ON THE COMPLEXITY OF INVERTING THE AUTOCORRELATION FUNCTION OF A FINITE INTEGER SEQUENCE, AND THE PROBLEM OF LOCATING n POINTS ON A LINE, GIVEN THE $\binom{n}{2}$ UNLABELLED DISTANCES BETWEEN THEM

By

Paul Lemke

and

Michael Werman

IMA Preprint Series # 453

October 1988

ON THE COMPLEXITY OF INVERTING THE AUTOCORRELATION FUNCTION OF A FINITE INTEGER SEQUENCE, AND THE PROBLEM OF LOCATING n POINTS ON A LINE, GIVEN THE $\binom{n}{2}$ UNLABELLED DISTANCES BETWEEN THEM

Paul Lemke*

and

Michael Werman†

Abstract. We show that a polynomial time algorithm exists to find all integer sequences with a given autocorrelation function, and we show that the same methods yield a polynomial bound on the number of different arrangements (up to translation and reflection) of n points on a line which generate a given multi-set of $\binom{n}{2}$ unlabelled distances between pairs of points.

1. Introduction. Finding a function given only certain autocorrelation information about it is a problem which arises in signal processing. Uniqueness of the solution and the difficulty of finding it are usually the issues. We investigate here the case where the data is a finite sequence of integers. We then look at the problem of locating n points on a line so that the $\binom{n}{2}$ distances between pairs of points correspond to some given set.

Much of the mathematics needed for our result has already been done in other contexts; we repeat some of it here for more self-containment in this context. References to other work are given when known.

Problem Definition. For a finite sequence of integers $b_0, b_1, \ldots b_n$ (which are usually constrained to be equal to ± 1, but which may be any integers of small absolute value for our purposes), the autocorrelation function $a(k)$ is defined by

$$a(k) = \sum_{i=0}^{n-k} b_i b_{i+k} \quad \text{for } k = 0, 1, \ldots, n. \tag{1}$$

We may then ask the question: Given the integer n and the values $a(0), a(1), \ldots, a(n)$, is the sequence $b_0, b_1, \ldots, b_n$ uniquely defined up to reversal of order? Is it easy to compute all such sequences? For the first question, we looked at the specific case of 0-1 sequences, that is, where we have $b_i \in \{0, 1\}$ for $i = 0, 1, \ldots, n$. We assume that the sequence does not begin or end with 0 to avoid the ambiguity associated with shifting the sequence (for example, for $n = 5$, 101100, 010110 and 001011 all have the same autocorrelation function). We also did not consider two sequences "different" if they differ

* Institute for Mathematics and its Applications, 514 Vincent Hall, University of Minnesota, Minneapolis, MN 55455

† Institute for Mathematics and its Applications, 514 Vincent Hall, University of Minnesota, Minneapolis, MN 55455

only by order reversal. Our computer search shows that for $n \leq 10$, no two different 0-1 sequences give rise to the same autocorrelation function. For $n = 11$, however, the sequences 110000110101 and 111000101001 have the same autocorrelation function, with values $6, 2, 2, 1, 2, 2, 2, 1, 1, 1, 1, 1$. However, there were only 234 such pairs among the 66,047 different strings with $n \leq 17$, and there are no sets of 3 or more different strings with the same autocorrelation function for $n \leq 17$. This data would seem to indicate that, while the solution is not always unique, it usually is, and is of only small multiplicity when it is not. We now show, in fact, that the multiplicity of solutions *is* polynomially bounded in the length of the sequence when the members of the sequence are restricted to belong to a fixed set of small integers. We also show that all solutions can be found in polynomial time.

We may assume without loss of generality that $b_n > 0$, $b_0 \neq 0$, and $\gcd(b_o, b_1, \ldots, b_n) = 1$. We may then define the polynomials $b(x)$ and $A(x)$ by:

$$b(x) \equiv \sum_{i=0}^{n} b_i x^i$$

$$\text{and } A(x) \equiv \frac{a(n)}{|a(n)|} \sum_{i=0}^{2n} a(|n-i|)x^i.$$

In general, if $f(x)$ is a polynomial in x of degree d with $f(0) \neq 0$, and with its highest degree term coefficient positive, define the "reciprocal" polynomial $\hat{f}(x)$ by:

$$\hat{f}(x) \equiv \frac{f(0)}{|f(0)|} x^d f(\frac{1}{x}).$$

Note that if $g(x) \equiv \hat{f}(x)$, then $\hat{g}(x) \equiv f(x)$ for each such polynomial $f(x)$. Equation (1) then becomes:

$$A(x) \equiv b(x)\hat{b}(x). \tag{2}$$

Given the polynomial $A(x)$, we then wish to find a polynomial $b(x)$ with integer coefficients satisfying (2), whose coefficients may also have to belong to a certain restricted set (such as $\{-1, 1\}$).

We will show that given a polynomial $p(x) \equiv \sum_{i=0}^{2n} c_i x^i$ with $c_{2n} > 0$, it is possible to find *all* polynomials $f(x)$ with integer coefficients and with positive highest degree term coefficient satisfying

$$f(x)\hat{f}(x) \equiv p(x), \tag{3}$$

in running time bounded by a polynomial function of $\max(n, \sum_{i=0}^{2n} |c_i|)$.

The algorithm to do this is essentially that of J. Rosenblatt and P.D. Seymour in [2], which deals with more general polynomials, although he does not prove a polynomial bound on the running time.

Algorithm. Factor $p(x)$ into polynomials which are irreducible over the integers, and which have the coefficient of their highest degree term positive. Arrange them into pairs $(p_1(x), \hat{p}_1(x))$, $(p_2(x), \hat{p}_2(x))$, $\cdots, (p_k(x), \hat{p}_k(x))$. If this cannot be done, there are no solutions. Otherwise, let I be the set of all $i \in \{1, 2, \ldots, k\}$ such that $p_i(x) \equiv \hat{p}_i(x)$ and let $J = \{1, 2, \ldots, k\} - I$. Compute all $2^{|J|}$ polynomials $f(x)$ such that

$$f(x) = \left(\prod_{j \in I} p_j(x)\right)\left(\prod_{j \in J} p_j(x)^{d_j} \hat{p}_j(x)^{1-d_j}\right)$$

with $d_j \in \{0, 1\}$ for $j \in J$. Sort out the resulting list to remove duplicates.

Proof that the algorithm works. If a solution to (3) exists, and if $f_1(x), f_2(x), \ldots, f_\ell(x)$ are the irreducible factors of $f(x)$ with positive highest degree term coefficients, then $\hat{f}_1(x)$, $\hat{f}_2(x), \ldots, \hat{f}_\ell(x)$ are the corresponding irreducible factors of $\hat{f}(x)$, so that the factors of $p(x)$ can be paired as indicated above. Since the pairings are unique up to their order, and the order of the polynomials within each pair, and since the above algorithm finds all distinct polynomials which are the product of ℓ polynomials, one from each of the ℓ pairs, it must find $f(x)$. Conversely, if $f(x)$ is a polynomial found by the algorithm, it is clear that $f(x)\hat{f}(x) \equiv p(x)$.

Proof that the algorithm is polynomial time. Polynomials can be factored into irreducible polynomials in time which is polynomial in n and depends only slightly on the coefficient size ([1]). Thus $p(x)$ can be factored in time which is polynomial in $\max(n, \sum_{j=0}^{2n} |c_j|)$, and clearly the polynomial multiplications and pairings can also be done in polynomial time. The remaining part of the proof is therefore to show that $2^{|J|}$ is bounded from above by a fixed power of $\sum_{j=0}^{2n} |c_j|$. This may be done by the method of C.J. Smyth in [3] in the context of Newman polynomials, modified for more general polynomials.

First, we need the inequality:

$$h_d \prod_{|r_j| \geq 1} |r_j| \geq \lambda, \tag{4}$$

where $h(x) = \sum_{j=0}^{d} h_j x^j$ is a polynomial with integer coefficients with $h_d > 0, h_0 \neq 0$, $h(x) \not\equiv \hat{h}(x)$, and with complex roots $r_1, r_2, \ldots, r_d$, and where $\lambda = 1.3247\ldots$ is the real root of the equation $x^3 - x - 1 = 0$. Eq. (4) follows immediately from an earlier result of Smyth ([4]) in which he uses arguments involving the power series of $\frac{h(x)}{\hat{h}(x)}$ to prove (4) when $h_d = 1$. We then need the inequality

$$|q_d| \prod_{|r_j| \geq 1} |r_j| \leq \left(\sum_{j=0} q_j^2\right)^{\frac{1}{2}}, \tag{5}$$

where $q(x) \equiv \sum_{j=0}^{d} q_j x^j$ is a polynomial with real coefficients and complex roots $r_1, r_2, \ldots, r_d$. This inequality is credited in [3] to W. Specht ([5]). It may be proven as

follows:

$$\frac{1}{2\pi}\int_0^{2\pi}|q(e^{i\theta})|^2d\theta=\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{j=0}^{d}q_je^{ij\theta}\right|^2d\theta$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\left|i\sum_{j=0}^{d}q_j\sin j\theta+\sum_{j=0}^{d}q_j\cos j\theta\right|^2d\theta$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\left(\left(\sum_{j=0}^{d}q_j\sin j\theta\right)^2+\left(\sum_{j=0}^{d}q_j\cos j\theta\right)^2\right)d\theta$$

$$=\frac{1}{2\pi}\sum_{\ell=0}^{d}\sum_{j=0}^{d}\int_0^{2\pi}q_\ell q_j(\sin\ell\theta\sin j\theta+\cos\ell\theta\cos j\theta)d\theta,$$

and since

$$\int_0^{2\pi}\sin\ell\theta\sin j\theta\ d\theta=\int_0^{2\pi}\cos\ell\theta\cos j\theta\ d\theta=0\text{ for }\ell\neq j,$$

the above becomes:

$$\frac{1}{2\pi}\int_0^{2\pi}|q(e^{i\theta})|^2d\theta=\sum_{j=0}^{d}q_j^2. \tag{6}$$

Since we have

$$\frac{1}{b-a}\int_a^b g(f(x))dx\leq g\left(\frac{1}{b-a}\int_a^b f(x)dx\right)$$

when g is any concave function, we derive from (6)

$$\frac{1}{\pi}\int_0^{2\pi}ln|q(e^{i\theta})|d\theta=\frac{1}{2\pi}\int_0^{2\pi}ln(|q(e^{i\theta})|^2)d\theta\leq ln\left(\frac{1}{2\pi}\int_0^{2\pi}|q(e^{i\theta})|^2\right)=ln\left(\sum_{j=0}^{d}q_j^2\right),$$

so that we have:

$$\frac{1}{2\pi}\int_0^{2\pi}ln|q(e^{i\theta})|d\theta\leq\frac{1}{2}ln\left(\sum_{j=0}^{d}q_j^2\right). \tag{7}$$

We now need to prove the identity

$$ln\left(q_d\prod_{|r_j|\geq1}|r_j|\right)=\frac{1}{2\pi}\int_0^{2\pi}ln|q(e^{i\theta})|d\theta. \tag{8}$$

We note that

$$\frac{1}{2\pi}\int_0^{2\pi}ln|q(e^{i\theta})|=\frac{1}{2\pi}\int_0^{2\pi}ln\left|q_d\prod_{j=1}^{d}(e^{i\theta}-r_j)\right|d\theta=ln|q_d|+\sum_{j=1}^{d}\frac{1}{2\pi}\int_0^{2\pi}ln|e^{i\theta}-r_j|d\theta. \tag{9}$$

Substituting $\theta = \phi + \phi_j$ in the integrand in the last term, where $\phi_j = arg(r_j)$ for $j = 1, 2, \ldots, d$ gives:

$$\int_0^{2\pi} ln|e^{i\theta} - r_j|d\theta = \int_{-\phi_j}^{2\pi-\phi_j} ln|e^{i\phi}e^{i\phi_j} - |r_j|e^{i\phi_j}|d\phi = \int_{-\phi_j}^{2\pi-\phi_j} ln(|e^{i\phi_j}||e^{i\phi} - |r_j||)d\phi$$
$$= \int_{-\phi_j}^{2\pi-\phi_j} ln|e^{i\phi} - |r_j||d\phi = \int_{-\phi_j}^{0} ln|e^{i\phi} - |r_j||d\phi + \int_0^{2\pi-\phi_j} ln|e^{i\phi} - |r_j||d\phi$$
$$= \int_{2\pi-\phi_j}^{2\pi} ln|e^{i\phi} - |r_j||d\phi + \int_0^{2\pi-\phi_j} ln|e^{i\phi} - |r_j||d\phi = \int_0^{2\pi} ln|e^{i\phi} - |r_j||d\phi =$$
$$\int_0^{2\pi} \frac{1}{2} ln(\sin^2\phi + (\cos\phi - |r_j|)^2)d\phi.$$

This last integral may be evaluated to give:

$$\int_0^{2\pi} \frac{1}{2} ln(\sin^2\phi + (\cos\phi - |r_j|)^2)d\phi = \begin{cases} 0 & \text{if } |r_j| \leq 1 \\ 2\pi ln|r_j| & \text{if } |r_j| \geq 1. \end{cases}$$

Substituting this result in (9) gives (8).

It is easier to use Specht's inequality in the form (7), and to use (8) with (4) to get

$$\frac{1}{2\pi} \int_0^{2\pi} ln|h(e^{i\theta})|d\theta \geq ln\lambda, \tag{10}$$

where $h(x)$ is any polynomial satisfying the constraints listed after (4). Since $p_\ell(x)$ and $\hat{p}_\ell(x)$ each satisfy these conditions for $\ell \in J$, and since from (8) we know that the left-hand side of (10) is always non-negative when $h(x)$ is any polynomial with integer coefficients, we have

$$\frac{1}{2} ln\left(\sum_{i=0}^{2n} c_i^2\right) \geq \frac{1}{2\pi} \int_0^{2\pi} ln|p(e^{i\theta})|d\theta$$
$$= \sum_{\ell=1}^{k} \frac{1}{2\pi} \int_0^{2\pi} ln|p_\ell(e^{i\theta})|d\theta + \sum_{\ell=1}^{k} \frac{1}{2\pi} \int_0^{2\pi} ln|\hat{p}_\ell(e^{i\theta})|d\theta$$
$$\geq 2|J|ln\lambda,$$

so that

$$|J| \leq \frac{1}{4ln\lambda} ln\left(\sum_{i=0}^{2n} c_i^2\right), \tag{11}$$

and since $\sum_{i=0}^{2n} c_i^2 \leq (\sum_{j=0}^{2n} |c_i|)^2$, we have

$$2^{|J|} \leq \left(\sum_{i=0}^{2n} |c_i|\right)^{\frac{ln2}{2ln\lambda}} = \left(\sum_{i=0}^{2n} |c_i|\right)^{1.23248\ldots},$$

and the proof is complete.

Clearly, $2^{|J|}$ represents an upper bound on the number of solutions to (3). The sum of the squares of the coefficients of $A(x)$ in (2) is bounded from above by $(\sum_{i=0}^{n} b_i^2)^2 \leq (\max_i b_i)^4(n+1)^2$, so that with (11) the general upper bound for a sequence of length ℓ and maximum term magnitude m is $(lm^2)^{\frac{ln2}{2lnλ}} = (lm^2)^{1.232482627...}$.

For the case of 0-1 sequences, this bound is somewhat larger than our data would suggest. This is not surprising, considering that only a small fraction of the polynomials $b(x)$ which satisfy (2) will have coefficients which all belong to the restricted set. Also, the polynomial $A(x)$ in (2) probably cannot have nearly as many non-self-reciprocal irreducible polynomial factors as is allowed by the bound on $|J|$ implied by (7). It might therefore be possible to get a better bound on the number of solutions to (1) derived above.

A simple lower bound for an upper bound for zero-one strings may be obtained by defining the polynomial $g(x)$ by:

$$g(x) \equiv \prod_{j=1}^{k}(1 + x^{\frac{3^k-1}{2}} + x^{3^k}),$$

and then letting $p(x) \equiv g(x)\hat{g}(x)$. Equation (3) will then have 2^k solutions with $f(x)$ a polynomial whose coefficients are 0 or 1, and the degree of $f(x)$ will be $\frac{3^{k+1}-1}{2}$. In terms of the length ℓ of the sequence, the number of solutions to (1) will then be approximately $\ell^{\frac{ln2}{ln3}} = \ell^{0.63092975...}$. This example is a modification of examples given by J.N. Franklin in [6] and later by Rosenblatt in [7] in the context of homometric sets (discussed later).

Another approach to obtaining an upper bound to the number of solutions to (1) for non-negative sequences $\{b_i\}$ is to solve the following problem: Given a polynomial $p(x)$ with non-negative integer coefficients, in how many different ways can $p(x)$ be factored into two polynomials with non-negative coefficients, as a function of $p(1)$? This is an interesting problem in its own right, and we currently do not have a polynomial upper bound. A lower bound for an upper bound can be obtained as follows: for any polynomial $f(x)$ with non-negative integer coefficients and $f(0) \neq 0$, define $q(f)$ to be the number of factorizations of $f(x)$ into 2 polynomials with non-negative integer coefficients, counting different orders of the factors as different, for example if $f(x) \equiv 2x + 2$ and $g(x) \equiv 12$, then $q(f) = 4$ and $q(g) = 6$. Now define $s(f)$ by

$$s(f) = \frac{ln(q(f))}{ln(f(1))},$$

so that $q(f) = f(1)^{s(f)}$. We would now like to know how large $s(f)$ can be. We may note that if $f(x)$ is a polynomial of degree d and if $p_k(x)$ is defined by:

$$p_k(x) \equiv \prod_{i=0}^{k} f(x^{(d+1)^i}),$$

then we have $p_k(1) = f(1)^{k+1}$, and it is not hard to show that $q(p_k) \geq q(f)^{k+1}$, so that $s(p_k) \geq s(f)$ for $k = 0, 1, 2, \ldots$. Thus it follows that there are infinitely many integers n such that for some polynomial $p(x)$, $p(1) = n$ and $q(p) \geq n^{s(f)}$. In particular, if $f(x) = x + 1$, then we have $s(f) = 1$, so that any upper bound on the number of non-negative factorizations of a polynomial $p(x)$ must grow at least linearly with $p(1)$. A better bound, due to S. Winograd ([8]), can be obtained by defining

$$f_k(x) \equiv (x^2 + x + 1)\Big(\prod_{i=0}^{k}(x^{2^i} + 1)\Big)\prod_{i=0}^{k}(x^{2^{i+1}} - x^{2^i} + 1).$$

We then enumerate the ways to partition the $2k+1$ factors of $f_k(x)$ given above into 2 sets so that the product of the factors in each set is a polynomial with non-negative coefficients. The factor x^2+x+1 may be assigned to either set, and, for fixed $j \in \{0, 1, \ldots, k\}$ the factors $x^2 - x + 1$, $x^4 - x^2 + 1, \ldots, x^{2^{j+1}} - x^{2^j} + 1$ are assigned to the same set, giving a product of $x^{2^{j+2}} + x^{2^{j+1}} + 1$, which has non-negative coefficients. The factors $x^{2^{j+2}} - x^{2^{j+1}} + 1$ and $x^{2^{j+1}} + 1$, whose product is also non-negative, are then assigned to the opposite set (if $j < k$). Finally, the factors $x + 1$, $x^2 + 1, \ldots, x^{2^j} + 1$, and the pairs of factors $x^{2^i} + 1$ and $x^{2^{i+1}} - x^{2^i} + 1$ (whose product is non-negative) for $i = j+2, j+3, \ldots, k$ may be arbitrarily assigned to either set. For each value of $j < k$ we have 2^k possible ways to assign the factors according to the above rule, 2^{k+1} if $j = k$, or $(k+2)2^k$ ways altogether, and each way gives a non-negative factorization of $f_k(x)$. We therefore have $q(f_k) \geq (k+2)2^k$, and since $f_k(1) = 3 \cdot 2^{k+1}$, we have

$$s(f_k) \geq \frac{ln((k+2)2^k)}{ln(3 \cdot 2^{k+1})}.$$

The value of the right side of (8) is maximized when $k = 15$, when we have $s(f_k) \geq 1.0854423$.

We may also ask how many non-negative polynomial divisors $f(x)$ a given polynomial $p(x)$ can have (regardless of whether the quotient $\frac{p(x)}{f(x)}$ is non-negative). The number is unbounded as a function of $p(1)$, since $x^n + 1$ has $x^d + 1$ as a divisor when n is odd and d divides n, and n can have arbitrarily many divisors. However, it may well be bounded by a polynomial function of $\max(p(1), deg(p))$, where $\deg(p)$ denotes the degree of $p(x)$. The same would then be true for $q(p)$, even if $q(p)$ turned out not to be bounded by a polynomial function of $p(1)$ alone.

Points on a Line. A problem in discrete geometry related to inverting autocorrelation functions is the problem of determining the position of n points on a line, given the $\binom{n}{2}$ unlabelled distances between all pairs of them. Precisely, we are given a multi-set S of $\binom{n}{2}$ non-negative real numbers and we wish to find real numbers $x_1, x_2, \ldots, x_n$ such that

$$\bigcup_{i>j}\{|x_i - x_j|\} = S. \tag{12}$$

We may then ask the questions:

1. If a solution exists, is it unique up to translation and reflection? If not, how many different solutions can there be as a function of n?

2. Is it possible to compute some or all solutions in polynomial time?

Question 1, in particular, has been given considerable attention in the literature, usually as a problem occurring in crystallography in which the points involved form a repeating unit of an infinite periodic set in $\mathbf{R}^d$. (In this context, equation (12) is modified by having the absolute value signs removed and the restriction $i > j$ dropped. The set S then consists of n^2 *differences* rather than $\binom{n}{2}$ distances.) If more than one solution exists for some S, the solution sets which share the same difference set are then called *homometric sets*. Work on finite homometric sets goes back at least as far as 1939 ([9]), with more recent results in [6], [2], and [7]. Here, we will limit our concerns to finite sets in $\mathbf{R}^1$.

To avoid ambiguity, we can require:

i) $x_1 \leq x_2 \leq \ldots \leq x_n$

ii) $x_1 = 0$

iii) $x_1, x_2, \ldots, x_n$ is lexicographically smaller than or equal to $x_n - x_n, x_n - x_{n-1}, \ldots, x_n - x_1$.

We will call the members of s $d_1, d_2, \ldots, d_{\binom{n}{2}}$, and assume that $d_1 \geq d_2 \geq \ldots \geq d_{\binom{n}{2}}$. Given these conditions, we have, in addition to: a) $x_1 = 0$, the equalities:

b) $x_n = d_1$

c) $x_n - x_2 = d_2$

d) $\sum_{i<j}(x_j - x_i) = \sum_{k=1}^{n}(2k - n - 1)x_k = \sum_{\ell=1}^{\binom{n}{2}} d_\ell$.

These equalities uniquely define the multi-set of points for $n \leq 4$. For $n = 5$, the 4 points other than x_3 are uniquely defined, and x_3 is then defined by the following lemma:

Lemma 1: If $x_1, x_2, \ldots, x_k, y$ and z are real numbers, and if the $\binom{k+1}{2}$ distances associated with the multi-set $\{x_1, x_2, \ldots, x_k, y\}$ are the same as the distances associated with the multi-set $\{x_1, x_2, \ldots, x_k, z\}$, then these two multi-sets are either equal, or equivalent under reflection and translation. Specifically, if we assume that $x_1 \leq x_2 \leq \ldots \leq x_k$, then we must have either $y = z$ or else $y + z = x_i + x_{k-i+1}$ for $i = 1, 2, \ldots, k$.

Proof: We assume that the above statement is true for $k \leq n - 1$ for some integer n, and we will then prove it for $k = n$.

Since the $\binom{n}{2}$ distances $\{|x_i - x_j|\}_{i>j}$ are contained in each set, the hypothesis of the

Lemma is equivalent to:

$$\bigcup_{i=1}^{n}\{|x_i - y|\} = \bigcup_{i=1}^{n}\{|x_i - z|\}. \tag{13}$$

Equating the maxima of these multi-sets gives

$$\left|\frac{x_n + x_1}{2} - y\right| + \frac{x_n - x_1}{2} = \left|\frac{x_n + x_1}{2} - z\right| + \frac{x_n - x_1}{2},$$

so that if $y \neq z$ we must have $y + z = x_1 + x_n$. We then must have:

$$\{|x_1 - y|\} \cup \{|x_n - y|\} = \{|x_1 - z|\} \cup \{|x_n - z|\} = \{d_m, |x_n - x_1 - d_m|\},$$

where d_m denotes the maximum of the multi-sets in (13). This implies:

$$\bigcup_{i=2}^{n-1}\{|x_i - y|\} = \bigcup_{i=2}^{n-1}\{|x_i - z|\},$$

which, using the inductive hypothesis, implies that if $y \neq z$, then $y + z = x_i + x_{n-i+1}$ for $i = 2, 3, \ldots, n-1$. Since we already know that $y + z = x_1 + x_n$, we are done. □ □

Thus for $n = 5$, x_3 can only go in one place subject to constraint *iii* once the other 4 points are defined, and so the arrangement of n points on a line is uniquely determined by their $\binom{n}{2}$ unordered distances up to translation and reflection for $n \leq 5$. For $n = 6$, however, we may use the example given earlier of the two binary sequences with the same autocorrelation function, by treating the 1's in the sequences as points on a line, with their positions in the sequence corresponding to their locations on the line. This gives the two point sets $0, 1, 6, 7, 9, 11$ and $0, 1, 2, 6, 8, 11$ which have the same distance multi-sets.

This example generalizes to the sets $X = \{0, 1, 5, 6, \ldots, n-3, n-2, n, n+1, n+3, n+5\}$ and $X \cup \{2\} \cup \{n+2\} - \{n+1\} - \{n+3\}$ for all $n \geq 6$.

In fact, if the coordinates of the points are known to be integers, the problem of finding the points, given the multiset of distances between them is equivalent to the problem of inverting the autocorrelation function. Specifically, the polynomial $p(x)$ in (3) is determined from S in (12) by defining:

$$p(x) \equiv \sum_{i=0}^{2k} c_{|i-k|} x^i, \tag{14}$$

where k is the maximum of S, $c_0 = n$, and, for $j = 1, 2, \ldots, k$, c_j is the number of members of S equal to j. A non-negative polynomial $f(x) = \sum_{i=0}^{k} f_i x^i$ which solves (3) will correspond to a multi-set $\{x_1, x_2, \ldots, x_n\}$ solving (12) by letting

$$\{x_1, x_2, \ldots, x_n\} = \bigcup_{i=0}^{k}\{i\}^{f_i},$$

that is, the point i occurs f_i times in the multi-set $\{x_1, x_2, \ldots, x_n\}$. Conversly, given a set of integers $\{x_1, x_2, \ldots, x_n\}$ satisfying (12) we translate the set so that its minimum is 0, and we can then go backwards to find a solution to (3). In the generalized example given above, we have $f(x) = (\frac{x^{n+1}-1}{x-1} - x)(x^5 - x^2 + 1)$ or $f(x) = (\frac{x^{n+1}-1}{x-1} - x)(x^5 - x^3 + 1)$.

Thus if the points are integers, we can find them in polynomial time in $\max(S \cup \{n\})$.

If we cannot assume that the points have integer coordinates, we can analyze the problem as follows:

We can normalize the distances in the given multi-set S by dividing equation (12) through by the maximum of S. Thus we can assume without loss of generality that $d_1 = 1$ (where $d_1, d_2, \ldots, d_{\binom{n}{2}}$ are the members of S, and are considered given and fixed for this analysis). We also assume that conditions i and ii apply to $x_1, x_2, \ldots, x_n$. We now construct a system L of linear equations with variables $y_1, y_2 \ldots, y_{\binom{n}{2}}$. The system contains $y_1 = 1$, $y_i + y_j = y_\ell$ whenever $d_i + d_j = d_\ell$, and $y_i = 0$ whenever $d_i = 0$. Clearly $y_i = d_i$ for $i = 1, 2, \ldots, \binom{n}{2}$ is a solution to this system. If it is the only solution, then d_i must be rational for each i since the data in L are rational. Otherwise, let L^* denote the set of solutions to L. When $y_i = d_i$ for $i = 1, 2, \ldots, \binom{n}{2}$, the equations:

(15a) $$x_1 = 0$$

(15b) $$\bigcup_{i=2}^{n} \bigcup_{j=1}^{i-1} \{x_i - x_j\} = \bigcup_{\ell=1}^{\binom{n}{2}} \{y_\ell\}$$

are equivalent to (12) plus requirements i and ii, and will therefore have the same number, clearly a finite number, of solutions. (This will be at most twice the number of solutions that the system has with requirement iii added). Assume that there are k distinct solutions to (12), i, and ii, and denote them by $(x_1^\ell, x_2^\ell, \ldots, x_n^\ell)$ for $\ell = 1, 2, \ldots, k$. For each ℓ, there must be a one-to-one function q_ℓ mapping $\{(i,j) | n \geq i > j \geq 1\}$ into $\{1, 2, \ldots, \binom{n}{2}\}$ such that

(16) $$x_i^\ell - x_j^\ell = d_{q_\ell(i,j)}.$$

The solution $(x_1^\ell, x_2^\ell, \ldots x_n^\ell)$ is then given explicitly in terms of q_ℓ by:

(17a) $$x_1^\ell = 0$$

(17b) $$x_i^\ell = d_{q_\ell(i,1)} \quad \text{for } i = 2, 3, \ldots, n.$$

The sets $\bigcup_{i=2}^{n} \{d_{q_\ell(i,1)}\}$ must therefore be different for each ℓ, since the corresponding solutions $(x_1^\ell, x_2^\ell, \ldots, x_n^\ell)$ are distinct by definition. Now, if $(y_1, y_2, \ldots, y_{\binom{n}{2}}) \in L^*$ and if we define $(\overline{x}_1^\ell, \overline{x}_2^\ell, \ldots, \overline{x}_n^\ell)$ by:

(18a) $$\overline{x}_1^\ell = 0$$

(18b) $$\overline{x}_i^\ell = y_{q_\ell(i,1)} \quad \text{for } i = 2,3,\ldots,n$$

then we will automatically have

(19) $$\overline{x}_i^\ell - \overline{x}_j^\ell = y_{q_\ell(i,j)} \quad \text{for } n \geq i > j \geq 1$$

since the equality

$$d_{q_\ell(i,1)} - d_{q_\ell(j,1)} = d_{q_\ell(i,j)} \quad \text{for } n \geq i > j \geq 2$$

is implied by (16) and (17b), which then implies

$$y_{q_\ell(i,1)} - y_{q_\ell(j,1)} = y_{q_\ell(i,j)} \quad \text{for } n \geq i > j \geq 2,$$

since these equalities must then be contained in the system L by definition. This, together with (18b), implies (19). It follows that $(\overline{x}_1^\ell, \overline{x}_2^\ell, \ldots, \overline{x}_n^\ell)$, as defined by (18a) and (18b), is a solution to (15a) and (15b) whenever $(y_1, y_2, \ldots, y_{\binom{n}{2}}) \in L^*$. We know that if $(y_1, y_2, \ldots, y_{\binom{n}{2}}) = (d_1, d_2, \ldots, d_{\binom{n}{2}})$, the vectors $(\overline{x}_1^\ell, \overline{x}_2^\ell \ldots, \overline{x}_n^\ell)$, $\ell = 1,2,\ldots,k$, will be distinct. It follows from continuity considerations that if

(20) $$|y_i - d_i| < \epsilon \quad \text{for } i = 1,2,\ldots,\binom{n}{2}$$

for some sufficiently small positive number ϵ, these k vectors will remain distinct. If we also have $(y_1, y_2, \ldots, y_{\binom{n}{2}}) \in L^*$, then we must have $y_i \geq 0$ for $i = 1,2,\ldots,\binom{n}{2}$, because if $d_i > 0$ for some i we can choose $\epsilon < d_i$. If $d_i = 0$, then the equality $y_i = 0$ will be contained in L, so that we have $y_i = 0$. Finally, since L^* is an affine space defined by rational equalities, rational points are dense in L^*, and since $(d_1, d_2, \ldots d_{\binom{n}{2}}) \in L^*$, for any given $\epsilon > 0$ we can choose rational numbers $y_1, y_2, \ldots, y_{\binom{n}{2}}$ such that $(y_1, y_2, \ldots, y_{\binom{n}{2}}) \in L^*$ and such that (20) holds. It follows that we can replace the members of S in (12) with non-negative rational numbers so that (12), with requirements i and ii, has at least as many distinct solutions as it did with the original S. This can then be multiplied through by a common denominator to obtain an all-integer problem, for which we can then use (14) to get the problem into the form of (3). Moreover, it seems that one can readily get it into that form in polynomial time, but unfortunately, the degree of the polynomial $p(x)$ in (3) can grow exponentially with n, even in the case where the solution to the system L is unique. This precludes carrying out the first step of the algorithm, which involves the factorization of $p(x)$. We therefore do not know at this time whether the points on a line problem can be solved in polynomial time for general positions. It seems it should be possible to do so, but it would, at least, require some modification of this method.

However, the bound on the *number* of solutions to (3) depends only on the sum of the absolute values of the coefficients of $p(x)$, which, for the points on a line problem, is always equal to n^2. It follows that the number of solutions does not exceed $\frac{1}{2}n^{2.465}$ for $n \geq 2$ (the factor $\frac{1}{2}$ coming about from the enforcement of requirement *iii*). As with the autocorrelation case, it seems that a better bound should exist. As for a lower bound for an upper bound we can use the example given previously for a lower bound for the

upper bound for zero-one sequences with a given autocorrelation function to show that there are infinitely many n such that there exists an example with at least $n^{0.6309}$ different arrangements of n points on a line with the same distance multi-sets.

Points in E^d. The above questions concerning points on a line can just as easily be asked about points in d dimensions, for $d \geq 2$. The examples for one dimension automatically apply to E^d to show that for $n \geq 6$ there exist two arrangements of n points in E^d which are different (not equivalent under translation, rotation, or reflection), but which have the same multi-set of $\binom{n}{2}$ distances. For E^2 an example exists for $n = 4$, which is clearly the smallest possible, namely $\{(0,0),(0,2),(1,0),(1,2)\}$ and $\{(0,0),(0,2),(1,0),(-1,0)\}$, which share the distance multi-set $\{1,1,2,2,\sqrt{5},\sqrt{5}\}$. Note that this violates Lemma 1 proven for E^1. In fact this example may be extended by adding any point x to both sets which is equidistant to the alternating points $(1,2)$ and $(-1,0)$. This method can be used successively to generate pairs of multi-sets of $d+2$ points in E^d which are different, span an affine space of d dimensions, and have the same distance multi-sets. We have not examined the problem of finding an upper bound to the number of different multi-sets of n points in E^d which share the same multi-set of distances, or of efficiently computing some or all of the point multi-sets, given the distance multi-sets. The question of how many different multi-sets of points can share the same *difference* multi-set, however (the *difference* problem is quite different from the *distance* problem for $d \geq 2$), might be approachable using the same methods as above, working with multi-variate polynomials instead of univariate ones.

References

[1] Susan Landau, *Factoring Polynomials Quickly*, Notices of the American Mathematical Society, Volume 34, Number 1, 1987, pp. 3–8.

[2] Joeseph Rosenblatt and Paul D. Seymour, *The Structure of Homometric Sets*, SIAM Journal on Algebra and Discrete Methods 3 (1982), pp. 343–350.

[3] C.J. Smyth, *Some Results on Newman Polynomials*, Indiana University Mathematics Journal 34 (1985), pp. 195–200.

[4] C.J. Smyth, *On the Product of the Conjugates Outside the Unit Circle of an Algebraic Integer*, Bulletin of the London Mathematical Society 3 (1971), pp. 169–175.

[5] W. Specht, *Abschätzungen der Wurzeln albebraischer Gleichungen*, Mathematische Zeitschift 52 (1949), pp. 310–321.

[6] Joel N. Franklin, *Ambiguities in the X-ray Analysis of Crystal Structures*, Acta Crystallographic A 30 (1974), pp. 698-702.

[7] Joeseph Rosenblatt, *Phase Retrieval*, Communications in Mathematical Physics 95 (1984), pp. 317–343.

[8] Shmuel Winograd, Personal Communication, July 1988.

[9] S. Piccard, *Sur les Ensembles de Distances des Ensembles de Points d'un Espace Euclidien*, Mémories de l'Université de Neuchatel, 13 (1939).