of A .*

Now we are ready to discuss the subject of optimal transfer and Brenier’s theorem.

4.2 Optimal transportation

Let μ, ν be probability measures on \mathbb{R}^n . A probability measure π on $\mathbb{R}^n \times \mathbb{R}^n$ is said to have *marginals* μ and ν when the following holds for all measurable $A, B \subseteq \mathbb{R}^n$:

$$\pi[A \times \mathbb{R}^n] = \mu[A], \text{ and } \pi[\mathbb{R}^n \times B] = \nu[B].$$

Let $\Pi(\mu, \nu)$ denote the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are μ and ν . Note that this set is always nonempty. For a given *cost function* $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ and $\pi \in \Pi(\mu, \nu)$, define the *total transportation cost* $I_c[\pi]$ as

$$I_c[\pi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y).$$

The *optimal transportation cost* between μ and ν is the value

$$\mathbb{I}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I_c[\pi].$$

Elements of $\Pi(\mu, \nu)$ are called *transference plans*. We are interested in transference plans induced by measurable maps, that is, plans of the form $\pi_T = (I \times T)\#\mu \in \Pi(\mu, \nu)$ where $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a measurable map. The total transportation cost associated with a transport map T is

$$I_c[T] = I_c[\pi_T] = \int_{\mathbb{R}^n} c(x, T(x)) d\mu(x).$$

A transport map T for which the cost is optimal, that is for which $I_c[\pi_T] = \mathbb{I}_c(\mu, \nu)$, is called an *optimal transport map*. The problem of minimizing $I_c[T]$ over the set of all transfer maps is known as the *Monge optimal transportation problem*. In general, such a

map is not guaranteed to exist. However, under some assumptions on μ, ν and c , it does exist.

The following important result is known as Brenier’s theorem and it lies at the heart of our construction.

Theorem 4.2.1 (A version of Brenier’s theorem, cf. Theorem 2.12 in (Villani 2003)). *Let μ, ν be probability measures on \mathbb{R}^n . Suppose μ is absolutely continuous (with respect to the Lebesgue measure) and the following holds:*

$$\int_{\mathbb{R}^n} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2} d\nu(y) < \infty.$$

Then there exists a convex function ϕ such that $\nabla \phi \# \mu = \nu$. Moreover, $\nabla \phi$ is the unique (i.e. uniquely determined μ -almost everywhere) function which pushes μ onto ν that is the gradient of a convex function.

Now we are ready to review Zahorski’s construction on the real line in the context of optimal transport theory.

4.3 Interpretation of Zahorski’s construction with respect to Brenier’s theorem

If μ and ν are two probability measures on \mathbb{R} with μ being absolutely continuous, the optimal transfer map (w.r.t. quadratic cost) from μ onto ν has a known closed form. It is

$$f = (\mathbf{cdf}_{\nu})^{-1} \circ \mathbf{cdf}_{\mu}$$

where \mathbf{cdf}_{μ} and \mathbf{cdf}_{ν} are respective cumulative distribution functions of μ and ν . This fact was already known to Hoeffding and Fréchet. For a proof and more details, see Section 2.2 in (Villani 2003).

The function f from the effective Zahorski construction described in Subsection 4.1 happens to be the optimal transfer map (with respect to the quadratic cost) from μ_M onto λ_1 .

Brenier’s theorem guarantees the existence of an optimal monotone map in higher dimensions too. However, it doesn’t give a closed form expression for such a map.

4.4 A generalised Zahorski construction for monotone maps

Let $A \subset [0, 1]^n$ be a G_{δ} null-set. The goal is to exhibit a monotone function $f_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which is differentiable at no point of A . Let μ_A be an absolutely continuous probability measure on \mathbb{R}^n such that it is differentiable at no point of A .

The main idea is that just like in the case of Zahorski’s construction on the real line, f_A can be defined as the optimal transfer map from μ_A onto λ_n . The following classical result is needed to show that such f_A is differentiable at no point of A .

Theorem 4.4.1 (Jacobian theorem for monotone maps, cf. Theorem A.2 in (McCann 1997)). *Let ϕ be a convex function on \mathbb{R}^n and suppose it is twice differentiable at $x \in \mathbb{R}^n$. Then*

$$\lim_{r \rightarrow 0} \frac{\lambda(\partial \phi(B_r(x)))}{\lambda(B_r(x))} = \det D^2 \phi(x).$$

Remark 4.4.2. Theorem 4.2.1 shows that there exists a unique (that is, μ_A almost everywhere unique) monotone function $f_A = \nabla \phi_A$, such that $f_A \# \mu_A = \lambda_n$, where ϕ_A is some convex function. Since μ_A is absolutely continuous, by Theorem 2.12 (iv) and Lemma 4.6 from (Villani 2003), $\lambda(\partial \phi_A(X)) = \lambda(\nabla \phi_A(X))$ for all Borel $X \subseteq \mathbb{R}^n$. Hence

$$\lim_{r \rightarrow 0} \frac{\lambda(\partial \phi_A(B_r(x)))}{\lambda(B_r(x))} = \lim_{r \rightarrow 0} \frac{\lambda(\nabla \phi_A(B_r(x)))}{\lambda(B_r(x))} = D_{\lambda} \mu(x)$$

for all x where the limit exists. By Theorem 4.4.1, $\nabla \phi_A$ is not differentiable at z when $D_{\lambda} \mu_A(z)$ does not exist.

4.5 An effective version of Brenier's theorem

Since we want to effectivize the generalized Zahorski construction described previously, we need to effectivize the most important part of it: Brenier's theorem. Given two computable probability measures, the optimal map between them can not, in general, be computable. The main reason for this is that an optimal map can be discontinuous. However, under some conditions on support of both measures, the optimal map is effectively monotone in the sense of the Definition 2.7.2. The following is our effective version of Brenier's theorem.

Theorem 1.1 (The effective Brenier theorem). *Let μ, ν be absolutely continuous computable probability measures on \mathbb{R}^n such that $\text{supp}(\mu) = [0, 1]^n$ and the support of ν is bounded. There exists a computable convex function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\nabla\phi$ is the optimal transport map from μ to ν .*

As it stated, this result, when used in the generalized Zahorski construction, is sufficient to prove the converse for the effective version of the Aleksandrov theorem from (Galicki 2015). To get the converse for the Theorem 5.2.1, we need to exhibit a computable map, not just an effectively monotone one. To make sure the resulting optimal transfer is computable, we will use the so called *regularity theory* for the Monge-Ampère equation.

4.6 The Monge-Ampère equation.

The Monge-Ampère equation is a famous fully nonlinear elliptic partial differential equation of the following form:

$$\det D^2\phi(x) = F(x, \phi(x), \nabla\phi(x)).$$

It is known that convex functions guaranteed to exist by Brenier's theorem are generalised solutions of the following particular case of the Monge-Ampère equation:

$$\det D^2\phi(x) = \frac{f(x)}{g(\nabla\phi(x))} \quad (3)$$

where g and f correspond to respective densities of target and source measures.

There is a regularity theory for equations of this type developed by Caffarelli and Urbas. Naturally, this theory provides regularity estimates for optimal maps too. For more information see (Urbas 1997). Since optimal maps can be discontinuous, the corresponding convex functions are not smooth and can be considered as solutions of (3) only in a specific generalised sense. It is outside of scope of this paper to discuss different types of generalised solutions. However, we will rely of the fact that when both measures are absolutely continuous (which is always the case in this paper), the optimal map is an Aleksandrov solution of (3).

We will need the following result.

Theorem 4.6.1. [cf. Theorem 4.13 in (Villani 2003)] *Let ϕ be an Aleksandrov solution of*

$$\det D^2\phi = h$$

in a bounded set Ω . If h is bounded from above and below by some positive constants, then $\phi \in C^{1,\alpha}$ for some universal exponent α .

This means that if we manage to construct μ_A in such a way that its density is bounded from above and below, then f_A will be Hölder continuous. Finally, note that a Hölder continuous effectively monotone function is computable.

5. Applications of the effective Brenier theorem

5.1 Converse of the effective Aleksandrov theorem

The following result is the converse to the effective Aleksandrov theorem (Theorem 20 in (Galicki 2015)).

Theorem 5.1.1. *Let $z \in \mathbb{R}^n$ not be computably random. There exists a computable convex function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ not twice-differentiable at z .*

Before proceeding to the proof, we need to state the following simple but useful proposition.

Proposition 5.1.2. *Let $x \in \mathbb{R}^n$ not be computably random. There is an absolutely continuous computable probability measure ν on \mathbb{R}^n such that $D_\lambda\nu(x)$ does not exist.*

Proof. First, suppose all coordinates of x are computably random. Then the Theorem 5.3(4) (Rute) shows existence of a computable absolutely continuous probability measure ν and a sequence of basic dyadic cubes $[\sigma_i]$ containing x such that $\nu([\sigma_j]) \geq 2^j \lambda([\sigma_j])$ for all j . $D_\lambda\nu(x)$ does not exist since every $[\sigma_i]$ is contained in an open ball centered at x with radius less than $\sqrt{n}d_j$ where d_j is the side length of $[\sigma_j]$.

Now suppose the first coordinate of x , $x_1 \in \mathbb{R}$, is not computably random. Let ν_1 be a computable absolutely continuous probability measure not differentiable at x_1 . Then the product measure $\nu_1 \otimes \lambda^{n-1}$, where λ^{n-1} is the Lebesgue measure on \mathbb{R}^{n-1} , is a computable absolutely continuous probability measure not differentiable at x . \square

Proof of Theorem 5.1.1. Let $z \in [0, 1]^n$ not be computably random. We may assume that every coordinate of z is computably random. For suppose z_i is not computably random for some i . There exists a computable convex function $u : \mathbb{R} \rightarrow \mathbb{R}$ not differentiable at z_i . Then the function $(x_1, \dots, x_n) \mapsto u(x_i)$ is a computable convex function from \mathbb{R}^n to \mathbb{R} not twice-differentiable at z .

Let μ be an absolutely continuous computable probability measure on \mathbb{R}^n such that $D_\lambda\mu(z)$ does not exist. Without loss of generality we may assume that μ is supported on $[0, 1]^n$.

Let f be the optimal transfer map from μ onto λ_n . We know it is not differentiable at z and by the effective Brenier theorem there is a computable convex function $u : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\nabla u = f$. \square

5.2 Converse of Theorem 3.4.1

In this subsection we sketch a proof of the following converse to the Theorem 3.4.1.

Theorem 5.2.1. *Suppose $z \in \mathbb{R}^n$ is not computably random. There exists a computable monotone function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ not differentiable at z .*

For a given computable martingale M , we define a computable probability measure μ_M on $[0, 1]^n$ by

$$\mu_M([\sigma]) = \lambda([\sigma]) \cdot M(\sigma) \text{ for all } \sigma \text{ with } |\sigma| = ns \text{ for some } s.$$

Lemma 5.2.2. *Let M be a computable martingale and let μ_M be the corresponding probability measure on $[0, 1]^n$. Let $z \in [0, 1]^n$ and let Z be the binary expansion of z . Suppose the following two conditions hold:*

- (P1) *the measure μ_M is absolutely continuous, not differentiable at z , and*
- (P2) *$0 < M(\sigma) < C$ for some fixed C and all σ .*

Then the optimal transport map from μ_M to λ_n is computable.

Proof. We know that there is a computable convex function ϕ such that $\nabla\phi$ is the optimal transfer map of μ onto λ . Define $h(x) = D_\lambda\mu_M(x)$. Then ϕ is an Aleksandrov solution of the following instance of the Monge-Ampère equation:

$$\det D^2 f = h.$$

Since h is bounded away from both 0 and ∞ , by Theorem 4.6.1, ϕ is $C^{1,\alpha}$ for some $\alpha > 0$. Since $\nabla\phi$ is a.e. computable and Hölder continuous, it must be computable. \square

The last bit needed to complete the proof of Theorem 5.2.1 is the following technical lemma concerned with the existence of martingale M satisfying the conditions of the previous lemma.

Lemma 5.2.3. *Suppose $z \in [0, 1]^n$ is not computably random. There does exist a computable martingale M that satisfies both (P1) and (P2) conditions.*

The proof of this lemma is a variation of the corresponding one-dimensional construction from the proof of Theorem 4.2 in (Freer et al. 2014).

Theorem 5.2.1 was the last part needed to prove Theorem 1.2 stated in the introduction.

Theorem 1.2. *Let $z \in \mathbb{R}^n$. The following are equivalent:*

1. z is computably random,
2. every computable monotone function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is differentiable at z , and
3. every computable convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice-differentiable at z .

Proof. (1) \implies (2) follows from Theorem 3.4.1. (2) \implies (1) follows from Theorem 5.2.1.

(1) \implies (3) follows from the effective version of Aleksandrov theorem from (Galicki 2015). (3) \implies (1) follows from Theorem 5.1.1. \square

5.3 Differentiability of computable measures

With the results we have proven so far, we can easily get the following characterisation of computable randomness in \mathbb{R}^n in terms of differentiability of absolutely continuous computable probability measures on $[0, 1]^n$.

Theorem 5.3.1. *Let $z \in [0, 1]^n$.*

z is computably random \iff every absolutely continuous computable probability measure on $[0, 1]^n$ is differentiable at z .

Proof \implies . Let μ be an absolutely continuous computable probability measure on $[0, 1]^n$ and let $z \in [0, 1]^n$. Suppose $D_\lambda\mu(z)$ does not exist.

Define a probability measure on \mathbb{R}^n by $\hat{\mu} = \frac{1}{2}(\mu + \lambda_n)$. It is absolutely continuous and its support is equal to $[0, 1]^n$.

Consider an optimal transport map $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ from $\hat{\mu}$ to λ_n , so that $\lambda_n = f\#\hat{\mu}$. By the effective Brenier theorem, we may assume f is an effective monotone map. By Theorem 4.4.1, f is not differentiable at z and by Theorem 3.4.1 and Remark 4.4.2, z is not computably random. \square

Proof \Leftarrow . This follows from Proposition 5.1.2. \square

This theorem can be reformulated as an effective version the Lebesgue differentiation theorem for functions that are densities

of computable absolutely continuous measures on $[0, 1]^n$. It is not known if there is a more natural (from computability theory point of view) characterisation of such functions.

An analogous characterisation of Schnorr randomness has been proven in (Rute 2012) (for measures) and (Pathak et al. 2014) (for functions on \mathbb{R}^n). In that case the corresponding class of effective functions is very natural ($L1$ -computability), while the class of corresponding probability measures is somewhat artificial.

5.4 Critical values of computable Lipschitz functions

Theorem 5.4.1. *Let $z \in \mathbb{R}^n$. The following are equivalent:*

1. z is computably random, and
2. z is not a critical value of any computable Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Proof \implies . Follows from Theorem 3.3.2. \square

Proof \Leftarrow . Suppose $z \in \mathbb{R}^n$ is not computably random. By Theorem 5.2.1, there is some computable monotone function $u : \mathbb{R}^n \rightarrow \mathbb{R}^n$ not differentiable at z . Then

$$f = (u + I)^{-1}$$

is a computable Lipschitz function. Theorem 3.1.3 implies that $f(u(z) + z) = z$ is a critical value of f . \square

5.5 Monge-Ampère equation

Robert McCann in (McCann 1997) demonstrated that the Monge-Ampère equation of the type associated with Optimal Transport holds true almost everywhere. The following result is an effective form of a somewhat weaker result (we fix the target measure to be the Lebesgue one).

Theorem 5.5.1. *Let $z \in \mathbb{R}^n$. The following are equivalent:*

1. z is computably random, and
2. for every computable absolutely continuous probability measure μ and every computable (Aleksandrov) solution f of

$$\det D^2 f = D_\lambda\mu \tag{4}$$

the following holds

$$\det D^2 f(z) = D_\lambda\mu(z). \tag{5}$$

Proof \implies . Let μ be a computable absolutely continuous probability measure and let f be a computable weak solution of (4) such that (5) doesn't hold. We know f is convex and by Theorem 4.4.1 f is not twice differentiable at z . Hence, by the effective Aleksandrov theorem, z is not computably random. \square

Proof \Leftarrow . Suppose $z \in \mathbb{R}^n$ is not computably random. Then there is a computable absolutely continuous probability measure μ on $[0, 1]^n$ not differentiable at z . We may assume $\text{supp}(\mu) = [0, 1]^n$ so that we can apply Theorem 1.1. It follows that there is a computable $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which is an Aleksandrov solution of (4) such that (5) doesn't hold. \square

6. Further directions

This work is only a first step towards an effective theory of Optimal Transport. The effective version of Brenier's theorem stated in this paper is not the last word on the subject of effectivizing Brenier's result either, as our result was formulated with specific applications in mind rather than generality.

We believe that proving *efficient* (with explicit bounds on time/space) versions of Brenier's theorem is of particular interest, both theoretical and practical.

The subject of connections between computable randomness and computable Lipschitz functions on \mathbb{R}^n is also not closed. Generalising Zahorski's construction to Lipschitz functions on \mathbb{R}^n and proving the converse of the Effective Rademacher's Theorem are two important open problems.

References

- G. Alberti and L. Ambrosio. A geometrical approach to monotone functions in \mathbb{R}^n . *Mathematische Zeitschrift*, 230:259–316, 1999.
- A. Aleksandrov. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad Univ. Ann. (Math. ser.)*, (6):3–35, 1939.
- J.-D. Benamou, B. Froese, and A. Oberman. A viscosity solution approach to the Monge-Ampère formulation of the optimal transportation problem. 2013. URL <http://arxiv.org/abs/1208.4873>.
- J.-D. Benamou, B. Froese, and A. Oberman. Numerical solution of the optimal transportation problem using the Monge-Ampère equation. *Journal of Computational Physics*, 260:107–126, 2014. ISSN 0021-9991.
- V. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007. ISBN 978-3-540-34513-8; 3-540-34513-2.
- V. Brattka, J. Miller, and A. Nies. Randomness and differentiability. *Transactions of the AMS*, 368:581–605, 2016. <http://arxiv.org/abs/1104.4465>.
- Y. Brenier. Polar decomposition and increasing rearrangement of vector fields. *C.R. Acad. Sci. Paris Sér. I Math.*, (305):805–808, 1987.
- R. Downey and D. Hirschfeldt. *Algorithmic randomness and complexity*. Springer-Verlag, Berlin, 2010. 855 pages.
- T. Fowler and D. Preiss. A simple proof of Zahorski's description of non-differentiability sets of Lipschitz functions. *Real Anal. Exchange*, 34(1):127–138, 2009. ISSN 0147-1937. URL <http://projecteuclid.org/getRecord?id=euclid.rae/1242738925>.
- C. Freer, B. Kjos-Hanssen, A. Nies, and F. Stephan. Algorithmic aspects of Lipschitz functions. *Computability*, 3(1):45–61, 2014.
- P. Gács. Uniform test of algorithmic randomness over a general space. *Theoret. Comput. Sci.*, 341(1-3):91–137, 2005. ISSN 0304-3975. doi: 10.1016/j.tcs.2005.03.054. URL <http://dx.doi.org/10.1016/j.tcs.2005.03.054>.
- A. Galicki. Randomness and differentiability of convex functions. 9136:196–205, 2015. doi: 10.1007/978-3-319-20028-6_20.
- A. Galicki and D. Turetsky. Differentiability and randomness in higher dimensions. 2014. Submitted.
- M. Hoyrup and C. Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Inform. and Comput.*, 207(7):830–847, 2009. ISSN 0890-5401. doi: 10.1016/j.ic.2008.12.009. URL <http://dx.doi.org/10.1016/j.ic.2008.12.009>.
- R. J. McCann. A convexity principle for interacting gasses. *Adv. Math.*, 128, 1:153–179, 1997.
- F. Mignot. Contrôle optimal dans les inéquations variationnelles elliptiques. *J. Funct. Anal.*, (22):130–185, 1976.
- G. Minty. Monotone nonlinear operators on a Hilbert space. *Duke Math J.*, (29):341–346, 1962.
- A. Nies. *Computability and randomness*, volume 51 of *Oxford Logic Guides*. Oxford University Press, Oxford, 2009. ISBN 978-0-19-923076-1. doi: 10.1093/acprof:oso/9780199230761.001.0001. URL <http://dx.doi.org/10.1093/acprof:oso/9780199230761.001.0001>. 444 pages. Paperback version 2011.
- N. Pathak, C. Rojas, and S. G. Simpson. Schnorr randomness and the Lebesgue differentiation theorem. *Proc. Amer. Math. Soc.*, 142(1):335–349, 2014. ISSN 0002-9939. doi: 10.1090/S0002-9939-2013-11710-7. URL <http://dx.doi.org/10.1090/S0002-9939-2013-11710-7>.
- M. Pour-El and I. Richards. *Computability in Analysis and Physics*. Springer-Verlag, Berlin-New York, 1988.
- D. Preiss. Differentiability of Lipschitz functions on banach spaces. *Journal of Functional Analysis*, 91(2):312–345, 1990. ISSN 0022-1236.
- D. Preiss and G. Speight. Differentiability of Lipschitz functions in lebesgue null sets. *Inventiones mathematicae*, 199(2):517–559, 2015. ISSN 0020-9910. doi: 10.1007/s00222-014-0520-5.
- G. A. D. Preiss and M. Csörnyei. Differentiability of Lipschitz functions, structure of null sets, and other problems. In *Proceedings of the International Congress of Mathematicians 2010*, pages 1379–1394. World Scientific Publishing, 2011.
- H. Rademacher. Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variablen und über die Transformation der Doppelintegrale. *Math. Ann.*, 79(1):340–359, 1919.
- R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Grundlehren der Mathematischen Wissenschaften 317. Springer-Verlag, 1997.
- J. Rute. Computable randomness and betting for computable probability spaces. To appear in *Mathematical Logic Quarterly*.
- J. Rute. Algorithmic randomness, martingales and differentiation. 2012. preprint.
- J. Urbas. Mass transfer problems. Lecture notes from a course given in Univ. Bonn. 1997.
- C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.
- C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 2009.
- K. Weihrauch. *Computable Analysis*. Springer, Berlin, 2000.
- Z. Zahorski. Sur l'ensemble des points de non-dérivabilité d'une fonction continue. *Bull. Soc. Math. France*, 74:147–178, 1946. ISSN 0037-9484.