The Annals of Applied Probability
1993, Vol. 3, No. 4, 1100–1111

THE INTEGRABILITY OF THE SQUARE EXPONENTIAL TRANSPORTATION COST[1]

BY M. TALAGRAND AND J. E. YUKICH

Université Paris VI and Ohio State University, and Lehigh University

Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be i.i.d. with the uniform distribution on $([0,1]^2, \|\ \|)$, where $\|\ \|$ denotes the Euclidean norm. Using a new presentation of the Ajtai–Komlós–Tusnády (AKT) transportation algorithm, it is shown that the square exponential transportation cost

$$\inf_{\pi} \sum_{i=1}^{n} \exp\left(\frac{\|X_i - Y_{\pi(i)}\|}{K(\log n/n)^{1/2}} \right)^2,$$

where π ranges over all permutations of the integers $1, \dots, n$, satisfies an integrability condition. This condition strengthens the optimal matching results of AKT and supports a recent conjecture of Talagrand. Rates of growth for the L_p transportation cost are also found.

1. Introduction and Statement of Results. Given two collections of points $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ in the unit square $([0,1]^2, \|\ \|)$, where $\|\ \|$ denotes the Euclidean norm, the Euclidean two-sample matching problem (or transportation cost problem) involves finding a perfect bipartite matching between the x and y points so as to minimize the sum of the edge lengths. The L_p, $1 \le p < \infty$, *transportation cost* between the x and y points equals

$$\inf_{\pi} n^{-1} \sum_{i=1}^{n} \|x_i - y_{\pi(i)}\|^p,$$

where the inf ranges over all the permutations π of the integers $1, 2, \dots, n$.

Letting $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ denote two sequences of i.i.d. random variables which have the uniform distribution λ on $[0,1]^2$, define the random transportation cost by

$$T_p(n) := \inf_{\pi} n^{-1} \sum_{i=1}^{n} \|X_i - Y_{\pi(i)}\|^p.$$

The following theorem concerning the exact rate of growth of $\mathbb{E}T_1(n)$ was proved by Ajtai, Komlós and Tusnády (AKT) (1984), who were the first to discover the depth of the transportation cost problem. They proved this result using the so-called transportation algorithm (described in Section 2), which involves a clever shifting of the sample points. Throughout, the notation $f(x) = \Theta(g(x))$ means that $f(x) = O(g(x))$ and $g(x) = O(f(x))$.

Received August 1992; revised March 1993.

[1] Research partially supported by NSF grants.

AMS 1991 *subject classifications.* Primary 05C70, 60C05; secondary 60D05.

Key words and phrases. Euclidean bipartite matching, transportation cost, subgaussian.

THEOREM A [AKT (1984)]. $\mathbb{E}T_1(n) = \Theta((\log n/n)^{1/2})$.

Subsequently, Shor (1985) combined the marriage lemma with the AKT transportation algorithm to provide an alternative proof of the upper bound of Theorem A; see Coffman and Lueker (1990) for an exposition of this approach. Talagrand (1993), following a different approach based on majorizing measures, also gave a proof of the upper bound of Theorem A. The approach of Talagrand (1993) solves several related matching problems and, among other things, yields the following exponential moment result, significantly strengthening the upper bound of Theorem A. For $1 \le i \le n$, let the components of X_i (respectively Y_i) be denoted by $X_{1,i}$ and $X_{2,i}$ (respectively $Y_{1,i}$ and $Y_{2,i}$).

THEOREM B [Talagrand (1993)]. *For $\beta < 1/4$, there is a constant $K := K(\beta)$ depending on β only, with the following property: with probability greater than $1 - n^{-2}$, there is a matching π such that*

$$\max_{i \le n} |X_{1,i} - Y_{1,\pi(i)}| \le K(\log n/n)^{1/2}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \exp\left(\frac{|X_{2,i} - Y_{2,\pi(i)}|}{K(\log n/n)^{1/2}} \right)^{\beta} \le K.$$

The main contribution of the present paper provides a new approach to the AKT transportation method. Rather than studying the displacement of a sample point X_i, $1 \le i \le n$, we instead consider the displacement of a *fixed* point $u \in [0,1]^2$. When conditioned appropriately, the successive displacements of the fixed point u become symmetric random variables, have zero correlation and also exhibit subgaussian character. This approach considerably clarifies the transportation method and removes a number of technical obstacles. As an illustration of its power, we will prove the following main result, which does not seem obtainable by the classical approach. As with Theorem B, this result provides a substantial strengthening of Theorem A, although in a different direction.

Assuming throughout that the random variables X and Y are defined on the probability space $(\Omega, \mathscr{A}, P)$, the main result shows that the "square exponential transportation cost" is integrable on large sets:

THEOREM 1. *There is a constant K such that for all $n \ge 1$ there is a subset A_n of Ω with $P(A_n^c) \le n^{-2}$ and*

$$\mathbb{E}\left(\inf_{\pi} \frac{1}{n} \sum_{i=1}^{n} \exp\left(\frac{\|X_i - Y_{\pi(i)}\|}{K(\log n/n)^{1/2}} \right)^2 \right) 1_{A_n} \le K.$$

REMARKS. (i) Notice that by the convexity of $\exp(x^2)$, Theorem 1 implies

$$\mathbb{E}\left(\exp\inf_{\pi}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\|X_i - Y_{\pi(i)}\|}{K(\log n/n)^{1/2}}\right)^2\right)1_{A_n} \le K$$

or, stated in terms of $T_2(n)$,

$$\mathbb{E}\left(\exp\frac{n}{K\log n}T_2(n)\right)1_{A_n} \le K.$$

Thus, $(n/\log n)T_2(n)$ has an exponential tail on large sets.

(ii) Using $\exp(x) \ge x^p/p!$, *we deduce rates in terms of* n *and* p for the L_p transportation cost: There is a constant K such that for all $p \ge 1$,

$$\mathbb{E}T_p(n) \le (Kp)^{p/2}(\log n/n)^{p/2}.$$

(iii) It is an interesting open question whether $(n/\log n)^{1/2}\mathbb{E}T_1(n)$ converges as $n \to \infty$.

Theorem 1 takes on added significance in connection with the deep matching result of Leighton and Shor (1986), who prove that the minimax matching length

$$\inf_{\pi}\max_{1\le i\le n}\|X_i - Y_{\pi(i)}\|$$

between the X and Y points is of order $n^{-1/2}\log^{3/4} n$. Although they actually obtain a.s. rates, we will only give a weak form of their result:

THEOREM C [Leighton and Shor (1986)].

$$\mathbb{E}\inf_{\pi}\max_{1\le i\le n}\|X_i - Y_{\pi(i)}\| = \Theta(n^{-1/2}\log^{3/4} n).$$

In Talagrand (1993) it is conjectured that the exponent 2 of Theorem 1 may actually be increased to 4. Were this conjecture true, it would contain the upper bounds of both the AKT and Leighton–Shor results. In this context notice that Theorem 1 and convexity imply the minimax estimate

$$\mathbb{E}\inf_{\pi}\max_{1\le i\le n}\|X_i - Y_{\pi(i)}\| = O(n^{-1/2}\log n).$$

In the sense of exponential integrability, Theorem 1 *thus interpolates between AKT and a weak form of Leighton–Shor.*

2. Proof of Theorem 1. The proof of Theorem 1 involves a special subdivision of the unit square $[0,1]^2$.

Begin by subdividing $[0,1]^2$ in a manner used first by Ajtai, Komlós and Tusnády (AKT) (1984) and subsequently by Shor (1985) and Shor and Yukich (1991). First divide $[0,1]^2$ in half vertically (with a vertical bisector) and linearly transform each half so that it has area equal to the fraction of sample points $\{X_i\}_{i=1}^n$ in it. When each half is transformed, the sample points in that half are similarly transformed. Then subdivide each of these halves horizontally and transform each to have area equal to the ratio of the number of transformed points contained within to the *total number of points*, n. This process generates four rectangles which partition $[0,1]^2$. Apply the procedure recursively, alternating vertical and horizontal divisions. Repeating this refining procedure until each region contains at most one point yields n nondegenerate rectangles. For any random sample $\{X_i\}_{i=1}^n$, each of the n transformed sample points is contained in a nondegenerate rectangle with area n^{-1}.

Take r with $2^{-r} = K(\log n/n)^{1/2}$, where here and henceforth K denotes a universal constant whose value may change from line to line. Repeatedly apply the preceding transformation until step $2r$. The resulting transformation, which we will call $T := T(X_1, \dots, X_n)$, generates 2^{2r} subrectangles. Let $\{\tilde{X}_i\}_{i=1}^n$ denote the collection of transformed sample points at step $2r$. Similarly, given the sample $Y_1, \dots, Y_n$, define a second transformation $T(Y_1, \dots, Y_n)$ and let $\{\tilde{Y}_i\}_{i=1}^n$ denote the transformed Y points at step $2r$.

For any random variable W and subset A of Ω, set $\mathbb{E}_A W := \mathbb{E}(W 1_A)$. To prove Theorem 1, it suffices by the convexity of $\exp(x^2)$ to show there exists a K such that for all n there are sets A_n and B_n, with $P(A_n^c) \le \frac{1}{2}n^{-2}$ and $P(B_n^c) \le \frac{1}{2}n^{-2}$, such that

$$\text{(2.1a)} \qquad \begin{aligned} &\mathbb{E}_{A_n}\left\{\frac{1}{n}\sum_{i=1}^n \exp\left(\frac{\|X_i - \tilde{X}_i\|}{K(\log n/n)^{1/2}}\right)^2\right\} \le K, \\ &\mathbb{E}_{B_n}\left\{\frac{1}{n}\sum_{i=1}^n \exp\left(\frac{\|Y_i - \tilde{Y}_i\|}{K(\log n/n)^{1/2}}\right)^2\right\} \le K \end{aligned}$$

and

$$\text{(2.1b)} \qquad \mathbb{E}_{A_n \cap B_n}\left\{\inf_{\pi} \frac{1}{n}\sum_{i=1}^n \exp\left(\frac{\|\tilde{X}_i - \tilde{Y}_{\pi(i)}\|}{K(\log n/n)^{1/2}}\right)^2\right\} \le K.$$

First step: Proof of (2.1a). Clearly, it suffices to prove the first inequality in (2.1a). Consider the sample point $X_1 \in [0,1]^2$ and its displacement according to the transformation $T := T(X_1, \dots, X_n)$. Let $D_j := D_j(X_1)$, $1 \le j \le 2r$, denote its displacement on the jth step of the recursion. When j has odd parity, D_j represents a horizontal displacement; when j has even parity, D_j represents a vertical displacement.

The approach of Theorem 1 would be rather obvious if it were true that the displacements of similar parity were independent. However, this is not the case. The following proposition, which is proved in the Appendix, actually shows that the displacements of similar parity are negatively correlated. (As a simple consequence of the negative correlation, we note in passing that the variance of the total horizontal displacement is bounded by

$$\mathbb{E}\left|\sum_{\substack{j\le 2r\\ j\,\mathrm{odd}}} D_j\right|^2 \le \sum_{j=1}^{2r} \mathbb{E}D_j^2 \le Kr/n,$$

since $\mathbb{E}D_j^2 \le 1/n$, by an easy computation. This estimate, together with the approach of Shor (1985), provides a rigorous proof of the upper bound of Theorem A using only the AKT transportation algorithm. In the original proof [AKT (1984)], the authors do not fully justify the fact that the correlations between the shifts D_j have a negligible effect.)

PROPOSITION 1. *For all* $1 \le i, j \le 2r$ *of similar parity*, $\mathbb{E}D_iD_j < 0$.

As might be expected, the negative correlation unfortunately leads to involved technical analysis when proving exponential integrability results and we will not pursue this line of investigation. *Instead, we consider the displacement of a fixed point* $u := (x, y) \in [0,1]^2$. As it turns out, this formulation bypasses the need to work with negatively correlated displacements and simplifies the analysis.

Given $u \in [0;1]^2$, let $D_j := D_j(u)$, $1 \le j \le 2r$, denote its displacement on the jth step of the transformation T. We note that D_3 is symmetric when conditioned with respect to D_1. Thus $\mathbb{E}(D_1D_3) = 0$. Similarly, for any odd m, the random variable D_m, when conditioned on the previous displacements $D_1,\dots,D_{m-2}$, is symmetric, and so $\mathbb{E}(D_1\dots D_m) = 0$. *The displacements of the fixed point* u *thus have zero correlation.*

The proof of (2.1a) will be accomplished with the aid of a few preliminary lemmas that describe the behavior of T on certain subsquares of $[0,1]^2$. First, define for a fixed $l \le r$ the set $\mathscr{S}_l$ consisting of the 2^{2l} dyadic squares:

$$\left[k2^{-l},(k+1)2^{-l}\right] \times \left[k'2^{-l},(k'+1)2^{-l}\right], \qquad 1 \le k, k' \le 2^l.$$

Next, define the event $A_n \subset \Omega$ to be the set of $\omega \in \Omega$ such that:
(i) for each dyadic square $S \in \mathscr{S}_l$, $1 \le l \le r$, the proportion $p_{l,S} := p_{l,S}(\omega)$ of sample points $X_1,\dots,X_n$ in the left half H of S satisfies

$$|p_{l,S} - \tfrac{1}{2}| \le K(l/n)^{1/2}2^{(r+l)/2}; \tag{2.2}$$

and
(ii) the proportion $p^*_{l,S}$ of sample points in the upper half of H and H^c satisfies (2.2).

LEMMA 1. $P(A_n^c) \le \frac{1}{2}n^{-2}$.

PROOF. Let $\mathrm{Bi}(m,p)$ denote a binomial random variable with parameters m and p. Notice that $p_{l,S} \overset{d}{=} \mathrm{Bi}(M,\frac{1}{2})/M$, where $M \overset{d}{=} \mathrm{Bi}(n,2^{-2l})$. Now apply standard binomial tail estimates as in Lemma 2.1 of Shor and Yukich (1991). □

The next lemma shows that the transformation T changes the aspect ratio of any dyadic square by a small amount. By aspect ratio, we mean the ratio of the longest side to the shortest.

LEMMA 2. *On the set A_n the aspect ratios of the transformed dyadic squares are uniformly bounded by a constant.*

PROOF. Consider a fixed dyadic square. T transforms it into a rectangle. After subdividing the rectangle vertically in half, the side is multiplied by a factor of $2p$ [or $(2p)^{-1}$]. However, on the set A_n condition (2.2) implies that on the lth subdivision,

$$|2p - 1| \le 2K(l/n)^{1/2} 2^{(r+l)/2}.$$

Thus, the aspect ratio is at worse multiplied by a factor of

$$\left(1 - 2K(l/n)^{1/2} 2^{(r+l)/2}\right)^{-1}.$$

On the following horizontal subdivision of the subrectangle, condition (2.2) again shows that the aspect ratio is at worst multiplied by the same factor.

After $2r$ such subdivisions the aspect ratio is at most

$$\left[\prod_{l=1}^{r} \left(1 - 2K(l/n)^{1/2} 2^{(r+l)/2}\right)^{-1}\right]^2.$$

It is easily verified that the above product is $O(1)$ since

$$\sum_{l=1}^{r} (l/n)^{1/2} 2^{(r+l)/2} = O(1). \qquad \square$$

The next lemma follows easily from Lemma 2 and the definition of aspect ratio.

LEMMA 3. *There is a constant K such that on the set A_n, the edge length of any transformed dyadic square $S \in \mathscr{S}_l$, $1 \le l \le r$, is bounded below by $K^{-1} 2^{-l/2}$ and above by $K 2^{-l/2}$.*

Therefore, when $\omega \in A_n$, the transformation T generates subrectangles of roughly similar proportions, allowing a precise estimation of the displacement of u, as shown by the next two lemmas. The first lemma shows that on A_n, the random variable $D_j(u)$, when conditioned on $D_1, \dots, D_{j-2}$, is a subgaussian random variable with parameter $1/n$; see Ledoux and Talagrand (1991) for a discussion of subgaussian random variables.

LEMMA 4. *There is a constant K such that for all j of odd parity, $1 \le j \le 2r$, for all $\beta > 0$ and $u \in [0,1]^2$,*

$$\left(\mathbb{E}\left(\exp \beta D_j(u) | D_1, \dots, D_{j-2}\right)\right) 1_{A_n} \le \exp K\beta^2/n.$$

PROOF. At the jth stage, the point u belongs to one of the 2^{j-1} transformed dyadic squares S belonging to $\mathscr{S}_{(j-1)/2}$. Assume without loss of generality that u belongs to the left half H of S. The square S contains M_j sample points, where $M_j \overset{d}{=} \mathrm{Bi}(n, 2^{-(j-1)})$, and H contains $\mathrm{Bi}(M_j, \frac{1}{2})$ points. Notice that the displacement $D_j(u)$, when conditioned upon $D_1, D_3, \dots, D_{j-2}$, equals $C(p_j - \frac{1}{2})L_j$, where p_j is the proportion of points in H, L_j is the length of the horizontal side of S and $C := C(u) \le 1$ is a constant prescribing the position of u relative to the bisector of S. It follows that $D_j(u)$, when conditioned on $D_1, D_3, \dots, D_{j-2}$, is equal in distribution to the random variable

$$CL_j/M_j \times \left(\mathrm{Bi}\left(M_j, \tfrac{1}{2}\right) - M_j/2\right).$$

Using the elementary inequality $\mathbb{E} \exp \lambda\varepsilon \le \exp \lambda^2/2$, where ε is a Bernoulli random variable satisfying $\Pr\{\varepsilon = 1\} = \Pr\{\varepsilon = -1\} = \frac{1}{2}$, together with independence, it follows that for all $\beta > 0$,

$$\mathbb{E}\left(\exp \beta D_j(u) | D_1, \dots, D_{j-2}\right) \le \exp C\beta^2 L_j^2/8M_j.$$

Now notice that if L'_j denotes the length of the vertical side of $S_{(j-1)/2}$, then

$$L_j L'_j = M_j/n.$$

By Lemma 3, $K^{-1}L_j \le L'_j \le KL_j$ on A_n so that

$$L_j^2 \le KM_j/n$$

on A_n. Therefore, by combining the above inequalities, the desired estimate

$$\mathbb{E}\left(\exp \beta D_j(u) | D_1, \dots, D_{j-2}\right) 1_{A_n} \le \exp K\beta^2/n$$

follows. □

The next lemma shows that the net horizontal displacement is subgaussian with parameter r/n.

LEMMA 5.

$$\mathbb{E}_{A_n} \exp\left(\beta \sum_{\substack{i \le 2r \\ i \text{ odd}}} D_i\right) \le \exp Kr\beta^2/n.$$

PROOF. Using conditional expectations, it follows that

$$\begin{aligned}
\mathbb{E}_{A_n} \exp\left(\beta \sum_{i \le 2r} D_i\right) &\le \mathbb{E}_{A_n} \mathbb{E}\left[\exp\left(\beta \sum_{i \le 2r} D_i\right) | D_1, \dots, D_{2r-2}\right] \\
&= \mathbb{E}_{A_n}\left[\exp\left(\beta \sum_{i \le 2r-2} D_i\right) \mathbb{E}\left(\exp \beta D_{2r} | D_1, \dots, D_{2r-2}\right)\right] \\
&\le \exp\left(K\beta^2/n\right) \mathbb{E}_{A_n} \exp\left(\beta \sum_{i \le 2r-2} D_i\right)
\end{aligned}$$

by Lemma 4. Now continue recursively. □

Using Chebyshev's inequality and choosing β appropriately, it follows that for all $t > 0$,

$$\Pr\left\{\left|\sum_{i\le 2r} D_i\right| > t,\, A_n\right\} \le 2\exp(-t^2 n/Kr).$$

A similar inequality holds for the net vertical displacement $\sum_{i\le 2r} D_i$, where i runs over indices of even parity. Next, let $D(u)$ denote the total displacement of the point u by the transformation T up to and including stage r. Since $D(u)$ is bounded by the sum of the net horizontal and vertical displacements of u, it follows that

$$\Pr\{|D(u)| > t,\, A_n\} \le 8\exp\{-t^2 n/Kr\}.$$

Using $\mathbb{E}X = \int_0^\infty P(X > t)\,dt$ for positive random variables X, it follows by standard arguments that there is a K such that for all $u \in [0,1]^2$,

$$\mathbb{E}_{A_n} \exp(nD^2(u)/rK) \le 2 \tag{2.3}$$

and, therefore, by Fubini's theorem,

$$\mathbb{E}_{A_n} \iint_{[0,1]^2} \exp(nD^2(u)/rK)\, du \le 2. \tag{2.4}$$

Next, let $\mathscr{S}_r := \{S_1, \dots, S_{2^{2r}}\}$ be the collection of the 2^{2r} dyadic squares of side 2^{-r} and let N_i, $1 \le i \le 2^{2r}$, denote the number of sample points in square S_i. Notice that T depends only on the value of the vector

$$\hat{N} := \langle N_1, \dots, N_{2^{2r}}\rangle.$$

Therefore, the total displacement of u may be written using the more suggestive notation $D(u) := D(u, \hat{N})$.

Consider now a fixed sample point X_1. We wish to show that (2.3) holds with u replaced by X_1. Since

$$\Pr\{X_1 \in S_i | \langle N_1, \dots, N_{2^{2r}}\rangle\} = N_i/n, \qquad i = 1, \dots, 2^{2r},$$

it follows that the conditional density of X_1 given $\hat{N}$ has the form

$$\sum_{i=1}^{2^{2r}} \frac{N_i}{n} 2^{2r} 1_{S_i}.$$

Now N_i/n is just the area of the transformed square $T(S_i)$, which by Lemma 3 is at most $K2^{-2r}$ on the set A_n. Thus, on the set A_n, the conditional density of X_1 given $\hat{N}$ is bounded by some constant K uniformly over $[0,1]^2$. Recognizing that A_n describes a set $\mathscr{A}_n$ of admissible values of $\hat{N}$, it follows that

$$\begin{aligned}\mathbb{E}_{A_n} \exp nD^2(X_1)/rK &= \mathbb{E}_{A_n}\Big(\mathbb{E}\big\{\exp nD^2(X_1, \hat{N})/rK\big\}\Big|\hat{N}\Big)\\ &= \sum_{\hat{k}\in\mathscr{A}_n} \mathbb{E}\Big(\exp nD^2\big(X_1, \hat{N} = \hat{k}\big)/rK \mid \hat{N} = \hat{k}\Big)\Pr(\hat{N} = \hat{k}).\end{aligned}$$

From the above we know that the conditional density of X_1 given $\hat{N} = \hat{k}$ is bounded by a constant K uniformly over $[0,1]^2$ whenever $\hat{k} \in \mathscr{A}_n$. It follows that

$$\mathbb{E}_{A_n} \exp nD^2(X_1)/rK \le K \sum_{\hat{k} \in \mathscr{A}_n} \left\{ \iint_{[0,1]^2} \exp nD^2(u, \hat{k})/rK\, du \right\} \Pr(\hat{N} = \hat{k})$$

$$\le K \mathbb{E}_{A_n} \iint_{[0,1]^2} \exp nD^2(u)/rK\, du$$

$$\le 2K$$

by (2.4). Since the above analysis holds for all sample points, we obtain

$$\mathbb{E}_{A_n} \frac{1}{n} \sum_{i=1}^{n} \exp nD^2(X_i)/rK \le 2K,$$

which is precisely (2.1a). This completes the first step. $\square$

Second step: *Proof of* (2.1b). It will suffice to show there is a K such that for all $\omega \in A_n$,

$$\inf_{\pi} \max_{1 \le i \le n} \|\tilde{X}_i - \tilde{Y}_{\pi(i)}\| / K(\log n/n)^{1/2} \le 1.$$

Let $\hat{X}_i$, $1 \le i \le n$ (respectively $\hat{Y}_i$, $1 \le i \le n$) be the transformed points after the completion of the recursion scheme based on the sample $X_1, \dots, X_n$ (respectively $Y_1, \dots, Y_n$). Notice that $\hat{X}_i$, $1 \le i \le n$ (respectively $\hat{Y}_i$, $1 \le i \le n$) is contained in a rectangle R_i (respectively Q_i) of area n^{-1}. Moreover, the rectangles $\{R_i\}_{i=1}^n$ and $\{Q_i\}_{i=1}^n$ form partitions of $[0,1]^2$. On the subset $A_n \subset \Omega$, the diameter of the rectangles $R_1, \dots, R_n$ is at most the diameter at the $r = \log(Kn/\log n)$ stage; that is, is at most $O(\log n/n)^{1/2}$. Thus, on A_n, it follows that

$$\|\tilde{X}_i - \hat{X}_i\| \le K(\log n/n)^{1/2}.$$

Similarly, there is a second subset $B_n \subset \Omega$, with $P(B_n^c) \le \frac{1}{2} n^{-2}$, such that on B_n,

$$\|\tilde{Y}_i - \hat{Y}_i\| \le K(\log n/n)^{1/2}.$$

It thus suffices to show for all $\omega \in A_n \cap B_n$,

$$\inf_{\pi} \max_{1 \le i \le n} \|\hat{X}_i - \hat{Y}_{\pi(i)}\| / K(\log n/n)^{1/2} \le 1. \tag{2.5}$$

It only remains to show for all $\omega \in A_n \cap B_n$ that there is a perfect matching π between the collections $\{R_i\}_{i=1}^n$ and $\{Q_i\}_{i=1}^n$. This is accomplished with the following lemma, which is a simple corollary of the marriage theorem and which first appeared in Shor (1985) and then Shor and Yukich (1991).

LEMMA 6. *Suppose that there are two partitions of* $[0,1]^2$ *into n rectangles of equal areas. Label the rectangles* $R_1, \dots, R_n$ *and* $Q_1, \dots, Q_n$. *Then there is a matching* π *between the* R_i*'s and the* Q_i*'s such that for all* $1 \le i \le n$, $R_i \cap Q_{\pi(i)} \neq \varnothing$.

Since $\|\hat{X}_i - \hat{Y}_{\pi(i)}\|$ is at most the sum of the diameters of the rectangle R_i and the associated rectangle $Q_{\pi(i)}$, the estimate (2.5) follows, completing the proof of Theorem 1. □

APPENDIX

We provide a proof of the fact that displacements of sample points are negatively correlated; see Proposition 1.

PROOF OF PROPOSITION 1. We will show that $\mathbb{E} D_1 D_3 < 0$. The approach for this special case can be modified to treat the general case. For simplicity of exposition, we will assume that n has even parity. The following arguments may be modified to treat the case of odd parity.

Throughout, take $S_1 := [0,1]^2$ and let S_3 be the dyadic square of side length $\frac{1}{2}$ which contains X_1. Let H_1 and H_3 denote the left halves of S_1 and S_3, respectively. Clearly, by symmetry it is enough to show that

$$\mathbb{E} D_1 D_3 1_{[X_1 \in H_1]} < 0.$$

Let Δ denote the distance between X_1 and the bisector of S_3 (prior to the transformation T): thus, Δ is uniformly distributed over $[0, \frac{1}{4}]$. Also, let M denote the number of sample points in H_1 *in addition to* X_1. Finally, for $i = 1, 3$ let

$$D_i^L := D_i 1_{[X_1 \in H_3]} 1_{[X_1 \in H_1]} \quad \text{and} \quad D_i^R := D_i 1_{[X_1 \in H_3^c \cap S_3]} 1_{[X_1 \in H_1]}$$

denote the displacement of X_1 according to its position in S_3.

Notice that when Δ and M are specified, then D_1^L and D_1^R are completely determined and in fact

$$D_1^R 1_{[\Delta = \delta, M = m]} = c D_1^L 1_{[\Delta = \delta, M = m]},$$

where $c := c(\delta) > 1$ depends only on δ. (This is true because if X_1 is in the right half of S_3, then it is closer to the bisector of S_1 and undergoes a larger displacement.) Symmetry considerations imply that

$$D_3^R 1_{[\Delta = \delta, M = m]} \stackrel{d}{=} D_3^L 1_{[\Delta = \delta, M = m]}$$

and, therefore,

$$D_1^R D_3^R 1_{[\Delta = \delta, M = m]} \stackrel{d}{=} -c D_1^L D_3^L 1_{[\Delta = \delta, M = m]}.$$

Thus, using conditional expectations and the independence of Δ and M we compute

$$\begin{aligned}
&\mathbb{E} D_1 D_3 1_{[X_1 \in H_1]} \\
&\quad = \iint \mathbb{E}\big(D_1 D_3 1_{[X_1 \in H_1]} | (\Delta, M)\big)\, dP_M\, dP_\Delta \\
&\quad = \tfrac{1}{2} \iint \mathbb{E}\big(D_1^L D_3^L + D_1^R D_3^R | (\Delta, M)\big)\, dP_M\, dP_\Delta \\
&\quad = \tfrac{1}{2} \iint \mathbb{E}\big(D_1^L D_3^L - c D_1^L D_3^L | (\Delta, M)\big)\, dP_M\, dP_\Delta \\
&\quad = \tfrac{1}{2} \int_0^{1/4} \sum_{m=0}^{n-1} (1-c) \mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = m)\big) \Pr\{M = m\}\, dP_\Delta(\delta).
\end{aligned}$$

Since $c > 1$, it remains to show that the integrand is positive for each choice of δ. The integrand has the form

$$\begin{aligned}
&\sum_{m=0}^{n/2-1} \mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = m)\big) \Pr\{M = m\} \\
&\qquad + \mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = n-m-1)\big) \Pr\{M = n-m-1\} \\
&\quad = \sum_{m=0}^{n/2-2} \Big(\mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = m)\big) \Pr\{M = m\} \\
&\qquad\qquad + \mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = n-m-2)\big) \Pr\{M = n-m-2\} \Big) \\
&\qquad + \mathbb{E}\big(D_1^L D_3^L | (\Delta = \delta, M = n-1)\big) \Pr\{M = n-1\},
\end{aligned} \tag{A.1}$$

since when $M = n/2 - 1$, there are equal numbers of points in H_1 and H_1^c, and so $D_1 = 0$. Observe that the last term in (A.1) is positive. Notice that for $m \le n/2 - 2$, D_1^L conditioned on $\{\Delta = \delta, M = m\}$ equals $C(\delta, m)$, where $C(\delta, m) < 0$ is a constant. Moreover, D_1^L conditioned on $\{\Delta = \delta, M = n - m - 2\}$ equals $-C(\delta, m)$. Therefore, the preceding sum becomes

$$\begin{aligned}
\left(\tfrac{1}{2}\right)^{n-1} \sum_{m=0}^{n/2-2} C(\delta, m) \Bigg[& \mathbb{E}\big(D_3^L | (\Delta = \delta, M = m)\big) \binom{n-1}{m} \\
& - \mathbb{E}\big(D_3^L | (\Delta = \delta, M = n-m-2)\big) \binom{n-1}{m+1} \Bigg].
\end{aligned}$$

The factor within brackets is a scalar multiple of the difference

$$(m+1) \mathbb{E}\big(D_3^L | (\Delta = \delta, M = m)\big) - (n-m-1) \mathbb{E}\big(D_3^L | (\Delta = \delta, M = n-m-2)\big),$$

which is negative, because

$$(m+1) \mathbb{E}\big(D_3^L | (\Delta = \delta, M = m)\big), \qquad 0 \le m \le n-2, \tag{A.2}$$

is increasing with m. To see this, fix m and define the random variables

$$U := \text{number of sample points in } S_3 \text{ in addition to } X_1$$

and

$$V := \text{number of sample points in } H_3 \text{ in addition to } X_1.$$

Then $U \stackrel{d}{=} \mathrm{Bi}(m, \frac{1}{2})$ and $V \stackrel{d}{=} \mathrm{Bi}(U, \frac{1}{2})$, where $\mathrm{Bi}(n, p)$ denotes a binomial random variable with parameters n and p. Since the width of the dyadic square S_3 after the transformation T is equal to $(m+1)/n$, the displacement D_3^L given $(\Delta = \delta, M = m)$ equals

$$C(\delta)\left(\frac{V+1}{U+1} - \frac{U-V}{U+1}\right)\frac{m+1}{n},$$

where $C(\delta)$ is a constant depending only on δ. Conditioning on U and noting $\mathbb{E}\mathbb{E}(2V - U|U) = 0$, it follows that

$$\mathbb{E}\big(D_3^L|(\Delta = \delta, M = m)\big) = \frac{C(\delta)}{n}\mathbb{E}\left(\frac{m+1}{U+1}\right).$$

Elementary calculations give

$$\mathbb{E}\left(\frac{m+1}{U+1}\right) = 2\left[1 - \left(\frac{1}{2}\right)^{m+1}\right],$$

and thus (A.2) is clearly increasing with m. Thus, $\mathbb{E}D_1D_3 < 0$, as desired. □

REFERENCES

AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259—264.

COFFMAN, E. G. and LUEKER, G. S. (1990). *Probabilistic Analysis of Packing and Partitioning Algorithms.* Wiley, New York.

LEIGHTON, T and SHOR, P. W. (1986). Tight bounds for minimax grid matching, with applications to the average case analysis of algorithms. In *Proceedings of the 18th ACM Symposium on Theory of Computing* 91–103. Association for Computing Machinery, New York.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, New York.

SHOR, P. W. (1985). Random planar matching and bin packing. Ph.D. thesis, MIT.

SHOR, P. W. and YUKICH, J. E. (1991). Minimax grid matching and empirical measures. *Ann. Probab.* **19** 1338–1348.

TALAGRAND, M. (1993). Matching theorems and empirical discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* To appear.

EQUIPE D'ANALYSE
ERA AUX CNRS NO. 754
UNIVERSITÉ PARIS VI
4 PLACE JUSSIEU
75230 PARIS CEDEX 05
FRANCE
AND
OHIO STATE UNIVERSITY
DEPARTMENT OF MATHEMATICS
COLUMBUS, OHIO 43210

DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015