

Transition to nestedness in multi- to one-dimensional optimal transport†

PIERRE-ANDRÉ CHIAPPORI¹, ROBERT MCCANN²
and BRENDAN PASS³

¹*Department of Economics, Columbia University, New York, NY, USA*
email: pc2167@columbia.edu

²*Department of Mathematics, University of Toronto, Toronto, Canada*
email: mccann@math.toronto.edu

³*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada*
email: pass@ualberta.ca

(Received 12 January 2018; revised 20 August 2018; accepted 23 August 2018; first published online 1 October 2018)

We study a one-parameter class of examples of optimal transport problems between a two-dimensional source and a one-dimensional target. Our earlier work identified a nestedness condition on the surplus function and marginals, under which it is possible to solve the problem semi-explicitly. In the family of examples we consider, we classify the values of parameters which lead to nestedness. In those cases, we derive an almost explicit characterisation of the solution.

Key words: Monge-Kantorovich problem, optimal transport, unequal dimensions, matching

2010 Mathematics Subject Classification: 49B15, 91B68

1 Introduction

Given probability measures μ and ν on domains $X \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^n$, and a surplus function $s: X \times Y \rightarrow \mathbb{R}$, the optimal transport problem of Monge and Kantorovich is to maximise

$$\int_{X \times Y} s(x, y) d\gamma(x, y) \tag{1.1}$$

among probability measures on $X \times Y$ whose marginals are μ and ν .

The problem has been a vibrant and dynamic field of investigation since the late 1980s and has a wealth of applications; it is extensively reviewed in [6–8]. Until recently, most research focused on the case where the source X and target Y have the same dimension, $m = n$. In this setting, conditions are known under which a unique solution exists and is concentrated on a graph $y = F(x)$ over X ; the solution can then be characterised by a Monge–Ampere-type partial

† The authors are grateful to Toronto’s Fields’ Institute for the Mathematical Sciences for its kind hospitality during part of this work, and to Boyang Wu for independently checking the simulations. They acknowledge partial support of RJM’s research by Natural Sciences and Engineering Research Council of Canada Grants No. 217006-08 and No. 217006-15. Chiappori gratefully acknowledges financial support from the NSF (Award No. 1124277). Pass is pleased to acknowledge support from Natural Sciences and Engineering Research Council of Canada Grant No. 412779-2012 and a University of Alberta start-up grant.

differential equation. Regularity of solutions is now fairly well understood. When both dimensions are one, closed form solutions can be derived for many choices of s , μ and ν .

On the other hand, the case where the dimensions m and n differ has thus far received relatively little attention. However, motivated by matching problems in economics, we recently initiated the study of multi- to one-dimensional problems, the case where $m > n = 1$ [1, 2]; for $m > n > 1$, see [4]. In particular, our recent work in this case identified a joint condition on s , μ and ν , called *nestedness*, under which the optimal map F can be obtained almost explicitly, and a regularity theory developed.

In the absence of nestedness, little is known about the structure of solutions. As in the equal dimensional setting, general conditions on s ensure that the solution is unique and concentrated on the graph of a function $F : X \rightarrow Y$, but examples in [5] imply that there exist smooth marginals μ and ν for which the solution is discontinuous, *unless* s takes an *index form*, $s(x, y) = b(I(x), y) + \alpha(x)$ for some functions $I : X \rightarrow \mathbb{R}$, $\alpha : X \rightarrow \mathbb{R}$, in which case the problem reduces to an optimal transport problem between unidimensional domains $I(X)$ and Y , with surplus b .

We expect optimal maps to often be discontinuous in the absence of nestedness; equivalently, we expect the Kantorovich potential on the high-dimensional side to not be C^1 . It is not clear what structure the singular set can be expected to exhibit, or whether the Kantorovich potential on the lower dimensional side will also exhibit singularities. In addition, though the concept of nestedness is fairly straightforward to check, and we expect it to hold on a fairly wide class of problems, there is so far a scarcity of fully worked examples where the condition has been checked, and, if it holds, solutions to the optimal transport problem obtained. Apart from simple examples worked out in [1, 2], we do not know of any examples of solutions to non-nested problems; a rich family of these would be useful for testing conjectures.

In this short note, we study a one-parameter family of models and classify for which parameter values nestedness holds. In the nested case, we characterise the solutions more or less explicitly, illustrating the utility of this concept. In the non-nested case, we do not have an explicit solution, but note that a singularity develops along a certain curve. We then present numerical solutions for three values of the parameter, two for which the model is nested and one for which it is not. In the nested case, the solutions agree qualitatively with our analytic solution, while in the non-nested case the presence of the singularity is confirmed. We hope that this class of examples can shed some light on the structure of solutions, and possibly be a useful test case for numerical methods.

In what follows, we will sometimes adopt terminology from the marriage matching problems which originally motivated our work in this area. In particular, we will sometimes refer to X as the set of wives and Y the set of husbands; under this interpretation, the level set $F^{-1}(y)$ of the optimal map identifies those wives among whom husband y is indifferent in equilibrium and are sometimes called iso-husband curves.

2 Background: nestedness

In this section, we recall the definition of nestedness and its implications for the structure of optimal transport plans. We assume throughout this section that the target Y is one dimensional, $n = 1$. The presentation is adapted from [2]; for ease of exposition, we do not always work under the most general possible conditions.

Given a surplus $s(x, y) \in C^2(X \times Y)$, we will assume that s is *non-degenerate*: $D_x s_y(x, y) \neq 0$ throughout $X \times Y$, where $D_x s_y(x, y)$ is the differential with respect to x of the partial derivative

$s_y(x, y) = \frac{\partial s}{\partial y}(x, y)$ with respect to y of the surplus. This condition, together with the implicit function theorem, implies that for each fixed $k \in s_y(X, Y)$, the level set

$$X_=(y, k) := \{x \in X : s_y(x, y) = k\}$$

is a C^1 -hypersurface of X . We note that each such level surface divides X into sub- and super-level sets, and the sub-level sets are increasing in k . We denote by $k(y)$ the (unique, under mild conditions on μ and ν [2]) value of k so that the mass of the sub-level set matches that of $(-\infty, y)$:

$$\mu(X_<(y, k)) = \nu((-\infty, y)), \tag{2.1}$$

where

$$X_<(y, k) := \{x \in X : s_y(x, y) \leq k\}.$$

We say that the model is nested if the sub-level sets selected in this way are nested; i.e. if

$$X_<(y_0, k(y_0)) \subseteq X_<(y_1, k(y_1)),$$

whenever $y_0 < y_1$.

Nestedness depends on the interaction between s , μ and ν , which is somewhat atypical in optimal transport; in fact, if s, μ, ν is nested for all choices of marginals μ and ν , then s must be of index form [2].

We show in [2] that nestedness is equivalent to the following two properties: (1) disjointness of $X_=(y_0, k(y_0))$ and $X_=(y_1, k(y_1))$ for any $y_0 \neq y_1$ and (2) the property that for each x there is *exactly one* y such that $x \in X_=(y, k(y))$. We then showed that this condition allows one to characterise the solution.

Theorem 1. *Assume that the model is nested. Then the mapping that takes each $x \in X_=(y, k(y))$ to y , for each $y \in Y$, is well defined. Furthermore, it is the unique optimal map.*

In the nested case, we showed that the function $k(y)$ coincides with the derivative of an s -convex function:

$$v'(y) = k(y), \text{ where } v(y) = \max_{x \in X} [s(x, y) - u(x)]$$

for some $u(x)$. Furthermore, u can be taken to be the s -conjugate of v , $u(x) = \max_{y \in Y} [s(x, y) - v(y)]$, and the pair (u, v) solves the Kantorovich dual problem:

$$\min \left(\int_X u(x) d\mu(x) + \int_Y v(y) d\nu(y) \right) \tag{2.2}$$

among all $(u, v) \in L^1(\mu) \times L^1(\nu)$ such that

$$u(x) + v(y) \geq s(x, y) \tag{2.3}$$

for all $(x, y) \in X \times Y$.

Heuristic description of nestedness: Nestedness and its connection to the dual problem (2.2) can be understood heuristically as follows. Suppose that γ solves the optimal transport problem and $u(x)$ and $v(y)$ solve the dual problem (2.2), it is then well known that the inequality (2.3) is

an equality γ almost everywhere. This then means that the function $y \mapsto s(x, y) - v(y) - u(x)$ is maximised at any y such that (x, y) is in the support of γ ; assuming differentiability, one must then have $s_y(x, y) = v'(y)$ at such points.

This suggests that for each y , the level set $F^{-1}(y)$ of the optimal map F (if it exists) should be contained in a level set of $x \mapsto s_y(x, y)$:

$$\{x : F(x) = y\} \subseteq X_=(y, v'(y)).$$

If this containment is actually an *equality*, then for each y knowing the level set $F^{-1}(y)$ is equivalent to knowing $k(y) := v'(y)$. Without knowing $v'(y)$, one might try to construct the optimal map by guessing that the particular level set of $x \mapsto s_y(x, y)$ corresponding to y is the one that splits the population proportionally (i.e. satisfies equation (2.1)), and that the corresponding sub- and super-level sets get matched to the intervals above and below y , respectively. This leads to a well-defined mapping $F : X \rightarrow Y$ as long as the level sets selected in this way for different choices of y do not intersect; this is essentially the nestedness condition, and Theorem 1 confirms that, when it holds, the mapping constructed as above is indeed optimal.

3 A class of examples

Consider matching uniform measure on the disk $\{x_1^2 + x_2^2 \leq 1\}$ with uniform measure on an arc, $\{(\cos(\theta), \sin(\theta)) : |\theta| \leq \theta_0\}$, to maximise the restriction of the bi-linear surplus from $\mathbb{R}^2 \times \mathbb{R}^2$, $s(x_1, x_2, \theta) = x_1 \cos(\theta) + x_2 \sin(\theta)$.

The following result classifies the values of the parameter θ_0 for which the model is nested.

Proposition 2 *The model is nested if and only if $\theta_0 \leq \frac{\pi}{4}$.*

The proof of the proposition requires the following lemma.

Lemma 3 *The optimal map $\theta = F(x_1, x_2)$ satisfies:*

- (1) $F(x_1, x_2)x_2 \geq 0$.
- (2) $F(x_1, x_2) = -F(x_1, -x_2)$ for all (x_1, x_2) .

(We note that an optimal map exists because the surplus is twisted, i.e. $\theta \mapsto D_x s(x, \theta)$ is injective. This condition is well known to guarantee unique, graphical solutions [6–8].)

Proof The symmetry in point (2) follows immediately from uniqueness of the solution, because the marginals and surplus are both symmetric under the transformation $(x_1, x_2, \theta) \mapsto (x_1, -x_2, -\theta)$. Turning to the first property, we have that if $(x_1, x_2, \theta = F(x_1, x_2))$ is in the support of the optimal measure, then so is $(x_1, -x_2, -\theta = F(x_1, -x_2))$. Applying s -monotonicity (e.g. [6, Theorem 1.38]) to these two points yields:

$$s(x_1, x_2, \theta) + s(x_1, -x_2, -\theta) \geq s(x_1, -x_2, \theta) + s(x_1, x_2, -\theta),$$

which reduces to $x_2 \sin(\theta) \geq 0$ (or, equivalently, $x_2 \theta \geq 0$). □

We now prove the proposition.

Proof The level curves of $(x_1, x_2) \mapsto s_\theta(x_1, x_2, \theta) = -x_1 \sin(\theta) + x_2 \cos(\theta)$ are straight lines parallel to the line segment joining the origin to $(\cos(\theta), \sin(\theta))$ (c.f. [3, Lemma 1.6]). The population splitting curve corresponding to θ has the form $-x_1 \sin(\theta) + x_2 \cos(\theta) = k(\theta)$, where $k(\theta)$ is to be determined from the population balance condition (2.1). Note that $k(0) = 0$. In fact, by rotational symmetry, $\mu(X_{\leq}(\theta, 0)) = \frac{1}{2}$ for all θ ; as $v(-\theta_0, \theta) = \frac{\theta - \theta_0}{2\theta_0} > \frac{1}{2}$ for $\theta > 0$, we must have $k(\theta) > 0$ for positive θ . □

Now, the mass balance condition is equivalent to $1 - \mu(X_{\leq}(\theta, k)) = A(k, \theta) \frac{1}{\pi} = \frac{\theta_0 - \theta}{2\theta_0} = 1 - v(-\theta_0, \theta)$, where $A(k, \theta)$ is the area of the super-level set $\{-x_1 \sin(\theta) + x_2 \cos(\theta) \geq k\}$; $k(\theta)$ is the unique solution to this equation. By rotational symmetry, we note that $A(k, \theta) = A(k)$ is in fact independent of θ .

Differentiating with respect to θ yields:

$$A'(k(\theta))k'(\theta) = -\frac{\pi}{2\theta_0}. \tag{3.1}$$

To determine $A(k)$, we note that by rotational symmetry, the distance from the line to the origin is k .

Now, note that $A(k)$ is the area of a wedge of the circle minus the area of a triangle. The height of the triangle is k , and its base length is $2\sqrt{1 - k^2}$. The angle of the wedge is given by $2 \arccos(k)$, and so its area is $\arccos(k)$. The area of the triangle is $k\sqrt{1 - k^2}$, so

$$A(k) = \arccos(k) - k\sqrt{1 - k^2}.$$

Differentiating, we get

$$A'(k) = -\frac{1}{\sqrt{1 - k^2}} - \sqrt{1 - k^2} + \frac{k^2}{\sqrt{1 - k^2}} = \frac{1}{\sqrt{1 - k^2}}(-1 - (1 - k^2) + k^2) = -2\sqrt{1 - k^2}. \tag{3.2}$$

Therefore, substituting into equation (3.1), we have

$$k'(\theta) = \frac{\pi}{4\theta_0\sqrt{1 - k^2}}. \tag{3.3}$$

Now, the lemma implies that we can find the complete solution to the problem by solving the sub-problem matching the top half of the disk, $\{x_1^2 + x_2^2 \leq 1, x_2 > 0\}$ to the top half of the arc, $\{(\cos(\theta), \sin(\theta)) : 0 \leq \theta \leq \theta_0\}$ and then reflecting. The overall problem will be nested provided that

- (1) The sub-problem is nested.
- (2) The level curves for the sub-problem corresponding to $\theta > 0$ do not intersect the x_1 axis (as in this case, they will meet a level curve corresponding to $-\theta$ as well).

We consider the second point first. This is equivalent to $x_0(\theta) \leq -1$, or equivalently, $k(\theta) \geq \sin(\theta)$, for all $\theta \in (0, \theta_0)$, where $(x_0(\theta), 0) = (-\frac{k(\theta)}{\sin(\theta)}, 0)$ is the intersection point between the population splitting level curve corresponding to θ and the x_1 axis. We show that this holds if and only if $\theta_0 \leq \frac{\pi}{4}$.

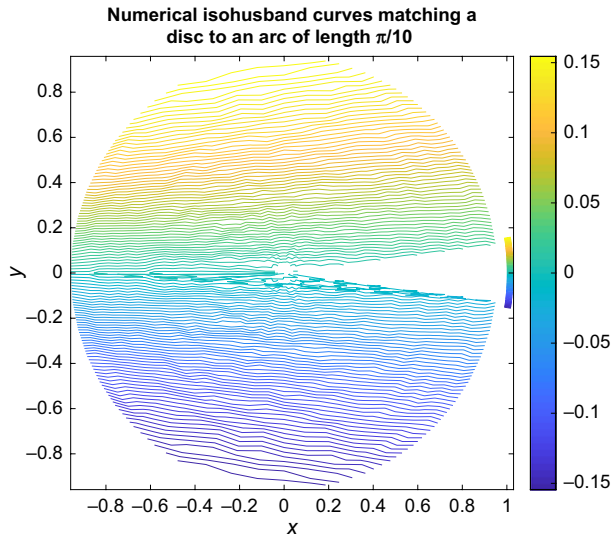


FIGURE 1. Numerical solution matching uniform measure on the disc to an arc of length $\pi/10$ ($\theta_0 = \pi/20$), resulting in a nested model. The coloured curves on the disc are the level sets of the matching function, corresponding to the point on the arc of the same colour.

Noting that $k > 0$ for all $\theta > 0$, equation (3.3) implies that if $\theta_0 \leq \frac{\pi}{4}$, we have

$$k'(\theta) > 1 > \cos(\theta)$$

for all $\theta > 0$; integrating yields $k(\theta) > \sin(\theta)$ as desired.

On the other hand, if $\theta_0 > \frac{\pi}{4}$, then $k'(\theta) < \cos(\theta)$ for small θ , so that $x_0(\theta) > -1$ for sufficiently small values of θ , implying that the model is not nested.

We turn now to the first point above: the nestedness of the sub-problem. Observe that as $x_0(\theta) > -1$, the population splitting level curves for the overall problem coincide with those for the sub-problem. The sub-problem is nested provided that $x_0(\theta)$ is a monotone decreasing function of θ , in which case since the slope $\tan(\theta)$ of the (affine) level curve is monotone increasing in θ , the fact that $x_0(\theta_1) \leq x_0(\theta_2)$ when $\theta_1 > \theta_2$ will imply that these two lines do not intersect in the upper half space $\{x_2 > 0\}$.

We have $-x'_0(\theta) \sin^2(\theta) = k'(\theta) \sin(\theta) - k(\theta) \cos(\theta)$. The right-hand side is 0 at $\theta = 0$; it suffices to show it is always non-negative. Note that its derivative is

$$k''(\theta) \sin(\theta) + k(\theta) \sin(\theta).$$

As k' is monotone increasing in k by equation (3.3) and k monotone increasing in θ as $k' > 0$, we have that $k'' > 0$, yielding the desired result.

As the proof shows, in the nested case one can derive an equation (3.3) yielding implicitly the relationship between k and θ , and consequently the level sets of the optimal map. We present in Figures 1–3 numerical solutions for three values of θ_0 , which agree with the qualitative properties of our analysis.

We note here that for all values of θ_0 , symmetry arguments imply that the level set $F^{-1}(0)$ is contained in the line segment $\{x_2 = 0\} \cap X$. In the first two simulations, the values of θ_0 are

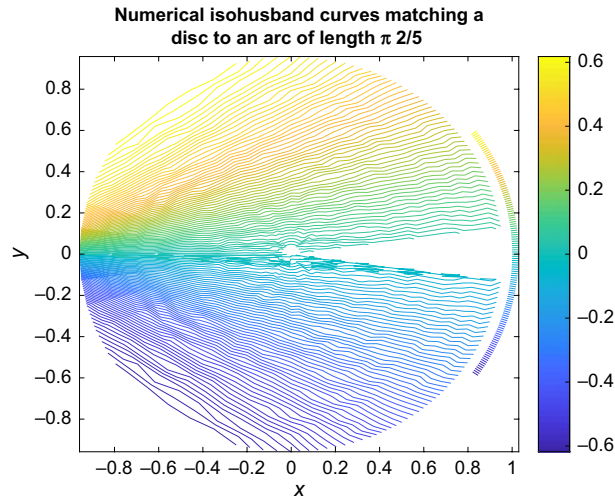


FIGURE 2. Numerical solution matching uniform measure on the disc to an arc of length $2\pi/5$ ($\theta_0 = \pi/5$), resulting in a nested model. The coloured curves on the disc are the level sets of the matching function, corresponding to the point on the arc of the same colour.

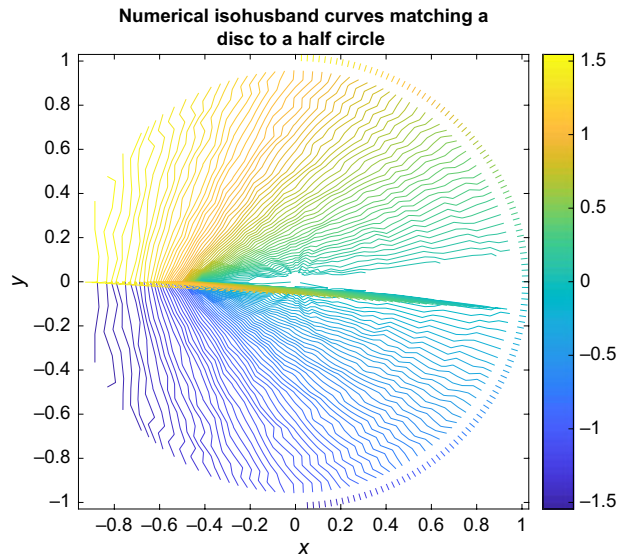


FIGURE 3. Numerical solution matching uniform measure on the disc to an arc of length π ($\theta_0 = \pi/2$), resulting in a non-nested model. The coloured curves on the disc are the level sets of the matching function, corresponding to the point on the arc of the same colour.

below the nestedness threshold of $\frac{\pi}{4}$. The population slitting level curves $F^{-1}(\theta)$ for $\theta > 0$ lie entirely above the x_1 axis and distinct level curves do not intersect. Note as well that the level curves corresponding to the same value of $\theta \neq 0$ in the two graphs are disjoint from each other, illustrating a multi-market testability property from [1] (see Proposition 18 and Corollary 20 in [1]).

In the third simulation, we take θ_0 above the nestedness threshold, so that the model is non-nested. As we show in [2], the procedure we describe in Section 2 above breaks down in this case. More precisely, if we find the $k = k(\theta)$ such that

$$\mu(X_{\leq}(\theta, k)) = \nu(-\theta_0, \theta), \quad (3.4)$$

it turns out that the resulting line $\{s_{\theta}(x, \theta) = k(\theta)\}$ intersects the line $\{x_2 = 0\} \cap X$ for small values of θ . Thus, the procedure to define a mapping F described in the previous section is not consistent in this case; it attempts to map the point of intersection to both θ and $-\theta$.

In Figure 3, the numerical solution is presented. One can see that for small values of θ , the level sets $F^{-1}(\theta)$ meets the level set $F^{-1}(-\theta)$ at a point on the line segment $\{x_2 = 0\} \cap X$ (note, however, that these true level curves are not the same level curves predicted by equation (3.4)). Thus, there is a discontinuity in the matching function along this line segment; wives on this segment are indifferent between two husbands, of the form θ and $-\theta$ for some $\theta > 0$.

Remark 4 *Even when the overall problem is non-nested, we expect, at least for $\theta_0 - \frac{\pi}{4}$ small, that the subproblem introduced in the proof of the proposition is nested. In this case, the equation characterising k for small θ is different from equation (3.3), as the mass splitting level curves for the top half of the disc will pass through the boundary $\{x_2 = 0\}$. Nestedness of the sub-problem will follow, provided one can show that $x_0(\theta)$ is monotone decreasing. As above, this will be the case if $k''(\theta) + k(\theta) > 0$. It is not hard to show that $k > 0$; convexity of k would imply the desired result. The equation governing k becomes quite complicated in this case, and verifying this convexity looks non-trivial.*

For general multi- to one-dimensional problems, it seems reasonable to ask whether the region X divides into several subregions X_i and that the model $(s, X_i, F(X_i))$ is nested, where F is the optimal map.

Remark 5 *One can bootstrap from equation (3.3) to infer that $k(\theta)$ is infinitely many times differentiable except near the boundary $\theta = \pm\theta_0$, where k approaches ± 1 , respectively, illustrating a regularity result (Theorem 16) from [2].*

Conflicts of interest

None.

References

- [1] CHIAPPORI, P.-A., MCCANN, R. J. & PASS, B. (submitted, 2016) Multidimensional matching. Preprint at arXiv:1604.05771.
- [2] CHIAPPORI, P.-A., MCCANN, R. J. & PASS, B. (2017) Multi- to one-dimensional optimal transport. *Commun. Pure Appl. Math.* **70**(12), 2405–2444.
- [3] GANGBO, W. & MCCANN, R. J. (2000) Shape recognition via Wasserstein distance. *Quart. Appl. Math.* **58**(4), 705–737.
- [4] MCCANN, R. J. & PASS, B. (submitted, 2018) Optimal transportation between unequal dimensions. Preprint at <https://arxiv.org/abs/1805.11187>.
- [5] PASS, B. (2012) Regularity of optimal transportation between spaces with different dimensions. *Math. Res. Lett.* **19**(2), 291–307.

- [6] SANTAMBROGIO, F. (2015) *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and their Applications, Vol. **87**, Birkhäuser/Springer, Cham.
- [7] VILLANI, C. (2003) *Topics in Optimal Transportation*. Graduate Studies in Mathematics, Vol. **58**, American Mathematical Society, Providence.
- [8] VILLANI, C. (2009) *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, Vol. **338**, Springer, New York.