ON BERNOULLI DECOMPOSITIONS FOR RANDOM VARIABLES, CONCENTRATION BOUNDS, AND SPECTRAL LOCALIZATION

MICHAEL AIZENMAN, FRANÇOIS GERMINET, ABEL KLEIN, AND SIMONE WARZEL

Abstract. As was noted already by A. N. Kolmogorov, any random variable has a Bernoulli component. This observation provides a tool for the extension of results which are known for Bernoulli random variables to arbitrary distributions. Two applications are provided here: *i.* an anti-concentration bound for a class of functions of independent random variables, where probabilistic bounds are extracted from combinatorial results, and *ii.* a proof, based on the Bernoulli case, of spectral localization for random Schrödinger operators with arbitrary probability distributions for the single site coupling constants. For a general random variable, the Bernoulli component may be defined so that its conditional variance is uniformly positive. The natural maximization problem is an optimal transport question which is also addressed here.

Contents

Date: June 29, 2007.

1. Introduction

This article has a twofold purpose. As a general observation it is noted that in any random variable one may find a Bernoulli component. A decomposition which is based on the above observation allows then to extend results which for systems of Bernoulli variables are available by combinatorial methods to systems of random variables of arbitrary distribution.

A Bernoulli decomposition of a real-valued random variable X is a representation of the form

$$X \stackrel{\mathcal{D}}{=} Y(t) \; + \; \delta(t)\, \eta \,, \tag{1.1}$$

where $Y(\cdot)$ and $\delta(\cdot) \geq 0$ are functions on $(0,1)$, the variable t is uniformly distributed in $(0,1)$, and η is a Bernoulli random variable taking the values $\{0,1\}$ with probabilities $\{1-p, p\}$ independently of t. The relation in (1.1) is to be understood as expressing equality of the distributions of the corresponding random variables.

Bernoulli decompositions are constructed here for arbitrary random variables of non-degenerate distributions. For certain purposes it is useful to have positive uniform conditional variance of the Bernoulli term, i.e.,

$$\inf_{t \in (0,1)} \delta(t) > 0 \,. \tag{1.2}$$

We present such a representation below and discuss related issues of optimality.

Two applications mentioned here: *i.* anti-concentration bounds for monotone, though not necessarily linear, functions of independent random variables, and *ii.* a proof, based on the Bernoulli case [BK], of spectral localization for random Schrödinger operators with arbitrary probability distributions for the single site coupling constants.

In the first application, we consider functions $\Phi(X_1, \ldots, X_N)$ of independent non-degenerate random variables $\{X_j\}$ whose distributions are either identical or, in a sense explained below, are of widths greater than some common $b_X > 0$. It is shown here that if for some $\varepsilon > 0$ the function satisfies

$$\Phi(\boldsymbol{u} + v\boldsymbol{e}_j) - \Phi(\boldsymbol{u}) \; > \; \varepsilon \tag{1.3}$$

for all $v \geq b_X$, all $\boldsymbol{u} \in \mathbb{R}^N$, and $j = 1, \ldots, N$, where $\boldsymbol{e}_j$ is the unit vector in the j-direction, then the following concentration bound applies:

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(\{\Phi(X_1, \ldots, X_N) \in [x, x+\varepsilon]\}\right) \; \leq \; \frac{C_X}{\sqrt{N}} \,, \tag{1.4}$$

with a constant $C_X < \infty$ which depends on the uniform bounds on the distributions of $\{X_j\}$. The proof employs the Bernoulli representation along with the combinatorial bounds of Sperner [S], and the more general LYM lemma [E].

The use of combinatorial estimates for concentration bounds first appeared in the context of Bernoulli variables in P. Erdös' variant of the Littlewood-Offord Lemma [Er]. The presence of a Bernoulli component in any random variable was noted implicitly in the work of A. N. Kolmogorov [Ko] where it was put to use in an improvement of the earlier concentration bounds of W. Doeblin and P. Lévy [DoL, Do] on linear functions of independent random variables. Initially, Kolmogorov did not extract the maximal benefit from the method by not connecting it with Sperner theory, and in particular the concentration bound in [Ko] includes an unnecessary logarithmic factor; the corresponding improvement was made by B. A. Rogozin [R1].

The bounds were further improved in a series of works, in particular [Es, K, R2] where use was also made of other methods. One may note here that perhaps quite naturally a general method like the Bernoulli decomposition is not optimized for specific applications. Nevertheless, it has the benefit of providing a simple perspective on a number of topics.

In our second application, we establish spectral localization for a broad class of continuum, alloy-type random Schrödinger operators (cf. (4.1)), building on a result of J. Bourgain and C. Kenig [BK] for the Bernoulli case. The model and the results are presented more explicitly in Section 4. The main point to be made here is that the understanding of spectral localization for the Bernoulli case can be extended through the Bernoulli decomposition to random operators with single site coupling parameters of arbitrary distribution (cf. Theorem 4.2).

2. Bernoulli decompositions for random variables

Randomness often is in the eyes of the beholder, as probability measures are used to express averages over specified sets of rather varied nature. However, it may be true that the most elementary model underlying the basic popular perception of probability is the simple 'coin toss', with two possible outcomes: heads or tails, which is modeled by a Bernoulli random variable: a binary variable equal to 1 with probability p and equal to 0 with probability $1-p$.

2.1. **The decomposition in two variants.**

The following statement assert that any real valued random variable has a Bernoulli component, which can even be chosen to be of uniformly positive variance.

Given a real random variable X by default we shall denote its probability distribution by μ and let $G:(0,1)\to(-\infty,\infty)$ be the function defined by

$$G(t) \;:=\; \inf\{u\in\mathbb{R}\,:\,\mu\left((-\infty,u]\right)\geq t\}\,. \tag{2.1}$$

One may observe that G is the 'inverse' distribution function of μ, which takes values in the essential range of X. It can alternatively be described by

$$G(t)\leq u \quad\Longleftrightarrow\quad \mu\left((-\infty,u]\right)\geq t\,, \tag{2.2}$$

and satisfies $\mu\left((-\infty,G(t)-\varepsilon]\right)<t\leq\mu\left((-\infty,G(t)]\right)$ for all $t\in\mathbb{R}$ and $\varepsilon>0$.

Theorem 2.1. *Let X be a non-degenerate real-valued random variable with a probability distribution μ. Then, for each $p\in(0,1)$, X admits a decomposition of the form:*

$$X \;\stackrel{\mathcal{D}}{=}\; Y_p(t) \;+\; \delta_p^+(t)\;\eta\,, \tag{2.3}$$

in the sense of equality of the corresponding probability distributions, where:

(1) *η and t are independent random variables, with η a binary variable taking the values $\{0,1\}$ with probabilities $\{1-p,p\}$, correspondingly, and t having the uniform distribution in $(0,1)$,*

(2) *$Y_p\,:\,(0,1)\mapsto(-\infty,\infty)$ is the monotone non-decreasing function*

$$Y_p(t) \;:=\; G((1-p)t)\,, \tag{2.4}$$

(3) *$\delta_p^+\,:\,(0,1)\mapsto[0,\infty)$ is the function*

$$\delta_p^+(t) \;:=\; G(1-p+pt)-G((1-p)t)\,, \tag{2.5}$$

(4) *for at least one value of* $p \in (0,1)$ *we have*

$$\beta^+(p,\mu) \ := \inf_{t\in(0,1)} \delta_p^+(t) > 0\,. \tag{2.6}$$

Some explicit expressions for $\beta^+(p,\mu)$ are mentioned in Remark 2.1 below. The Bernoulli component of the measure is not a uniquely defined notion, and other representations similar to (2.3) but with different distributions for the conditional variance of the Bernoulli component, i.e., for $\delta(t)$, can also be obtained. In the following construction its uniform positivity may be lost but one gains the feature that the range of values which δ assumes reaches up to the diameter of the support of the measure μ.

Theorem 2.2. *Let X be a non-degenerate real-valued random variable with probability distribution μ. Then, for each $p \in (0,1)$, X admits a decomposition of the form:*

$$X \ \stackrel{\mathcal{D}}{=} \ Y_p(t) \ + \ \delta_p^-(t)\ \eta \tag{2.7}$$

where t, η and the function Y_p are as in Theorem 2.1, satisfying the above (1) and (2), but instead of (3) and (4) the following holds

(3') $\delta_p^- : (0,1) \mapsto [0,\infty)$ *is the non-increasing function:*

$$\delta_p^-(t) \ := G(1-pt) - G((1-p)t)\,, \tag{2.8}$$

(4') *for any $x_- < x_+$ and $p_\pm > 0$ such that*

$$\mathbb{P}\left(\{X \le x_-\}\right) \ge p_- \quad \textit{and} \quad \mathbb{P}\left(\{X > x_+\}\right) > p_+\,, \tag{2.9}$$

at the particular value $p = \frac{p_+}{p_-+p_+}$ we have

$$\mathbb{P}_t\left(\left\{\delta_p^-(t) > x_+ - x_-\right\}\right) \ \ge \ p_- + p_+\,, \tag{2.10}$$

where the probability is with respect to the uniform random variable t.

In the proofs we employ two versions of what is called here the *Pac-Man* algorithm for the construction of a joint distribution ρ of a pair of random variables, of the form $\{Y_1(t), Y_2(t)\}$, whose marginal probability measures, ρ_1, ρ_2, satisfy

$$\mu \ = \ (1-p)\,\rho_1 \ + \ p\,\rho_2\,. \tag{2.11}$$

The representations (2.3) and (2.7) correspond to letting:

$$\begin{aligned} Y_p(t) \ &:= \ Y_1(t) \\ \delta_p^\pm(t) \ &:= \ Y_2(t) - Y_1(t)\,. \end{aligned} \tag{2.12}$$

The two Theorems will be proven in reverse order.

Proof of Theorem 2.2: We start by recalling the known observation that for any continuous function $\phi \in C(\mathbb{R})$:

$$\int_0^1 \phi(G(s))\,ds \ = \ \int_{\mathbb{R}} \phi(x)d\mu(x) \ \ . \tag{2.13}$$

This relation allows to represent, in terms similar to (2.7), as

$$X \ \stackrel{\mathcal{D}}{=} \ G(t)\,, \tag{2.14}$$

with t the random variable with the uniform distribution in $[0,1]$.

Extending the above representation, we now define a pair of coupled random variables through the following functions of $t \in [0,1]$:

$$\begin{aligned} Y_1(t) &:= G\left((1-p)t\right) \\ Y_2(t) &:= G\left(1-pt\right) \end{aligned} \tag{2.15}$$

As is the case of G in (2.14), the functions Y_1 and Y_2 are made into random variables by assigning to them the joint probability distribution which is induced by Lebesgue measure on $[0,1]$. Their marginal distributions satisfy (2.11), since for any continuous function $\phi \in C(\mathbb{R})$

$$\begin{aligned} (1-p)\int_0^1 \phi(Y_1(t))\,dt \quad &+ \quad p\int_0^1 \phi(Y_2(t))\,dt \\ &= (1-p)\int_0^1 \phi(G((1-p)t))\,dt \;+\; p\int_0^1 \phi(G(1-pt))\,dt \\ &= \left[\int_0^{1-p} + \int_{1-p}^1\right]\phi(G(s))\,ds \;=\; \int \phi(x)d\mu(x) \end{aligned} \tag{2.16}$$

where the last equality is by (2.13).

By (2.16) the random variable seen on the right side of (2.7) has the same distribution as X. The statement (2) readily follows from the definition (2.15) and (2.12).

For a proof of (4') we note that (2.9) is equivalent to

$$G(p_-) \;\le\; x_- \qquad \text{and} \qquad G(1-p_+) \;>\; x_+\,. \tag{2.17}$$

This implies $\delta_p^+(t) > x_+ - x_-$ for all $t \le p_+ + p_-$, and hence (2.10) holds true. □

In the above proof, one may regard the functions $Y_1(t)$ and $Y_2(t)$ defined by (2.15) as describing the motion of a pair of markers which move along $\mathbb{R}$ consuming the μ-measure at the steady rates of $(1-p)$ and p, correspondingly. The markers leap discontinuously over intervals of zero μ-measure and, conversely, linger at points of positive mass. Their motion invokes the image of a linear version of the *Pac-Man* game, and hence we shall refer to the construction by this name. Whereas in the above construction the Pac-Men move towards each other, we shall next use the Pac-Man algorithm with one marker chasing the other.

Proof of Theorem 2.1. For the representation (2.3) we shall employ the following variant of (2.15):

$$\begin{aligned} Y_1(t) &:= G\left((1-p)t\right) \\ Y_2(t) &:= G\left(1-p+pt\right) \end{aligned} \tag{2.18}$$

In this case, both Y_1 and Y_2 are monotone non-decreasing in t and

$$Y_1(t) \le G(1-p) \le G(1-p+0) \le Y_2(t) \tag{2.19}$$

for all $t \in (0,1)$, where $G(1-p+0) = \lim_{\varepsilon\downarrow 0} G(1-p+\varepsilon)$. Moreover, for any $T \in (0,1)$ we have the lower bound

$$\beta^+(p,\mu) \ge \min\left\{G(1-p) - Y_1(T), Y_2(T) - G(1-p+0)\right\}, \tag{2.20}$$

since

$$\delta_p^+(t) \ge \begin{cases} G(1-p) - Y_1(T) & \text{if } 0 < t \le T, \\ Y_2(T) - G(1-p+0) & \text{if } T \le t < 1\,. \end{cases} \tag{2.21}$$

For a sufficient condition for the uniform positivity of $\delta_p^+(t) = Y_2(t) - Y_1(t)$ let us consider the arrival/departure times:

$$\begin{aligned} T_1 &= \inf\{t \in (0,1) \,:\, Y_1(t) = G(1-p)\} && \text{(arrival time of } Y_1)\,, \\ T_2 &= \sup\{t \in (0,1) \,:\, Y_2(t) = G(1-p+0)\} && \text{(departure time of } Y_2)\,. \end{aligned}$$

The times T_1, T_2 are non-random and depend on p and μ only. If

$$T_1 > T_2\,, \tag{2.22}$$

then for each $T \in (T_2, T_1)$ we have

$$\beta^+(p,\mu) \geq \min\left\{G(1-p) - Y_1(T), Y_2(T) - G(1-p+0)\right\} > 0. \tag{2.23}$$

The collection of $p \in (0,1)$ such that (2.22) is not empty whenever the support of the measure includes more than one point. □

Remark 2.1. (i) *Explicit lower bounds on* β^+. For the Bernoulli decomposition which is presented in Theorem 2.1 (i.e., based on the 'chasing Pac-Men' algorithm), an expression for the lower bound $\beta^+(p,\mu)$ in terms of the distribution function of μ is given in (2.34) below. A simple lower bound can be obtained in terms of just the "half-time" points for the two markers. i.e., from (2.20) with $T = \frac{1}{2}$:

$$\beta^+(p,\mu) \;\geq\; \min\left\{ \left[G(1-p) - G\left(\frac{1-p}{2}\right)\right], \left[G\left(\frac{2-p}{2}\right) - G(1-p)\right]\right\}. \tag{2.24}$$

This shows that for continuous measures μ one has $\beta^+(p,\mu) > 0$, i.e., (2.6), for any $p \in (0,1)$.

If the support of μ consists of exactly two points the representation (2.7) is trivially available, though at a unique value of $p \in (0,1)$. If the support of μ contains more than two points, there exists at least one $\hat{x} \in \mathbb{R}$ such that

$$\begin{aligned} &\mu\left((-\infty, x]\right) < \mu\left((-\infty, \hat{x}]\right) \quad \text{if} \quad x < \hat{x}\,, \\ &0 < \mu\left((-\infty, \hat{x})\right) \leq \mu\left((-\infty, \hat{x}]\right) < 1\,. \end{aligned} \tag{2.25}$$

At the particular value $p = 1 - \mu\left((-\infty, \hat{x})\right)$ we then have $G(1-p) = \hat{x}$ and

$$\begin{aligned} &\beta^+(p,\mu) \geq \min\left\{\hat{x} - G((1-p)t), G(1-p+pt) - \hat{x}\right\} > 0 \\ &\qquad \text{for each } t \text{ such that} \quad \frac{\mu\left(\{\hat{x}\}\right)}{p} < t < 1\,. \end{aligned} \tag{2.26}$$

(ii) *An alternative form.* For another form of a Bernoulli decomposition, with a binary random variable $\sigma = \pm 1$, let

$$\sigma \;=\; 2\eta - 1 \qquad \text{and} \qquad W \;=\; Y_p + \tfrac{1}{2}\delta_p^+\,. \tag{2.27}$$

When such a substitution is made in (2.3) the two resulting functions $W(t)$ and $\delta_p^+(t)$ are monotone non-decreasing in t and $\delta_p^+(\cdot)$ is constant over each interval of constancy of $W(\cdot)$. It follows that the value of $\delta_p^+(t)$ can be expressed in terms of $W(t)$, and thus one obtains a representation of the form:

$$X \;\stackrel{\mathcal{D}}{=}\; W + b(W)\sigma\,, \tag{2.28}$$

with W and σ independent random variables, and $b(\cdot)$ a measurable function which is determined by μ and p.

(iii) *Some precedents.* As was commented above, the Bernoulli decomposition of Theorem 2.2 with $p = 1/2$ has appeared already in a work of A. N. Kolmogorov [Ko]. For random variables with values in $\mathbb{Z}$, the representation (2.3) of Theorem 2.1 is related to the somewhat similar representation (though with $\delta = 0, 1$, not necessarily positive) which D. McDonald showed to be useful for the analysis of local limit theorems for integer random variables ([M]).

2.2. **Optimality of the Pac-Man algorithm.** In applications of the decomposition it is desirable to maximize the conditional variance of the binary term. We shall now address related questions from an *optimal transport* perspective, and in particular establish optimality, in a certain limited sense, of the 'chasing Pac-Men' construction.

In addition to the explicit choices presented in Theorems 2.1 and 2.2 there are other possibilities for a Bernoulli decomposition of the form (2.3). With a change of variables as in (2.12), such representations can alternatively be expressed in terms of joint distributions of the variables Y_1, Y_2 with the properties listed in the following definition.

Definition 2.1. A $(1-p, p)$ Bernoulli decomposition of a probability measure μ on $\mathbb{R}$ is a probability measure $\rho(dY_1\, dY_2)$ on $\mathbb{R}^2$ whose marginals ρ_1 and ρ_2 satisfy:

$$(1-p)\,\rho_1 \;+\; p\,\rho_2 \;=\; \mu\,. \tag{2.29}$$

This concept can of course be easily generalized to variables with values in $\mathbb{R}^d$, or $\mathbb{C}$. For real variables the defining condition (2.29) is conveniently expressed in terms of the distribution functions, as

$$(1-p)\,F_1(x) \;+\; p\,F_2(x) \;=\; F(x) \tag{2.30}$$

where $F(x) = \mu((-\infty, x])$, and $F_j(x) = \rho(\{Y_j \le x\})$ for $j = 1, 2$.

For each Bernoulli decomposition of a probability measure on $\mathbb{R}$ we denote:

$$\left\{ \begin{array}{c} \beta^{\#} \\ \beta_* \end{array} \right\} (p, \rho) \;:=\; \mathrm{ess}_\rho \left\{ \begin{array}{c} \sup \\ \inf \end{array} \right\} (\mathrm{Y}_2 - \mathrm{Y}_1)\,. \tag{2.31}$$

Theorem 2.3. *For any $(1-p, p)$, among all the Bernoulli decomposition of a given probability measure μ on $\mathbb{R}$:*

1. *The minimal conditional variation $\beta_*(p, \rho)$ is maximized by the 'chasing Pac-Men' algorithm which is presented in the proof of Theorem 2.1, i.e. for any Bernoulli decomposition*
$$\beta_*(p, \rho) \;\le\; \beta^+(p, \mu) := \operatorname{ess\,inf}_{t\in(0,1)} \delta_p^+(t)\,, \tag{2.32}$$
where $\operatorname{ess\,inf}_{t\in(0,1)}$ *yields the same value as* $\inf_{\mathrm{t}\in(0,1)}$.
2. *The maximal conditional variation $\beta^{\#}(p, \rho)$ is maximized by the 'colliding Pac-Men' algorithm of Theorem 2.2, for which $\beta^{\#}(p, \rho)$ equals the diameter of the essential support of μ.*

The equality: $\mathrm{essinf}_{\mathrm{t}\in[0,1]}\delta_{\mathrm{p}}^+(\mathrm{t}) = \inf_{\mathrm{t}\in[0,1]}\delta_{\mathrm{p}}^+(\mathrm{t})$ is a simple consequence of the left-continuity property of the chasing Pac-Men algorithm, where $Y_j(t) = Y_j(t-0)$ and hence also $\delta^+(t) = \delta^+(t-0)$.

To prove (2.32) let us first establish a helpful expression for $\beta^+(p, \mu)$. Denoting by F_j^+ the distribution functions corresponding to Y_1 and Y_2 of (2.18) we have:

Lemma 2.1. *For the 'chasing Pac-Men' construction, of Theorem 2.1:*

$$F_1^+(x) = \frac{1}{1-p} \min\{F(x),\, 1-p\}\ , \quad F_2^+(x) = \frac{1}{p} \max\{F(x)+p-1,\, 0\} \tag{2.33}$$

and

$$\beta^+(p,\mu) \ = \ \sup\left\{b \in \mathbb{R} \,:\, F_1^+(x) \geq F_2^+(x+b) \quad \text{for all } x \in \mathbb{R}\right\} . \tag{2.34}$$

Proof. The statements (2.33) follow directly from the definition of the Pac-Man process (2.18). In the derivation of (2.34), we shall use the fact that for all $t \in (0,1)$ and $\varepsilon > 0$:

$$F_j^+(Y_j^+(t) - \varepsilon) < t \ \leq F_j^+(Yj^+(t)) \tag{2.35}$$

Let $S := \sup\left\{b \in \mathbb{R} \,:\, F_1^+(x) \geq F_2^+(x+b) \quad \text{for all } x \in \mathbb{R}\right\}$. Clearly, for any $u > S$, there is $x \in \mathbb{R}$ such that

$$F_1(x) \ < \ F_2(x+u)\,. \tag{2.36}$$

It follows that for any $t \in (F_1(x), F_2(x+u))$:

$$Y_2(t) \ \leq \ x+u \qquad \text{and} \qquad Y_1(t) \ > \ x\,, \tag{2.37}$$

and therefore $\delta^+(t) = Y_2(t) - Y_1(t) \ \leq \ u$. Thus: $\inf_{t\in(0,1)} \delta_p^+(t) \leq S$.

For the converse direction, let us note that due to the monotonicity of F the condition on b in (2.34) is satisfied by all $u < S$. Thus, if $u < S$, then, for all $x \in \mathbb{R}$:

$$F_2^+(x+u) \ \leq \ F_1^+(x-0)\,, \tag{2.38}$$

and hence for any $t \in (0,1)$:

$$F_2^+(Y_1^+(t)+u) \ \leq \ F_1^+(Y_1^+(t)-0) \leq \ t\,, \tag{2.39}$$

which implies that $Y_1^+(t) + u \leq Y_2^+(t)$. Therefore

$$\inf_{t\in(0,1)} \left(Y_2^+(t) - Y_1^+(t)\right) \ \geq \ u\,. \tag{2.40}$$

It follows that $\inf_{t\in[0,1]} \delta_p^+(t) \geq S$, which completes the proof of (2.34). □

Proof of Theorem 2.3 . The second assertion is an elementary consequence of (2.8). To prove (1) we shall show that for any $b > \beta^+(p,\mu)$ it is also true that $b > \beta_*(p,\rho)$.

The condition (2.30) readily implies that $(1-p)\,F_1(u) \leq F(u)$, or $F_1(x) \ \leq \min\left\{(1-p)^{-1}F(x),\, 1\right\}$, and hence

$$\begin{aligned} F_1(x) \ &\leq \ F_1^+(x) \\ F_2(x) \ &\geq \ F_2^+(x)\,. \end{aligned} \tag{2.41}$$

Now, by Lemma 2.1, for any $b > \beta^+(p,\mu)$ there exist some $t, u \in \mathbb{R}$, such that $F_1^+(u) \ = t \ < \ F_2^+(u+b)$ and therefore, due to (2.41), also

$$F_1(u) \ \leq t \ < \ F_2(u+b)\,. \tag{2.42}$$

Eq. (2.42) means that $\rho(\{Y_1 \leq u\}) \leq t$ and $\rho(\{Y_2 > u+b\}) < 1-t$. Since the probabilities of the two events add to less than 1 the complement of their union is of positive probability, and this implies:

$$\rho(\{Y_2 - Y_1 \leq b\}) \ > \ 0\,, \tag{2.43}$$

and hence $b > \beta_*(p,\rho)$. This concludes the proof of (2.32). □

Remark 2.2. The idea of seeking optimal joint realizations of random variables with constrained marginals has allowed to present a wide range of analytical results from a common 'optimal transport' perspective (see, e.g., [V]). The most familiar variants of the problem concern couplings which *minimize* a distance function between the two coupled variables. As our discussion demonstrates, it may also be of interest to seek couplings which *maximize* the difference between the two variables with constrained marginals.

3. Concentration Bounds

We shall now demonstrate how the Bernoulli decomposition yields probabilistic bounds from combinatorial results. If there is any novelty in this section it is in the formulation of the bounds for the non-linear case, as the two main ideas were noted before in the context of linear functions: P. Erdös [Er] observed that concentration bounds for linear functions of Bernoulli variables can be derived from the combinatorial theory of E. Sperner [S], and B. A. Rogozin [R1] has used the Bernoulli decomposition of A. N. Kolmogorov [Ko] for the further extension of these bounds to arbitrary random variables.

First, we present some essentially known results of Sperner theory; in the second subsection these results will be combined with the Bernoulli decomposition to yield concentration bounds for functions of independent random variables.

3.1. **Probabilistic Sperner Estimates.** The configuration space $\{0,1\}^N$ for a collection of Bernoulli random variables $\boldsymbol{\eta} = \{\eta_1, ..., \eta_N\}$ is partially ordered by the relation defined by:

$$\boldsymbol{\eta} \prec \boldsymbol{\eta}' \quad \Longleftrightarrow \qquad \text{for all} \;\; i \in \{1, ..., N\} : \quad \eta_i \leq \eta_i' \,. \tag{3.1}$$

A set $\mathcal{A} \subset \{0,1\}^N$ is said to be an *antichain* if it does not contain any pair of configurations which are compatible in the sense of "$\prec$". The original Sperner Lemma states that for any such set: $|\mathcal{A}| \leq \binom{N}{[\frac{N}{2}]}$. A more general result is the LYM inequality for antichains (cf. [An]):

$$\sum_{\boldsymbol{\eta} \in \mathcal{A}} \frac{1}{\binom{N}{|\boldsymbol{\eta}|}} \;\leq\; 1 \,, \tag{3.2}$$

where $|\boldsymbol{\eta}| = \sum \eta_j$.

The LYM inequality has the following probabilistic implication.

Lemma 3.1. *Let $\{\eta_j\}$ be independent copies of a Bernoulli random variable η with*

$$\mathbb{P}(\{\eta = 1\}) = p \,, \qquad \mathbb{P}(\{\eta = 0\}) = q := 1 - p \,, \tag{3.3}$$

where $p \in (0,1)$. Then for any antichain $\mathcal{A} \subset \{0,1\}^N$:

$$\mathbb{P}(\{\boldsymbol{\eta} \in \mathcal{A}\}) \;\leq\; \frac{\Theta}{\sigma_\eta \sqrt{N}} , \tag{3.4}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$, $\sigma_\eta = \sqrt{pq}$ is the standard deviation of η, and Θ is an independent constant which does not exceed $2\sqrt{2}$.

Proof. Let A_k be the subset of $\mathcal{A}$ consisting of configurations with $|\boldsymbol{\eta}| = k$. Then:

$$\mathbb{P}(\{\boldsymbol{\eta} \in \mathcal{A}\}) = \sum_{k=0}^{N} p^k q^{N-k} \, |\mathcal{A}_k| = \sum_{k=0}^{N} b(k; N, p) \frac{|\mathcal{A}_k|}{\binom{N}{k}} \leq \max_{k=0,1,\ldots,N} b(k; N, p), \tag{3.5}$$

where $b(k;N,p) := p^k q^{N-k}\binom{N}{k}$ is the binomial distribution, and the inequality is by (3.2). The maximum of $b(k;N,p)$ over k, which is known to occur near $k = pN$ (cf. [F, Theorem 1 on p. 140]) yields (3.4). $\square$

The bound (3.4) has the virtue of being valid for all N; for $N \to \infty$ it holds with a smaller constant which tends to the asymptotic value $\Theta \to 1/\sqrt{2\pi}$ (implied by (3.5) and Stirling's formula).

Following is an extension of Lemma 3.1 to the case of non-identically distributed random variables.

Lemma 3.2. *Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$, where $\{\eta_j\}$ are independent Bernoulli random variables with possibly different values of p_j, and set*

$$\alpha := \min_{j=1,2,\ldots,N} \min\{p_j, 1-p_j\} \in (0, 1/2]\,. \tag{3.6}$$

Then, for any antichain $\mathcal{A} \subset \{0,1\}^N$:

$$\mathbb{P}\{\boldsymbol{\eta} \in \mathcal{A}\} \le \frac{\widetilde{\Theta}}{\alpha\sqrt{N}}\,, \tag{3.7}$$

where $\widetilde{\Theta}$ is an independent constant which does not exceed 4.

The proof gives us the chance to introduce the technique of 'double sampling'.

Proof. We start from the observation that any Bernoulli variable η with parameter p_η as in (3.3) may be decomposed in terms of two independent Bernoulli variables χ and ξ as

$$\eta \stackrel{\mathcal{D}}{=} \xi\,\chi\,, \tag{3.8}$$

with $p_\xi\, p_\chi = p_\eta$.

By the definition of α, eq. (3.6), $p_j \in [\alpha, 1-\alpha]$ for all $j = 1, 2, \ldots, N$. Hence the variables $\boldsymbol{\eta}$ may be represented as in (3.8) with independent identically distributed (iid) Bernoulli variables $\{\chi_j\}$ with common $p_\chi := 1-\alpha$. We abbreviate this representation as $\boldsymbol{\xi}\,\boldsymbol{\chi} := (\xi_1\chi_1, \ldots, \xi_N\chi_N)$. Evaluating the probability by first conditioning on the values of $\boldsymbol{\xi}$, one has

$$\mathbb{P}\{\boldsymbol{\eta} \in \mathcal{A}\} = \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\xi}\,\boldsymbol{\chi} \in \mathcal{A}\,|\,\boldsymbol{\xi}\right\}\right] \tag{3.9}$$

For specified values of the variables $\boldsymbol{\chi}$, the event $\mathcal{A}$ depends only on the values of χ_j with j in the set $J_{\boldsymbol{\xi}} := \{j\,:\,\xi_j \neq 0\}$, and as such it is an antichain in $\{0,1\}^{J_{\boldsymbol{\xi}}}$. Bounding its conditional probability by Lemma 3.1 we obtain

$$\mathbb{P}\left\{\boldsymbol{\xi}\,\boldsymbol{\chi} \in \mathcal{A}\,|\,\boldsymbol{\xi}\right\} \le \min\left\{1, \frac{\Theta}{\sigma_\chi\sqrt{|J_{\boldsymbol{\xi}}|}}\right\}\,, \tag{3.10}$$

where $\sigma_\chi = \sqrt{\alpha(1-\alpha)}$ is the common standard deviation of χ_j.

To conclude the proof of (3.7) it remains to estimate the expected value of the right hand side of (3.10), where $|J_{\boldsymbol{\xi}}| = \sum_{j=1}^N \xi_j$. Noting that $\mathbb{E}(\xi_j) = p_{\xi_j} = p_j/(1-\alpha) \ge \alpha/(1-\alpha)$, we see that the mean satisfies:

$$\mathbb{E}(|J_{\boldsymbol{\xi}}|) \;\ge\; \frac{\alpha N}{1-\alpha}\,. \tag{3.11}$$

The event $\{|J_{\boldsymbol{\xi}}| \leq \alpha N/2(1-\alpha)\}$ is of exponentially small probability, as can be seen by a standard large deviation estimate for independent variables. It then readily follows that

$$\mathbb{E}\left(\min\left\{1, \frac{\Theta}{\sigma_\chi\sqrt{|J_{\boldsymbol{\xi}}|}}\right\}\right) \leq \frac{\widetilde{\Theta}}{\alpha\sqrt{N}}, \tag{3.12}$$

with a constant for which elementary estimates yield $\widetilde{\Theta} \leq 4$. □

Remark 3.1. The above notions and results have natural extensions to integer valued independent random variables, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_N)$, whose configuration space, $\mathbb{Z}^N$, is also partially ordered by the natural extension of the relation (3.1). The Bernoulli decomposition (2.7) can be used for an extension of the probabilistic bound of Lemma 3.2 to this more general case. One way to derive the general statement is through the application of the bound (3.7) to the conditional probability for the Bernoulli component, as in the arguments which appear below. Alternatively, one may note that the statement directly follows from Theorem 3.1 which is presented in the next section.

For completeness it should be added that in addition to the anti-concentration upper bounds it is of interest to know the asymptotic behavior. That is covered by known results, such as is presented in Engel [E, Theorem 7.2.1]:

$$\lim_{N\to\infty} \sigma_\mu\sqrt{2\pi N}\left\{\max_{\mathcal{A}\subset\{0,1,\ldots,k\}^N \text{ antichain}} \mathbb{P}\{\boldsymbol{\tau}\in\mathcal{A}\}\right\} = 1, \tag{3.13}$$

which amounts to a 'local' central limit theorem (CLT).

3.2. **Concentration Bounds for Functions of Independent Random Variables.** We shall now employ the Bernoulli decomposition of Section 2, along with the results presented in the previous subsection, for an upper bound on the concentration probability

$$Q_Z(\xi) := \sup_{x\in\mathbb{R}} \mathbb{P}\left(\{Z \in [x, x+\xi]\}\right) \tag{3.14}$$

for random variables of the form

$$Z = \Phi(X_1, X_2, \ldots, X_N), \tag{3.15}$$

where $\{X_j\}$ are independent random variables.

Theorem 3.1. *Let $\boldsymbol{X} = (X_1, \ldots, X_N)$ be a collection of independent random variables whose distributions satisfy, for all $j \in \{1, ..., N\}$:*

$$\mathbb{P}\left(\{X_j \leq x_-\}\right) \geq p_- \quad \text{and} \quad \mathbb{P}\left(\{X_j > x_+\}\right) > p_+ \tag{3.16}$$

at some $p_\pm > 0$ and $x_- < x_+$, and $\Phi : \mathbb{R}^N \mapsto \mathbb{R}$ a function such that for some $\varepsilon > 0$

$$\Phi(\boldsymbol{u} + v\boldsymbol{e}_j) - \Phi(\boldsymbol{u}) > \varepsilon \tag{3.17}$$

for all $v \geq x_+ - x_-$, all $\boldsymbol{u} \in \mathbb{R}^N$, and $j = 1, \ldots, N$, with $\boldsymbol{e}_j$ the unit vector in the j-direction. Then, the random variable Z which is defined by (3.15) obeys the concentration bound

$$Q_Z(\varepsilon) \leq \frac{4}{\sqrt{N}}\sqrt{\frac{1}{p_+} + \frac{1}{p_-}}, \tag{3.18}$$

where 4 can also be replaced by the constant $\widetilde{\Theta}$ of (3.7).

Proof. We start by selecting $p \in (0,1)$ by the condition $p = \frac{p_+}{p_+ + p_-}$. Next, we represent the variables $\{X_j\}$ using Theorem 2.2:

$$\boldsymbol{X} \overset{\mathcal{D}}{=} \boldsymbol{Y}(\boldsymbol{t}) + \boldsymbol{\delta}(\boldsymbol{t})\,\boldsymbol{\eta} := \Big(Y_{p,1}(t_1) + \delta^-_{p,1}(t_1)\,\eta_1, \ldots, Y_{p,N}(t_N) + \delta^-_{p,N}(t_N)\,\eta_N \Big)\,, \tag{3.19}$$

with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ a collection of iid Bernoulli variables taking values $\{0,1\}$ with probability $\{1-p, p\}$. From (2.10) one may conclude that for all $j \in \{1, \ldots, N\}$:

$$\mathbb{P}_t(\{\delta^-_{p,j}(t) \geq x_+ - x_-\}) \geq p_+ + p_-\,. \tag{3.20}$$

We express the probability of the event $\{Z \in [x, x+\varepsilon]\}$ through first conditioning on the $\{t_j\}$ variables. For all $x \in \mathbb{R}$:

$$\mathbb{P}\left(\{Z \in [x, x+\varepsilon]\}\right) = \mathbb{E}\left[\mathbb{P}\left(\mathcal{A}_{\boldsymbol{t}} \,\middle|\, \boldsymbol{t}\right)\right] \tag{3.21}$$

where

$$\mathcal{A}_{\boldsymbol{t}} := \left\{ \boldsymbol{\eta} \in \{0,1\}^N \,:\, \Phi(\boldsymbol{Y}(\boldsymbol{t}) + \boldsymbol{\delta}(\boldsymbol{t})\,\boldsymbol{\eta}) \in [x, x+\varepsilon] \right\}\,. \tag{3.22}$$

By virtue of (3.17), the set $\mathcal{A}_{\boldsymbol{t}}$ is an antichain in its dependence on $\{\eta_j\}_{j \in J_{\boldsymbol{t}}}$ with $J_{\boldsymbol{t}} := \{j \,:\, \delta_j(t_j) \geq x_+ - x_-\}$. Lemma 3.1 thus yields

$$\mathbb{P}\left\{\mathcal{A}_{\boldsymbol{t}} \,\middle|\, \boldsymbol{t}, \{\eta_j\}_{j \notin J_{\boldsymbol{t}}}\right\} \leq \min\left\{1, \frac{\Theta}{\sigma_\eta \sqrt{|J_{\boldsymbol{t}}|}}\right\} \tag{3.23}$$

with $\sigma_\eta = \sqrt{p(1-p)}$. We conclude by the large-deviation argument used in the proof of Lemma 3.2. Using (3.20) the expected value of $|J_{\boldsymbol{t}}| = \sum_{j=1}^N 1_{\{j \,:\, \delta_j(t_j) \geq x_+ - x_-\}}$ is bounded below:

$$\mathbb{E}\left(|J_{\boldsymbol{t}}|\right) \geq (p_+ + p_-)N\,. \tag{3.24}$$

Therefore $\{|J_{\boldsymbol{t}}| \leq \frac{1}{2}(p_+ + p_-)N\}$ is a large deviation event and its probability is exponentially bounded. Elementary estimates lead to

$$\mathbb{E}\left(\min\left\{1, \frac{\Theta}{\sigma_\eta \sqrt{|J_{\boldsymbol{t}}|}}\right\}\right) \leq \frac{\widetilde{\Theta}}{\sqrt{N}} \sqrt{\frac{1}{p_+} + \frac{1}{p_-}}\,, \tag{3.25}$$

with the same constant $\widetilde{\Theta}$ as in (3.7). □

Remark 3.2. (i) *A simpler proof for iid variables.* For iid non-degenerate random variables $X_1, \ldots, X_N$ the theorem has a simpler proof using the binary decomposition of Theorem 2.1; there is no need for the large deviation argument. The constants in the theorem will then depend on the value of p and its corresponding lower bound in (2.6).

(ii) *The linear case.* For linear functions,

$$Z = \Phi(X_1, \ldots, X_N) = \sum_{j=1}^N X_j\,, \tag{3.26}$$

concentration inequalities as in (3.18) go back to W. Doeblin, P. Lévy [DoL, Do], P. Erdös [Er] (for the Bernoulli case, where it reduces to the Littlewood-Offord problem), A. N. Kolmogorov [Ko], B. A. Rogozin [R1], H. Kesten [K] and C. G. Esseen [Es]. In this case, sharper inequalities than (3.18) are known, e.g. [R3],

$$Q_Z(\varepsilon) \leq \Theta\,\varepsilon \left[\sum_{j=1}^N \varepsilon_j^2 \left(1 - Q_{X_j}(\varepsilon_j)\right)\right]^{-1/2} \tag{3.27}$$

where Θ is some constant. A recent application of the discrete case of the concentration bounds is found in [TV].

(iii) *An extension.* As it is already true for (3.27), the statement of Theorem 3.1 has an immediate extension to functions which in some variables are monotone increasing and in some are monotone decreasing, satisfying the natural analog of (3.17). For this extension, one only needs to replace p_+ and p_- in (3.18) by $\hat{p} = \min\{p_+, p_-\}$.

(iv) *Sperner bounds from concentration inequalities.* In the proof of Theorem 3.1 concentration bounds were deduced from the probabilistic Sperner estimate (3.4). For antichains in the multiset $S = \{0, 1, \ldots, K\}^N$ the implication can also be established in the opposite direction. For that, one may use the fact that in such a multiset for any antichain $\mathcal{A}$ there is a function $\Phi : S \mapsto \mathbb{R}$ which satisfies the 'representation condition' (in the terminology of [E])

$$\Phi(\boldsymbol{u} + \boldsymbol{e}_j) - \Phi(\boldsymbol{u}) \;\geq\; 1 \tag{3.28}$$

and for which $\Phi(\boldsymbol{u}) = 0$ if and only in $\boldsymbol{u} \in \mathcal{A}$.

4. An Application to Random Schrödinger Operators

As a demonstration of a possible uses of the elementary observations which are made in this article, let us present the case of spectral localization under random iid single site potential for an arbitrary probability distribution.

The (continuum) Anderson Hamiltonian is the random Schrödinger operator given by

$$H_{\boldsymbol{\omega}} = -\Delta + V_{\boldsymbol{\omega}} \quad \text{on} \quad \mathrm{L}^2(\mathbb{R}^d), \tag{4.1}$$

with

$$V_{\boldsymbol{\omega}}(x) = \sum_{\xi \in \mathbb{Z}^d} \omega_\xi \, u(x - \xi), \tag{4.2}$$

where

(1) $u(\cdot)$, the single site potential, is a nonnegative bounded measurable function on $\mathbb{R}^d$ with compact support, uniformly bounded away from zero in a neighborhood of the origin,
(2) $\boldsymbol{\omega} = \{\omega_\xi\}_{\xi \in \mathbb{Z}^d}$ is a family of independent identically distributed random variables, whose common probability distribution μ satisfies: $\{0, M\} \in \operatorname{supp} \mu \subset [0, M]$, for some $M > 0$.

The random operator $H_{\boldsymbol{\omega}}$ is a function of $\boldsymbol{\omega}$, and as such it is defined over a probability space which is invariant under the ergodic action of the group of $\mathbb{Z}^d$ translations. The induced maps on this operator valued function are implemented by unitary translations.

Ergodicity considerations carry the implication that there exist fixed subsets of $\mathbb{R}$ so that the spectrum of the self-adjoint operator $H_{\boldsymbol{\omega}}$, as well as its pure point (pp), absolutely continuous (ac), and singular continuous (sc) components, are equal to these fixed sets with probability one (c.f. [P, KuS, KiM]). In the case of the random potential (4.2), the positivity of $u(\cdot)$ and the support properties of μ imply that

$$\sigma(H_{\boldsymbol{\omega}}) \stackrel{as}{=} [0, \infty)\,. \tag{4.3}$$

Although definitions of localization may come in several flavors, they all include (or imply) spectral localization (i.e., pure point spectrum), as given in the following definition.

Definition 4.1. A self-adjoint operator H on $\mathrm{L}^2(\mathbb{R}^d)$ is said to exhibit spectral localization in a closed interval $I \subset \mathbb{R}$ if $\sigma(H) \cap I \neq \emptyset$ and the corresponding spectral projection $P_I(H)$ is given by a countable sum of orthogonal projections on proper eigenspaces.

This property is clearly invariant under translations. The defining condition is equivalent to the requirement that for a spanning set of vectors the spectral measure is pure-point within I. The set of $\boldsymbol{\omega}$ for which this holds for the random operator $H_{\boldsymbol{\omega}}$ is known to be measurable.

In the one-dimensional case the continuous Anderson Hamiltonian has been long known to exhibit spectral localization in the whole real line for any non-degenerate μ, i.e. when the random potential is not constant [GoMP, DSS]. In the multi-dimensional case, localization at the bottom of the spectrum is already known at great, but nevertheless not all-inclusive, generality; cf. [St, Kl, BK] and references therein. The Bernoulli decomposition presented here allows to prove localization for general non-degenerate single site distributions μ.

More explicitly, the simplest case to deal with, for the different approaches which yield proofs of localization, has been when the single site probability distribution is absolutely continuous with bounded derivative. The absolute continuity condition can be relaxed to Hölder continuity of μ, both in the approach based on the multiscale analysis which was introduced in [FrS] and is discussed in [Kl], and in the one based on the fractional moment method of [AM, AE+]. (The basis in the former case is an improved analysis of the Wegner estimate, which can be found in [St, CHK].) However, techniques relying on the regularity of μ seem to reach their limit with log-Hölder continuity. In particular, until recently the Bernoulli random potential had been beyond the reach of analysis in more than one dimension. For that extreme case, i.e., of $H_{\boldsymbol{\omega}}$ with $\mu\{1\} = \mu\{0\} = \frac{1}{2}$, localization at the bottom of the spectrum was recently proven by Bourgain and Kenig [BK]. A crucial step in the analysis of [BK] is the estimation of the probabilities of energy resonances using Sperner's Lemma, i.e., the $p = \frac{1}{2}$ version of (3.4).

The point which we would like to make here is that the Bernoulli decomposition of random variables enables one to turn the latter result of Bourgain and Kenig [BK] into a tool for a general proof of localization at the edge of the spectrum for arbitrary non-degenerate μ.

First, the Bourgain and Kenig [BK] analysis needs to be extended to Schrödinger operators which incorporate an additional background potential $U \in L^\infty(\mathbb{R}^d)$, and for which the variances of the Bernoulli terms are uniformly positive, thought not necessarily uniform. More explicitly, the class is broadened to include operators of the form

$$H_{\boldsymbol{\eta}} = -\Delta + U(x) + \sum_{\xi \in \mathbb{Z}^d} \eta_\xi \, b_\xi \, u(x-\xi)\,, \tag{4.4}$$

where $u(\cdot)$ is as in (4.2), satisfying the above condition (1), but instead of (2):

(2') $\boldsymbol{\eta} = \{\eta_\xi\}_{\xi \in \mathbb{Z}^d}$ are iid Bernoulli random variables taking the values $\{0,1\}$ with probabilities $\{1-p, p\}$, and the coefficients $\{b_\xi\}_{\xi\in\mathbb{Z}^d}$ satisfy

$$0 < b_- \leq b_\xi \leq b_+ < \infty \qquad \text{for all } \xi \in \mathbb{Z}^d, \tag{4.5}$$

and

(3) $U \in L^\infty(\mathbb{R}^d)$ satisfies, for all $x \in \mathbb{R}^d$: $0 \le U(x) \le U_+ < \infty$.

Due to the presence of the background potential U the spectrum of $H_{\boldsymbol{\eta}}$ need not be deterministic, i.e., equal to some fixed set with probability one. For our main purpose it would suffice to restrict attention to U for which the spectrum of $H_{\boldsymbol{\eta}}$ is almost surely $[0,\infty)$. Such restriction is not included in the following statement but instead there is a caveat in the conclusion.

The extended BK result, whose proof is presented in [GK], is:

Theorem 4.1. *Given a function $u(\cdot)$ as above, and: $p \in (0,1)$, $b_\pm > 0$ and $U_+ < \infty$, there exist $E_0 > 0$ such that any random operator $H_{\boldsymbol{\eta}}$ of the form* (4.4)*, satisfying conditions (1), (2') and (3), for otherwise arbitrary external potential U, with probability one, either exhibits spectral localization in $[0,E_0]$ or $\sigma(H_{\boldsymbol{\eta}}) \cap [0,E_0]=\emptyset$.*

Theorem 2.1 allows now to deduce the following general statement from the above non-trivial Bernoulli result.

Theorem 4.2. *Let $H_{\boldsymbol{\omega}} = -\Delta + V_{\boldsymbol{\omega}}$ be a Schrödinger operator with the random potential given by* (4.2)*, satisfying the above conditions (1) and (2). Then for some $E_0 > 0$ the operator $H_{\boldsymbol{\omega}}$, with probability one, exhibits spectral localization in $[0,E_0]$.*

Proof. The Bernoulli decomposition (2.3) allows to write the coefficients in the random potential in the form:

$$\boldsymbol{\omega} \overset{\mathcal{D}}{=} \left\{ Y_+(t_\xi) + \delta_p^+(t_\xi)\, \eta_\xi \right\}_{\xi\in\mathbb{Z}^d} , \tag{4.6}$$

with $\boldsymbol{t} = \{t_\xi\}_{\xi\in\mathbb{Z}^d}$ a family of independent random variables which are uniformly distributed in $(0,1)$, Y_+ and δ_p^+ the functions defined in (2.12) in terms of the distribution function of μ, and $\boldsymbol{\eta} = \{\eta_\xi\}_{\xi\in\mathbb{Z}^d}$ a family of iid Bernoulli variables, independent of $\boldsymbol{t}$, which take values in $\{0,1\}$ with probabilities $\{1-p,p\}$ for some $p \in (0,1)$ such that (2.6) holds.

As a consequence, the random operator can be written as:

$$H_{\boldsymbol{\omega}} \overset{\mathcal{D}}{=} -\Delta + U_{\boldsymbol{t}} + V_{\boldsymbol{t},\boldsymbol{\eta}} =: H_{\boldsymbol{t},\boldsymbol{\eta}} \tag{4.7}$$

where

$$U_{\boldsymbol{t}}(x) := \sum_{\xi\in\mathbb{Z}^d} Y_+(t_\xi)\, u(x-\xi) \quad \text{and} \quad V_{\boldsymbol{t},\boldsymbol{\eta}}(x) := \sum_{\xi\in\mathbb{Z}^d} \delta_p^+(t_\xi)\, \eta_\xi\, u(x-\xi)\,, \tag{4.8}$$

and the following bounds hold

$$\begin{aligned} &0 \le U_{\boldsymbol{t}}(x) \le U_+ := M \left\| \sum_{\xi\in\mathbb{Z}^d} u(\cdot - \xi) \right\|_\infty < \infty\,, \\ &0 < b_- := \inf_{t\in(0,1)} \delta_p^+(t) \le b_+ := M < \infty\,. \end{aligned} \tag{4.9}$$

This implies that when conditioned on the values of $\boldsymbol{t}$ the operator $H_{\boldsymbol{t},\boldsymbol{\eta}}$ is of the form (4.4), with p, U_+ and $b_\pm$ independent of $\boldsymbol{t}$. Thus, by Theorem 4.1 there exists $E_0 > 0$ such that when conditioned on $\boldsymbol{t}$ with probability one, $H_{\boldsymbol{t},\boldsymbol{\eta}}$ either exhibits spectral localization or has no spectrum in $[0,E_0]$. However, the latter is excluded (almost surely, also with respect to the conditional probability) by (4.3) and Fubini. □

Remark 4.1. In addition to the spectral localization it is also of interest to establish the existence of uniform localization length, i.e., to prove that all eigenfunctions ϕ of $H_{\boldsymbol{\omega}}$ with eigenvalue in $[0, E_0]$ satisfy

$$\int_{|x-y|\leq\frac{1}{2}} |\phi(y)|^2 \, dy \leq C_\phi \, e^{-2|x|/\ell} \quad \text{for all } x \in \mathbb{R}^d . \tag{4.10}$$

This can be accomplished in the following two ways, for which the details are presented in [GK].

To establish uniform localization length under the hypotheses of Theorem 4.2 one may use the Bernoulli decomposition (4.6) *before* performing the multiscale analysis which is behind the proof of Theorem 4.1. The multiscale analysis is then executed for the random Schrödinger operator $H_{\boldsymbol{t},\boldsymbol{\eta}}$ in (4.7), in such a way that all events in the analysis are jointly measurable in $\boldsymbol{t}$ and $\boldsymbol{\eta}$.

An alternative proof of Theorem 4.2, which yields also uniform localization length, can be based on the concentration bound of Theorem 3.1. Namely, the Bourgain-Kenig proof can be extended to arbitrary single site probability distribution μ, with the probabilities of energy resonance estimated by the concentration bound instead of by Sperner's Lemma as in [BK] (see [GK]).

Acknowledgements

We thank the Oberwolfach center for hospitality at a meeting where the four-way collaboration started, and the Isaac Newton Institute where some of the work was done. We also thank B. Sudakov for an instructive review of the recent results in Sperner's theory, S. Molchanov for alerting us to relevant references, and M. Cranston for many helpful discussions of results in probability theory. This work was supported in parts by the NSF grants DMS-0602360 (MA), DMS-0457474 (AK) and DMS-0701181 (SW), and a Rothschild Fellowship at INI (MA).

References

[AE+] Aizenman, M., Elgart, A., Naboko, S., Schenker, J., Stolz, G.: Moment analysis for localization in random Schrödinger operators. Inv. Math. **163**, 343–413 (2006)

[AM] Aizenman, M., Molchanov, S.: *Localization at large disorder and at extreme energies: an elementary derivation.* Commun. Math. Phys. **157**, 245–278 (1993)

[An] Anderson, I.: *Combinatorics of finite sets.* Corrected reprint of the 1989 edition. Dover Publications Inc. Mineola, NY, 2002

[BK] Bourgain, J., Kenig, C.: *On localization in the continuous Anderson-Bernoulli model in higher dimension.* Invent. Math. **161**, 389–426 (2005).

[CHK] Combes, J.M., Hislop, P.D., Klopp, F.: *An optimal Wegner estimate and its application to the continuity of the integrated density of states for random Schrödinger operators*, preprint math-ph/0605030.

[DSS] Damanik, D., Sims, R., Stolz, G.: *Localization for one dimensional, continuum, Bernoulli-Anderson models.* Duke Math. J. **114**, 59–100 (2002).

[Do] Doeblin, W.: *Sur les sommes d'un grand nombre de variables aléatoires indpendantes*, Bull. Sci. Math. **63** 23–64 (1939).

[DoL] Doeblin, W, Lévy, P.: *Calcul des probabilités. Sur les sommes de variables aléatoires indépendantes à dispersions bornées inférieurement*, C.R. Acad. Sci. **202** 2027-2029 (1936).

[E] Engel, K.: *Sperner theory.* Encyclopedia of Mathematics and its Applications **65**. Cambridge University Press, Cambridge, 1997.

[Er] Erdös, P.: *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51**, 898–902 (1945).

[Es] Esseen, C. G.: *On the concentration function of a sum of independent random variables* Z. Wahrscheinlichkeitstheorie verw. Geb. **9**, 290-308 (1968).

[F] Feller, W.: *An introduction to probability theory and its applications.* Vol. I. 2nd ed. John Wiley and Sons, Inc., New York, 1957.

[FrS] Fröhlich, J., Spencer, T.: *Absence of diffusion in the Anderson tight binding model for large disorder or low energy.* Commun. Math. Phys. **88**, 151–184 (1983)

[GK] Germinet, F., Klein, A.: In preparation.

[GoMP] Gol'dsheid, Ya., Molchanov, S., Pastur, L.: *Pure point spectrum of stochastic one dimensional Schrödinger operators.* Funct. Anal. Appl. **11**, 1–10 (1977).

[K] Kesten, H.: *A sharper form of the Doeblin-Lévy-Kolmogorov-Rogozin inequality for concentration functions.* Math. Scand. **25**, 133-144 (1969).

[KiM] Kirsch, W., Martinelli, F.: *On the ergodic properties of the spectrum of general random operators.* J. Reine Angew. Math. **334**, 141-156 (1982).

[Kl] Klein, A.: Multiscale analysis and localization of random operators. In *Random Schrodinger operators: methods, results, and perspectives.* Panorama & Synthèse, Société Mathématique de France. To appear.

[Ko] Kolmogorov, A. N.: *Sur les propriétés des fonctions de concentration de M. P. Lévy.* Ann. Inst. H. Poincaré **16**, 27-34 (1958-60).

[KuS] Kunz, H., Souillard, B.: *Sur le spectre des operateurs aux differences finies aleatoires.* Commun. Math. Phys. **78**, 201–246 (1980).

[M] McDonald, D.: *A local limit theorem for large deviations of sums of independent, non-identically distributed random variables.* Ann. Probab. **7**, 526–531 (1979).

[P] Pastur, L.: *Spectral properties of disordered systems in the one-body approximation*, Commun. Math. Phys. **75**, 179–196 (1980).

[R1] Rogozin, B. A.: *An Estimate for Concentration Functions*, Theory Probab. Appl., **6**, 94-97 (1961) Russian original: Teoriya Veroyatnost. i Prilozhen. **6** 103–105 (1961).

[R2] Rogozin, B. A.: *An integral-type estimate for concentration functions of sums of independent random variables* Dokl. Akad. Nauk SSSR **211**, 1067-1070 (1973) [in Russian].

[R3] Rogozin, B. A.: Concentration function of a random variable. In *Encyclopaedia of Mathematics*, M. Hazewinkel (ed.), Springer 2002.

[S] Sperner, E.: *Ein Satz über Untermengen einer endlichen Menge.* Math. Zeit. **27**, 544–548 (1928).

[St] Stollmann, P.: *Caught by disorder: bound states in random media.* Boston: Birkhäusser, 2001.

[TV] Tao, T., Vu, V.: *On Random* ± 1 *Matrices: Singularity and Determinant.* Random Structures and Algorithms **28**, 1–23 (2006).

[V] Villani, C.: Topics in Optimal Transportation. AMS 2003.

(Aizenman) Princeton University, Departments of Mathematics and Physics, Princeton, NJ 08544, USA

E-mail address: aizenman@princeton.edu

(Germinet) Université de Cergy-Pontoise, Laboratoire AGM, UMR CNRS 8088, Département de Mathématiques, Site de Saint-Martin, 2 avenue Adolphe Chauvin, 95302 Cergy-Pontoise cedex, France

E-mail address: germinet@math.u-cergy.fr

(Klein) University of California, Irvine, Department of Mathematics, Irvine, CA 92697-3875, USA

E-mail address: aklein@uci.edu

(Warzel) Princeton University, Department of Mathematics, Princeton, NJ 08544, USA

E-mail address: swarzel@princeton.edu