DISCREPANCY THEOREMS AND REPRESENTING DIAGONAL MATRICES

D. H. J. POLYMATH

1. Introduction

Let X be a finite set and let $\mathcal{A}$ be a collection of subsets of X. For every red/blue colouring κ of X, the $\mathcal{A}$-*discrepancy* of κ is defined to be the maximum over all $A \in \mathcal{A}$ of the difference between the number of red points in A and the number of blue points in A.There are many important problems of the following kind: given a set-system $\mathcal{A}$, obtain bounds for the smallest $\mathcal{A}$-discrepancy of any red/blue colouring of the ground set X.

One of the best-known results of this kind is due to Roth. It deals with the case where X is the set $\{1, 2, \dots, n\}$ and $\mathcal{A}$ is the set of all subsets of X that are arithmetic progressions. The smallest possible discrepancy was shown by Roth to be at least $cn^{1/4}$, and this bound was proved to be tight by Matousek and Spencer.

Later, Lovász gave a different proof of Roth's lower bound using semidefinite programming. The purpose of this paper is to give a third proof, related to Lovász's proof but not quite the same.

To explain the relationship, we need to say a little bit more about the semidefinite programming approach. So let us begin by stating a rather abstract criterion that gives a lower bound for discrepancy. The first step is to reformulate the problem analytically. A simple reformulation is to replace the red/blue colouring by a function f from X to $\{-1, 1\}$. Then the quantity we are trying to minimize is

$$\mathrm{disc}_{\mathcal{A}}(f) = \max_{A \in \mathcal{A}} |\sum_{x \in A} f(x)|.$$

Once we have reformulated the problem in this way, it is natural to seek a lower bound that is valid for all functions f, and not just ± 1-valued functions. Of course, the lower bound will have to depend on f, so a natural idea is to try to prove a bound of the form $\mathrm{disc}_{\mathcal{A}}(f) \geq \|f\|$ for some suitably chosen norm. Since eventually we would like a uniform bound for all ± 1-valued functions, it is also natural if all such functions have the same norm.

It is also natural to try to prove the result by means of some kind of averaging argument, and to use a Euclidean norm for the lower bound. Moreover, one would like to replace the modulus by a modulus squared. With all these thoughts in mind, it is natural to try to find non-negative coefficients λ_A and weights $b(x)$ and to seek a bound of the form

$$\sum_{A\in\mathcal{A}}\lambda_A|\sum_{x\in A}f(x)|^2\geq\sum_{x\in X}b(x)|f(x)|^2.$$

If we have such a bound and f is ± 1-valued, then

$$\sum_{A\in\mathcal{A}}\lambda_A|\sum_{x\in A}f(x)|^2\geq\sum_{x\in X}b(x),$$

which implies by averaging that $\mathrm{disc}_{\mathcal{A}}(f)^2$ is at least $\sum_{x\in X}b(x)/\sum_{A\in\mathcal{A}}\lambda_A$.

The basic idea behind the semidefinite programming method is to observe that

$$Q_\lambda : f\mapsto\sum_{A\in\mathcal{A}}\lambda_A|\sum_{x\in A}f(x)|^2$$

is a positive semidefinite quadratic form on $\mathbb{R}^X$, and that the desired inequality will be true if the quadratic form

$$Q_{\lambda,b} : f\mapsto\sum_{A\in\mathcal{A}}\lambda_A|\sum_{x\in A}f(x)|^2-\sum_{x\in X}b(x)|f(x)|^2$$

is also positive semidefinite. So for a given choice of coefficients λ_A our problem is reduced to a *semidefinite programming problem*, or SDP for short: that is, we would like to maximize the sum $\sum_x b(x)$ subject to the positive semidefiniteness of the quadratic form $Q_{\lambda,b}$. A great deal is known about SDPs. In particular, there are efficient algorithms for solving them, a fact that has many ramifications in extremal combinatorics and theoretical computer science.

It is straightforward to show that if $Q_{\lambda,b}$ is positive semidefinite, then we can extend the discrepancy result to vector-valued functions. More precisely, if each $f(x)$ is a vector in some Euclidean space H, then we have the lower bound

$$\sum_{A\in\mathcal{A}}\lambda_A\|\sum_{x\in A}f(x)\|^2\geq\sum_{x\in X}b(x)\|f(x)\|^2.$$

In particular, it follows that if the $f(x)$ are unit vectors, then $\mathrm{disc}_{\mathcal{A}}(f)$ (now defined to be the maximum of $\|\sum_{x\in A}f(x)\|$ over all sets $A\in\mathcal{A}$) is again at least the square root of $\sum_{x\in X}b(x)/\sum_{A\in\mathcal{A}}\lambda_A$. Interestingly, the converse is true as well: if $\mathrm{disc}_{\mathcal{A}}(f)$ is at least t whenever each $f(x)$ is a unit vector in some Hilbert space H, then we can find coefficients

λ_A and $b(x)$ such that $\sum_{x\in X} b(x)/\sum_{A\in\mathcal{A}}\lambda_A = t^2$ and the quadratic form $Q_{\lambda,b}$ is positive semidefinite.

The approach taken in this paper is different, but it is also based on proving an inequality of the form

$$\sum_{A\in\mathcal{A}}\lambda_A|\sum_{x\in A} f(x)|^2 \geq \sum_{x\in X} b(x)|f(x)|^2.$$

This we do as follows. For each set $A\in\mathcal{A}$, let us also write A for its characteristic function. That is, $A(x)=1$ if $x\in A$ and 0 otherwise. If $u,v:X\to\mathbb{C}$, then we write $u\otimes v$ for the matrix with xy entry equal to $u(x)v(y)$. Suppose we can find coefficients λ_{AB} such that $\sum_{A,B\in\mathcal{A}}\lambda_{AB}A\otimes B = D$, where D is a diagonal matrix with $D(x,x)=b(x)$ for every $x\in X$. Then by averaging

$$\sum_x b(x)|f(x)|^2 = \langle Df, f\rangle = \sum_{A,B\in\mathcal{A}}\lambda_{AB}\sum_{x\in A} f(x)\sum_{y\in B} f(y).$$

Therefore, there exist sets $A,B\in\mathcal{A}$ such that

$$|\sum_{x\in A} f(x)\sum_{y\in B} f(y)| \geq \frac{\sum_x b(x)}{\sum_{A,B\in\mathcal{A}}|\lambda_{AB}|},$$

from which it follows that $\mathrm{disc}_{\mathcal{A}}(f)^2 \geq \sum_x b(x)/\sum_{A,B\in\mathcal{A}}|\lambda_{AB}|$.

To convert this observation into a proof, one must find an efficient way of representing a diagonal matrix as a linear combination of rank-1 matrices $A\otimes B$, where A and B belong to the set $\mathcal{A}$. In this paper, we shall show how to do so when $X=\{1,2,\dots,n\}$ and $\mathcal{A}$ is the set of all subsets of X that are arithmetic progressions.

We remark that it is possible to show that this approach works and gives a lower bound of t if and only if the following stronger discrepancy-type statement holds. Let f and g be two functions from X to a Euclidean space H, and suppose that $\langle f(x), g(x)\rangle = 1$ for every $x\in X$. Then there exist $A,B\in\mathcal{A}$ such that $\langle\sum_{x\in A} f(x), \sum_{y\in B} g(y)\rangle \geq t^2$. Note also that if $g=f$ then we recover the statement that was equivalent to the success of the SDP approach, so this approach is less likely to work. However, if it works, then it gives a stronger result, and searching for an efficient representation of a diagonal matrix may be easier than proving that a certain matrix is positive semidefinite.

2. Roth's discrepancy theorem

Now let us specialize to the case that will interest us in this paper. That is, let $X=\{1,2,\dots,n\}$ and let $\mathcal{A}$ be the set of arithmetic subprogressions of X. We shall let D be

the identity matrix: that is, we shall take $b(x) = 1$ for every x. Also, we shall identify X with $\mathbb{Z}_N$.

For each $r \in \mathbb{Z}_N$ let ω_r be the function $\omega_r(x) = e^{2\pi irx/N} = \omega^{rx}$, where $\omega = e^{2\pi i/N}$. Our starting point is the following very simple lemma. We write $\mathbb{E}_r$ to stand for $N^{-1}\sum_r$.

Lemma 2.1. $I_N = \mathbb{E}_r \omega_r \otimes \overline{\omega_r}$.

Proof. Let us evaluate the right-hand side at (x, y). We obtain

$$\mathbb{E}_r \omega^{rx}\omega^{-ry} = \mathbb{E}_r \omega^r(x-y) = \delta_{xy},$$

which proves the lemma. □

That was not a completely sensible proof: it is straightforward to show that $I_n = \mathbb{E}_r u_r \otimes \overline{u_r}$ for any orthonormal basis (u_r) (with respect to the inner product $\langle f, g\rangle = \mathbb{E}_x f(x)\overline{g(x)}$).

Next, we show that each function ω_r can be decomposed as a linear combination of arithmetic progressions with small common difference. For this we need a standard lemma.

Lemma 2.2. *Let $0 < m \leq N$ and let $\alpha \in \mathbb{R}$. Then there exists $0 < d \leq m$ such that $\|\alpha d\| \leq m^{-1}$.*

For each r let us fix a d that satisfies the conclusion of the lemma with $\alpha = r/N$. From the lemma it follows that for every r we can find $0 < d \leq m$ such that $|1 - \omega^{rd}| \leq 2\pi/m$. Let us now define P to be the interval $[-m/4\pi, m/4\pi]$ (meaning the set of all residues in $\mathbb{Z}_N$ that lie in this interval when they are considered as real numbers). For each $0 < d \leq m$ let P_d be the mod-N progression dP. Note that the diameter of P_d is at most $m^2/2\pi$: this will be important to us later. Let π_d be the characteristic measure of P_d. That is, $\pi_d(x) = N/|P_d|$ when $x \in P_d$ and 0 otherwise.

In the next lemma, we define the convolution of two functions f and g by the formula

$$f * g(x) = \mathbb{E}_{y+z=x} f(y)g(z).$$

Lemma 2.3. *Suppose that $|1 - \omega^{rd}| \leq 2\pi/m$. Then $\omega_r * \pi_d = \lambda_{r,d}\omega_r$ for some real number $\lambda_{r,d}$ that lies between 1/2 and 1.*

Proof. Let us evaluate the left-hand side at x. We have

$$\begin{aligned} \omega_r * \pi_d(x) &= \mathbb{E}_y \omega^{r(x-y)}\pi_d(y) \\ &= \omega^{rx}\mathbb{E}_{y\in P_d}\omega^{-ry}. \end{aligned}$$

Since P_d is symmetric, $\lambda_{r,d} = \mathbb{E}_{y\in P_d}\omega^{-ry}$ is real. Note also that for every $y \in P_d$ we have $y = sd$ for some s of modulus at most $m/4\pi$, from which it follows that $|1-\omega^{-ry}| \leq (m/4\pi)|1-\omega^{-rd}| \leq 1/2$. Therefore, $\lambda_{r,d} \geq 1/2$ as claimed. It is trivial that $\lambda_{r,d} \leq 1$ since $\|\pi_d\|_1 = 1$. □

Writing $\mu_{r,d}$ for $\lambda_{r,d}^{-1}$, we may rewrite ω_r as $\mu_{r,d}\omega_r * \pi_d$. Now $\omega_r * \pi_d(x)$ can also be written as $\mathbb{E}_y\omega^{ry}\pi_d(x-y)$. Writing $\pi_{d,y}(x)$ for $\pi_d(x-y)$, we can write this as $\mathbb{E}_y\omega^{ry}\pi_{d,y}(x)$. That is, $\omega_r * \pi_d = \mathbb{E}_y\omega^{ry}\pi_{d,y}$, so $\omega_r = \mu_{r,d}\mathbb{E}_y\omega^{ry}\pi_{d,y}$. Note that $\pi_{d,y}$ is the characteristic measure of the arithmetic progression $P_{d,y} = P_d + y$, so we can also write this as $\mu_{r,d}(N/|P|)\mathbb{E}_y\omega^{ry}P_{d,y}$.

Since, $\omega_r = \mu_{r,d}(N/|P|)\mathbb{E}_y\omega^{ry}P_{d,y}$ for each r, and $I_n = \mathbb{E}_r\omega_r \otimes \overline{\omega_r}$, this gives us the decomposition

$$I_n = \frac{N^2}{|P|^2}\mathbb{E}_r\mu_{r,d}^2\mathbb{E}_{y,z}\omega^{r(y-z)}P_{d,y}\otimes P_{d,z}.$$

Our next task is to show that this decomposition is efficient.

To do this, let us fix d and work out the sum over y and z of the absolute values of the coefficients of $P_{d,y}\otimes P_{d,z}$ in the decomposition. Recall that $d = d(r)$: for each d with $0 < d \leq m$ let us write $R(d)$ for the set of r such that $d(r) = d$. Since always choose d such that $\|dr/N\| \leq m^{-1}$, it follows that $|R(d)| \leq 3N/m$ for every d.

First we rewrite the decomposition itself as

$$I_n = \frac{1}{N|P|^2}\sum_r \mu_{r,d(r)}^2 \sum_{y,z}\omega^{r(y-z)}P_{d(r),y}\otimes P_{d(r),z}.$$

Now it becomes easier to see that the coefficient of $P_{d,y}\otimes P_{d,z}$ is

$$\frac{1}{N|P|^2}\sum_{r\in R(d)}\mu_{r,d}^2\omega^{r(y-z)},$$

so the sum in question is

$$\frac{N}{|P|^2}\mathbb{E}_{y,z}|\sum_{r\in R(d)}\mu_{r,d}^2\omega^{r(y-z)}|,$$

Let $\nu(r) = \mu_{r,d}^2$ if $r \in R(d)$ and 0 otherwise. Then this expression is equal to

$$\frac{N^2}{|P|^2}\mathbb{E}_{y,z}|\mathbb{E}_r\nu(r)\omega^{r(y-z)}| = \frac{N^2}{|P|^2}\mathbb{E}_y|\mathbb{E}_r\nu(r)\omega^{ry}|.$$

Now

$$\begin{aligned}(\mathbb{E}_y|\mathbb{E}_r\nu(r)\omega^{ry}|)^2 &= N^{-2}(\sum_y|\mathbb{E}_r\nu(r)\omega^{ry}|)^2\\ &= N^{-2}\|\hat{\nu}\}_1^2\\ &\leq N^{-1}\|\hat{\nu}\|_2^2\\ &= N^{-1}\|\nu\|_2^2\\ &\leq 4N^{-2}|R(d)|.\end{aligned}$$

Therefore, since $|P| \geq m/16$ and $|R(d)| \leq 3N/m$, the coefficients have absolute values summing to at most $256(N/m)^2.2N^{-1/2}m^{-1/2} = 512N^{3/2}m^{-5/2}$. Since the set of possible d has size m, the sum of the absolute values of *all* coefficients is at most $512(N/m)^{3/2}$.

This gives us a discrepancy bound of $2^{-9/2}m^{3/4}N^{-1/4}$ if we take $\mathcal{A}$ to be the set of all arithmetic progressions of the form $P_{d,y}$. However, although these are arithmetic progressions in the $\mathbb{Z}_N$ sense, they do not necessarily become arithmetic progressions when we identify $\mathbb{Z}_N$ with $\{1, 2, \ldots, N\}$.

This is where the diameter estimate comes in. Since the diameter of P_d is at most $m^2/2\pi$, we know that each $P_{d,y}$ can wrap around at most once, provided that $m \leq \sqrt{N}$. If it wraps around once, then we can split it into two genuine arithmetic progressions, and at least half the discrepancy must occur on one of them. Thus, by taking $m = \sqrt{N}$ we obtain a bound of size $cN^{3/8}N^{-1/4} = cN^{1/8}$.

3. Improving the bound

Our next aim is to get from a bound of $cN^{1/8}$ to a bound of $cN^{1/4}$. We begin by giving the basic thought behind the improvement.

In the argument above, we bounded $\|\hat{\nu}\|_1$ above by $N^{1/2}\|\hat{\nu}\|_2$. Now this bound is sharp only if $\hat{\nu}$ is evenly spread throughout $\mathbb{Z}_N$, but that is not the case. Speaking rather roughly, ν is a bit like the characteristic function of $R(d)$, which is a bit like an arithmetic progression of length N/m, which means that $\hat{\nu}$ should be a bit like an arithmetic progression of length m, which means that $\|\hat{\nu}\|_1/\|\hat{\nu}\|_2$ should be more like $m^{1/2}$ than $N^{1/2}$. So if we can make these thoughts precise, then it is reasonable to aspire to an improvement by a factor $(m/N)^{1/2}$ to the sum of the coefficients in the decomposition, which would improve the

discrepancy bound by a factor of $(N/m)^{1/4}$, taking it from $cm^{3/4}N^{-1/4}$ to $cm^{1/2}$. Encouragingly, this is exactly the Roth bound when m is at its maximum of $c\sqrt{n}$, and it gives us at least some information for all smaller (but non-constant) values of m.

A standard trick that will enable us to avoid unwanted logarithmic factors is to work not with characteristic functions of arithmetic progressions but with similar functions that have absolutely summable Fourier coefficients. The following lemma gives us the main technical fact that we need.

Lemma 3.1. *Let $m \leq N/2$ and let $f_{d,m} : \mathbb{Z}_N \to \mathbb{C}$ be defined by the formula $f_{d,m}(dx) = \max\{1 - |x|/m, 0\}$ for every x with $|x| \leq N/2$. Then $\|\hat{f}_{d,m}\|_1 = 1$.*

Proof. Let P be the mod-N arithmetic progression $\{d, 2d, \dots, md\}$, and also the characteristic function of that arithmetic progression, and similarly for $-P$. Then

$$P*(-P)(xd) = \mathbb{E}_y P(y)P(y-xd) = N^{-1}|P\cap(P+xd)| = N^{-1}\max\{m-|x|, 0\} = N^{-1}mf_{d,m}(dx).$$

We also know that

$$\|P * (-P)\|_1 = \sum_r \hat{P}(r)\overline{\hat{P}(r)} = \|\hat{P}\|_2^2 = \|P\|_2^2 = N^{-1}m.$$

It follows that $\|\hat{f}_{d,m}\|_1 = 1$, as claimed. □

Note that $\hat{f}$ is a non-negative real-valued function, since $\hat{f}(r) = |\hat{P}(r)|^2$ for every r. This fact will also be useful to us.

The next lemma is a variant of Lemma 2.3.

Lemma 3.2. *Let $d, r \in \mathbb{Z}_N$. Then $\omega_r * \pi_d * \pi_d = \mu_{r,d}\omega_r$ for some non-negative real number $\mu_{r,d}$. Moreover, if $|1 - \omega^{rd}| \leq 2\pi/m$, then $1/4 \leq \mu_{r,d} \leq 1$.*

Proof. The second part of the lemma follows directly from Lemma 2.3, applied twice. In general, we know in Lemma 2.3 that $\lambda_{r,d}$ is real, so the positivity of $\mu_{r,d}$ follows from the fact that $\mu_{r,d} = \lambda_{r,d}^2$. □

The next lemma is a simple consequence of Lemmas 2.2 and 3.2. For the next few lemmas, all sums over d will be from $d = 1$ to $d = m$.

Lemma 3.3. *Let $s = \lceil N/m \rceil$, and for each d with $0 < d \leq m$ and for each $r \in \mathbb{Z}_N$ let $\alpha_{r,d} = f_{1,2s}(rd)$. Let ω_r be the function $x \mapsto \omega^{rx}$. Then there is a real number $\lambda_r \geq 1/32$ such that*

$$\lambda_r\omega_r \otimes \overline{\omega_r} = \sum_d \alpha_{r,d}(\omega_r * \pi_d * \pi_d) \otimes (\overline{\omega_r} * \pi_d * \pi_d).$$

Proof. By Lemma 2.2 there exists a positive $d \leq m$ such that $\|rd/N\| \leq m^{-1}$. It follows that $f_{1,2s}(rd) = \alpha_{r,d} \geq 1/2$.

For this d we know that $|1-\omega^{rd}| \leq 2\pi/m$, so by Lemma 3.2 we can conclude that

$$\alpha_{r,d}(\omega_r * \pi_d * \pi_d) \otimes (\overline{\omega_r} * \pi_d * \pi_d) = \alpha_{r,d}\mu_{r,d}^2 \omega_r \otimes \overline{\omega_r}$$

for some real constant $\mu_{r,d} \geq 1/4$. For all other d, we have the same conclusion, but this time all we can guarantee is that $\alpha_{r,d}\mu_{r,d}^2 \geq 0$. Summing over all d and setting $\lambda_r = \sum_d \alpha_{r,d}\mu_{r,d}^2$, we obtain the result stated. □

We are now ready for the main estimate, after which the proof will be more or less finished.

Lemma 3.4. *There exist coefficients λ_r, all at least 1/16, such that the matrix $\mathbb{E}_r \lambda_r \omega_r \otimes \overline{\omega_r}$ can be expressed as a linear combination*

$$\sum_{0<d\leq m} \sum_{x,y} \gamma_{d,x,y} P_{d,x} \otimes P_{d,y}$$

where each $P_{d,x}$ and $P_{d,y}$ is an arithmetic progression mod N of diameter at most m^2, and $\sum_{d,x,y} |\gamma_{d,x,y}| \leq N/4m$.

Proof. Lemma 3.3 gives us the decomposition

$$\begin{aligned} \mathbb{E}_r \lambda_r \omega_r \otimes \overline{\omega_r} &= \mathbb{E}_r \sum_d \alpha_{r,d}(\omega_r * \pi_d * \pi_d) \otimes (\overline{\omega_r} * \pi_d * \pi_d) \\ &= \frac{N}{|P|^2} \sum_r \sum_d \alpha_{r,d}(\omega_r * \pi_d * P_d) \otimes (\overline{\omega_r} * \pi_d * P_d). \end{aligned}$$

Let us write Q_d for the function $\pi_d * P_d$. Now $\omega_r * Q_d = \mathbb{E}_x \omega^{rx} Q_{d,x}$, where $Q_{d,x}(y) = Q_d(y-x) = Q_d(x-y)$. So we can rewrite our decomposition as

$$\frac{N}{|P|^2} \sum_r \sum_d \alpha_{r,d} \mathbb{E}_{x,y} \omega^{r(x-y)} Q_{d,x} \otimes Q_{d,y} = \frac{1}{N|P|^2} \sum_r \sum_d \alpha_{r,d} \sum_{x,y} \omega^{r(x-y)} Q_{d,x} \otimes Q_{d,y}.$$

The coefficient of $Q_{d,x} \otimes Q_{d,y}$ in this decomposition is

$$\frac{1}{N|P|^2} \sum_r \alpha_{r,d} \omega^{r(x-y)},$$

so the sum of the absolute values of these coefficients is

$$\begin{aligned}\frac{1}{N|P|^2}\sum_d\sum_{x,y}|\sum_r \alpha_{r,d}\omega^{r(x-y)}| &= \frac{N}{|P|^2}\sum_d \mathbb{E}_{x,y}|\sum_r \alpha_{r,d}\omega^{r(x-y)}| \\ &= \frac{N}{|P|^2}\sum_d \mathbb{E}_x|\sum_r \alpha_{r,d}\omega^{rx}| \\ &= \frac{N}{|P|^2}\sum_d\sum_x |\mathbb{E}_r\alpha_{r,d}\omega^{rx}| \\ &= \frac{N}{|P|^2}\sum_d \|\hat{\alpha}_d\|_1,\end{aligned}$$

where we define α_d by the formula $\alpha_d(r) = \alpha_{r,d}$.

Now $\alpha_d(r) = f_{1,2s}(dr) = f_{d^{-1},2s}(r)$ for every r. Therefore, by Lemma 3.1 we have that $\|\hat{\alpha}_d\|_1 = 1$ for every $d \leq m$. It follows that the sum of the absolute values of the coefficients is $Nm/|P|^2 \leq N/4m$. □

To complete the proof, we first observe that the matrix $\mathbb{E}_r\lambda_r\omega_r \otimes \overline{\omega_r} - I_N/16$ is positive semidefinite. That is for the simple reason that it is equal to $\mathbb{E}_r(\lambda_r - 1/16)\omega_r \otimes \overline{\omega_r}$, and we have ensured that each λ_r is at least $1/16$. Therefore, by Lemma 3.4, given any function $f : \mathbb{Z}_N \to \mathbb{C}$, we have some pair $(Q_{d,x}, Q_{d,y})$ such that $|\sum_z Q_{d,x}(z)f(z)\sum_w Q_{d,y}(w)\overline{f(w)}| \geq m\|f\|_2^2/4$. (This number $m/4$ is the ratio of the trace $N/16$ of $I_N/16$ to the upper bound $N/4m$ for the sum of the coefficients in the decomposition.) It follows that we can find x such that $|\sum_z Q_{d,x}(z)f(z)| \geq m^{1/2}\|f\|_2/2$. Since $Q_{d,x}$ is an average of characteristic functions of translates of the arithmetic progression P_d, it follows that we can find some x such that $|\sum_{z\in P_d+x} f(z)| \geq m^{1/2}\|f\|_2/2$. And finally, since P_d has diameter at most N (assuming that $m \leq c\sqrt{N}$, it follows that we can find an arithmetic progression of length m and common difference at most m on which f has discrepancy at least $m^{1/2}\|f\|_2/4$. In particular, if f is ± 1-valued, then we obtain a lower bound for the AP-discrepancy of $m^{1/2}/4$. Once again, this argument is valid when $m \leq c\sqrt{N}$, so we obtain Roth's bound of $cN^{1/4}$.