Ergod. Th. & Dynam. Sys. (2002), **22**, 1181–1189
DOI: 10.1017/S0143385702000718
© *2002 Cambridge University Press*
Printed in the United Kingdom

Greedy is not so bad

STEVEN KALIKOW
9308 Cherry Hill Road, Apartment 124, College Park, MD 20740, USA
(e-mail: 72604.3435@compuserve.com)

(*Received* 10 *October* 2000 *and accepted in revised form* 2 *December* 2001)

Abstract. Let Ω be a metric space. There is a natural metric that can be put on the space of all measures on Ω called the Wasserstein metric. It is obtained by taking the infimum over all couplings of the two measures of the average distance between two points. There is a simple coupling called the greedy coupling for the two measures, but operations researchers will assure you that it is terrible to use the value obtained by the greedy coupling as a substitute for the Wasserstein distance between two measures. The purpose of this paper, however, is to show that if all that you are interested in is the topology obtained by the metric, the value you get from the greedy coupling is good enough.

0. *Introduction*

Ω is a finite metric space, with a metric $\mathbf{d}$, such that all distances are less than or equal to one. μ and λ are two measures on Ω.

Definition. A *coupling* of μ and λ is a measure on $\Omega \times \Omega$ such that the projections of that measure on the axes are μ and λ.

Definition. The *Wasserstein metric* (see [**STR**]) on probability measures on Ω is defined by minimizing over all couplings ν of μ and λ;

$$\Delta(\nu) = \int_{\Omega\times\Omega} \mathbf{d}(p,q)\, d\nu.$$

(In the language of probability theory, $\Delta(\nu)$ can be written as $E_\nu\, d(x,y)$, where x, y is chosen in $\Omega \times \Omega$ according to ν. In this paper, Ω is finite so $\Delta(\nu) = \sum_{p,q\in\Omega\times\Omega} \mathbf{d}(p,q)\nu(p,q)$.)

There is an algorithm for obtaining a coupling that minimizes this integral. It is called the Hungarian algorithm. However, it is a complicated algorithm. In this paper, we discuss a very natural coupling called the greedy coupling. If ν is the greedy coupling and we compute $\Delta(\nu)$, then the value we get will generally be larger than the Wasserstein distance. We ergodic theorists are interested in the Wasserstein metric because there is a

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

special case of the Wasserstein metric, called the dbar metric, which is essential for ergodic theory. However, I have never seen an example in ergodic theory where we actually need to compute that distance precisely. All we are ever interested in is the topology induced by that metric. The purpose of this paper is to establish that, for topological purposes, we might as well use the value obtained by the greedy algorithm instead of the Wasserstein metric. More precisely we show that there is a real valued function f such that:

(1) $f(x)$ goes to zero as x goes to zero;

(2) if the Wasserstein distance between between two measures is ε then the matrix you obtain from using the greedy algorithm is less than $f(\varepsilon)$.

The function we use for f is $f(x) = 4(x^{\ln(2)/\ln(3)})$. However, if you could obtain a much easier argument for another function obeying (1), then we would regard this as an improvement on this paper regardless of how slowly your function goes to zero, because the precise f you use is irrelevant for our purposes.

1. *Construction of the greedy coupling in a specific case*

Here we are considering the case where S_1 and S_2 are two subsets of Ω with the same number of points, say k, μ is the measure assigning measure $1/k$ to every point in S_1 and 0 to every other point, and λ is the measure assigning $1/k$ to every point in S_2 and 0 to every other point. A greedy coupling for this case is defined as follows.

Choose $p_1 \in S_1$ and $p_2 \in S_2$ to minimize the distance between any two such points. If the minimum is obtained by more than one pair, choose any such pair (i.e. there may be more than one greedy coupling). Assign measure $1/k$ to the ordered pair (p_1, p_2). We refer to this as 'marrying off p_1 with p_2'. Now take $p_3 \in S_1 - \{p_1\}$ and $p_4 \in S_2 - \{p_2\}$ to minimize the distance between any two such points and marry off p_3 with p_4. Continue this procedure until all points in S_1 are married off with points in S_2.

2. *Proof of (2) for a specific case*

Let $\mathbf{d}$ be the given metric on Ω, $S_1 = \{p_1, p_2, p_3, \ldots, p_k\}$ and $S_2 = \{q_1, q_2, q_3, \ldots, q_k\}$ and consider an arbitrary coupling of S_1 and S_2.

Temporary assumption. We will temporarily assume that our arbitrary coupling takes on the form of a function (i.e. for any p_1, p_2, p_3; $p_2 \neq p_3$ it is impossible for both (p_1, p_2) and (p_1, p_3) to have positive measures). This assumption will be released later.

Our arbitrary coupling will assign $1/k$ to each of the following ordered pairs

$$(p_1, q_1), (p_2, q_2), \ldots, (p_k, q_k)$$

where these pairs are arranged so that

$$\mathbf{d}(p_1, q_1) \leq \mathbf{d}(p_2, q_2) \leq \cdots \leq \mathbf{d}(p_k, q_k).$$

From here on in these couples will be called the arbitrary couples. If (p, q) is an arbitrary couple, p will be regarded as the arbitrary husband and q will be regarded as the arbitrary wife (regard p and q to be distinct, but allow distances to be zero).

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

Permute the points of S_1 so that $S_1 = \{g_1, g_2, g_3, \ldots, g_k\}$ and permute the points in S_2 so that $S_2 = \{h_1, h_2, h_3, \ldots, h_k\}$ in such a way that the greedy coupling assigns a measure $1/k$ to each of the following pairs

$$(g_1, h_1), (g_2, h_2), \ldots, (g_k, h_k)$$

where these pairs are arranged so that

$$\mathbf{d}(g_1, h_1) \leq \mathbf{d}(g_2, h_2) \leq \cdots \leq \mathbf{d}(g_k, h_k).$$

From here on in these couples will be called the greedy couples.

Let

$$\varepsilon = \left(\frac{1}{k}\right) \sum \mathbf{d}(p_i, q_i).$$

Our goal is to establish that

$$\left(\frac{1}{k}\right) \sum \mathbf{d}(g_i, h_i) < 4(\varepsilon^{\ln(2)/\ln(3)}).$$

Comment 1. Fix an n. We will now show that we can assume that k is divisible by 2^n. For example, suppose $k = 5$ and we would like k to be divisible by 64. Five is not divisible by 64 so we have a problem. Here is how we settle the problem. We assume we have an arbitrary and greedy coupling on Ω. $64 \times 5 = 320$. We create a new space Ω' with subsets S_1' and S_2' each having 320 points, 64 points corresponding to each of p_1, p_2, p_3, p_4, p_5, q_1, q_2, q_3, q_4, q_5. For any two of these 64 points we define their distance to be the distance between the corresponding points and, in particular, this means that their distance is defined to be zero if they correspond to the same point. Now we simple replace each p_i, q_i, g_i, and h_i with all their corresponding points to get a new arbitrary coupling and greedy coupling. Note that our new greedy coupling is indeed a greedy coupling. The reader can show that if we can establish the result for this new pair of couplings it follows that it holds for our original couplings because the new greedy and new arbitrary couplings yield the exact same values as our old greedy and arbitrary coupling.

Comment 2. Hence we assume that k is divisible by 2^n. Using this assumption we will obtain a formula which involves n. Comment 1 implies that the formula holds for all n.

LEMMA 1.

$$\mathbf{d}(g_1, h_1) \leq \mathbf{d}(p_1, q_1), \mathbf{d}(g_2, h_2) \leq \mathbf{d}(p_3, q_3), \mathbf{d}(g_3, h_3) \leq \mathbf{d}(p_5, q_5), \mathbf{d}(g_4, h_4) \leq \mathbf{d}(p_7, q_7), \ldots.$$

Proof. We will just verify that $\mathbf{d}(g_3, h_3) \leq \mathbf{d}(p_5, q_5)$ and the reader will easily be able to generalize.

$\{g_1, h_1, g_2, h_2\}$ contains four points, so they are members of at most four arbitrary couples. When these arbitrary couples are removed, by the method that (g_3, h_3) is chosen, $\mathbf{d}(g_3, h_3)$ is at least as small as any other arbitrary couple. Hence there are at most four possible values for i, such that $\mathbf{d}(p_i, q_i) < \mathbf{d}(g_3, h_3)$. □

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

LEMMA 2.

$$\left(\frac{1}{k}\right)\sum_{1}^{k/2}\mathbf{d}(g_i, h_i) \leq \frac{\varepsilon}{2}.$$

Proof. By Lemma 1,

$$\left(\frac{1}{k}\right)\sum_{1}^{k/2}\mathbf{d}(g_1, h_1) \leq \left(\frac{1}{k}\right)\sum_{1}^{k/2}\mathbf{d}(p_{2i-1}, q_{2i-1}) \leq \left(\frac{1}{k}\right)\sum_{1}^{k/2}\mathbf{d}(p_{2i}, q_{2i})$$

and the sum of the latter two expressions is ε. □

Definition. We refer to the greedy pairs

$$\left\{(g_i, h_i) : i \leq \frac{k}{2}\right\}$$

as the 'old' pairs and

$$\left\{(g_i, h_i) : i > \frac{k}{2}\right\}$$

as the 'new' pairs. An old (new) pair g_i, h_i are called an old (new) husband and wife, respectively.

Example. Let a_0 be a new husband. Let b_0 be his arbitrary wife. Is b_0 new? Assume that she is not. Let a_1 be her greedy husband (i.e. (a_1, b_0) is an old greedy couple). Let b_1 be his arbitrary wife. Is she new? Suppose not. Let a_2 be her greedy husband. Let b_2 be his arbitrary wife. Is she new? Suppose so. Then $\{a_0, b_0, a_1, b_1, a_2, b_2\}$ will be referred to as a family and a_0, b_2 will be called the father and mother of the family, respectively.

Definition. $\{a_0, b_0, a_1, b_1, \ldots, a_i, b_i\}$ is called a *family* with a father a_0 and a mother b_i iff a_0, b_i are the only new members of the set, each a_j, b_j is an arbitrary couple, and each a_{j+1}, b_j is a greedy couple.

Comment. The families are disjoint and cover all $S_1 \cup S_2$. The number of families is the number of new fathers, namely $k/2$.

Definition. For a family F, let the *value* of F, denoted $V(F)$, be the distance between its father and mother.

LEMMA 3. *The sum of the values of the families is less than or equal to* $3\varepsilon k/2$.

Proof. Let $\{a_0, b_0, a_1, b_1, \ldots, a_i, b_i\}$ be a family. By the triangle inequality, the value of the family is less than or equal to

$$\mathbf{d}(a_0, b_0) + \mathbf{d}(b_0, a_1) + \mathbf{d}(a_1, b_1) + \mathbf{d}(b_1, a_2) + \cdots + \mathbf{d}(a_i, b_i)$$

so

the sum of the values of the families

$\leq$ the sum of the distances between the arbitrary couples

$+$ the sum of the distances between the old greedy couples

$$\leq \varepsilon k + \frac{\varepsilon k}{2}.$$

□

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

LEMMA 4.

$$\left(\frac{1}{k}\right)\sum_{k/2+1}^{3k/4}\mathbf{d}(g_i,h_i)\le\frac{3\varepsilon}{4}.$$

Proof. Let $F_1, F_2, F_3 \ldots F_{k/2}$ be a listing of the families with non-decreasing values. By the same reasoning as Lemma 1:

$$\begin{aligned}\mathbf{d}(g_{k/2+1},h_{k/2+1})\le V(F_1),\mathbf{d}(g_{k/2+2},h_{k/2+2})\le V(F_3),\ldots,\mathbf{d}(g_{3k/4},h_{3k/4})\\ \le V(F_{(k/2)-1})\end{aligned}$$

so the result follows by Lemma 3 and the reasoning of Lemma 2. □

LEMMA 5. *For all $i \le n$,*

$$\left(\frac{1}{k}\right)\sum_{(1-1/2^{i-1})k+1}^{(1-1/2^{i})k}\mathbf{d}(g_i,h_i)\le 3^{i-1}\frac{\varepsilon}{2^i}.$$

Proof by induction. When proving this for $i = 3$, for instance, define all g_i, h_i such that $i \le 3k/4$ to be second-order old, all g_i, h_i such that $i > 3k/4$ to be second-order new, define families again analogously so that every family has a second-order new father and mother and proceed as in the proof of Lemma 4. □

LEMMA 6.

$$\left(\frac{1}{k}\right)\sum_{i=1}^{k}\mathbf{d}(g_i,h_i)\le\varepsilon\frac{3^n}{2^n}+\frac{1}{2^n}.$$

Proof. Since all distances are less than or equal to one,

$$\left(\frac{1}{k}\right)\sum_{(1-1/2^{n})k+1}^{k}\mathbf{d}(g_i,h_i)\le\frac{1}{2^n}.$$

Hence

$$\left(\frac{1}{k}\right)\sum_{1}^{k}\mathbf{d}(g_i,h_i)\le\varepsilon\left(\frac{1}{2}+\frac{3}{4}+\frac{9}{8}+\cdots+\frac{3^{n-1}}{2^n}\right)+\frac{1}{2^n}\le\frac{\varepsilon 3^n}{2^n}+\frac{1}{2^n}.$$ □

Proof of (2). By Comment 2 and Lemma 6, all we need to prove is that for every ε there exists a non-negative integer n such that

$$\frac{\varepsilon 3^n}{2^n}+\frac{1}{2^n}<4(\varepsilon^{\ln(2)/\ln(3)}).$$

Choose n so that $3^n \le 1/\varepsilon < 3^{n+1}$. Then $\varepsilon^{\ln(2)/\ln(3)} > 1/2^{n+1}$ and hence

$$\frac{\varepsilon 3^n}{2^n}+\frac{1}{2^n}\le\frac{1}{2^n}+\frac{1}{2^n}<4(\varepsilon^{\ln(2)/\ln(3)}).$$ □

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

3. *The general case without the temporary assumption*

Ω is a finite metric space $\{p_1, p_2, p_3, \ldots, p_n\}$ with a metric $\mathbf{d}$ and two measures on it μ and λ. μ assigns points $p_1, p_2, p_3, \ldots, p_n$ probabilities $r_1, r_2, r_3, \ldots, r_n$, respectively, and λ assigns points $p_1, p_2, p_3, \ldots, p_n$ probabilities $s_1, s_2, s_3, \ldots, s_n$, respectively. We are given an arbitrary coupling† ν which assigns each p_i, p_j the probabilities $\nu(p_i, p_j)$ such that fixing i,

$$\sum_j \nu(p_i, p_j) = r_i$$

and fixing j,

$$\sum_i \nu(p_i, p_j) = s_j.$$

We are about to define a coupling of the two measures g, called the greedy coupling, which assigns each pair a probability $g(p_i, p_j)$ such that fixing i,

$$\sum_j g(p_i, p_j) = r_i$$

and fixing j,

$$\sum_i g(p_i, p_j) = s_j.$$

Our goal is to define the greedy coupling for this case and reduce this to the case in §1 with the temporary assumption and, thereby, prove that if $\varepsilon = \Delta(\nu)$ then

$$\Delta(g) < 4(\varepsilon^{\ln(2)/\ln(3)}).$$

4. *Construction of the greedy coupling in general*

To write out a general definition of the greedy coupling would make this paper hard to read, so we will just define it in an example and the reader will easily be able to generalize the definition.

Example. Ω contains three points p_1, p_2, p_3. $\mathbf{d}$ is defined by

$$\mathbf{d}(p_1, p_2) = \tfrac{1}{3}, \quad \mathbf{d}(p_2, p_3) = \tfrac{1}{5} \quad \text{and} \quad \mathbf{d}(p_1, p_3) = \tfrac{1}{6},$$

μ is defined by

$$\mu(p_1) = \tfrac{1}{3}, \quad \mu(p_2) = \tfrac{1}{2}, \quad \mu(p_3) = \tfrac{1}{6}$$

and λ is defined by

$$\lambda(p_1) = \tfrac{1}{10}, \quad \lambda(p_2) = \tfrac{7}{10}, \quad \lambda(p_3) = \tfrac{1}{5},$$

and we now show how to find a greedy coupling between them.

Step 1. Minimize the distance between any ordered pair of points to which μ and λ give positive measures. That minimum is zero and it is achieved by three pairs (p_1, p_1), (p_2, p_2), and (p_3, p_3).

† We need not concern ourselves with the possibility that the set of all such couplings may turn out to be empty, because we actually construct one here, namely the greedy coupling. However it is a little ridiculous to quote the greedy coupling when demonstrating the existence of a coupling because you can just use the product measure, $\nu = \mu \times \lambda$.

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

Step 2. You should be indifferent as to which of these three pairs you choose. In general, when there is a tie like this, the pair you choose may determine which coupling you end up with so there can be more than one greedy coupling. It does not matter. The results of this paper hold for any of them. Here we will choose (p_1, p_1).

Step 3. $\mu(p_1) = \frac{1}{3}, \lambda(p_1) = \frac{1}{10}$. The minimum of these two numbers is $\frac{1}{10}$ so we choose to let $g(p_1, p_1) = \frac{1}{10}$.

Step 4. We just married off $\frac{1}{10}$ of the μ measure of p_1 with $\frac{1}{10}$ of the λ measure of p_1 so we now should note what is left unmarried from μ and what is left unmarried from λ.

The still unmarried part of μ assigns measures $\frac{7}{30}$ to p_1, $\frac{1}{2}$ to p_2, $\frac{1}{6}$ to p_3. The still unmarried part of μ assigns measure 0 to p_1, $\frac{7}{10}$ to p_2, $\frac{1}{5}$ to p_3.

Step 5. Look at the points which μ still gives positive measure to and the points which ν still gives positive measure to and minimize the distance. The minimum is still zero and it is achieved by (p_2, p_2) and (p_3, p_3). Here we pick the pair (p_2, p_2).

Step 6. The minimum of $\frac{1}{2}$ and $\frac{7}{10}$ is $\frac{1}{2}$ so we choose to let $g(p_2, p_2) = \frac{1}{2}$.

Conclusion. Ordinarily we would continue this procedure until the two measures are completely married off. In this particular example, we need not continue because we are already done. The rest of the coupling is determined. Our final greedy measure is

$$g(p_1, p_1) = \tfrac{1}{10}, \quad g(p_2, p_2) = \tfrac{1}{2}, \quad g(p_3, p_3) = \tfrac{1}{6},$$
$$g(p_1, p_3) = \tfrac{1}{30}, \quad g(p_1, p_2) = \tfrac{1}{5}.$$

Proof of (2) in the general case without the temporary assumption. Standard continuity arguments allow us to assume all the probabilities $g(p_i, p_j)$ and $\nu(p_i, p_j)$ (and hence all $\mu(p_i)$ and $\lambda(p_j)$) to be rational numbers. Make that assumption. Let k be a common denominator for all these numbers. We now construct a new space Ω' containing $2k$ points, the first k of which we will call S_1 and the second k of which we will call S_2. Let μ' be the measure which assigns measure zero to all S_2 and assigns a measure $1/k$ to each member of S_1, and λ' be the measure that assigns measure zero to all S_1 and $1/k$ to each member of S_2. If $\mu(p_i)$ is a/k then let a of the points of S_1 correspond to p_i and if $\nu(p_i) = b/k$ let b of the points in S_2 correspond to p_i. In this manner, every point in Ω' will correspond to a point in Ω. The metric $\mathbf{d}'$ on Ω' is defined by letting let $\mathbf{d}'(p_i, p_j)$ be the distance between the point corresponding to p_i with the point corresponding to p_j for any p_i, p_j (i.e. zero for two points corresponding to the same point.)

We now describe how to marry off S_1 with S_2 in two ways to form g' and ν'. If $g(p_i, p_j)$ is c/k, marry off c of the points corresponding to p_i in S_1 with c of the points corresponding to p_j in S_2. In other words, for the c points $p_1, p_2, p_3, \ldots, p_c$ in S_1 and the c points $q_1, q_2 q_3, \ldots, q_c$ in S_2 that we are referring to, let $g'(p_i, q_i) = 1/k$. This completely defines g'. Similarly, define ν' by marrying off d points of S_1 corresponding to p_i with d points of S_2 corresponding p_j whenever $\nu(p_i, p_j) = d/k$. Note that g' is a greedy coupling

$$\sum_{i,j} g'(p_i, p_j)\mathbf{d}'(p_i, p_j) = \sum_{i,j} g(p_i, p_j)\mathbf{d}'(p_i, p_j) = \Delta(g)$$

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

and

$$\sum_{i,j} v'(p_i, p_j)\mathbf{d}'(p_i, p_j) = \sum_{i,j} v(p_i, p_j)\mathbf{d}'(p_i, p_j) = \Delta(v)$$

so we have reduced the general case without the temporary assumption to the specific case with the temporary assumption.

5. *Not sharp?*

To test whether or not this bound is sharp, we obtained a metric space and sets S_1 and S_2 and obtained a greedy coupling which was constructed to yield a particularly bad greedy match. Fix n and let S be the space

$$\left\{0, \frac{1}{(3^n)}, \frac{2}{(3^n)}, \frac{3}{(3^n)}, \ldots, 1\right\}$$

with standard metric and sets $S_1 = \{a/3^n) : a \text{ even}\}$ and $S_2 = \{a/3^n) : a \text{ odd}\}$.

Assigning measures $1/((3^n) + 1)$ to all points $\{0, 1/(3^n), 2/(3^n), 3/(3^n), \ldots, 1\}$ and then obviously the Wasserstein distance between S_1 and S_2 is

$$\varepsilon = \frac{1}{3^n}$$

obtained by marrying $a/3^n$ with $(a+1)/3^n$ for 'a' any even number.

This map is also a greedy coupling, but it is not the only greedy coupling. Here is another one that does much worse. Fix an odd a other than 3^n. Let b be the largest power of three dividing a (note that one is a power of three). If a/b is 1 mod 3, marry $a/3^n$ with $(a+b)/3^n$. If a/b is 2 mod 3, marry $a/3^n$ with $(a-b)/3^n$. Then marry one to zero. If the reader complains that this is cheating because we did not have to use this greedy coupling, we could have forced this to be the only greedy coupling by moving all points an arbitrarily small amount.

For large n, this coupling obtains a greedy distance of about

$$\left(\frac{2}{3\log 3}\right)\varepsilon\left(\log\left(\frac{1}{\varepsilon}\right)\right).$$

Unfortunately this is much less than the upper bound my proof obtained of

$$4(\varepsilon^{\ln(2)/\ln(3)}).$$

We leave the following questions for the reader to consider.

Question 1. Can you get an example where a greedy coupling does worse than the one in the above example?

Question 2. Can you force the upper bound to be close to $(2/(3\log 3))\varepsilon(\log(1/\varepsilon))$?

Question 3. If we allow an upper bound f obeying (1) that perhaps goes to zero slowly, can the proof be made much simpler?

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press

Acknowledgement. I am grateful to the referee for all the assistance he gave in correcting this paper.

REFERENCE

[STR] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. Wiley, New York, 1991.

https://doi.org/10.1017/S0143385702000718 Published online by Cambridge University Press