*The Annals of Probability*
1994, Vol. 22, No. 2, 919–959

# THE TRANSPORTATION COST FROM THE UNIFORM MEASURE TO THE EMPIRICAL MEASURE IN DIMENSION $\geq 3$

BY M. TALAGRAND[1]

*Université Paris VI and Ohio State University*

Consider two independent sequences $(X_i)_{i \leq n}$ and $(X_i')_{i \leq n}$ that are independent and uniformly distributed over $[0,1]^d$, $d \geq 3$. Under mild regularity conditions, we describe the convex functions $\varphi$ such that, with large probability, there exists a one-to-one map $\pi$ from $\{1, \dots, n\}$ to $\{1, \dots, n\}$ for which

$$\sum_{i \leq n} \frac{1}{n} \varphi\left(\frac{X_i - X'_{\pi(i)}}{n^{-1/d} K_\varphi}\right) \leq 1,$$

where $K_\varphi$ depends on $\varphi$ only.

**1. Introduction.** Consider an integer $d \geq 1$. Consider $m$ points $(X_i)_{i \geq m}$ that are independent and uniformly distributed over the set $[0,1]^d$. How far is the set $\{X_1, \dots, X_m\}$ from being uniformly spread over $[0,1]^d$? The way this will be measured is by considering the transportation cost from the empirical measure $m^{-1} \sum_{i \leq m} \delta_{X_i}$ to the uniform measure $\lambda_d$ on $[0,1]^d$ when the cost of transporting a unit mass from $x$ to $y$ is given by $\varphi(x-y)$, where $\varphi$ is a function on $\mathbb{R}^d$. The notion of transportation cost will be explained formally in Section 2. For the purpose of this introduction we will chose an essentially equivalent formulation that avoids all technicalities. Consider another independent uniform sample $(X_i')_{i \leq m}$ independent of $(X_i)_{i \leq m}$. How close are the sets $\{X_1, \dots, X_m\}$ and $\{X_1', \dots, X_m'\}$? Here closeness will be measured as

$$\inf \frac{1}{m} \sum_{i \leq m} \varphi\big(X_i - X'_{\pi(i)}\big), \tag{1.1}$$

where the infimum is computed over all one-to-one maps $\pi$ from $\{1, \dots, m\}$ to itself. [As it turns out, (1.1) is exactly the transportation cost of the measure $m^{-1} \sum_{i \leq m} \delta_{X_i}$ to the measure $m^{-1} \sum_{i \leq m} \delta_{X_i'}$].

The depth of the topic was discovered by Ajtai, Komlós and Tusnády [1], who prove that when $d = 2$ and $\varphi(x) = \|x\|$, then (1.1), with high probability, is of order $m^{-1/2} (\log m)^{1/2}$. Another landmark result is due to Leighton and Shor [2]. They show that if $d = 2$ and $\varphi(x) = 0$ for $\|x\| \leq K m^{-1/2} (\log m)^{3/4}$, then (1.1) is zero with high probability (there, as in the rest of this paper, $K$

Received September 1992.
[1]Work partially supported by an NSF grant.
*AMS* 1991 *subject classification*. Primary 60D05.
*Key words and phrases.* Optimal matchings, transportation cost, empirical measure.

denotes a universal constant, not necessarily the same at each occurrence). Further equally remarkable results were obtained by Shor [4]. The case $d = 2$ is certainly the deepest, and, despite considerable efforts, it is still far from being completely understood [7].

When $d = 3$ (or more), the situation is simpler. The method of [1] shows that if $\varphi(x) = \|x\|$, then with high probability (1.1) is of order $m^{-1/d}$. This method was adapted by Shor and Yukich [6] to show that (1.1) is zero with high probability when $\varphi(x) = 0$ for $\|x\| \le K(d)(\log m/m)^{1/d}$, where $K(d)$ depends on $d$ only.

There is no reason why the function $\varphi$ should depend only on the distance to the origin. Actually, in certain natural situations, one must give a different weight to the different coordinates; for example, consider the function $\varphi$ given by

$$\varphi\big((x_1,x_2,x_3)\big) = \begin{cases} +\infty, & \text{if } |x_1| \ge Km^{-1/3}, \\ \sqrt{x_2^2 + x_2^2}, & \text{if } |x_1| < Km^{-1/3}. \end{cases}$$

Then a special case of a result of [3] shows that, with high probability, (1.1) is of order $m^{-1/3}$ (for a suitable $K$). (A similar situation has also been considered in dimension 2 in [4], [5] and [7].)

The purpose of the present paper is to present a result that, in dimension $d \ge 3$, will contain the three results mentioned above, and many others. This result is essentially the final word on the question, as it provides (under a mild regularity condition) an essentially complete description of the functions $\varphi$ for which, with high probability, the transportation cost (1.1) is less than 1. (It is a proof when $d \ge 3$ of the "ultimate matching conjecture" of [7].) For the sake of simplicity, we state and prove the result only when $d = 3$; no extra ideas are required when $d > 3$. For a Borel subset $A$ of $\mathbb{R}^3$, we denote by $|A|$ its Lebesgue measure.

THEOREM 1.1. *There exists a universal constant K with the following property. Consider a convex function $\varphi$ on $\mathbb{R}^3$ (which possibly takes some infinite values) with $\varphi(0) = 0$. We assume that the following hold:*

$$\text{if } t > 0, \text{ then } |\{\varphi \ge t\}| \ge \log t. \tag{1.2}$$

$$\varphi(\pm x_1, \pm x_2, \pm x_3) = \varphi(x_1, x_2, x_3). \tag{1.3}$$

$$\text{if } \varphi(x) \ge 2^{40}, \text{ then } \varphi\Big(\frac{4x}{5}\Big) \le \frac{1}{4}\varphi(x). \tag{1.4}$$

$$\varphi(2^5,0,0), \varphi(0,2^5,0), \varphi(0,0,2^5) < 2^{39}. \tag{1.5}$$

*Then, with probability greater than or equal to $1 - Km^{-2}$, the following occur.*

(a) *The transportation cost of the empirical measure $m^{-1}\sum_{i \le m}\delta_{X_i}$ to $\lambda_3$ is less than or equal to 1 when the cost of moving a unit mass from x to y is measured by $\varphi((x-y)/Km^{-1/3})$.*

(b) *We have*

$$\inf_{\pi} \frac{1}{m} \sum_{i \le m} \varphi\left(\frac{X_i - X'_{\pi(i)}}{Km^{-1/3}}\right) \le 1. \tag{1.6}$$

COMMENTS (a) In the condition "with probability greater than or equal to $1 - Km^{-2}$," the exponent $-2$ can be replaced by any other (with a different constant $K$).

(b) Certainly (1.5) is very mild. If $\varphi$ is finite in a neighborhood of the origin, the function $\varphi(x/K_\varphi)$ satisfies (1.5) for some $K_\varphi$ large enough. Condition (1.3) is satisfied in the most natural examples. While we have not checked all the details, we feel that no essential new ideas are needed to remove the condition. However, this would create technical complications. These would obscure the otherwise rather natural line of approach and we have decided not to pursue this direction. Condition (1.4) fails for the most natural example, $\varphi(x) = \|x\|$. However, it does hold on the natural examples where (1.2) is tight (from which the case $\varphi(x) = \|x\|$ can be recovered). Apparently this condition is purely technical, and there is not much motivation to try to supress it.

(c) Condition (1.2) is the essential point. If one replaces this condition by $|\{\varphi \le t\}| \ge L^{-1} \log t$, for some constant $L > 0$, only trivial modifications to the proof we present are needed to show that the conclusion still holds (with different constants). It must be pointed out that this condition is *necessary*. (The remarkable feature of Theorem 1.1 is that the condition is also sufficient.) More precisely, if (1.6) holds, with constant $K_1$ in the denominator, for $t$ large enough, we must have

$$|\{\varphi \le t\}| \ge \frac{\log t}{KK_1^3}, \tag{1.7}$$

where $K$ is universal. To show (1.7), let us consider $t$, and the largest $m$ such that $\log m$ is less than $L|\{\varphi \le t\}|$, where $L$ is a parameter that will be adjusted later. Then (for $t$ large enough) $|\{\varphi \le t\}|$ is at least $D(t)/K$, where $D(t)$ is the diameter of $\{\varphi \le t\}$. Thus (for $t$ large enough) $D(t) \le (K/L) \log m$ is much smaller than $m^{1/3}$. It then follows that one can find a subset $F$ of $[0,1]^3$, such that the sets $x + 2K_1 m^{-1/3}\{\varphi \le t\}$, for $x \in F$, are all disjoint and contained in $[0,1]^3$ where $\operatorname{card} F \ge m/K_1 \log m$. If we choose $L = KK_1^3$, where $K$ is large enough, it is easy to see that, with probability greater than or equal to $\frac{1}{2}$ we will find some $x \in F$ such that

$$\begin{aligned}&\operatorname{card}\{i \le m;\ X_i \in x + K_1 m^{-1/3}\{\varphi \le t\}\}\\ &\qquad > \operatorname{card}\{i \le m;\ X'_i \in x + 2K_1 m^{-1/3}\{\varphi \le t\}\}.\end{aligned}$$

It then follows that whatever the choice of $\pi$ there exists $i$ such that $X'_{\pi(i)} \notin X_i + K_1 m^{-1/3}\{\varphi \le t\}$, that is,

$$\varphi\left(\frac{X_i - X'_{\pi(i)}}{K_1 m^{-1/3}}\right) \ge t.$$

Now, if (1.6) occurs, we have $t \le m$ so that

$$|\{\varphi \le t\}| \ge \frac{\log m}{L} \ge \frac{\log t}{L}.$$

(d) The strange constants that appear in this statement arise from the fact that no effort has been made to optimize these constants, and that we have rather tried to give the simplest possible proof.

Let us now give a few examples.

EXAMPLE 1.2. Consider the function $\varphi(x) = -1 + \exp\|x\|^3$. Then, for some constant $K$, the function $\varphi(x/K)$ satisfies (1.2)–(1.5). Thus, with probability greater than or equal to $1 - m^{-2}$, we can find a one-to-one map $\pi$ such that

$$\sum_{i \le m} \frac{1}{m} \exp m \frac{\|X_i - X'_{\pi(i)}\|^3}{K} \le 2. \tag{1.8}$$

In particular, we have

$$\max \|X_i - X'_{\pi(i)}\| \le K m^{-1/3} (\log m)^{1/3}$$

(the result of Shor and Yukich [6]). We also have (since $\exp x^3 - 1 \ge x$ for $x \ge 1$)

$$\sum_{i \le m} \|X_i - X'_{\pi(i)}\| \le K m^{2/3},$$

so that (1.8) also improves upon the result of Ajtai, Komlós and Tusnády [1].

EXAMPLE 1.3. Consider the function $\varphi$ given by

$$\varphi(x_1, x_2, x_3) = \begin{cases} +\infty, & \text{if } |x_1| > 1, \\ -1 + \exp\left(x_2^2 + x_3^2\right), & \text{if } |x_1| \le 1. \end{cases}$$

Then, for some constant $K$, the function $\varphi(x/K)$ satisfies (1.2) to (1.5).

EXAMPLE 1.4. Consider the function $\varphi$ given by

$$\varphi(x_1, x_2, x_3) = \begin{cases} +\infty, & \text{if } x_1^2 + x_2^2 \ge 1, \\ -1 + \exp|x_3|, & \text{if } x_1^2 + x_2^2 \le 1. \end{cases}$$

Then, for some constant $K$, the function $\varphi(x/K)$ satisfies (1.2)–(1.5).

While Theorem 1.1 gives, for $d = 3$ (or any $d > 3$), a rather complete solution to the problem of the transportation cost of the empirical measure to the uniform measure, the constant $K$ involved in (1.6) grows rather fast with $d$. The problem of understanding the correct order of growth of this constant in function of $d$ is not solved (and the methods of the present paper would need major improvements to attack this question). So far, progress has been made only when $\varphi(x) = \|x\|$; see [8].

Let us now discuss the methods of this paper. There exist essentially two approaches to this type of problem. The first originates in [1] where an explicit transportation scheme (of considerable interest in itself) is developed. This is the method pursued in [6]. The second method is more indirect; it uses "duality" and, in the end, reduces the problem to the evaluation of the supremum of the empirical process over a certain class of functions. This method was apparently introduced in [2] and is developed in [7]. As of today it seems that the duality method, while less direct and more complicated, is also more powerful. For example, while Yukich has recently succeeded in proving (1.8) using the Ajtai–Komlós–Tusnády scheme, this scheme seems powerless to prove results where one requires a tight control on one component, such as in Examples 1.3 and 1.4. The present paper is, of course, based on the duality method. For the convenience of the reader, everything that is needed about the transportation cost and the use of duality is proved in Section 2. The duality method involves certain classes of functions. It is rather difficult to work with these classes, since their definition is complicated. The purpose of Section 3 is to prove that these classes do satisfy some more manageable properties that will be sufficient for our purposes. In Section 4, we learn how to decompose a function in these classes into simpler pieces, and the final computations are performed in Section 5. The scheme of proof is somehow similar to that of [3], but the proof is made harder by the fact that the present result is much more tight.

It might be useful to provide some help to the reader. The level of technicality is low in Section 2 and increases to reach its peak in Section 5. We hope that the reader will have no difficulty in grasping the overall scheme of the proof. Unfortunately the relative simplicity of the approach is spoiled by the apparently unavoidable fact that in the main computation, much energy must be devoted to controlling lower-order terms. This is a great pity, because the possibility of controlling the main terms is a rather beautiful phenomenon that is buried deep in the proof of Theorem 5.1. This phenomenon is the heart of this paper, since, once it is understood, completing the proof is simply a matter of hard work. This phenomenon (under a slightly different form) appears again in the proof of Proposition 5.8. The reader unwilling to penetrate the entire paper, but wishing to understand why Theorem 1.1 does not reduce to a mere technicality (however complicated), should try to read the proof of this proposition.

**2. Transportation cost.** Consider a compact space $K$ and a continuous function $\psi$ on $K \times K$. The number $\psi(x,y)$ represents the "cost" of transporting one unit of mass from $x$ to $y$. Consider two probability measures $\mu$ and $\nu$ on $K$. The transportation cost of $\mu$ to $\nu$ is defined as the infimum (which is also the minimum) of the integral $\int \psi(x,y)\, d\theta(x,y)$, where $\theta$ belongs to the class of probability measures on $K \times K$ with first and second marginals equal, respectively, to $\mu$ and $\nu$. The reason for this is better understood in the case where $\mu$ (resp. $\nu$) is supported by a finite set $F$ (resp. $G$). In that case, an integral $\int \psi(x,y)\, d\theta(x,y)$

represents the cost incurred when, for all $x \in F$ and $y \in G$, one transports mass $\theta(\{x,y\})$ from $x$ to $y$, and the transportation cost is the minimum cost one can achieve by choosing $\theta$ optimally. A case of special interest is when $\operatorname{card} F = \operatorname{card} G = n$, and when

$$\mu = \frac{1}{n} \sum_{x \in F} \delta_x \quad \text{and} \quad \nu = \frac{1}{n} \sum_{y \in G} \delta_y.$$

In that case, if $\theta$ is a probability on $K \times K$ with marginals $\mu$ and $\nu$ the matrix $(\theta(\{x,y\}))_{x \in F,\, y \in G}$ is bistochastic. A celebrated theorem of Birkoff states that the extreme points of the set of bistochastic matrices are permutation matrices; thus, the transportation cost is obtained at one permutation matrix and is thus the minimum of $(1/n)\sum_{x \in F} \psi(x, \pi(x))$ over one-to-one maps $\pi$ from $F$ to $G$.

How to compute the transportation cost? Depending on one's taste or background, one could appeal to the duality theorem of linear programming or to results of Strassen on measures with given marginal. As, however, the proof requires only a few lines, we will give it. The basic observation is as follows.

LEMMA 2.1. *Consider a compact $K$, and denote by $1_K$ the function identically equal to 1. Consider a linear subspace $X$ of $\mathcal{C}(K)$ containing $1_K$. Consider a linear functional $\xi$ on $X$. Assume that $\xi(1_K) = 1$ and that, whenever $f \in X$, $f \geq 0$, we have $\xi(f) \geq 0$. Then*

$$\forall f \in X, \qquad |\xi(f)| \leq \|f\|. \tag{2.1}$$

PROOF. Suppose that $f \in X$, $f(t) \leq 1$ for all $t \in K$. Then $1_K - f \geq 0$ and $1_K - f \in X$, so that

$$1 - \xi(f) = \xi(1_K) - \xi(f) = \xi(1_K - f) \geq 0,$$

and thus $\xi(f) \leq 1$. If $f(t) \geq -1$ for all $t$ in $K$, using the above for $-f$ gives $\xi(f) \geq -1$. In particular, $|\xi(f)| \leq 1$ if $\|f\| \leq 1$, and the result follows by homogeneity. □

The use of Lemma 2.1 is that condition (2.1) allows us, by the Hahn–Banach theorem, to extend $\xi$ to a norm 1 linear functional on $\mathcal{C}(K)$. Since $\xi(1_K) = 1$, by Riesz's representation theorem, there exists a probability measure $\theta$ on $K$ such that

$$\forall f \in X, \qquad \xi(f) = \int f \, d\theta.$$

Consider now a continuous function $\psi$ on $K \times K$, and a real $a$. When is it possible to find a probability $\theta$ on $K \times K$, with given marginal $\mu$ and $\nu$, respectively, such that $a = \int \psi d\theta$? First we observe a necessary condition. If $b \in \mathbb{R}$ and $f, g \in \mathcal{C}(K)$, by integrating with respect to $\theta$, we see that

$$(2.2)\quad \big(\forall\, x,\, y \in K,\ b\psi(x,y) + f(x) + g(y) \geq 0\big) \Rightarrow ab + \int f\, d\mu + \int g\, d\nu \geq 0.$$

The remarkable fact is that this condition is sufficient. To see it, we consider the linear subspace $X$ of $\mathcal{C}(K \times K)$ that consists of the functions $b\psi(x,y) + f(x) + g(y)$, for $b \in \mathbb{R}$, and $f, g \in \mathcal{C}(K)$, and the linear functional $\xi$ on $X$ given by

$$\xi\big(b\psi(x,y) + f(x) + g(y)\big) = ab + \int f\, d\mu + g\, d\nu.$$

We observe that (2.2) is the hypothesis of Lemma 2.1; we then apply this lemma as in its discussion.

By homogeneity, (2.2) is equivalent to the following conditions together:

$$(2.3)\quad \big(\forall\, x,y \in K,\ \psi(x,y) + f(x) + g(y) \geq 0\big) \quad \Rightarrow \quad a + \int f\, d\mu + \int g\, d\nu \geq 0$$

$$(2.4)\quad \big(\forall\, x,y \in K,\ -\psi(x,y) + f(x) + g(y) \geq 0\big) \quad \Rightarrow \quad -a + \int f\, d\mu + \int g\, d\nu \geq 0.$$

For simplicity, let us assume that $K$ is finite. (In the case where $\mu$ and $\nu$ are supported by finite sets, this is not a loss of generality). Then (2.3) is equivalent to saying that $a + \int f\, d\mu + \int g\, d\nu \geq 0$ when $f$ is the smallest possible, that is,

$$f(x) = \sup_y \big( -\psi(x,y) - g(y)\big) = -\inf_y \big(\psi(x,y) + g(y)\big),$$

so that (2.3) means

$$(2.5)\qquad a \geq \sup_{g \in \mathcal{C}(K)} \left( \int \inf_y \big(\psi(x,y) + g(y)\big)\, d\mu(x) - \int g\, d\nu \right).$$

In a similar fashion, (2.4) means

$$(2.6)\qquad a \leq \inf_{g \in \mathcal{C}(K)} \left( \int \sup_y \big(\psi(x,y) - g(y)\big)\, d\mu(x) + \int g\, d\nu \right).$$

In particular, the smallest possible value of $a$ is given by the right-hand side of (2.5).

We now reformulate this result in the case where $\mu$ and $\nu$ give equal weight to a finite number of points.

PROPOSITION 2.2. *Consider points $x_1, \dots, x_m, y_1, \dots, y_{m'}$ and the measures $\mu = (1/m)\sum_{i \leq m} \delta_{x_i}$ and $\nu = (1/m')\sum_{j \leq m'} \delta_{y_j}$. Then the transportation cost from $\mu$ to $\nu$ is*

$$\sup \left( \frac{1}{m} \sum_{i \leq m} \inf_{j \leq m'} \big(\psi(x_i, y_j) + a_j\big) - \frac{1}{m'} \sum_{j \leq m'} a_j \right),$$

*where the supremum is taken over all $m'$-tuples $a_1, \dots, a_{m'}$.*

We want to study the transportation cost from the empirical measure $(1/m) \times \sum_{i \le m} \delta_{X_i}$ to the uniform measure $\lambda_3$ on $[0, 1]$. For that purpose, it will be convenient to break the transportation process into several steps. We denote by $T_\psi(\mu, \nu)$ the transportation cost from $\mu$ to $\nu$ with cost function $\psi$.

LEMMA 2.3. *Consider three probabilities $\mu$, $\nu$ and $\eta$. Then*

$$T_{\psi_1}(\mu, \eta) \le T_\psi(\mu, \nu) + T_\psi(\nu, \eta),$$

*where*

$$\psi_1(x, y) = \inf_z \big(\psi(x, z) + \psi(z, y)\big).$$

PROOF. We give the proof only in the case we need, that is, where $\mu$ and $\nu$ are supported by a finite set $F$. Then we can find a measure $\theta$ on $F \times F$ with marginals $\mu$ and $\nu$ such that

$$\sum_{x, y \in F} \psi(x, y)\theta\Big(\big\{(x, y)\big\}\Big) = T_\psi(\mu, \nu).$$

Consider a measure $\theta'$ on $F \times K$ with marginal $\nu$ and $\eta$. We can write

$$\theta' = \sum_{x \in F} \nu\big(\{x\}\big)\delta_x \otimes \theta_x,$$

where $\theta_x$ is a probability on $K$, and where $\sum_{x \in F} \nu(\{x\})\theta_x = \eta$. We choose $\theta'$ such that $\int_{F \times K} \psi \, d\theta' = T_\psi(\nu, \eta)$.

Consider the probability $\theta''$ on $F \times K$ given by

$$\theta'' = \sum_{x \in F} \delta_x \otimes \left( \sum_{y \in F} \theta\Big(\big\{(x, y)\big\}\Big)\theta_y \right).$$

It is elementary to see that it has marginals $\mu$ and $\eta$, respectively.

We have

$$\int_{F \times K} \psi_1 \, d\theta'' = \sum_{x \in F} \sum_{y \in F} \int_K \theta\Big(\big\{(x, y)\big\}\Big)\psi_1(x, z)\, d\theta_y(z).$$

Since $\psi_1(x, z) \le \psi(x, y) + \psi(y, z)$ we have

$$\begin{aligned}\int \psi_1 \, d\theta'' &\le \sum_{x \in F} \sum_{y \in F} \psi(x, y)\theta\Big(\big\{(x, y)\big\}\Big) \\ &\quad + \sum_{x \in F} \sum_{y \in F} \int \psi(y, z)\theta\Big(\big\{(x, y)\big\}\Big) d\theta_y(z) \\ &\le T_\psi(\mu, \nu) + \int \psi \, d\theta' = T_\psi(\mu, \nu) + T_\psi(\nu, \eta).\end{aligned}$$

□

In this paper, $K$ will be a subset of $\mathbb{R}^3$, and the transportation cost $\psi(x,y)$ will be $\varphi(x-y)$, where $\varphi$ is a convex function. (In that case, the transportation cost $T_\psi$ will be simply denoted by $T_\varphi$). We observe that, by convexity of $\varphi$, we have

$$2\varphi\left(\frac{x-y}{2}\right) \leq \varphi(x-z)+\varphi(z-y),$$

so that the function $\psi_1$ of Lemma 2.3 satisfies $\psi_1(x,y) \geq 2\varphi((x-y)/2)$.

To study the transportation cost from the empirical measure $(1/m)\sum_{i\leq m}\delta_{X_i}$ to the uniform measure $\lambda_3$ on $[0,1]^3$, we proceed in three steps. We consider a parameter $K_0$; this parameter will be adjusted later. It is a universal constant. We consider the largest $n$ such that $K_0 2^{3n} \leq m$, and the grid $G'_n$ defined as

$$G'_n = \left\{(u_j 2^{-n})_{j\leq 3};\ u_j \in \mathbb{N},\ 0 \leq u_j \leq 2^n - 1\right\}.$$

For $i \leq m$, we consider the largest point $Z_i$ on $G'_n$ that is smaller than $X_i$ (when $\mathbb{R}^3$ is ordered by the cone $\mathbb{R}^{+3}$). The points $Z_i$ are i.i.d. uniform on $G'_n$. Consider the uniform measure $\nu'$ on $G'_n$

Consider a convex function $\xi$ on $\mathbb{R}^3$. Assume that

$$\|x\| \leq \frac{\sqrt{3}}{2^n} \quad \Rightarrow \quad \xi(x) \leq \frac{1}{3}. \tag{2.7}$$

Then it should be clear that

$$T_\xi\left(\frac{1}{m}\sum_{i\leq m}\delta_{X_i}, \frac{1}{m}\sum_{i\leq m}\delta_{Z_i}\right) \leq \frac{1}{3},\ T_\xi(\nu',\lambda_3) \leq \frac{1}{3}.$$

Thus, if we have

$$T_\xi\left(\frac{1}{m}\sum_{i\leq m}\delta_{Z_i}, \nu'\right) \leq \frac{1}{3}, \tag{2.8}$$

it follows from Lemma 2.3 that

$$T_{\xi'}\left(\frac{1}{m}\sum_{i\leq m}\delta_{Z_i}, \lambda_3\right) \leq 1,$$

where $\xi'(x) = 4\xi(x/4)$.

The point is thus to prove (2.8). To simplify the notation, it is more convenient to rescale. We consider the grid

$$G_n = \left\{(u_j)_{j\leq 3} \in \mathbb{N}^3;\ 0 \leq u_j \leq 2^n - 1\right\}.$$

We consider the uniform measure $\nu$ on $G_n$. From Proposition 2.2, we see that for points $(Z_i)_{i\le m}$ of $G_n$ we have

$$
(2.9)\qquad \begin{aligned} &T_\varphi\left(\frac{1}{m}\sum_{i\le m}\delta_{Z_i},\nu\right)\\ &\quad=\sup_{\overline{a}}\left(\frac{1}{m}\sum_{i\le m}\inf_{u\in G_n}\big(\varphi(Z_i-u)+a(u)\big)-2^{-3n}\sum_{u\in G_n}a(u)\right),\end{aligned}
$$

where the supremum is taken over all choices of $\overline{a}=(a(u))_{u\in G_n}$. For such a family $\overline{a}$, we define the function $h_{\overline{a}}$ on $G_n$ by

$$
h_{\overline{a}}(w)=\inf_{u\in G_n}\big(\varphi(w-u)+a(u)\big).
$$

Thus $h_{\overline{a}}(u)\le a(u)$. We set

$$
Eh_{\overline{a}}=2^{-3n}\sum_{u\in G_n}h_{\overline{a}}(u),\qquad E\,\overline{a}=2^{-3n}\sum_{u\in G_n}a(u).
$$

Observe that $Eh_{\overline{a}}\le E\,\overline{a}$.

We can now rewrite (2.9) as

$$
T_\varphi\left(\frac{1}{m}\sum_{i\le m}\delta_{Z_i},\nu\right)=\sup_{\overline{a}}\left(\frac{1}{m}\sum_{i\le m}\big(h_{\overline{a}}(Z_i)-Eh_{\overline{a}}\big)+Eh_{\overline{a}}-E\,\overline{a}\right).
$$

To prove Theorem 1.1, it will suffice, by the preceding analysis, to prove the following statement.

THEOREM 2.4. *There exists a universal constant $K_0$ with the following property. If $m\ge K_0 2^{3n}$, and if the sequence $(Z_i)_{i\le m}$ is i.i.d. uniform on $G_n$, then, with probability greater than or equal to $1-K_0 2^{-6n}$, the following occurs. Given any $\overline{a}=(a(u))_{u\in G_n}$, we have*

$$
\frac{1}{m}\sum_{i\le m}\big(h_{\overline{a}}(Z_i)-Eh_{\overline{a}}\big)\le\max\left(\frac{1}{3},E\,\overline{a}-Eh_{\overline{a}}\right).
$$

(Indeed, one then chooses $n$ as the largest for which $2^{3n}\le K_0 m$, so that $m^{1/3}\le K2^n$.)

**3. Geometry.** Consider a number $\Delta>0$. In this section we study the class $\mathcal{S}(\Delta)$ of functions on $G_n$ given by

$$
\mathcal{S}(\Delta)=\{h_{\overline{a}};\ E\,\overline{a}-Eh_{\overline{a}}\le\Delta\}.
$$

The problem is that it is not clear a priori what functions of $\mathcal{S}(\Delta)$ look like; the purpose of this section is to describe a property of this class which we can manipulate more easily than the definition. This property is related to a sequence $(\mathcal{P}_k)$ of partitions of $G_n$. This sequence is closely related to the geometry of $\varphi$, and our first task is to construct it.

For a subset $A$ of $\mathbb{R}^3$, we recall that we denote by $|A|$ the Lebesgue measure of $A$. For $k \geq 0$, we define a number $b_k$ by

$$|\{\varphi \leq b_k\}| = 2^{k+39}. \tag{3.1}$$

It follows from (1.2) that

$$\log b_k \leq 2^{k+39}. \tag{3.2}$$

For $K \geq 1$ and $j \leq 3$, we consider the smallest integers $n_j(k) \in \mathbb{Z}$ such that

$$\{\varphi \leq b_k\} \subset \prod_{j\leq 3} \left[ -2^{n_j(k)+6}, 2^{n_j(k)+6} \right].$$

This implies that

$$\varphi(2^{n_1(k)+6}, 0, 0) \geq b_k, \qquad \varphi(0, 2^{n_2(k)+6}, 0) \geq b_k, \qquad \varphi(0, 0, 2^{n_3(k)+6}) \geq b_k.$$

By condition (1.5), this implies that $n_j(k) \geq 0$ for $k \geq 0$ and $j \leq 3$.

Also, by condition (1.3) and the definition of $n_j(k)$ we have

$$\varphi(2^{n_1(k)+5}, 0, 0) \leq b_k, \qquad \varphi(0, 2^{n_2(k)+5}, 0) \leq b_k, \qquad \varphi(0, 0, 2^{n_3(k)+5}) \leq b_k.$$

If we set

$$Z_k = \prod_{j\leq 3} \left[ -2^{n_j(k)}, 2^{n_j(k)} \right],$$

by convexity of $\varphi$, we have

$$2^3 Z_k \subset \frac{2^5}{3} Z_k \subset \{\varphi \leq b_k\} \subset 2^6 Z_k. \tag{3.3}$$

Let us observe that

$$2^{k+39} = |\{\varphi \leq b_k\}| \leq 2^{18}|Z_k| = 2^{39} 2^{\sum_{j\leq 3} n_j(k)},$$

so that

$$\sum_{j\leq 3} n_j(k) \geq k. \tag{3.4}$$

Another useful property is that

$$Z_{k+1} \subset 2^{27} Z_k. \tag{3.5}$$

Indeed, by (3.3) we have

$$\left(\frac{2^5}{3}\right)^3 |Z_{k+1}| \leq |\{\varphi \leq b_{k+1}\}| = 2^{k+40},$$

so that

$$|Z_{k+1}| \leq 27 \cdot 2^{k+25} \leq 2^{k+30}.$$

By (3.4), we have $|Z_{k+1}| \leq 2^{12}|Z_k|$, and this implies the result since $Z_k \subset Z_{k+1}$.

We set $m_j(0) = 0$ for $j \leq 3$. For $j \leq 3$ and $k \geq 1$, we set $m_j(k) = \min(n, n_j(k))$. We consider the following subset of $G_n$:

$$R_k = \prod_{j \leq 3} \{0, \dots, 2^{m_j(k)} - 1\}.$$

Thus, $R_0 = \{(0,0,0)\}$.

For $k \geq 0$, we consider the partition $\mathcal{P}_k$ of $G_n$ by sets $u + R_k$, where

$$u = \left(u_1 2^{m_1(k)}, u_2 2^{m_2(k)}, u_3 2^{m_3(k)}\right), \qquad 0 \leq u_j \leq 2^{n-m_j(k)} \text{ for } 0 \leq j \leq 3.$$

We will say that two such sets $u + R_k$ and $u' + R_k$ are $q$-adjacent if they satisfy $|u_j - u_j'| \leq q$ for each $j \leq 3$. We will say adjacent instead of 1-adjacent.

We note that the sequence $(\mathcal{P}_k)$ of partitions of $G_n$ decreases. Let us now state a crucial property.

LEMMA 3.1. *Consider $k \geq 1$. Consider $x \in [0, 2^n - 1]^3$. Then we can find $A \in \mathcal{P}_k$ such that $A \subset x + Z_k$.*

PROOF. If $A$ is the element of $\mathcal{P}_k$ such that $x \in \operatorname{conv} A$, then $A \subset Z_k + x$, since $\operatorname{conv} A$ is a translate of $Z_k/2$. □

We now turn to the basic result. For $k \geq 0$, we set $s(k) = \sum_{j \leq 3} m_j(k)$.

THEOREM 3.2. *Consider a function $f \in \mathcal{S}(\Delta)$. Then we can find a partition $(B_k)_{k \geq 0}$ of $G_n$ such that $B_k$ is a union of sets of $\mathcal{P}_k$, and for each $C \in \mathcal{P}_k$, $C \subset B_k$, we can find a number $z(C)$ such that the following properties hold:*

$$\sum_{k \geq 0} \sum_{C \in \mathcal{P}_k, C \subset B_k} 2^{s(k)} z(C) \leq K 2^{3n} \Delta. \tag{3.6}$$

$$k \geq 0, C \in \mathcal{P}_k, C \subset B_k \Rightarrow b_k \leq z(C) < b_{k+1}. \tag{3.7}$$

*If $C$ and $C'$ belong to $\mathcal{P}_k$, $k \geq 0$, and are adjacent, and if $C \subset B_k$, then*

$$u \in C,\ u' \in C' \quad \Rightarrow \quad |f(u) - f(u')| \leq z(C). \tag{3.8}$$

*(Observe that this condition holds in particular when $C' = C$.)*

PROOF.
*Step* 1. By definition of $\mathcal{S}(\Delta)$, we can find a family $\overline{a} = (a(u))_{u \in G_n}$ such that

$$\forall\, u \in G_n, \qquad f(u) = \min\{a(w) + \varphi(u - w);\ w \in G_n\}, \tag{3.9}$$

$$\sum_{u \in G_n} \big(a(u) - f(u)\big) \le 2^{3n}\Delta. \tag{3.10}$$

In particular, for each $w$ in $G_n$, we have

$$f(u) \le a(w) + \varphi(u - w).$$

This shows that if we define

$$\widehat{f}(w) = \max\left\{f(u) - \varphi(w - u),\ u \in G_n\right\}, \tag{3.11}$$

we have $\widehat{f}(w) \le a(w)$. By (3.10) we have

$$\sum_{w \in G_n} \left(\widehat{f}(w) - f(w)\right) \le \sum_{w \in G_n} \big(a(w) - f(w)\big) \le 2^{3n}\Delta. \tag{3.12}$$

*Step* 2. Consider $C$ in $\mathcal{P}_k$. We define

$$y(C) = \min_{u \in C} \widehat{f}(u) - \max_{u \in C} f(u).$$

We consider a universal constant $q$, to be determined later. We set

$$x(C) = \max\Big(0, \max\big\{y(C');\ C' \in \mathcal{P}_k,\ C' \text{ is } q\text{-adjacent to } C\big\}\Big).$$

For $k \ge 1$, we define $D_k$ as the union of all the sets $C \in \mathcal{P}_k$ for which $x(C) \ge b_k/2$. We observe that $D_l = \emptyset$ for $l$ large enough; we set

$$B_k = D_k \setminus \bigcup_{l > k} D_l.$$

We set $B_0 = G_n \setminus \cup_{l \ge 1} D_l$.
*Step* 3. We prove that

$$\sum_{k \ge 1} \sum_{C \subset B_k, C \in \mathcal{P}_k} 2^{s(k)} x(C) \le K 2^{3n}\Delta. \tag{3.13}$$

If $C \in \mathcal{P}_k, C \subset B_k$ and $x(C) > 0$, by definition of $x(C)$, we can choose $\theta(C) \in \mathcal{P}_k$ that is $q$-adjacent to $C$, and such that $x(C) = y(\theta(C))$. Since $C$ is $q$-adjacent to $\theta(C)$, there can be at most $(2q+1)^3$ sets $C' \in \mathcal{P}_k$ for which $\theta(C') = \theta(C)$. We define

$$L_k = \bigcup \left\{\theta(C);\ C \in \mathcal{P}_k,\ C \subset B_k\right\},$$

and we observe that $y(C) \ge b_k/2$ whenever $C \in \mathcal{P}_K$ and $C \subset L_k$.

We see that

$$\sum_{C \subset B_k, C \in \mathcal{P}_k} 2^{s(k)} x(C) \le (2q+1)^3 \sum_{C \subset L_k, C \in \mathcal{P}_k} 2^{s(k)} y(C)$$
$$\le (2q+1)^3 \sum_{u \in L_k} [\widehat{f}(u) - f(u)],$$

where the last inequality follows from the definition of $y(C)$ and the fact that each $C \in \mathcal{P}_k$ has cardinality $2^{s(k)}$. Thus, to prove (3.13), it suffices to show that the sets $(L_k)_{k \ge 1}$ are disjoint.

Consider $k > 1$, $C \in \mathcal{P}_k$, $C \subset L_k$. If $C' \in \mathcal{P}_k$ is $q$-adjacent to $C$, we have $x(C') \ge y(C)$ by definition of $x(C')$. Thus $x(C') \ge b_k/2$, and hence $C' \subset D_k$. Consider now $k_1 < k, C_1, C_1' \in \mathcal{P}_{k_1}$, $C_1 \subset C$ and denote by $C'$ the element of $\mathcal{P}_k$ that contains $C_1'$. If $C_1$ and $C_1'$ are $q$-adjacent, then $C$ and $C'$ are $q$-adjacent, so that, as shown, $C' \subset D_k$ and thus, by definition of $B_{k_1}$, $C_1' \not\subset B_{k_1}$. Thus $C_1$ cannot be $q$-adjacent to any $C_1' \in \mathcal{P}_{k_1}$, $C_1' \subset B_{k_1}$. In particular $C_1 \not\subset L_{k_1}$. Thus $L_k$ and $L_{k_1}$ are disjoint.

*Step* 4. If $C \in \mathcal{P}_k$ and $C \subset B_k$, $k \ge 1$, we set $z(C) = \min(2x(C), b_{k+1}) \ge b_k$. If $C \in \mathcal{P}_0$ and $C \subset B_0$, we set $z(C) = b_0$. Thus (3.6) follows from (3.13), and (3.7) holds by construction. It remains to prove (3.8). With the notation of (3.8), we prove in this step that

$$f(u') \le f(u) + z(C).$$

By (3.9), we can find $w$ in $G_n$ such that

$$f(u) = a(w) + \varphi(u - w). \tag{3.14}$$

We consider the number

$$b = \max\bigl(f(u') - f(u), \varphi(u - w)\bigr).$$

Since we want to prove that $f(u') - f(u) \le z(C)$, we can assume that $b \ge b_k$. We set $U = \{\varphi \le b\}$. Since $C$ and $C'$ are adjacent, we have, by (3.3),

$$u - u' \in 2Z_k \subset \frac{U}{4}.$$

Since $\varphi(w - u) \le b$, we have $w \in u + U$, so that $w \in u' + 5U/4$ (and $u' \in w + 5U/4$). Thus we have

$$\frac{w + u'}{2} \in u' + \frac{5U}{8}, \qquad \frac{w + u'}{2} \in w + \frac{5U}{8}. \tag{3.15}$$

We set

$$V = \frac{w + u'}{2} + \frac{U}{8}.$$

By (3.15) we have

$$V \subset \left(u' + \frac{3U}{4}\right) \cap \left(w + \frac{3U}{4}\right).$$

We now recall that by condition (1.4) we have $\varphi(x) \le b/4$ for $x \in 3U/4$. Thus, for $v \in u' + 3U/4$, we have, by definition of $\widehat{f}$

$$\widehat{f}(v) \ge f(u') - \varphi(u' - v) \ge f(u') - \frac{b}{4}.$$

By (3.9), for $v \in w + 3U/4$, we have

$$f(v) \le a(w) + \varphi(w - v) \le a(w) + \frac{b}{4}.$$

Thus, for $v \in V$ and $v \in G_n$, we have, by (3.14),

$$\begin{aligned} \min_{v\in V} \widehat{f}(v) - \max_{v\in V} f(v) &\ge f(u') - a(w) - \frac{b}{2} \\ &\ge f(u') - f(u) + \varphi(u - w) - \frac{b}{2} \\ &\ge \max\big(f(u') - f(u), \varphi(u - w)\big) - \frac{b}{2} \\ &= \frac{b}{2}. \end{aligned} \tag{3.16}$$

We consider the largest integer $l \ge 0$ such that $b \ge b_l$. Since $b \ge b_k$, we have $l \ge k$. Also $b < b_{l+1}$. By (3.3) and Lemma 3.1, $V$ contains a set $A$ of $\mathcal{P}_l$, and by (3.16) we have $y(A) \ge b/2$.

Since $b < b_{l+1}$, by definition of $Z_l$ we have $U \subset 2^6 Z_{l+1}$ so that $U \subset 2^{33} Z_l$ by (3.5). We recall that $u' - u \in U/4$, so that

$$V \subset u' + \frac{3U}{4} \subset u + U \subset u + 2^{33} Z_l.$$

This shows that if $D \in \mathcal{P}_l$ is the unique set of $\mathcal{P}_l$ that contains $C$, then $D$ and $A$ are $2^{33}$-adjacent. Thus, if we have taken $q = 2^{33}$, we have $x(D) \ge y(A) \ge b/2 \ge b_l/2$. If we had $l > k$, we would then have $D \in D_l$; but this is impossible since $C \in B_k$; thus $l = k$. Then $D = C$ and $b \le 2x(C)$. Also, $b \le b_{l+1}$, so that $b \le z(C)$.

*Step* 5. Now we have to prove that

$$f(u) \le f(u') + z(C).$$

The argument is almost identical to that of Step 4, exchanging $u$ and $u'$. The one difference is that now we have

$$u - u' \in \frac{U}{4}, \qquad w \in u' + U \subset u + \frac{5U}{4},$$

so that

$$\frac{u + w}{2} \subset u + \frac{5U}{8} \subset u + U. \qquad \square$$

**4. Decomposing functions of $\mathcal{S}(\Delta)$.** Consider the class $\mathcal{S}'(\Delta)$ of functions on $G_n$ that satisfy conditions (3.6)–(3.8). The only property of the class $\mathcal{S}(\Delta)$ that we will use is that $\mathcal{S}(\Delta) \subset \mathcal{S}'(\Delta)$. We now show that in the study of $\mathcal{S}'(\Delta)$ there is no loss of generality to assume that

$$\sum_{j \le 3} n_j(k) = k. \tag{4.1}$$

(This will simplify the notation.) For that purpose, we observe that for $j \le 3$, we can find nondecreasing sequences $(n_j'(l))_l$ such that $\sum_{j \le 3} n_j(l) = l$, and that, for each $k \ge 1$ there exists $l, l \ge k$, such that $n_j(k) = n_j'(l)$ for $j \le 3$. We can then consider the sets

$$R_l' = \prod_{j \le 3} \{0, \dots, 2^{\min(n, n_j'(l))} - 1\}$$

and the partition $\mathcal{P}_k'$ of $G_n$ by translates of $R_l'$. Certainly $\mathcal{P}_k'$ is finer than $\mathcal{P}_k$, so that the class $\mathcal{S}'(\Delta)$ increases if we replace the sequence $(\mathcal{P}_k)$ by the sequence $(\mathcal{P}_k')$. This concludes the argument, and from now on we assume that (4.1) holds.

Consider $f \in \mathcal{S}'(\Delta)$ and a partition $(B_k)_{k \ge 0}$ of $G_n$ that satisfies conditions (3.6)–(3.8). Consider the function $f'$ defined as follows: if $C \in \mathcal{P}_k$ and $C \subset B_k$, then $f'$ is constant on $C$ and its value is the average value of $f$ on $C$. Set $f'' = f - f'$. Thus

$$C \in \mathcal{P}_k,\ C \subset B_k,\ u \in C \quad \Rightarrow \quad |f''(u)| \le z(C). \tag{4.2}$$

We now study $f'$. We denote by $k_0$ the largest integer such that $\max_{j \le 3} n_j(k_0) \le n$. We will decompose $f'$ as

$$f' = \sum_{0 \le k \le k_0} f_k.$$

If $\mathcal{A}_k$ denotes the algebra on $G_n$ generated by $\mathcal{P}_k$, we set

$$f_{k_0} = E(f \mid \mathcal{A}_{k_0})$$

and, for $0 \le k < k_0$, we set

$$f_k = E(f \mid \mathcal{A}_k) - E(f \mid \mathcal{A}_{k+1})$$

We now study the functions $f_k$. A first simple observation is that if $C \in \mathcal{P}_k$, $k < k_0$, and if $C \subset \cup_{l > k} B_l$, then $f_k$ is zero on $C$, since $f'$ is constant on the element $C' \in \mathcal{P}_{k+1}$, $C' \supset C$.

We denote by $(e_j)_{j \le 3}$ the canonical basis of $\mathbb{R}^3$.

LEMMA 4.1. *Consider $C$ and $C'$ in $\mathcal{P}_k$. Assume that $C' = C + e_j 2^{n_j(k)}$, and that neither $C$ nor $C'$ are contained in $\cup_{l > k} A_l$. Denote by $c$ (resp. $c'$) the average value that $f$ takes on $C$ (resp. $C'$). Then*

$$|c - c'| \le \sum_{l \le k} 2^{-k + l + n_j(k) - n_j(l)} \sum \{z(D);\ D \in \mathcal{P}_l,\ D \subset (C \cup C') \cap B_l\}.$$

PROOF. To simplify the notation, we assume $j = 1$. Consider the projection $W$ of $C$ and $C'$ on $\mathbb{R}^2$, that is,

$$W = \{w \in \mathbb{R}^2,\ \exists\, u_1 \in \mathbb{R},\ (u_1, w) \in C\}$$
$$= \{w \in \mathbb{R}^2,\ \exists\, u_1 \in \mathbb{R},\ (u_1, w) \in C'\}.$$

For $w \in W$, we set

$$c(w) = 2^{-n_1(k)} \sum \{f(u_1, w);\ (u_1, w) \in C\},$$

and we define $c'(w)$ in a similar way. For $w \in W$, we consider the collection $Z(w)$ of elements $D$ such that, for some $l \le k$, $D \in \mathcal{P}_l$ and $D \subset B_l \cap (C \cup C')$, and such that $D$ contains at least a point of the type $(u_1, w)$. It should be clear from (3.8) that if $(u_1, w) \in C$ and $(u_2, w) \in C'$, we have

$$|f(u_1, w) - f(u_2, w)| \le \sum \{z(D);\ D \in Z(w)\}.$$

We average this inequality over the $2^{n_1(j)}$ possible values of $u_1$ and $u_2$, and we get

$$\big|c(w) - c'(w)\big| \le \sum \{z(D);\ D \in Z(w)\}.$$

We now sum these inequalities over the $2^{n_2(l) + n_3(l)}$ values of $w$. Thus we get

$$\left| \sum_{w \in W} c(w) - \sum_{w \in W} c'(w) \right|$$
$$\le \sum \{2^{n_2(l) + n_3(l)} z(D);\ l \le k,\ D \in \mathcal{P}_l,\ D \subset (C \cup C') \cap B_l\}.$$

To obtain the result, we remember that $n_2(l) + n_3(l) = l - n_1(l)$ and we divide both sides by $\operatorname{card} W = 2^{n_2(k) + n_3(k)} = 2^{k - n_1(k)}$. $\square$

Since the function $f_k$ is constant on each element of $\mathcal{P}_k$, it is more natural to see it as a function on the grid

$$H_k = \prod_{j \le 3} \{0, \dots, 2^{n - n_j(k)} - 1\}.$$

To simplify notations, we set, for $l \ge 0$,

$$\alpha_l(f) = 2^{s(l)} \sum \{z(C);\ C \in \mathcal{P}_l,\ C \subset B_l\}.$$

Thus by (3.6) we have $\sum_{l \ge 0} \alpha_l(f) \le K 2^{3n} \Delta$.

LEMMA 4.2. *Consider $k < k_0$ and the unique $i \le 3$ such that $n_i(k + 1) = n_i(k) + 1$. Then, for $l \le k$ and $v \in H_k$, we can find numbers $a_{i,l,k}(v)$ with the* *following properties*:

$$\forall\, v \in H_k, |f_k(v)| \le \sum_{l \le k} a_{i,l,k}(v); \tag{4.3}$$

$$\forall\, l \le k, \sum_{v \in H_k} a_{i,l,k}(v) \le 2^{-k+n_i(k)-n_i(l)} \alpha_l(f); \tag{4.4}$$

$$\forall\, v \in H_k, \forall\, l \le k, \qquad a_{i,l,k}(v) \le 2^{n_i(k)-n_i(l)} b_{l+1} \le 2^{k-l} b_{l+1}. \tag{4.5}$$

PROOF. To each $v \in H_k$ corresponds $C_v \in \mathcal{P}_k$. We denote by $C'_v$ the unique element of $\mathcal{P}_{k+1}$ that contains $C_v$. Thus $C'_v = C_v \cup C_{v'}$, where either $v = v' + e_i$ or $v = v' - e_i$.

Since $E(f_k \mid \mathcal{A}_{k+1}) = 0$ we have $f_k(v) + f_k(v') = 0$, so that

$$f_k(v) = \frac{f_k(v) - f_k(v')}{2} = \frac{E(f \mid \mathcal{A}_k)(v) - E(f \mid \mathcal{A}_k)(v')}{2}. \tag{4.6}$$

We set, for $l \le k$,

$$c_{i,l,k}(v) = \tfrac{1}{2} \sum \{z(D);\ D \in \mathcal{P}_l,\ D \subset C'_v \cap B_l\},$$
$$a_{i,l,k}(v) = 2^{-k+l+n_i(k)-n_i(l)} c_{i,l,k}(v).$$

Thus, by Lemma 4.1 and (4.6) we have

$$|f_k(v)| \le \sum_{l \le k} a_{i,l,k}(v). \tag{4.7}$$

To prove (4.4), we observe that one given number $z(C)$ can contribute to at most two different $c_{i,l,k}(v)$ and we note that $\alpha(l) = l$ for $l \le k_0$. Condition (4.5) follows from the fact that $z(D) \le b_{l+1}$ for $D \in \mathcal{P}_l$, $D \subset B_l$, and that $2c_{i,l,k}(v)$ is the sum of at most $2^{k-l+1}$ such terms. □

In the sequel, it will be useful to know that the numbers $b_l$ increase fast enough. We prove that

$$b_{l+1} \ge 4 b_l. \tag{4.8}$$

Indeed, by (1.4) we have

$$\left|\left\{\varphi \le \frac{b_{l+1}}{4}\right\}\right| \ge \left(\frac{4}{5}\right)^3 \left|\{\varphi \le b_{l+1}\}\right| \ge \frac{1}{2} 2^{(l+1)+39}$$
$$= 2^{l+39} = \left|\{\varphi \le b_l\}\right|.$$

For a function $h$ on $H_k$, and $j \le 3$, we define

$$\Delta_j h(v) = |h(v + e_j) - h(v)|,$$

when $v, v+e_j \in H_k$. When $v \in H_k$ and $v+e_j \notin H_k$, we define $\Delta_j h(v) = 0$.

LEMMA 4.3. *Consider $k < k_0$ and $j \leq 3$, and assume that $n_j(k+1) = n_j(k)$. Then, for $l \leq k$ and $v \in H_k$, we can find numbers $a_{j,l,k}(v)$ with the following properties*:

$$\forall\, v \in H_k, \qquad \Delta_j f_k(v) \leq \sum_{l \leq k} a_{j,l,k}(v); \tag{4.9}$$

$$\forall\, v \in H_k, \qquad a_{j,l,k}(v) \leq 2^{n_j(k)-n_j(l)+1} b_{l+1} \leq 2^{k-l+1} b_{l+1}; \tag{4.10}$$

$$\forall\, l < k, \quad \sum_{v \in H_k} a_{j,l,k}(v) \leq 2^{-k+n_j(k)-n_j(l)+2} \alpha_l(f); \tag{4.11}$$

$$\sum_{v \in H_k} a_{j,k,k}(v) \leq 2^{-k+4} \sum_{r \geq k} 4^{-(r-k)} \alpha_r(f). \tag{4.12}$$

PROOF. Consider $v \in H_k$, such that $v' = v + e_j \in H_k$. Consider the elements $C$ and $C'$ of $\mathcal{P}_k$ that correspond to $v$ and $v'$, respectively.

*Case a.* If $C$ is contained in $\cup_{l>k} B_l$, then $f_k(v) = 0$, so that

$$\Delta_j f_k(v) = |f_k(v')|.$$

*Case b.* If $C'$ is contained in $\cup_{l>k} B_l$, the same conclusions holds.

*Case c.* Neither $C$ nor $C'$ is contained in $\cup_{l>k} B_l$. We recall that $f_k = g_k - g_{k+1}$, where $g_l = E(f \mid \mathcal{A}_l)$.

We set

$$a'_{j,l,k}(v) = 2^{-k+l+n_j(k)-n_j(l)} \sum \left\{ z(D); D \in \mathcal{P}_l,\ D \subset B_l \cap (C \cup C') \right\}. \tag{4.13}$$

By Lemma 4.1, we have

$$|g_k(v) - g_k(v')| \leq \sum_{l \leq k} a'_{j,l,k}(v).$$

We denote by $E$ (resp. $E'$) the element of $\mathcal{P}_{k+1}$ that contains $C$ (resp. $C'$). We set, for $l \leq k+1$,

$$a''_{j,l,k}(v) = 2^{-(k+l)+l+n_j(k+1)-n_j(l)} \sum \left\{ z(D);\ D \in \mathcal{P}_l;\ D \subset B_l \cap (E \cup E') \right\}. \tag{4.14}$$

We observe that, since neither $C$ nor $C'$ is contained in $B_{k+1}$, we have $a''_{j,k+1,k}(v) = 0$. Thus, it follows from Lemma 4.1 that

$$|g_{k+1}(v) - g_{k+1}(v')| \leq \sum_{l \leq k} a''_{j,l,k}(v).$$

Thus we have

$$(4.15) \qquad |f_k(v) - f_k(v')| \leq \sum_{l \leq k} \left(a''_{j,l,k}(v) + a'_{j,l,k}(v)\right).$$

We now define $a_{j,l,k}(v)$ as follows. If neither $C$ nor $C'$ is contained in $\cup_{r>k} B_r$ or if $l < k$, we set $a_{j,l,k}(v) = a'_{j,l,k}(v) + a''_{j,l,k}(v)$.

If either $C$ or $C'$ is contained in $\cup_{r>k} B_r$, we set $a_{j,k,k}(v) = 2b_{k+1}$.

Thus, (4.9) follows from (4.15), (3.7) and (3.8). To prove (4.10) it suffices to note (as in the proof of Lemma 4.1) that [since $n_j(k+1) = n_j(k)$]

$$a'_{j,l,k}(v), a''_{j,l,k}(v) \leq 2^{n_j(k) - n_j(l) + 1} b_{l+1}.$$

To prove (4.11), we observe (as in the proof of Lemma 4.1) that

$$\sum_{v \in H_k} a'_{j,l,k}(v) \leq 2^{-k+1+n_j(k)-n_j(l)} \alpha_l(f),$$

$$\sum_{v \in H_k} a''_{j,l,k}(v) \leq 2^{-k+1+n_j(k)-n_j(l)} \alpha_l(f).$$

[We note that a given $D \subset \mathcal{P}_l$ and $D \subset B_l$ can contribute to $a'_{j,l,k}(v)$ [resp. $a''_{j,l,k}(v)$] for up to two (resp. four) different values of $v$.]

We turn to the proof of (4.12). Let us denote by $N$ the number of points $v$ for which either $C$ or $C'$ is contained in $\cup_{r>k} B_r$. We have

$$\sum_{v \in H_k} a_{j,k,k}(v) \leq 2^{-k+2} \alpha_k(f) + 2N b_{k+1}.$$

For $r > k$, set

$$q_r = \operatorname{card}\{D \in \mathcal{P}_r,\ D \subset B_r\}.$$

For each $D \in \mathcal{P}_r$ there can be at most $2^{s(r)-k+1}$ sets $C$ of $\mathcal{P}_k$ for which either $C$ or $C'$ belong to $D$. Thus we have $N \leq \sum_{r>k} 2^{s(r)-k+1} q_r$. We recall that

$$2^{s(r)} \sum_{D \in \mathcal{P}_r, D \subset B_r} z(D) \leq \alpha_r(f).$$

Since $z(D) \geq b_r$ for $D \subset B_r$, we have $2^{s(r)} q_r b_r \leq \alpha_r(f)$.

Thus we have, using (4.8),

$$N \leq 2^{-k+1} \sum_{r>k} \frac{\alpha_r(f)}{b_r} \leq \frac{2^{-k+1}}{b_{k+1}} \sum_{r>k} 4^{-(r-k-1)} \alpha_r(f).$$

This completes the proof. □

We now turn to the study of $f_{k_0}$. In that case, it will suffice to use considerably less information than we had in Lemmas 4.2 and 4.3.

LEMMA 4.4. *For* $j \le 3$ *we have*

$$\sum_{v \in H_{k_0}} \Delta_j f_{k_0}(v) \le K 2^{3n - k_0 + n_j(k_0)} \Delta.$$

PROOF. We can suppose $n_j(k_0) < n$, since otherwise the left-hand side is zero. Consider $v \in H_{k_0}$ such that $v' = v + e_j \in H_{k_0}$. Consider the elements $C$ and $C'$ of $\mathcal{P}_{k_0}$ that correspond to $v$ and $v'$, respectively. If, for some $l \ge k_0$, we have $C \subset D \in \mathcal{P}_l$ and $D \subset B_l$, then

$$|f_{k_0}(v) - f_{k_0}(v')| \le z(D). \tag{4.16}$$

The same inequality occurs if instead $D \supset C'$. If neither $C$ nor $C'$ is contained in $\cup_{l \ge k_0} B_l$ then, by Lemma 4.1, we have

$$\Delta_j f_{k_0}(v) = |f_{k_0}(v) - f_{k_0}(v')| \le \sum_{l < k_0} a'_{j,l}(v), \tag{4.17}$$

where

$$a'_{j,l}(v) = 2^{-k_0 + l + n_j(k_0) - n_j(l)} \sum \{z(D);\ D \in \mathcal{P}_l;\ D \subset B_l \cap (C \cup C')\}.$$

By summation of (4.16) and (4.17) over all the values of $v$, we have

$$\sum_{v \in H_{k_0}} \Delta_j f_{k_0}(v) \le 2^{-k_0} \left[ \sum_{l < k_0} 2^{n_j(k_0) - n_j(l) + 1} \alpha_l(f) + \sum_{l \ge k_0} \alpha_l(f) \right].$$

The result follows since $n_j(l) \ge 0, \Sigma_{l \ge 0} \alpha_l(f) \le K 2^{3n} \Delta$. □

Unfortunately we will not be able to study the functions $f_k$ directly. To each of these functions we will apply a series of "reductions." We now describe the general procedure. This procedure is based on the elementary equality

$$h_1 y_1 + h_2 y_2 = \frac{h_1 + h_2}{2}(y_1 + y_2) + \frac{h_1 - h_2}{2}(y_1 - y_2). \tag{4.18}$$

Consider numbers $q_1, q_2, q_3 \ge 0$ and the grid

$$H = H(q_1, q_2, q_3) = \prod_{j \le 3} \{0, 1, \dots, 2^{q_j} - 1\}.$$

Consider $i \le 3$ such that $q_i \ge 1$. Consider the grid $H' = H(q'_1, q'_2, q'_3)$, where $q'_j = q_j$ if $j \ne i$ and $q'_i = q_i - 1$. Consider the map $Q$ from $H$ to $H'$ that sends $u = (u_1, u_2, u_3)$ to $Q(u) = (u'_1, u'_2, u'_3)$ where $u'_j = u_j$ if $j \ne i$ and $u'_i = [u_i/2]$. For $v \in H'$, we write $Q^{-1}(v) = \{v^-, v^+\}$, where the $i$-th component of $v^-$ is less than the $i$-th component of $v^+$.

With a function $h$ on $H$, we associate two functions $h^+$ and $h^-$ on $H'$ given by

$$h^+(v) = \tfrac{1}{2}\big(h(v^+) + h(v^-)\big), \qquad h^-(v) = \tfrac{1}{2}\big(h(v^+) - h(v^-)\big).$$

It follows from (4.18) that, given numbers $(Y_u)_{u \in H}$, we have

$$\sum_{u \in H} h(u) Y_u = \sum_{v \in H'} h^+(v) Y_v^+ + \sum_{v \in H'} h^-(v) Y_v^-, \tag{4.19}$$

where $Y_v^+ = Y_{v^+} + Y_{v^-}$ and $Y_v^- = Y_{v^+} - Y_{v^-}$.

The consequence of (4.19) that we will use is that, for a class $\mathcal{C}$ of functions on $H$, we have

$$\sup_{h \in \mathcal{C}} \left| \sum_{u \in H} h(u) Y_u \right| \le \sup_{h \in \mathcal{C}^+} \left| \sum_{v \in H'} h(v) Y_v^+ \right| + \sup_{h \in \mathcal{C}^-} \left| \sum_{v \in H'} h(v) Y_v^- \right|, \tag{4.20}$$

where

$$\mathcal{C}^+ = \{h^+;\ h \in \mathcal{C}\}, \qquad \mathcal{C}^- = \{h^-;\ h \in \mathcal{C}\}.$$

To make good use of this formula, it is essential to be able to describe $\mathcal{C}^+$ and $\mathcal{C}^-$. This is the purpose of the next two lemmas.

LEMMA 4.5. *Suppose that*

$$\forall\, u \in H, \qquad \Delta_i h(u) \le \sum_{l \le k} a_l(u).$$

*Then the following hold*:

(a)
$$\forall\, v \in H', \qquad |h^-(v)| \le \sum_{l \le k} d_l(v),$$

*where*

$$\forall\, l \le k,\ \sum_{v \in H'} d_l(v) \le \tfrac{1}{2} \sum_{u \in H} a_l(u), \tag{4.21}$$

$$\forall\, l \le k, \qquad \max_{v \in H'} d_l(v) \le \tfrac{1}{2} \max_{u \in H} a_l(u); \tag{4.22}$$

(b)
$$\forall\, v \in H', \qquad \Delta_i h^+(v) \le \sum_{l \le k} c_l(v),$$

*where*

$$\forall\, l \le k,\ \sum_{v \in H'} c_l(v) \le \sum_{l \le k} a_l(u), \tag{4.23}$$

$$\forall\, l \le k, \qquad \max_{v \in H'} c_l(v) \le 2 \max a_l(u). \tag{4.24}$$

PROOF. (a) This is obvious, since $|h^-(v)| = \frac{1}{2}\Delta_i h(v^-)$.
(b) This follows from the inequality

$$\left|\frac{b_+ + b_-}{2} - \frac{a_+ + a_-}{2}\right| \le \left|\frac{b_+ - b_-}{2}\right| + |b_- - a_+| + \left|\frac{a_+ - a_-}{2}\right|. \qquad \square$$

LEMMA 4.6. *Suppose that for some $j \neq i$ we have*

$$\forall\, u \in H, \qquad \Delta_j h(u) \le \sum_{l \le k} a_l(u).$$

*Then*

$$\forall\, v \in H', \qquad \Delta_j h^+(v), \Delta_j h^-(v) \le \sum_{l \le k} d_l(v),$$

*where*

$$\forall\, l \le k, \qquad \sum_{v \in H'} d_l(v) \le \frac{1}{2} \sum_{u \in H} a_l(u), \tag{4.25}$$

$$\max_{v \in H'} d_l(v) \le \max_{u \in H} a_l(u). \tag{4.26}$$

PROOF. This is a consequence of the formula

$$\left|\tfrac{1}{2}\big(h(v^+) \pm h(v^-)\big) - \tfrac{1}{2}\big(h(w^+) \pm h(w^-)\big)\right|$$

$$\le \tfrac{1}{2}\left|h(w^+) - h(v^+)\right| + \tfrac{1}{2}\left|h(w^-) - h(v^-)\right|. \qquad \square$$

**5. Probability.** First, let us introduce a definition. Consider a finite set $U$. We will say that a sequence $(Y_u)_{u \in U}$ of r.v.'s belongs to $\mathcal{B}(a)$ if it is generated in the following manner. For $m = a \operatorname{card} U$ (which is assumed to be an integer) there exists an i.i.d. sequence $(Z_j)_{j \le m}$ of r.v.'s that are uniformly distributed over $U$ and such that

$$\forall\, u \in U, \qquad Y_u = \operatorname{card}\{j \le m;\, Z_j = u\} - a. \tag{5.1}$$

An obvious, but important observation is that, if we denote by $\mathcal{V}$ a partition of $U$ in sets of equal cardinality $k$ and if for $V \in \mathcal{V}$ we set $Y_V = \sum_{u \in V} Y_u$, then the sequence $(Y_V)_{V \in \mathcal{V}}$ belongs to $\mathcal{B}(ka)$.

In particular, if $(Z_i)_{i \le m}$ is i.i.d. uniform over $G_n = \{0, \dots, 2^n - 1\}^3$, the sequence $(Y_u^0)_{u \in G_n}$ belongs to $\mathcal{B}(c)$, where $c = m2^{-3n}$ and where

$$Y_n^0 = \operatorname{card}\{i \le m;\, Z_i = u\} - m2^{-3n}.$$

To prove Theorem 2.4, it suffices to prove the following statement.

THEOREM 5.1. *There exists a universal constant $K_0$ with the following property. If $m \geq K_0 2^{3n}$, then with probability greater than or equal to $1 - K_0 2^{-6n}$, for each $\Delta \geq \frac{1}{3}$ and each $f \in \mathcal{S}(\Delta)$, we have*

$$\left| \sum_{u \in G_n} f(u) Y_u^0 \right| \leq m\Delta.$$

Indeed, consider $\overline{a} = (a(u))_{u \in G_n}$ and the function $f(u) = h_{\overline{a}}(u) - Eh_{\overline{a}}$. Since $\sum_{u \in G_n} f(u) = 0$, we have

$$\sum_{i \leq m} f(Z_i) = \sum_{u \in G_n} f(u) Y_u^0. \tag{5.2}$$

Set $\Delta' = E\overline{a} - Eh_{\overline{a}}$. By definition of $\mathcal{S}(\Delta')$, we have $f \in \mathcal{S}(\Delta')$. Set $\Delta = \max(\Delta', \frac{1}{3})$, so that $\Delta \geq \frac{1}{3}$ and $f \in \mathcal{S}(\Delta)$. By Theorem 5.1 and (5.2), we have

$$\frac{1}{m} \left| \sum_{i \leq m} f(Z_i) \right| \leq \Delta,$$

so that

$$\frac{1}{m} \sum_{i \leq m} f(Z_i) \leq \max\left(\frac{1}{3}, E\overline{a} - Eh_{\overline{a}}\right).$$

To prove Theorem 5.1, we will use the decomposition

$$f = f'' + f' = f'' + \sum_{0 \leq k \leq k_0} f_k,$$

which is introduced at the beginning of Section 4. We will write

$$\left| \sum_{u \in G_n} f'(u) Y_u^0 \right| \leq \sum_{0 \leq k \leq k_0} \left| \sum_{u \in G_n} f_k(u) Y_u^k \right|.$$

In this formula the function $f_k$ is viewed as a function on $G_n$. We prefer to think of this function as a function on $H_k$, so we write

$$\left| \sum_{u \in G_n} f_k(u) Y_u^0 \right| = \left| \sum_{v \in H_n} f_k(v) Y_v^0 \right|,$$

where $(Y_v^k)_{v \in H_k} \in \mathcal{B}(2^k c)$. What we are going to show is the following fact. With probability greater than or equal to $1 - K2^{-6n}$, for each $\Delta \geq 0$ and each $f \in \mathcal{S}(\Delta)$, we have

$$\sum_{0 \leq k \leq k_0} \left| \sum_{v \in H_k} f_k(v) Y_v^k \right| \leq K 2^{3n} (\Delta + 1) \sqrt{c}. \tag{5.3}$$

The interesting fact that makes the theory nontrivial is that in this series the term that contributes the most depends on $f$, and that it would not work to take the supremum over $f \in \mathcal{S}(\Delta)$ before making the summation.

The proof of (5.3) will take most of this section. Once it is completed, we will use (4.2) to prove a similar bound for the term $|\sum_{u \in G_n} f''(u) Y_u^0|$. This is simpler, yet nontrivial. (The reader is certainly advised not to follow the logical order of the proof and to start developing understanding of the situation in that simpler setting.) Once these two bounds are obtained, the proof of Theorem 5.1 will follow. Indeed, since $m = c2^{-3n}$, we can choose $K_0$ such that

$$\Delta \geq \tfrac{1}{3}, c \geq K_0 \quad \Rightarrow \quad K2^{3n}(\Delta + 1)\sqrt{c} \leq m\Delta.$$

We now start the preparation of the proof of (5.3). The first task is to investigate some elementary properties of a sequence $(|Y_i|)_{i \in U}$ in $\mathcal{B}(a)$.

Given numbers $(Y_u)_{u \in U}$, where $\operatorname{card} U = N$, we denote by $(Y_i^*)_{i \leq N}$ the nondecreasing rearrangement of the sequence $(Y_u)_{u \leq N}$, that is,

$$Y_i^* = \sup\left\{t \geq 0; \operatorname{card}\{u \in U; |Y_u| \geq t\} \geq i\right\}.$$

We observe that

$$\begin{aligned} \sum_{i \leq r} Y_i^* &= \sup\left\{\sum_{u \in I} |Y_u|; \operatorname{card} I = r\right\} \\ &\leq 2 \sup\left\{\left|\sum_{u \in J} Y_u\right|; \operatorname{card} J \leq r\right\}. \end{aligned} \tag{5.4}$$

LEMMA 5.2. *Consider a sequence $(Y_i)_{i \leq N} \in \mathcal{B}(a)$. Consider $r \in \mathbb{N}, 1 \leq r \leq N$ and $s \in \mathbb{R}, s \geq 1$. Then, if $\log(eN/r) \leq a$, we have*

$$P\left(\sum_{i \leq r} Y_i^* \geq 4sr\sqrt{a \log \frac{eN}{r}}\right) \leq 2\left(\frac{r}{eN}\right)^{sr} \leq \left(\frac{1}{eN}\right)^{s}.$$

PROOF. Consider a subset $J$ of $\{1, \dots, N\}$, with $\operatorname{card} J \leq r$. Then, if the sequence $(Z_j)$ is as in (5.1), we have

$$\sum_{i \in J} Y_i = \operatorname{card}\{j \leq N; Z_j \in J\} - m\frac{\operatorname{card} J}{N}. \tag{5.5}$$

Rather than using the classical bounds on the tail of the binomial law, we will provide a simple derivation of the weaker result we need. If a r.v. $W$ satisfies $P(W = 1 - b) = b, P(W = -b) = 1 - b$ for some $0 \leq b \leq 1$, we have , for $|\lambda| \leq 1$,

$$E \exp \lambda W = b \exp\left[\lambda(1 - b)\right] + (1 - b)\exp(-\lambda b) \leq 1 + \lambda^2 b \leq \exp(\lambda^2 b),$$

since $|e^x - 1 - x| \le x^2$ for $|x| \le 1$. Thus, for $b = \operatorname{card} J/N$, by independence of the sequence $(Z_j)$ and by (5.1), we have

$$E \exp \lambda \sum_{i \in J} Y_i \le \exp(m\lambda^2 b) = \exp(\lambda^2 a \operatorname{card} J) \le \exp(\lambda^2 ar),$$

so that

$$P\left(\left|\sum_{i \in J} Y_i\right| \ge st\right) \le 2\exp(-\lambda ts + \lambda^2 ar).$$

If $t \le 2ra$, we take $\lambda = t/2ra$, to get, if $s \ge 1$,

$$P\left(\left|\sum_{i \in J} Y_i\right| \ge st\right) \le 2\exp\left(-\frac{t^2 s}{4ra}\right).$$

If $\log(eN/r) \le a$, we can then take $t = 2r\sqrt{a \log(eN/r)}$, to get

$$P\left(\left|\sum_{i \in J} Y_i\right| \ge 2sr\sqrt{a \log \frac{eN}{r}}\right) \le 2\left(\frac{r}{eN}\right)^{rs}.$$

Thus,

$$P\left(\sum_{i \le r} Y_i^* \ge 2sr\sqrt{a \log \frac{eN}{r}}\right) \le 2\sum_{j \le r} \binom{N}{j}\left(\frac{r}{eN}\right)^{rs} \le 2\left(\frac{r}{eN}\right)^{r(s-1)}.$$

Replacing $s$ by $s + 1$ proves the result. □

LEMMA 5.3. *Consider a sequence $(Y_i)_{i \le N} \in \mathcal{B}(a)$. Consider an integer $\tau$ such that $2^\tau \le a$. We denote by $l_0$ the largest integer less than or equal to $\tau$ such that $2^{2^{l_0}} \le N$. For $1 \le l \le l_0$, we set $r_l = [2^{-2^l} N]$, and we set $r_0 = N$. Then, for all $s \ge 1$, the following occur with probability greater than or equal to $1 - (eN)^{-s+1}$:*

(a) *If $l_0 = \tau$, (i.e., $2^{2^\tau} < N$), then for each family of numbers $(h(i))_{i \le N}$ we have*

$$\sum_{i \le N} |h(i)|\,|Y_i| \le Ks\sqrt{a}\left(\sum_{l \le l_0} 2^{l/2} \sum_{i \le r_l} h^*(i) + r_\tau 2^{\tau/2} \max_{i \le N} |h(i)|\right).$$

(b) *If $l_0 < \tau$, (i.e., $2^{2^\tau} > N$), then for each family $(h(i))_{i \le N}$ of numbers we have*

$$\sum_{i \le N} |h(i)|\,|Y_i| \le Ks\sqrt{a}\left(\sum_{l \le l_0} 2^{l/2} \sum_{i \le r_l} h^*(i)\right).$$

PROOF. We define

$$I(0) = \{i \le N; |Y_i| \le Y_{r_1}^*\}.$$

For $1 \le l < l_0$, we define

$$I(l) = \{i \le N; Y^*_{r_{l+1}} \ge |Y_i| > Y^*_{r_l}\},$$

and finally we define

$$I(l_0) = \{i \le N; |Y_i| > Y^*_{r_{l_0}}\}.$$

Since card $I_l \le r_l$ we have $\sum_{i \in I(l)} |h(i)| \le \sum_{i \le r_l} h^*(i)$.
Thus we have

$$\sum_{i \notin I(l_0)} |h(i)|\,|Y_i| = \sum_{0 \le l < l_0} \sum_{i \in I(l)} |h(i)|\,|Y_i| \le \sum_{0 \le l < l_0} Y^*_{r_{l+1}} \sum_{i \le r_l} h^*(i). \tag{5.6}$$

We also have

$$\sum_{i \in I(l_0)} |h(i)|\,|Y_i| \le Y^*_1 \sum_{i \le r_{l_0}} h^*(i) \tag{5.7}$$

and

$$\sum_{i \in I(l_0)} |h(i)|\,|Y_i| \le \max_{i \in N} |h(i)| \sum_{i \le r_{l_0}} Y^*_i. \tag{5.8}$$

It follows from Lemma 5.2 that, for $l \le l_0$, we have

$$\sum_{i \le r_l} Y^*_i \le K s r_l \sqrt{a} 2^{l/2}, \tag{5.9}$$

with probability greater than or equal to $1 - 2(1/eN)^s$.

We observe that (5.9) implies $Y^*_{r_l} \le K s \sqrt{a} 2^{l/2}$. Then (a) follows from (5.6), (5.8) and (5.9), since $2l_0(1/(eN))^s \le (1/(eN))^{s-1}$.

To prove (b), we note that by Lemma 5.2 again, with $r = 1$, we have

$$P\left(Y^*_1 \ge 4s\sqrt{a \log eN}\right) \le 2\left(\frac{1}{eN}\right)^s.$$

Since $l_0 < \tau$, the definition of $l_0$ shows that $N \le 2^{2^{l_0+1}}$ so that $\log eN \le K2^{l_0}$. Finally, this occurs with probability greater than or equal to $1-(l_0+1)(1/(eN))^s \ge 1 - 2(1/(eN))^{s-1}$. □

We will apply these results to the study of the function $f_k$. The basic method is to use (4.20) to reduce the dimension of the grid. When one dimension has been reduced enough that it disappears, the grid becomes two-dimensional. After some more reductions, the grid becomes one-dimensional. Thus, the first task is to study a one-dimensional grid.

Consider the class $\mathcal{C}(q,D)$ of functions defined on $H(q) = \{0,\dots,2^q - 1\}$ that satisfy

$$\sum_{0 \le i < 2^q} h(i) = 0, \qquad \sum_{0 \le i < 2^q - 2} \Delta h(i) \le D, \tag{5.10}$$

where $\Delta h(i) = |h(i+1) - h(i)|$.

LEMMA 5.4. *Consider a sequence $(Y_i)_{i \le 2^q} \in \mathcal{B}(a)$, where $a \ge 2q$. Consider $r > 0$. With probability greater than or equal to $1 - e^{-s+2}$, the following occurs. For all $D$, we have*

$$\sup_{h \in \mathcal{C}(q,D)} \left| \sum_{0 \le i < 2^q} h(i) Y_i \right| \le Ks\sqrt{a}2^{q/2} D. \tag{5.11}$$

COMMENT. The order of the quantifiers is essential. The same event must occur for all $D$.

PROOF OF LEMMA 5.4. It is useful to think of $H = H(q)$ as $H(q,0,0)$. If $H' = H(q-1)$, we are thus in a position to use (4.20). As we pointed out, both sequences $(Y_v^+)_{v \in H}$, and $(Y_v^-)_{v \in H'}$ belong to $\mathcal{B}(2a)$. The second important fact is that by Lemma 4.5(b) we have $h^+ \in \mathcal{C}(q-1, D)$. By (4.20), we have

$$\begin{aligned} \sup_{h \in \mathcal{C}(q,D)} \left| \sum_{0 \le i < 2^q} h(i) Y_i \right| \le & \sup_{h \in \mathcal{C}(q-1,D)} \left| \sum_{0 \le i < 2^{q-1}} h(i) Y_i^+ \right| \\ & + \sup_{h \in \mathcal{C}(q-1,D)} \left| \sum_{0 \le i < 2^{q-1}} h^-(i) Y_i^- \right|. \end{aligned} \tag{5.12}$$

We have $\sum_{0 \le i < 2^{q-1}} |h^-(i)| \le D$. We apply Lemma 5.4 (b), taking as $\tau$ the smallest integer for which $q \le 2^\tau$. Thus, with probability greater than or equal to $1 - (1/(e2^{q-1}))^{s-1}$, for each value of $D$, the last term is less than or equal to $Ks\sqrt{2a2^q}$.

We observe that there is nothing to prove if $s \le 2$. If $s > 2$, to prove (5.11), we reiterate (5.12) and we observe that (since $s > 2$)

$$\sum_{l \le q} \left(2^l (q-l)\right)^{1/2} \le K2^{q/2}, \qquad \sum_{1 \le l \le q} \left(\frac{1}{e2^{l-1}}\right)^{s-1} \le e^{-s+2}.$$

In this argument we use the fact that if $q = 0$, then $\mathcal{C}(0, D)$ consists only of the function zero, so that, for $q = 1$, the first term on the right-hand-side of (5.12) is zero. □

Having understood what happens on grids of dimension 1, we can now turn to grids of dimension 2. We consider the grid

$$H = H(q_1, q_2) = H(q_1, q_2, 0) = \{0, \dots, 2^{q_1} - 1\} \times \{0, \dots, 2^{q_2} - 1\}.$$

We consider the class $\mathcal{C}(q_1, q_2, D_1, D_2)$ of functions $h$ on $H$ that satisfy

$$\sum_{v \in H} h(v) = 0 \quad \text{and} \quad \sum_{v \in H} \Delta_j h(v) \le D_j \quad \text{for } j = 1, 2.$$

LEMMA 5.5. *Consider a sequence* $(Y_u)_{u\in H} \in \mathcal{B}(a)$, *where* $a \geq 2(q_1+q_2)$ *and* $s \geq 0$. *Then with probability greater than or equal to* $1-q_1q_2e^{-s+2}$ *we have that, for all values of* $D_1$ *and* $D_2$,

$$\sup_{h\in\mathcal{C}(q_1,q_2,D_1,D_2)}\left|\sum_{u\in H} h(u)Y_u\right| \leq Ks\sqrt{a}(q_1+q_2)^2\max\left(2^{q_1/2-q_2/2}D_1, 2^{q_2/2-q_1/2}D_2\right). \tag{5.13}$$

PROOF. For $l \leq q_1$, we consider the grid $R_l = H(q_1-l, q_2)$, and, for numbers $A_1$ and $A_2$, we consider the class $\mathcal{C}(q_1-l, q_2, A_1, A_2)$ of functions $h$ on $R_l$ that satisfy

$$\sum_{u\in R_l} h(u) = 0, \qquad \sum_{u\in R_l}|h(u)| \leq A_1 \qquad \sum_{u\in R_l}\Delta_2 h(u) \leq A_2.$$

To apply (4.20), we observe (as is shown in particular by Lemmas 4.5 and 4.6) that, for $h \in \mathcal{C}(q_1,q_2,D_1,D_2)$, we have

$$h^+ \in \mathcal{C}\left(q_1-1, q_2, D_1, D_2/2\right), \qquad h^- \in \mathcal{C}_1\left(q_1-1, q_2, D_1/2, D_2/2\right).$$

Thus we have, by (4.20),

$$\begin{aligned}\sup_{h\in\mathcal{C}(q_1,q_2,D_1,D_2)}\left|\sum_{u\in H} h(u)Y_u\right| &\leq \sup_{h\in\mathcal{C}(q_1-1,q_2,D_1,D_2/2)}\left|\sum_{u\in R_1} h(u)Y_u^+\right| \\ &\quad + \sup_{h\in\mathcal{C}(q_1-1,q_2,D_1/2,D_2/2)}\left|\sum_{u\in R_1} h(u)Y_u^-\right|,\end{aligned} \tag{5.14}$$

and we observe that the sequences $(Y_u^+)_{u\in R_1}$ and $(Y_u^-)_{u\in R_1}$ belong to $\mathcal{B}(2a)$. Now, we iterate the use of (5.14). We get

$$\begin{aligned}\sup_{h\in\mathcal{C}(q_1,q_2,D_1,D_2)}\left|\sum_{u\in H} h(u)Y_u\right| &\leq \sum_{1\leq l\leq q_1}\sup_{h\in\mathcal{C}_1(q_1-l,q_2,D_1/2,2^{-l}D_2)}\left|\sum_{u\in R_l} h(u)Y_u^l\right| \\ &\quad + \sup_{h\in\mathcal{C}(0,q_2,D_1,2^{-q_1}D_2)}\left|\sum_{u\in R_{q_1}} h(u)Y_u^*\right|,\end{aligned} \tag{5.15}$$

where $(Y_u^l)_{u\in R_l} \in \mathcal{B}(2^l a)$ and $(Y_u^*)_{u\in R_{q_1}} \in \mathcal{B}(2^{q_1}a)$.

The last term of (5.15) is of the type considered in Lemma 5.4; so, with probability greater than or equal to $1 - e^{-s+2}$, for all values of $D_1$ and $D_2$, it is at most

$$Ks\sqrt{a2^{q_1}}2^{q_2/2}2^{-q_1}D_2 = Ks\sqrt{a}2^{q_2/2-q_1/2}D_2. \tag{5.16}$$

To understand the contribution of the terms in the sum in the right-hand side of (5.15), we need the following lemma. □

LEMMA 5.6. *Suppose $a \geq 2(q_1+q_2)$, and that $(Y_u)_{u \in H} \in \mathcal{B}(a)$. Consider $s > 0$. Then with probability greater than or equal to $1-q_1e^{-s+2}$ we have, for all values of $A_1$ and $A_2$,*

$$\sup_{h \in \mathcal{C}_1(q_1,q_2,A_1,A_2)} \left| \sum_{u \in H} h(u)Y_u \right| \tag{5.17}$$
$$\leq ks\sqrt{a}(q_1+q_2)\max\left(2^{q_1/2-q_2/2}A_1, 2^{q_2/2-q_1/2}A_2\right).$$

Before we prove Lemma 5.6, we finish the proof of Lemma 5.5. We apply (5.17) with $q_1 - l$ rather than $q_1$, $2^l a$ rather than $a$, $A_1 = D_1$ and $A_2 = 2^{-l}D_2$. Then the right-hand side of (5.17) is at most

$$Ks\sqrt{a}2^{l/2}(q_1+q_2)\max\left(2^{q_1/2-q_2/2-l/2}D_1, 2^{q_2/2-q_1/2+l/2}2^{-l}D_2\right).$$

This is independent of $l$. Thus, with probability greater than or equal to $1 - q_1q_2e^{-s+2}$ the sum in (5.15) is at most

$$Kq_1(q_1+q_2)s\sqrt{a}\max\left(2^{q_1/2-q_2/2}D_1, 2^{q_2/2-q_1/2}D_2\right),$$

and this finishes the proof of Lemma 5.5. □

PROOF OF LEMMA 5.6. We use again the same reduction procedure. For $l \leq n_2$, we consider the grid $S_l = H(q_1, q_2 - l)$. For numbers $B_1$ and $B_2$, we consider the class $\mathcal{C}_{1,2}(q_1, q_2 - l, B_1, B_2)$ of functions $h$ on $S_l$ that satisfy

$$\sum_{u \in S_l} h(u) = 0, \qquad \sum_{u \in S_l} |h(u)| \leq \min(B_1, B_2).$$

Using (4.20) together with Lemmas 4.5 and 4.6, we see that

$$\sup_{h \in \mathcal{C}_1(q_1,q_2,A_1,A_2)} \left| \sum_{u \in H} h(u)Y_u \right| \leq \sup_{h \in \mathcal{C}_1(q_1,q_2-1,A_1/2,A_2)} \left| \sum_{u \in S_1} h(u)Y_u^+ \right|$$
$$+ \sup_{h \in \mathcal{C}_{1,2}(q_1,q_2-1,A_1/2,A_2/2)} \left| \sum_{u \in S_1} h(u)Y_u^- \right|.$$

Iteration of this relation yields

$$\sup_{h \in \mathcal{C}_1(q_1,q_2,A_1,A_2)} \left| \sum_{u \in H} h(u) Y_u \right| \le \sum_{l \le q_2} \sup_{h \in \mathcal{C}_{1,2}(q_1,q_2-l,2^{-l}A_1,A_2)} \left| \sum_{u \in S_l} h(u) Y_u^l \right| \tag{5.18}$$
$$+ \sup_{h \in \mathcal{C}_1(q_1,0,2^{-q_2}A_1,A_2)} \left| \sum_{u \in S_{q_2}} h(u) Y_u^* \right|,$$

where the sequences $(Y_u^l)_{u \in S_l}$ belong to $\mathcal{B}(2^l a)$, and $(Y_u^*)_{u \in S_{q_2}}$ to $\mathcal{B}(2^{q_2}a)$. As in the proof of Lemma 5.5, we see that with probability greater than or equal to $1 - e^{-s+2}$, for each values of $A_1$ and $A_2$, the last term is at most

$$Ks\sqrt{a2^{q_2}}2^{q_1/2}\left(2^{-q_2}A_1\right) = Ks\sqrt{a}2^{q_1/2-q_2/2}A_1.$$

To evaluate the other terms, we note that if $h \in \mathcal{C}_{1,2}(q_1, q_2 - l, 2^{-l}A_1, A_2)$,

$$\left| \sum_{u \in S_l} h(u) Y_u^l \right| \le \max_{u \in S_l} |Y_u^l| \sum_{u \in S_l} |h(u)| \tag{5.19}$$
$$\le \max_{u \in S_l} |Y_u^l| \min\left(2^{-l}A_1, A_2\right).$$

Using Lemma 5.2 for $r = 1$, we see that

$$\max_{u \in S_l} |Y_u^l| \le Ks\sqrt{a2^l}(q_1 + q_2), \tag{5.20}$$

with probability greater than or equal to $1 - 2(e2^{q_1+q_2-l})^{-s} \ge 1 - 2e^{-s}$.

It is easy to see that

$$\sum_l 2^{l/2} \min\left(2^{-l}A_1, A_2\right) = \sum_l \min\left(2^{-l/2}A_1, 2^{l/2}A_2\right)$$
$$\le K\sqrt{A_1A_2}$$
$$\le K\max\left(2^{q_1/2-q_2/2}A_1, 2^{q_2/2-q_1/2}A_2\right).$$

This completes the proof. □

When applying this reduction method to grids of dimension 1 and 2, the reader must have observed an interesting difference. In dimension 1 the last term dominates, while in dimension 2 all the terms are of equal weight. This is what make the dimension 2 case important and difficult. In particular, the kind of methods used in the previous two lemmas cannot yield the Ajtai–Komlos–Tusnády theorem [one obtains a factor $\log n$ rather than the correct $(\log n)^{1/2}$]. In dimension 3 (or more) the situation again is different, since then the *first* term dominates (which makes the approach possible).

The factor $(q_1 + q_2)^2$ in Lemma 5.5 is not optimal (neither is the bound on the probability of failure). The reader could wonder how we can use a crude result to obtain an optimal one. The answer is that Lemma 5.5 will be used as an easy way to show that certain terms are of smaller order. A typical case is Proposition 5.7. When controlling the main terms, however, we will not be permitted to have any extraneous factor such as this $(q_1+q_2)^2$. The origin of that term is twofold. First, one factor $q_1+q_2$ comes from the phenomenon, mentioned above, that in dimension 2 the terms of the reduction in the proof of Lemma 5.5 are of the same order. This factor will disappear by itself in dimension 3. The other factor is rooted in the use of (5.19) and (5.20). In order to remove this factor, we will have to use the information provided by Lemmas 4.2 and 4.3 rather than cruder information on $\sum_{u \in H_k} \Delta_j f(u)$.

First, let us dispose of $f_{k_0}$.

PROPOSITION 5.7. *Consider a sequence $(Y_u)_{u \in H_{k_0}}$ that belongs to $\mathcal{B}(c2^{k_0})$. Then with probability greater than or equal to $1 - Kn^2 2^{-10n}$ we have*

$$\forall \Delta > 0, \qquad \sup_{f \in S(\Delta)} \left| \sum_{u \in H_{k_0}} f_{k_0}(u) Y_u \right| \le K 2^{3n} \Delta \sqrt{c} n^3 2^{-n/2}.$$

COMMENT. The factor $2^{-n/2}$ is what makes the factor $n^3$ unimportant.

PROOF OF PROPOSITION 5.7. There is no loss of generality to assume that $n_3(k_0) = n$. We use Lemma 4.4 to see that we can apply Lemma 5.5 with $a2^{k_0}$ rather than $a, q_j = n - n_j(k_0)$ and $D_j = K 2^{3n} 2^{-k_0 + n_j(k_0)} \Delta$ for $j = 1, 2$. We obtain the result by taking $s = 10n$ and by plugging these values into (5.13), and observing that

$$-\frac{k_0}{2} + \frac{n_1(k_0) + n_2(k_0)}{2} = -\frac{n_3(k_0)}{2} = -\frac{n}{2}. \qquad \square$$

We now turn to the center of the proof of (5.3), the study of the contribution of $f_k$, $k < k_0$. To simplify notation, we think of $k$ as fixed.

Consider $q_1, q_2, q_3$ and, for $j \le 3$ and $l \le k$, consider numbers $(c_{j,l})$ and $(d_{j,l})$. Consider a subset $J$ of $\{1, 2, 3\}$. On the grid $H = H(q_1, q_2, q_3)$ we consider the class $\mathcal{C}_J(q_1, q_2, q_3, (c_{j,l}), (d_{j,l}))$ of functions $h$ that satisfy $\sum_{u \in H} h(u) = 0$ and for which there exist numbers $a_{j,l}(u)$ for $j \le 3, l \le k$ and $u \in H$ such that

$$j \in J \quad \Rightarrow \quad \forall\, u \in H, \qquad |h(u)| \le \sum_{l \le k} a_{j,l}(u), \tag{5.21}$$

$$j \notin J \quad \Rightarrow \quad \forall\, u \in H, \qquad \Delta_j h(u) \le \sum_{l \le k} a_{j,l}(u), \tag{5.22}$$

and such that the numbers $a_{j,l}(u)$ satisfy

$$\forall\, l \le k, \qquad \sum_{u \in H} a_{j,l}(u) \le c_{j,l}, \tag{5.23}$$

$$\forall\, l \le k, \qquad \max_{u \in H} a_{j,l}(u) \le d_{j,l}. \tag{5.24}$$

To understand these conditions, we observe that the content of Lemmas 4.2 and 4.3 is as follows. Consider the unique $i \le 3$ such that $n_i(k+1) = n_i(k) + 1$, and set $J = \{i\}$. Set

$$d_{j,l} = 2^{k-l+1} b_{l+1}. \tag{5.25}$$

Set

$$c_{j,l} = 2^{-k+n_j(k)-n_j(l)} \beta_l(f), \tag{5.26}$$

where $\beta_l(f) = 16 \sum_{r \ge l} 4^{-|r-l|} \alpha_r(f)$. [The purpose of the introduction of $\beta_l(f)$ is to merge conditions (4.11) and (4.12), in order to simplify notation.] Then

$$f_k \in \mathcal{C}_J\big(n - n_1(k), n - n_2(k), n - n_3(k), (c_{j,l}), (d_{j,l})\big). \tag{5.27}$$

We now turn to the reduction procedure. Consider $i \notin J$ and such that $q_i > 0$. Set $H' = H(q'_1, q'_2, q'_3)$, where $q'_j = q_j$ if $j \ne i$, $q'_i = q_i - 1$. We write (4.20) as

$$\sup_{h \in \mathcal{C}} \left| \sum_{u \in H} h(u) Y_u \right| \le \sup_{h \in \mathcal{C}^+} \left| \sum_{u \in H'} h(u) Y_u^+ \right| + \sup_{h \in \mathcal{C}^-} \left| \sum_{u \in H'} h(u) Y_u^- \right|. \tag{5.28}$$

It follows from Lemmas 4.5 and 4.6 that if

$$\mathcal{C} = \mathcal{C}_J\big(q_1, q_2, q_3, (c_{j,l}), (d_{j,l})\big),$$

we have

$$\mathcal{C}^+ \subset \mathcal{C}_J\big(q'_1, q'_2, q'_3, (c'_{j,l}), (d'_{j,l})\big),$$

where

$$\forall\, l \le k, \qquad c'_{i,l} = c_{i,l}, \tag{5.29}$$

$$\forall\, l \le k,\ \forall\, j \ne i, \qquad c'_{j,l} = \tfrac{1}{2} c_{j,l}, \tag{5.30}$$

$$\forall\, l \le k, \qquad d'_{i,l} = 2 d_{i,l}, \tag{5.31}$$

$$\forall\, l \le k,\ \forall\, j \ne i, \qquad d'_{j,l} = d_{j,l}. \tag{5.32}$$

Also, we have

$$\mathcal{C}^- \subset \mathcal{C}_{J \cup \{i\}}\big(q'_1, q'_2, q'_3, (c'_{j,l}), (d'_{j,l})\big)$$

for the same sequences $c'_{j,l}$ and $d'_{j,l}$ as above. [Observe that by doing this we give up a factor 2 in (5.29) and (5.31); the loss of this factor is actually unimportant, and doing this simplifies notations.]

We note that, as usual, if $(Y_u)_{u \in H} \in \mathcal{B}(a)$, then $(Y_u^+)_{u \in H'}, (Y_u^-)_{u \in H'} \in \mathcal{B}(2a)$.

Having applied (5.28) once, we iterate its application. We follow the rule that, when reducing a term such as the left-hand side of (5.28), we always reduce the value of $q_j$ for the smallest $j$ such that $q_j > 1$, $j \notin J$. The purpose of this rule is that it insures at most $3n$ different terms can be created in the decomposition. We do not decompose a term if it is of either of the following types:

Type I. One of the $q_j$ is zero.
Type II. $J = \{1, 2, 3\}$.

Consider the left-hand side of (5.28), in the case where the class $\mathcal{C}$ is the class of (5.27). The decomposition procedure previously described shows that this quantity is bounded by the sum of terms of type I or II, and we evaluate the contribution of each term. We start with the easy part, that is, the terms of type I, for which we will obtain a bound similar to that of Proposition 5.8. Let us examine one term of type I. It concerns a class $\mathcal{C}_0$ of functions on the grid $H_0 = H(q_1, q_2, q_3)$, where one of the numbers $q_j$ is zero. For specificity, we assume that $q_3 = 0$. For $j \le 3$, set $p_j = n - n_j(k) - q_j$; this is the number of times the coordinate of rank $j$ has been reduced. Set $p = p_1 + p_2 + p_3$. If, from the information we have on $\mathcal{C}_0$ [from (5.27) and iteration of (5.29)–(5.32)], we keep only the information on $\sum_{u \in H_0} |h(u)|$ and $\sum_{u \in H_0} \Delta_j h(u)$, and if in the bounds for these quantities we replace $n_j(l)$ by 0 and $\beta_l(f)$ by $K\Delta 2^{3n}$ (as was done in the proof of Lemma 4.4), we see that, for $j = 1, 2$,

$$(5.33)\qquad \begin{aligned} j \in J \quad &\Rightarrow \quad \sum_{u \in H_0} |h(u)| \le K 2^{-k + n_j(k) + p_j - p} \Delta 2^{3n}, \\ j \notin J \quad &\Rightarrow \quad \sum_{u \in H_0} |\Delta h_j(u)| \le K 2^{-k + n_j(k) + p_j - p} \Delta 2^{3n}. \end{aligned}$$

Certainly $J$ must contain at least one element. If $J$ contains exactly one element, we appeal to Lemma 5.6 with $s = 10n$ to see that if the sequence $(Y'_v)_{v \in H_0}$ belongs to $\mathcal{B}(a)$, where $a \ge 2(q_1 + q_2)$, then, with probability greater than or equal to $1 - Kn2^{-10n}$,

$$(5.34)\qquad \sup_{h \in \mathcal{C}_0} \left| \sum_{v \in H_0} h(v) Y_v \right| \le K\sqrt{a} n^2 \max(U_1, U_2),$$

where

$$\begin{aligned} U_1 &= \Delta 2^{3n} 2^{-k + n_1(k) + p_1 - p} 2^{[n - n_1(k) - p_1]/2 - [n - n_2(k) - p_2]/2} \\ &= \Delta 2^{3n} 2^{-k + [n_1(k) + n_2(k)]/2 + (p_1 + p_2)/2 - p}. \end{aligned}$$

In a similar fashion, $U_2$ is bounded by the same quantity. Since $a = c2^{k+p}$, the left-hand side of (5.34) is bounded by

$$K\sqrt{c} n^2 2^{-k/2 + [n_1(k) + n_2(k)]/2 - p_3/2} \Delta 2^{3n}.$$

Since $q_3 = 0$, we have $p_3 = n - n_3(k)$, so this quantity is $K\sqrt{c} n^2 2^{-n/2} \Delta 2^{3n}$.

If $J = \{1, 2\}$ one can observe that a condition on $\sum |h(u)|$ such as (5.23) is actually stronger than the corresponding condition on $\Delta_j h(u)$. Then the above proof works. Alternatively, one can observe that what we need has been established in the proof of Lemma 5.6.

In summary, with probability greater than or equal to $1 - Kn^3 2^{-10n}$, the contribution of all the terms of type I will not exceed $K\sqrt{c}n^3 2^{-n/2}\Delta 2^{3n}$.

We now turn to the elements of type II. Such an element concerns a class $\mathcal{C}_0$ of functions on a grid $H(q_1, q_2, q_3)$. Consider $p_j = n - n_j(k) - q_j$, the number of times that the component of rank $j$ has been reduced. Set $p = p_1 + p_2 + p_3$. We will say that the term is of type II($p$). (The choice of $s$ in particular will depend upon $p$.) There are at most $3p$ terms of type II($p$).

From (5.27) and (5.29)–(5.32), it should be clear that a function $h$ in $\mathcal{C}_0$ satisfies

$$\forall j \le 3,\ \forall u \in H_0,\qquad |h(u)| \le \sum_{l \le k} a_{j,l}(u), \tag{5.35}$$

where

$$\forall j \le 3,\ \forall l \le k,\qquad \sum_{u \in H_0} a_{j,l}(u) \le 2^{-p+p_j} c_{j,l} \le 2^{-p+p_j-k+n_j(k)-n_j(l)}\beta_l(f) \tag{5.36}$$

$$\forall j \le 3,\ \forall l \le k,\qquad \max_{u \in H_0} a_{j,l}(u) \le 2^{p_j} 2^{n_j(k)-n_j(l)} b_{l+1} \le 2^{p+n_j(k)-n_j(l)} b_{l+1}. \tag{5.37}$$

Consider a sequence $(Y'_v)_{v \in H_0}$ that belongs to $\mathcal{B}(a)$, where $a = 2^{k+p}c$. To bound $|\sum_{u \in H_0} h(u) Y_u|$, we use Lemma 5.3. We have

$$N = \operatorname{card} H_0 = 2^{3n-p-k}.$$

The first task is to bound $\sum_{i \le r_\gamma} h^*(i)$ for any integer $\gamma$, where $r_\gamma = [N2^{-2^\gamma}]$.

Consider a subset $U$ of $H_0$ with $\operatorname{card} U \le r_\gamma$. Using (5.35), we see that, for any $j \le 3$,

$$\sum_{u \in U} |h(u)| \le \sum_{0 \le l \le k} \sum_{u \in H_0} a_{j,l}(u). \tag{5.38}$$

Also, if $\rho \ge 0$ is an integer (to be determined later), we can write, for any $j \le 3$,

$$\sum_{u \in U} |h(u)| \le \sum_{l \ge \rho} \sum_{u \in H_0} a_{j,l}(u) + \sum_{u \in U} \sum_{l < \rho} a_{j,l}(u). \tag{5.39}$$

Using (5.36), we have

$$\sum_{l \ge \rho} \sum_{u \in H_0} a_{j,l}(u) \le \sum_{\rho \le l \le k} 2^{-p+p_j-k+n_j(k)-n_j(l)}\beta_l(f) \le 2^{-p+p_j-k+n_j(k)-n_j(\rho)} \sum_{\rho \le l \le k} \beta_l(f). \tag{5.40}$$

By (5.37), we have

$$\forall u \in H_0,\qquad \sum_{l < \rho} a_{j,l}(u) \le \sum_{l < \rho} 2^{p+n_j(k)-n_j(l)} b_{l+1} \le 2^{p+1} b_\rho, \tag{5.41}$$

where we have used (4.8) in the last inequality.

Since

$$\operatorname{card} U \le r_\gamma \le N2^{-2^\gamma} = 2^{3n-p-k}2^{-2^\gamma},$$

we see that the last term of (5.39) is at most

$$\begin{aligned} 2^{3n-k+1}2^{-2^\gamma} b_\rho &\le 2^{3n-k+1}2^{-2^\gamma}\exp(2^{\rho+39}) \\ &\le 2^{3n-k+1}2^{2^{\rho+40}-2^\gamma}, \end{aligned}$$

by (3.2). In particular, if $\rho + 41 \le \gamma$, this is at most

$$2^{3n-k+1-2^{\gamma-1}}. \tag{5.42}$$

We now choose $\rho$. We set $\theta = [41 + p/6]$. If $\gamma \le \theta$, we take $\rho = 0$, and by (5.38) and (5.40) we have

$$\sum_{u\in U} |h(u)| \le 2^{-p+p_j-k+n_j(k)} \sum_{0\le l\le k} \beta_l(f). \tag{5.43}$$

If $\gamma > \theta$, we take $\rho = \gamma - \theta$. In that case, by (5.39), (5.40) and (5.42) we have

$$\sum_{u\in U} |h(u)| \le 2^{-p+p_j-k+n_j(k)-n_j(\gamma-\theta)} \sum_{\gamma-\theta\le l\le k} \beta_l(f) + 2^{3n-k+1-2^{\gamma-1}}. \tag{5.44}$$

We now choose $j$. We observe that for $k' \le k$ we have

$$\sum_{j\le 3} p_j + n_j(k) - n_j(k') = p + k - k',$$

so that there exists $j \le 3$ such that

$$p_j + n_j(k) - n_j(k') \le \tfrac{1}{3}(p + k - k').$$

Thus (taking $k' = 0$) from (5.43) it follows that, if $\gamma \le \theta$, we have

$$\sum_{u\in U} |h(u)| \le 2^{-2p/3-2k/3} \sum_{0\le l\le k} \beta_l(f), \tag{5.45}$$

while, if $\gamma > \theta$, from (5.44) it follows (taking $k' = \gamma - \theta$) that

$$\sum_{u\in U} |h(u)| \le 2^{-2p/3-2k/3-\gamma/3+\theta/3} \sum_{\gamma-\theta\le l\le k} \beta_l(f) + 2^{3n-k+1-2^{\gamma-1}}. \tag{5.46}$$

Since these inequalities hold whenever $\operatorname{card} U \le r_\gamma$, (5.45) and (5.46) provide bounds for $\sum_{i\le r_\gamma} h^*(i)$.

Consider now the largest integer $\tau$ such that $2^\tau \le c2^{k+p}$. There is no loss of generality to assume that $c \ge 2^{42}$, so that $\tau \ge k + p + 42$. By (5.35), (5.41) (used for $\rho = k + 1$) and (3.2) we have

$$\max_{u\in H_0} |h(u)| \le 2^{p+1} b_{k+1} \le 2^{p+1}\exp(2^{k+40}).$$

Thus, when $2^{2^\tau} \le N$ we have, since $\tau \ge k + p + 42 \ge k + 42$,

$$\begin{aligned} r_\tau 2^{\tau/2} \max_{u \in H_0} |h(u)| &\le N 2^{-2^\tau + \tau/2 + p + 1 + 2^{k+41}} \\ &\le N 2^{p+1+\tau/2-2^{\tau-1}} \\ &\le 2^{3n - k/2 + p/2 - 2^{k+p-1}}, \end{aligned}$$

since $\tau/2 - 2^{\tau-1} \le (k+p)/2 - 2^{k+p-1}$.

We now appeal to Lemma 5.3. With probability greater than or equal to $1 - (eN)^{-s+1}$, whether we are in case (a) or (b) of this lemma, we have

$$\sum_{u \in H_0} |h(u)||Y_u| \le Ks\sqrt{c}2^{(p+k)/2}\big(A(p,k) + B(p,k)\big),$$

where

$$\begin{aligned} A(p,k) &= \sum_{\gamma < l_0} 2^{\gamma/2 - 2p/3 - 2k/3 - \gamma/3 + \theta/3} \sum_{\gamma - \theta \le l \le k} \beta_l(f), \\ B(p,k) &= \sum_{\gamma \ge \theta} 2^{3n - k + 2 + \gamma/2 - 2^{\gamma-1}} + 2^{3n - k/2 + p/2 - 2^{k+p-1}} \end{aligned}$$

and where $l_0$ is the largest integer for which $2^{2^{l_0}} \le N$. [We have used the fact that the first term of the right-hand side of (5.46) dominates the right-hand side of (5.45).]

If we exchange the summations in the definition of $A(p,k)$, we see that if we set

$$l_1 = \min(l + \theta, l_0) \le \theta + \min(l, l_0),$$

we have

$$\begin{aligned} A(p,k) &\le 2^{-2p/3 - 2k/3 + \theta/3} \sum_{l \le k} \beta_l(f) \left( \sum_{\gamma \le l_1} 2^{\gamma/6} \right) \\ &\le K 2^{2p/3 - 2k/3 + \theta/3} \sum_{l \le k} 2^{l_1/6} \beta_l(f). \end{aligned}$$

Since, by definition of $l_0$, we have $2^{l_0} \le 3n$, we have $l_0 \le \log_2 3n$. Thus, recalling the value of $\theta$, we have, replacing $l_1$ by its bound above, that

$$2^{(p+k)/2} A(p,k) \le K 2^{-p/12} \sum_{l \le k} 2^{(1/6)\min(l, \log_2 3n) - k/6} \beta_l(f). \tag{5.47}$$

We now define $s = s(k,p)$ as follows. If both $k, p \le n$, we take $s(k,p) = 10$. We then have

$$(eN)^{-s+1} \le \left(2^{3n-k-p}\right)^{-s+1} \le (2^n)^{-s+1} \le 2^{-9n}.$$

If either $k > n$ or $p > n$, we take $s = 10n$. Then

$$(eN)^{-s+1} \le e^{-9n} \le 2^{-9n}.$$

Since there are at most $3p$ grids of type II($p$), we see that with probability greater than or equal to $1 - n^3 2^{-9n}$, for each $k \le n$ and each $p$, the contribution all the terms of type II($p$) is at most

$$K\sqrt{c}ps(p,k)2^{(p+k)/2}\big(A(p,k) + B(p,k)\big). \tag{5.48}$$

We now observe the following:

$$\sum_{k \ge l} 2^{(1/6)\min(l, \log_2 3n) - k/6} \le \sum_{k \ge l} 2^{(l-k)/6} \le K;$$

$$n \sum_{k \ge n} 2^{(1/6)\min(l, \log_2 3n) - k/6} \le n \sum_{k \ge n} 2^{\log_2 3n - k/6} \le Kn2^{\log_2 3n - n/6} \le K;$$

$$\sum_{p \ge 0} p2^{-p/12} < \infty;$$

$$n \sum_{p \ge n} p2^{-p/2} \le n^2 2^{-n/12} \le K.$$

We also observe that, by trivial bounds,

$$B(p,k) \le K\big(2^{3n - k - 2^{p/6}} + 2^{3n - 2^{p+k-2}}\big).$$

It follows from (5.47) and these relations (and interchanging the sums) that the sum of the terms (5.48) for $k, p \ge 0$ is bounded by

$$K\bigg(2^{3n} + \sum_{l \le 0} \beta_l(f)\bigg),$$

but since

$$\beta_l(f) = 4 \sum_{r \ge l} 4^{-|r-l|} \alpha_r(f),$$

we have

$$\sum_{l \ge 0} \beta_l(f) \le K \sum_{l \ge 0} \alpha_l(f) \le K2^{3n}\Delta.$$

We have completed the proof of (5.3). To complete the proof of Theorem 5.1, it remains to prove the following proposition.

PROPOSITION 5.8. *With probability greater than or equal to* $1 - K2^{-6n}$, *the following occurs. For any* $\Delta > 0$, *any partition* $(B_k)_{k \ge 0}$ *of* $G_n$, *such that each* $B_k$ *is union of sets of* $\mathcal{P}_k$, *and any family of numbers* $z(C)$, $C \in \mathcal{P}_k$, $C \subset B_k$, *that satisfies*

$$k \ge 0, C \in \mathcal{P}_n, C \subset B_k \quad \Rightarrow \quad z(C) \le b_{k+1}, \tag{5.49}$$

$$\sum_{k \ge 0} \sum_{C \subset B_k} 2^{s(k)} z(C) \le K2^{3n}\Delta \tag{5.50}$$

*we have*

$$\sum_{k\geq 0}\sum_{\substack{C\subset B_k\\ C\in\mathcal{P}_n}}\left(\sum_{u\in C}|Y_u^0|\right)z(C)\leq K2^{3n}\sqrt{c}(\Delta+1). \tag{5.51}$$

Indeed, once Proposition 5.8 is proved, by (4.2) we have

$$\left|\sum_{u\in G_n}f''(u)Y_u^0\right|\leq K2^{3n}\sqrt{c}(\Delta+1).$$

Thus, combining with (5.3), we see that with probability greater than or equal to $1-K2^{-6n}$, whenever $f\in\mathcal{S}(\Delta)$, we have

$$\left|\sum_{u\in G_n}f(u)Y_u^0\right|\leq K2^{3n}\sqrt{c}(\Delta+1),$$

from which, as already mentioned, Theorem 5.1 follows.

We start the proof of Proposition 5.8. Set $\varrho=41$. Consider the largest integer $\tau$ such that $2^\tau\exp(2^{\tau+\rho})\leq 2^{3n}$. For $k\leq\tau$, set $m_k=[2^{3n-s(k)}\exp(-2^{k+\rho})]$. We observe that $\tau$ is of order $\log n$, so that, for $n$ large enough, $\tau\leq k_0$. For $k>\tau$, set $m_k=1$. The key to Proposition 5.8 is the following statement.

LEMMA 5.9. *With probability greater than or equal to* $1-K2^{-6n}$, *the following occurs. Consider* $k\geq 0$, *and consider any set* $B\subset G_n$ *that is the union of at most* $m_k$ *atoms of* $\mathcal{P}_k$. *Then*

$$\sum_{u\in B}|Y_u^0|\leq Km_k2^{s(k)}\sqrt{c}. \tag{5.52}$$

PROOF. The proof of this lemma is a variation on the proof of Lemma 5.2. We recall from the proof of this lemma that if $J\subset G_n$ and $|\lambda|\leq 1$, we have

$$E\exp\lambda\sum_{i\in J}Y_i^0\leq\exp\left(\lambda^2c\operatorname{card}J\right). \tag{5.53}$$

Consider now the r.v. $W=\max_B\sum_{u\in B}|Y_u^0|$, where the supremum is taken over all families of sets $B$ that are the union of at most $m_k$ atoms of $\mathcal{P}_k$. We have

$$W\leq 2\max_{J\in\mathcal{J}}\left|\sum_{u\in J}Y_u^0\right|, \tag{5.54}$$

where $\mathcal{J}$ is the family of subsets of $G_n$ that are contained in the union of at most $m_k$ atoms of $\mathcal{P}_k$. If we combines (5.53), (5.54) and the crude (but effective) inequality $\exp\max(x_i)\leq\sum\exp x_i$, we get that, for $|\lambda|\leq 1$, we have

$$E\exp\lambda\frac{W}{2}\leq\operatorname{card}\mathcal{J}\exp\left(\lambda^2c2^{s(k)}m_k\right).$$

Now,

$$\operatorname{card} \mathcal{J} \le 2^{2^{s(k)}m_k}\binom{2^{3n-s(k)}}{m_k} \le 2^{2^{s(k)}m_k}\left(\frac{e2^{3n-s(k)}}{m_k}\right)^{m_k}.$$

For $k \le \tau$, we have $e2^{3n-s(k)}/m_k \le \exp(K2^k) = \exp(K2^{s(k)})$, so that card $\mathcal{J} \le \exp(K2^{s(k)}m_k)$. For $k \ge \tau$, we have

$$2^{3n-s(k)} \le 2^{3n-s(\tau)} = 2^{3n-\tau} \le \exp(2^{\tau+\rho}) \le \exp\bigl(2^{s(k)+\rho}\bigr),$$

so that again card $\mathcal{J} \le \exp K2^{s(k)}m_k$. Thus in any case, taking $\lambda = 1/\sqrt{c}$, we have

$$E\exp\frac{W}{2\sqrt{c}} \le \exp K2^{s(k)}m_k.$$

Thus, we have

$$P(W \ge 2\sqrt{c}t) \le \exp\bigl(-t + K2^{s(k)}m_k\bigr).$$

If we observe that $2^{s(k)}m_k \ge n/K$, the result follows by taking $t = 2K2^{s(k)}m_k$. □

We now go back to the proof of (5.51). We assume that the event of Lemma 5.10 occurs. Given $0 \le k \le \tau$, we consider a subset $B_k'$ of $B_k$ that is the union of at most $m_k$ atoms of $\mathcal{P}_k$, and for which $\sum_{u \in B_k'}|Y_u^0|$ is as large as possible. By (5.49), we have

$$\sum_{\substack{C \subset B_k' \\ C \in \mathcal{P}_k}}\left(\sum_{u \in C}|Y_u^0|\right)z(C) \le Km_k2^{s(k)}\sqrt{c}b_{k+1}. \tag{5.55}$$

For $k \le \tau$ let $B_k'' = B_k \backslash B_k'$, and for $k > \tau$ let $B_k'' = B_k$. It should be clear that for any atom $C$ of $\mathcal{P}_k$ contained in $B_k''$, one has

$$\sum_{u \in C}|Y_u^0| \le K2^{s(k)}\sqrt{c},$$

so that

$$\sum_{\substack{C \subset B_k'' \\ C \in \mathcal{P}_k}}\left(\sum_{u \in C}|Y_u^0|\right)z(C) \le K\sum_{C \subset B_k}2^{s(k)}z(C). \tag{5.56}$$

If we recall that $2^{s(k)}m_k \le 2^{3n}\exp(-2^{k+\rho})$ and $b_{k+1} \le \exp(2^{k+\rho-1})$, the result then follows from (5.50), by summation of the inequalities (5.55) and (5.56).

## REFERENCES


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

[2] LEIGHTON, T. and SHOR, P. W. (1989). Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9** 161–187.

[3] RHEE, W. T. and TALAGRAND, M. (1992). Matching random subsets of the cube with a tight control on one coordinate. *Ann. Appl. Probab.* **2** 695–713.

[4] SHOR, P. W. (1985). Random planar matching and bin packing. Ph. D. dissertation, MIT.

[5] SHOR, P. W. (1991). How to pack better than best fit: tight bounds for average-case on line bin packing. In *Proc. 32nd Annual Symp. on Foundations of Computer Science* 752–759. IEEE Computer Society Press, New York.

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

[7] TALAGRAND, M. (1994). Matching theorems and empirical discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* **7** 455–537.

[8] TALAGRAND, M. (1992). Matching random samples in many dimensions. *Ann. Appl. Probab.* **2** 846–856.



EQUIPE D'ANALYSE—TOUR 56
E. R. A. AU C. N. R. S. NO. 754
UNIVERSITÉ PARIS VI
4 PLACE JUSSIEU
75230 PARIS CEDEX 05
FRANCE

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
231 W. 18TH AVENUE
COLUMBUS, OHIO 43210-1174