NEEDLES IN EXPONENTIAL HAYSTACKS I

Notes of Joel Spencer
Cornell Probability School 2012

1 The Goal, and Discussion

Theorem 1.1 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$ with all $|\vec{r}_i|_\infty \leq 1$. Then there exists $\vec{x} = (x_1, \ldots, x_n) \in R^n$ with all $x_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot \vec{x}| \leq K\sqrt{n} \text{ for all } 1 \leq i \leq n \tag{1}$$

Here K is an absolute constant*, not dependent on n.*

`Remarks:` This problem originated with a question of Paul Erdős. What is the smallest $f(n)$ such that: Given n sets $S_1 \ldots, S_n \subseteq \{1, \ldots, n\}$ there exists a "two coloring" $\chi : \{1, \ldots, n\} \to \{-1, +1\}$ such that for *every* set S_i,

$$|\sum_{j \in S_i} \chi(j)| \leq f(n) \tag{2}$$

(The standard notation, which will not be needed here, is that the family $\mathcal{F} = \{S_1, \ldots, S_n\}$ has *discrepency* at most $f(n)$.) The result $f(n) \leq K_1\sqrt{n}$ (plus the more general Theorem 1.1) was first proven by this writer many years ago. To make this a special case of Theorem 1.1 simply let A be the $n \times n$ incidence matrix of points and sets, $a_{ij} = 1$ if $j \in S_i$, else $a_{ij} = 0$. Then $\vec{r}_i$ is the i-th row vector. A coloring χ corresponds to the vector $\vec{x} = (x_1, \ldots, x_n)$ with $x_j = \chi(j)$. In these notes, however, we will look at the more general setting.

2 Algorithms

In my original proof no algorithm was given. That is, given the $\vec{r}_1 \ldots, \vec{r}_n$ how do you *find in polynomial time* the $\vec{x}$. Indeed, I long conjectured that no such algorithm existed, that it would be equivalent to some other "hard" problems. However, a few years ago Nikhil Bansal gave an algorithmic proof using semidefinite programming. I will present here yet another approach, due to Shachar Lovett and Raghu Meka, both postdocs at IAS. Their approach uses a restricted Brownian motion, though to make it algorithmic one replaces the continuous process with a discrete walk.

3 Sets and Points

In this presentation the number of vectors and the number of dimensions is the same. However, this is only to facilitate the discussion. There are results for any number m of vectors in any dimensional space n. Indeed, as we go through the proof we find that we will be including more general results.

4 Erdos Magic

What about a randomly chosen $\vec{x}$? Let r_{ij} denote the j-th coordinate of $\vec{r}_i$ so that

$$\vec{x} \cdot \vec{r}_i = \sum_{j=1}^{n} x_j r_{ij} \tag{3}$$

Let the x_j be i.i.d. random variables, each $x_j = \pm 1$ uniformly. As all $|r_{ij}| \leq 1$ one can show (for any particular i)

$$\Pr[|\sum_{j=1}^{n} x_j r_{ij}| > K\sqrt{n}] < 2e^{-K^2/2} \tag{4}$$

(That is, the variable $X = \sum_{j=1}^{n} x_j r_{ij}$ has a tail bounded by a Gaussian.) Now select

$$K = [2\ln(2n)]^{1/2} \tag{5}$$

This makes the probability in (4) less than n^{-1}. For $1 \leq i \leq n$ let BAD_i denote the event that $|\vec{x} \cdot \vec{r}_i| > K\sqrt{n}$. All the BAD_i have probability less than n^{-1} so the probability of $\vee_i BAD_i$ is less than one. That means that with positive probability *none* of the BAD_i hold. Now comes the Erdős Magic – an event with positive probability must have an existence, there must (not maybe!) exist a $\vec{x}$ so that none of the BAD_i hold. This shows

Theorem 4.1 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$ with all $|\vec{r}_i|_\infty \leq 1$. Then there exists $\vec{x} = (x_1, \ldots, x_n) \in R^n$ with all $x_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot \vec{x}| \leq K\sqrt{n} \text{ for all } 1 \leq i \leq n \tag{6}$$

Here K is given by (5).

Of course, this does *not* imply Theorem 1.1 as K here is not an absolute constant. Indeed, it yields a viewpoint for the stronger Theorem 1.1. Pick K some large constant, say $K = 10$. Then with a random $\vec{x}$ the probability that $|\vec{r}_i \cdot \vec{x}| > K\sqrt{n}$ is very small. But there are n different $\vec{r}_i$ so there are very likely to be some *outliers*, some i with $|\vec{r}_i \cdot \vec{x}|$ bigger than $10\sqrt{n}$. The Lovett Meka approach will be randomized but in a way that removes the outliers.

5 Floating Colors

This idea goes way back but I like to feel that it originated with Jozsef Beck, at Rutgers. To each vertex j we attach a real variable x_j and we set $\vec{x} = (x_1, \ldots, x_n)$. We initialize by setting $\vec{x} \leftarrow \vec{0}$, that is, all $x_j \leftarrow 0$. Throughout the procedure we will always require that all $x_j \in [-1, +1]$. At the very very end all $x_j \in \{-1, +1\}$ and this will be our desired $\vec{x}$ for Theorem 1.1. We may conceptually think of moving $\vec{x}$ around continuously in $[-1, +1]^n$ until the coordinates reach the boundary. In the actual algorithm this is done in discrete steps.

There turns out to be nothing special with starting at $\vec{x} \leftarrow \vec{0}$. We shall actually show the stronger result:

Theorem 5.1 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$ with all $|\vec{r}_i|_\infty \leq 1$. Let $\vec{z} = (z_1, \ldots, z_n) \in R^n$ with all $z_j \in [-1, +1]$. Then there exists $\vec{x} = (x_1, \ldots, x_n) \in R^n$ with all $x_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \leq K\sqrt{n} \text{ for all } 1 \leq i \leq n \tag{7}$$

Here K is an absolute constant*, not dependent on n, nor on the initial $\vec{z}$.*

6 Mopping Up

Note to the Reader: You may well skip Mopping Up I,II,III at first reading and jump to the main parts.

6.1 Mopping Up I

In our actual process we will freeze x_j when $|x_j| > 1 - \epsilon$. Taking ϵ bigger doesn't affect the existence proof but it does quite greatly affect the speed of the algorithm. We shall set (constants not concerning us)

$$\epsilon = \frac{1}{10 \ln n} \tag{8}$$

We use a probability result:

Theorem 6.1 *Let x_j, $1 \leq j \leq n$, satisfy $1 - \epsilon < |x_j| < 1$. Let $\overline{x}_j$, $1 \leq j \leq n$ be independent random variables with $\Pr[\overline{x}_j = 1] = (1+x_j)/2$, $\Pr[\overline{x}_j = -1] = (1 - x_j)/2$. Let $|r_j| \leq 1$ for $1 \leq j \leq n$. Then*

$$\Pr[|\sum_{j=1}^{n} r_j(\overline{x_j} - x_j)| \geq 4\sqrt{n}] = o(n^{-1}) \tag{9}$$

We omit the proof of this large deviation result (don't worry about the 4, it is just some suitably large constant), but the idea is that $\overline{x_j} - x_j$ has mean zero and variance $O(\epsilon)$ so that $\sum_{j=1}^{n}(\overline{x_j} - x_j)$ has mean zero and variance $O(n\epsilon)$ and so, if it were Gaussian, we would be looking at the probability of it being $\Omega(\epsilon^{-1/2})$ standard deviation off the mean.

From this immediately follows:

Theorem 6.2 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$ with all $|\vec{r}_i|_\infty \leq 1$. Let x_j, $1 \leq j \leq n$, satisfy $1 - \epsilon < |x_j| < 1$ and let $\vec{x} = (x_1, \ldots, x_n)$. Then there exists $\vec{y} = (y_1, \ldots, y_n)$ with all $y_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot (\vec{y} - \vec{x})| \leq 4\sqrt{n} \text{ for all } 1 \leq i \leq n \tag{10}$$

Proof: We randomly select y_j with distribution $\overline{x_j}$. From Theorem 6.1 the probability that (refh) fails for any pariticular i is $o(n^{-1})$ and so with probability $1 - o(1)$ the y_j will satisfy Theorem 6.2. From Erdős Magic this implies the existence of the y_j. It also gives a randomized algorithm to find them.

6.2 Mopping Up II

As the process continues more and more of the x_j will become frozen. This part allows the algorithm to stop when the number of unfrozen variables becomes sufficiently small. Here we shall set (the 100 being a suitable large constant)

$$m = \frac{n}{100 \ln n} \tag{11}$$

Again we have a large deviation result:

Theorem 6.3 *Let $x_j \in [-1, +1]$, $1 \leq j \leq m$. Let $\overline{x}_j$, $1 \leq j \leq n$ be independent random variables with $\Pr[\overline{x}_j = 1] = (1 + x_j)/2$, $\Pr[\overline{x}_j = -1] = (1 - x_j)/2$. Let $|r_j| \leq 1$ for $1 \leq j \leq n$. Then*

$$\Pr[|\sum_{j=1}^{m} r_j(\overline{x_j} - x_j)| \geq \sqrt{n}] = o(n^{-1}) \tag{12}$$

As before, we outline the large deviation result. Now each $r_j(\overline{x_j} - x_j)$ has mean zero and variance at most one so that $\sum_{j=1}^{m} r_j(\overline{x_j} - x_j)$ has mean zero and variance at most m. If it were Gaussian then (12) would represent the probability of being at least $(n/m)^{1/2}$ standard deviations off the mean. As with Theorem 6.2 we continue with:

Theorem 6.4 *Let $\vec{r}_i \in R^m$, $1 \leq i \leq n$ with all $|\vec{r}_i|_\infty \leq 1$. Let $\vec{x} = (x_1, \ldots, x_m) \in [-1, +1]^m$ Then there exists $\vec{y} = (y_1, \ldots, y_m) \in R^m$ with all $y_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{y})| \leq \sqrt{n} \text{ for all } 1 \leq i \leq n \tag{13}$$

Proof: We randomly select y_j with distribution $\overline{x_j}$. From Theorem 6.3 the probability that (refh) fails for any pariticular i is $o(n^{-1})$ and so with probability $1 - o(1)$ the y_j will satisfy Theorem 6.2. From Erdős Magic this implies the existence of the y_j. It also gives a randomized algorithm to find them.

6.3 Mopping Up III:

We shall in the main part of the argument find a $\vec{x} \in [-1, +1]$ with $|x_j| > 1-\epsilon$ for all but at most m coordinates j. From Theorems 6.2, 6.4 we'll then be able to use standard randomized rounding to find $\vec{y} \in \{-1, +1\}$ such that all

$$|\vec{r_i} \cdot (\vec{x} - \vec{y})| \leq 5\sqrt{n} \tag{14}$$

This will only add 5 to the constant K in Theorem 1.1, so it would remain an absolute constant.

7 Phase I

In Phase I the algorithm starts with $\vec{x} \leftarrow \vec{z}$. At any stage we define

$$L_i = n^{-1/2}\vec{r_i} \cdot (\vec{x} - \vec{z}) \tag{15}$$

so that $L_i \leftarrow 0$ initially. We say that coordinate j is *frozen* if $|x_j| > 1 - \epsilon$. The other coordinates j are called *floating*. At each *step* there will be a (small) movement of $\vec{x}$ and a corresponding movement of the L_i. Once a coordinate is frozen it doesn't move. Phase I is concluded when the number of floating coordinates becomes at most $\frac{n}{2}$. It will be done in such a way that the final values of $|L_i|$ are *all* within a bounded range. All the main arguments appear in Phase I. For convenience we state the result:

Theorem 7.1 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$ with all $|\vec{r}_i|^2 \leq n$. Let $\vec{z} \in [-1, +1]^n$. Then there exists $\vec{x} = (x_1, \ldots, x_n) \in R^n$ with*

1. *All $x_j \in [-1, +1]$.*
2. *$|x_j| < 1 - \epsilon$ for at most $\frac{n}{2}$ values $1 \leq j \leq n$*

such that

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \leq 11\sqrt{n} \text{ for all } 1 \leq i \leq n \tag{16}$$

7.1 The Random Step

At the core of the argument, we now define a random step. We have a $\vec{x}$ and corresponding L_i. Let BIG denote those $\frac{n}{4}$ coordinates i with the biggest $|L_i|$. Let $SMALL$ denote the other i, $1 \leq i \leq n$. (In case of ties we could arbitrarily so that BIG is of this size. Actually, since we'll have variables with continuous distributions, ties will only occur at zero.) Now we define a vector space $V \subset R^n$. It consists of those $\vec{y} = (y_1, \ldots, y_n)$ with the following properties:

1. When x_j is frozen $y_j = 0$.
2. *Critical!!* When $i \in BIG$, $\vec{r}_i \cdot \vec{y} = 0$.
3. (A convenience) $\vec{y} \cdot \vec{x} = 0$

Let d denote the dimension of V so that $d \geq n - \frac{n}{2} - \frac{n}{4} - 1$. For simplicity we'll say $d \geq \frac{n}{4}$, the constants not mattering much. Let $\vec{g}$ be a Standard Gaussian on V, as defined below. Let δ be very small, as discussed later. The basic step is then

$$\vec{x} \leftarrow \vec{x} + \delta \vec{g} \tag{17}$$

7.2 Standard Gaussian

Let $V \subset R^n$ be a subspace of dimension d. Let $\vec{b}_1, \ldots, \vec{b}_d$ be an orthonormal basis. Then by Standard Gaussian we mean

$$\vec{g} = \frac{1}{\sqrt{d}}[n_1 \vec{b}_1 + \ldots n_d \vec{b}_d] \tag{18}$$

where the $n_1, \ldots, n_d$ are independent standard Gaussians. This distribution does not depend on the choice of orthonormal basis. The normalization gives that $E[|\vec{g}|^2] = 1$. The distribution is directionless. For any $\vec{v} \in V$, $\vec{v} \cdot \vec{g}$ has a Gaussian distribution with mean zero and standard deviation $d^{-1/2}|\vec{v}|$. For any $\vec{w} \in R^n$ let $\vec{v}$ be its projection onto V. (That is, write $\vec{w} = \vec{v} + \vec{z}$ with $\vec{v} \in V$, $\vec{z} \in V^{\perp}$.) Then $\vec{w} \cdot \vec{g}$ has the same distribution as $\vec{v} \cdot \vec{g}$ and so is Gaussian with mean zero and standard deviation at most $d^{-1/2}|\vec{w}|$.

7.3 Progress

Property (3) and Pythagorus gives that $|\vec{x}|^2$ increases at each step by $\delta^2 |\vec{g}|^2$. The chi-square distribution of $|\vec{g}|^2$ is tightly concentrated about 1 so we will hand wave and think of $|\vec{x}|^2$ as increasing by δ^2 each step. The initial $|\vec{x}|^2 = |\vec{z}|^2 \geq 0$ and at the end of Phase I we must have $|\vec{x}|^2 \leq n$ so the number of steps will be bounded (with very high probability) by

$$T^{max} \sim 1.01 n \delta^{-2} \tag{19}$$

7.4 Brownian Motion

If we think of $\delta \to 0$ we find $\vec{x}$ moving with a Brownian Motion, movements restricted to V. Simultaneously, the L_i will be making a 1-dimensional Brownian Motion. At various times a coordinate will become frozen or the order of the $|L_i|$ will change so that BIG changes and V changes so $\vec{x}$ moves with a shifting Brownian motion.

7.5 Making Small Steps Big

(If you don't care about the speed of the algorithm simply take $\delta = n^{-50}$ or some such and skip this section.) The movement (17) depends on the choice of δ. The bigger δ is the bigger the steps and so the faster the algorithm will run. The limiting factor is the possibility that some x_j will jump over the barrier at ± 1. Let $\vec{w}_j$ be the unit vector with one in the j-th position and $\vec{v}_j$ its projection onto V so that the change in x_j is $\delta \cdot |\vec{v}_j| d^{-1/2} N$ where N is standard normal. The total number of steps (indeed, through the entire process) is going to be at least $\Omega(n)$ (just in one step) and will be $O(n^{10})$ even

by the roughest estimates so lets set δ so that the chance of jumping over the barrier is $o(n^{-10})$. (As is usually the case, large deviation bounds are exponential so the difference between making it $o(n^{-1})$ and $o(n^{-10})$ is only a constant factor which, at least here, we don't care about.) To jump over the boundary we would need $|\delta d^{-1/2} N| \geq \epsilon$ which would require $|N| \geq d^{1/2}\epsilon\delta^{-1}$. By making the right hand side at least, say, $10\sqrt{\ln n}$ the probability becomes $O(n^{-50})$. As $d \geq \frac{n}{4}$ we then just need $\frac{1}{2}\sqrt{n}\frac{1}{100\ln n}\delta^{-1} \geq 10\sqrt{\ln n}$ so we can take $\delta = K\sqrt{n}\ln^{-3/2} n$. Then (19) becomes that the number of steps is $O(\ln^3 n)$, only polylog! Indeed, the time for the entire algorithm (assuming unit time for generating random standard normals) is only n times a polylog factor.

7.6 The Dot Products Take a Walk

Now lets fix i, $1 \leq i \leq n$, and look at the progress of $L_i := n^{-1/2}\vec{x}\cdot\vec{r}_i$ during the first phase. We use only that the Euclidean Norm $|\vec{r}_i| \leq \sqrt{n}$. Setting $\vec{u} = n^{-1/2}\vec{r}_i$ for convenience we now have $|\vec{u}| \leq 1$ and $L_i = \vec{x}\cdot\vec{u}$. OK, so when $\vec{x} \leftarrow \vec{x} + \delta\vec{g}$ we get

$$L_i \leftarrow L_i + \delta\vec{g}\cdot\vec{u} \tag{20}$$

We know that the change in L_i will have a Gaussian Distribution with mean zero. We don't know the standard deviation as it depends on the size of the projection of $\vec{u}$ onto V and V depends on all sorts of history. But it won't matter – we have a martingale! That is, let $L_i(s)$, $0 \leq s \leq T^{max}$ be the value of L_i at the end of step s. Here we define $L_i(0) = 0$. Further, if phase I ends before step T^{max} we simply keep $L_i(s)$ constant (a nice artifact when you don't know a priori how long the martingale will go) until $s = T^{max}$. At each stage $L_i(s+1) - L(s)$, conditional on the state of the system at time s, is Gaussian with mean zero and standard deviation $\delta\tau$ with $\tau \leq d^{-1/2}$. (This includes $\tau = 0$ when Phase I has actually ended.) Since we define Phase I to end when d slips below $\frac{n}{4}$ we have $\tau \leq 2n^{-1/2}$. (`Remark:` The factor 2 reflects the additional conditions placed on the walk. The more conditions placed the more variance (at least, as an upper bound) in the walk. There is a tradeoff here which comes from selecting the size of BIG and also in determining when Phase I is to end. In this presentation however, we do not attempt to optimize that tradeoff.) We have a general large deviation martingale result:

Theorem 7.2 *Let $0 = Z_0, Z_1, \ldots, Z_T = Z$ be a martingale. Set $\Delta_s = Z_s - Z_{s-1}$. Assume every Δ_s is a Gaussian with zero mean and variance at most σ^2. Let $a > 0$.*

$$\Pr[Z > a(T\sigma^2)^{1/2}] < e^{-a^2/2} \tag{21}$$

and

$$\Pr[|Z| > a(T\sigma^2)^{1/2}] < 2e^{-a^2/2} \tag{22}$$

Further, let $MAXZ$ denote the maximum $|Z_s|$, $0 \leq s \leq T$. Then

$$\Pr[MAXZ > a(T\sigma^2)^{1/2}] < 2e^{-a^2/2} \tag{23}$$

(`Remark:` This is quite natural if you think of Z in the worst case as being Gaussian with variance $T\sigma^2$.) By symmetry the bound on the upper tail of Z_T can be applied to the lower bound, so (22) will follow directly from (21). Now (23) follows by a standard trick. Define a new walk Z_s^* which is the same as Z_s except that if some first $|Z_s| > a(T\sigma^2)^{1/2}$ the walk Z_s^* stays where it is. This is still a martingale with the same properties and so (22) applies to it. But then $MAXZ > a(T\sigma^2)^{1/2}$ if and only if $|Z_T^*| > a(T\sigma^2)^{1/2}$. Now we attack (21) via Chernoff bounds. The conditional $E[e^{\lambda\Delta_s}]$ is $\exp[\lambda^2\tau^2/2]$ where τ^2 is the variance of Δ_s and so is at most $\exp[\lambda^2\sigma^2/2]$. Thus $E[e^{\lambda Z}] \leq \exp[\lambda^2(T\sigma^2)/2]$ and

$$\Pr[Z > a(T\sigma^2)^{1/2}] = \Pr[e^{\lambda Z} > e^{\lambda a(T\sigma^2)^{1/2}}] < e^{\lambda^2(T\sigma^2)/2 - a\lambda(T\sigma^2)^{1/2}} \qquad (24)$$

from which Theorem 7.2 follows by setting $\lambda = a(T\sigma^2)^{-1/2}$.

In our case T is given by (19) and we take $\sigma^2 = 4n^{-1}\delta^2$. Thus in calculating $T\sigma^2$ the factors of δ cancel out, which is natural as we were thinking of a discrete approximation to Brownian motion. Also the factors of n cancel out, which reflects our (well, Lovett and Meka's!) perspecacious scaling in defining the L_i. So $T\sigma^2 = 4.04$. The probability that the walk ever reaches, say, 10 is pretty small. Theorem 7.2 gives

$$\Pr[|L_i| > 10 \text{ at any time during Phase I }] < 0.1 \qquad (25)$$

Actually, this doesn't look so great. After all, the whole point was to eliminate the outliers and here we may have outliers. But we are ready for the brilliancy.

7.7 The Brilliancy

We cannot (perhaps more accurately, we do not) say anything about the dependency amongst the L_i. By linearity of Expectation

$$E[|\{i : 1 \leq i \leq n, |L_i| > 10 \text{ at any time during Phase I }\}|] < 0.1n \qquad (26)$$

This number is nonnegative. Thus it is less than $0.2n$ at least half the time. Our randomized algorithm (for Phase I) is to run as described. It a failure if more than $0.2n$ of the $|L_i|$ every reach 10. It is also a failure if the process doesn't stop in at most T^{max} steps or if at some time the barrier at $[-1, +1]$ is jumped over, or if some L_i moves by more than one in one step, all of which have $o(1)$ probability. Otherwise it is a success. So we have positive (bounded from zero, independent of n) chance of success. But, you still say, so what???

We go back to (2) in the definition of V, the conditions on the walk. When $i \in BIG$, V is perpendicular to $\vec{r}_i$ and therefore the value L_i does not change. Suppose we had success in Phase I as defined above. At most $\frac{n}{5}$ of the $|L_i|$ ever reached 10. But at every step we are freezing BIG, the top $\frac{n}{4}$ values. That means that every L_i for which $|L_i|$ ever reached 10 was frozen for the remainder of Phase I. But a single step is at most one, indeed $o(1)$. So when $|L_i|$ reached 10 it didn't go past 11 in that one step and then it was frozen. From which we conclude: *None* of the $|L_i|$ *ever* reach 11!! This concludes Theorem 7.1, the heart of the result.

8 The General Argument

We work the algorithm in Phases. The $\vec{r}_i \in R^n$, $1 \le i \le n$ are fixed throughout. In Phase s we begin with a $\vec{z} \in R^n$ with at most $n2^{1-s}$ coordinates j floating and end with a $\vec{x} \in R^n$ with at most $n2^{-s}$ coordinates j floating. Each L_i has then moved $n^{-1/2}\vec{r}_i \cdot (\vec{x} - \vec{z})$. From our discussions in section 6 we can restrict attention to $1 \le s \le K \ln\ln n$. Further, we can set $m = 2^{1-s}$ and consider only m coordinates (the floating ones and, for convenience, perhaps some others), the other coordinates not moving. Restricting to these m coordinates the $\vec{r}_i$ now have $|\vec{r}_i|^2 \le m$. (Interestingly, this is the only point in the proof where $|\vec{r}_i|_\infty \le 1$ is used.) We now generalize Theorem (7.1) as follows.

Theorem 8.1 *Let $m = 2^{1-s}$ with $1 \le s \le 10 \ln\ln n$. Let $\vec{r}_i \in R^m$, $1 \le i \le n$ with all $|\vec{r}_i|^2 \le m$. Let $\vec{z} = (z_1, \ldots, z_m)$ with all $z_j \in [-1, +1]$. Then there exists $\vec{x} = (x_1, \ldots, x_m) \in R^m$ with*

1. *All $x_j \in [-1, +1]$.*
2. *$|x_j| < 1 - \epsilon$ for at most $\frac{m}{2}$ values $1 \le j \le m$*

such that

$$|\vec{r}_i \cdot (\vec{z} - \vec{x})| \le K_s\sqrt{m} \text{ for all } 1 \le i \le n \tag{27}$$

where K_s satisfies (30) below.

We outline the argument, emphasizing the distinctions with Theorem 7.1. We start with $\vec{x} = \vec{z}$ and we move it in steps until at most $\frac{m}{2}$ coordinates are floating. For convenience we redefine

$$L_i = m^{-1/2}\vec{r}_i \cdot (\vec{z} - \vec{x}) \tag{28}$$

Thus L_i starts at 0 and we want $|L_i| \le K_s$ at the termination of Phase s. Observe the Theorem 8.1 is now Theorem 7.1 with an arbitrary starting point and n replaced by m *except* that the number of vectors $\vec{r}_i$ is not m by $m2^{s-1}$. Again, we give a single step, moving a small distance from $\vec{x}$. We let BIG denote those $\frac{m}{4}$ coordinates i with the largest $|L_i|$. We define a vector space $V \subset R^m$. It consists of those $\vec{y} = (y_1, \ldots, y_m)$ with the same Properties (1,2,3) as before. Let d denote the dimension of V so that $d \ge m - \frac{m}{2} - \frac{n}{4} - 1$. For simplicity we'll say $d \ge \frac{m}{4}$, the constants not mattering much. The walk for any single L_i follows the identical argument, with n replaced by m. The distinction comes at (25) where now we select K_s so that

$$\Pr[|L_i| > K_s \text{ at any time during Phase } s\ ] < 0.1\frac{m}{n} \tag{29}$$

Now the expected number of i such that $|L_i| > K_s$ at any time during Phase s is less than $0.1m$. This is less than $0.2m$ at least half the time. When this occurs, and the $o(1)$ probability failures do not occur, the algorithm ends with $|L_i| \le K_s + 1$ for *all* i.

It remains to check what the cost of replacing (25) by the more stringent (29) is. As in the earlier calculation we have $T\sigma^2 = 4.04$. So for Theorem 7.2 to yield (29) we need

$$2e^{-K_s^2/2(4.04)} \leq 0.1 \cdot 2^{1-s} \tag{30}$$

For such K_s we then have Theorem 8.1.

Indeed, this is a cost, as $K_s = \Theta(\sqrt{s})$. However, it is dwarfed by the gain we have achieved by replacing $\sqrt{n}$ with $\sqrt{m} = \sqrt{n}2^{(1-s)/2}$. When we reexpress Theorem 8.1 is terms of $\sqrt{n}$, and consider the movement of the original $L_i = n^{-1/2}\vec{r}_i \cdot (\vec{x} - \vec{z})$, we see that in Phase s all of the L_i move by at most $K_s 2^{(1-s)/2} = \Theta(\sqrt{s}2^{-s/2}$.

Now the Phases mesh, the final $\vec{x}$ from Phase $s-1$ becomes the initial $\vec{x}$ in Phase s. For each $1 \leq i \leq n$, consider the movement of L_i during the phases. The final $|L_i|$ is at most the sum of the absolute values of its movements during the phases. This sum is bounded by $\sum_s K_s 2^{(1-s)/2}$ which is a convergent sum, approaching some absolute constant K. It is for this K that our goal Theorem 5.1 applies.

9 References

The original paper with this result was:

J. Spencer, Six Standard Deviations Suffice
Trans. Amer. Math. Soc., 289 (1985), 679-706.

This is presented in my book

The Probabilistic Method
John Wiley (1992), 2nd ed. (2000), 3rd ed. (2008) (with N. Alon)

The argument of Nikhil Bansal, using semidefinite programming, can be found on ArXiv

ttp://arxiv.org/abs/1002.2259

The Lovett-Meka approach, basically what we have described here, can also be found on ArXiV

http://arxiv.org/abs/1203.5747