ON THE SHORTEST PATH THROUGH A NUMBER OF POINTS

S. VERBLUNSKY

1. **Introduction.** Let O denote a bounded open set in the x, y plane, with boundary of zero measure. There is a least absolute positive constant C with the following property: given $\epsilon>0$, there is an $N=N(O, \epsilon)$ such that, if $n>N$, then any n points in O are on a path of length less than $(C+\epsilon)(n\cdot mO)^{1/2}$. Here, mO is the measure of O. The path is not necessarily contained in O if O is not convex.

A set O of the kind specified is contained in a finite number of squares of total area arbitrarily near to mO. The above assertion is easily seen to be a consequence of the following. Given $\epsilon>0$, there is an $N(\epsilon)$ such that, if $n>N(\epsilon)$, then any n points in a unit square are on a path of length less than $(C+\epsilon)n^{1/2}$.

A simple argument (see §2) will show that $C\leqq 2$. On the other hand, it has been pointed out by Fejes (Math. Zeit. vol. 46 (1940) p. 85) that $C\geqq C_0$ where $C_0=2^{1/2}3^{-1/4}$. For we can arrange a large number n of points in a unit square so as to form a lattice of equilateral triangles, and a shortest path through the points is then seen to be asymptotically equal to $C_0n^{1/2}$. That $C\leqq(2.8)^{1/2}$ is a consequence of the following theorem.

THEOREM I. *Given n points in a unit square, there is a path through the points of length less than* $2+(2.8n)^{1/2}$.

This theorem can be deduced from the following theorem.

THEOREM II. *Given ν points $P_i(x_i y_i)$ in the strip $0\leqq y\leqq 1$ with $x_1 \leqq x_2\leqq \cdots \leqq x_\nu$, there is a path whose first point is P_1, whose last point is P_j say, which contains the ν points, and whose length does not exceed $x_j-x_1+.7\nu$.*

By examining the case in which the points P_i are $(0, 0)$, $(3^{-1/2}, 1)$, $(2\cdot 3^{-1/2}, 0)$, $\cdots$, we see that, in Theorem II, $.7$ cannot be replaced by a number less than $3^{-1/2}$.

2. **Proof that $C\leqq 2$.** The simple argument which implies that $C\leqq 2$ is as follows. Divide the square into m strips of breadth m^{-1}. Let A_iB_i $(i=1, \cdots, m)$ denote the longitudinal central lines of these strips. Then the path

$$p;\quad A_1B_1B_2A_2A_3\cdots$$

Presented to the Society, December 29, 1950; received by the editors October 31, 1950.

is of length $m+1-m^{-1}$. Each of the n points is at a distance not greater than $m^{-1}/2$ from p. We connect it to p by a segment of length not greater than $m^{-1}/2$. Then, if we count each segment twice, we obtain a path which contains the n points and is of length not exceeding $m+1+(n-1)m^{-1}$. It suffices to choose $m=1+[n^{1/2}]$, where $[x]$ denotes the greatest integer less than x.

3. **Theorem II implies Theorem I.** Turning now to the deduction of Theorem I from Theorem II, we suppose that the square has its sides parallel to the coordinate axes. Divide the square into m strips of breadth m^{-1}, parallel to the x axis. Consider any one of these strips which contains at least one, say ν, of the n points. By Theorem II (with a change of scale), there is a path containing the ν points of length not exceeding

$$d + .7\nu m^{-1}, \qquad d = \xi_l - \xi_f \geqq 0,$$

where ξ_f, ξ_l are the abcissae of the first and last points on the path. If $\nu=1$, we can put $\xi_f=\xi_l$.

The number of such strips is, say, $\mu \leqq m$. We enumerate *these* strips in order of increasing y. Call the path constructed in strip i the path i. Connect the last point of path 1 with the last point of path 2 by a segment; connect the first point of path 2 with the first point of path 3 by a segment; and so on. We thus obtain a path which contains the n points, and is of length not exceeding

$$\sum d_i + .7m^{-1} \sum \nu_i + s$$

where s is the sum of the $\mu-1$ segments which have been introduced.

The length of each segment does not exceed the sum of its projections on the coordinate axes. If we project the $\mu-1$ segments on the y axis, the amount of overlapping does not exceed $(\mu-2)m^{-1}$. Hence the sum of the projections on the y axis does not exceed

$$1 + (\mu - 2)m^{-1} \leqq 2 - 2m^{-1}.$$

The projection of the first segment on the x axis can be added to d_1 if $\xi_{l_2} > \xi_{l_1}$ to produce a modified term which does not exceed $1-\xi_{f_1}$; it can be added to d_2 if $\xi_{l_2} < \xi_{l_1}$ to produce a modified term which does not exceed $1-\xi_{f_2}$. The projection of the second segment on the x axis can be added to d_2 (or to the modified d_2) if $\xi_{f_3} < \xi_{f_2}$ to produce a term which does not exceed 1; it can be added to d_3 if $\xi_{f_3} > \xi_{f_2}$ to produce a modified term which does not exceed ξ_{l_3}; and so on. Hence the length of the path does not exceed

$$m + 2 - 2m^{-1} + .7nm^{-1}.$$

Choose $m=1+[(.7n)^{1/2}]$, and the theorem follows.

4. Reduction of proof of Theorem II to the lemma. Let $P_2', \cdots, P_\nu'$ denote a permutation of $P_2, \cdots, P_\nu$. The only paths we need consider are those consisting of $\nu-1$ segments

$$(1) \qquad P_1P_2', \; P_2'P_3', \cdots, P_{\nu-1}'P_\nu'.$$

Such paths will be called *admissible paths* through $P_1, P_2, \cdots, P_\nu$. We denote the path (1), as well as its length, by the symbol $P_1P_2' \cdots P_\nu'$. The assertion of Theorem II is that there is an admissible path $P_1 \cdots P_j$ through $P_1, P_2, \cdots, P_\nu$ of length not exceeding $x_j-x_1+.7\nu$.

On grounds of continuity, we may suppose that $x_1<x_2< \cdots <x_\nu$, provided that we prove

$$P_1 \cdots P_j < x_j - x_1 + .7\nu.$$

Call this result Theorem II′. It is a consequence of the following lemma.

LEMMA. *Given ν points $P_i(x_i, y_i)$ with*

$$x_1 < x_2 < \cdots < x_\nu, \qquad 0 \leqq y_i \leqq 1,$$

no two of which are at a distance less than .35, *there is an admissible path $P_1 \cdots P_j$ through $P_1, P_2, \cdots, P_\nu$ of length less than $x_j-x_1+.7\nu$.*

For, assuming the lemma, we need only consider the case where at least one distance is less than .35, and this case of Theorem II′ can be proved by induction. It is true for $\nu=2$. Suppose then that $\nu>2$. By hypothesis, there is a pair of points at a distance less than .35, say P, Q with $x_P<x_Q$. Then $Q\neq P_1$. Consider the $\nu-1$ points other than Q. By the inductive hypothesis, there is an admissible path $P_1 \cdots P_j$ through them such that

$$P_1 \cdots P_j < x_j - x_1 + .7(\nu - 1).$$

If $P_j=P$, then

$$\begin{aligned} P_1 \cdots P_jQ &< x_j - x_1 + .7(\nu - 1) + .35 \\ &< x_Q - x_1 + .7\nu, \end{aligned}$$

and $P_1 \cdots P_jQ$ is a path as required. If $P_j\neq P$, consider the symbol obtained from $P_1 \cdots P_j$ by inserting Q immediately after P. The corresponding path is of length less than

$$P_1 \cdots P_j + PQ + QP < x_j - x_1 + .7\nu,$$

and is a path as required.

5. **Proof of the lemma.** The idea of the proof of the lemma is as follows. We attempt to construct successive portions of an admissible path. If

$$P_1P_2 < x_2 - x_1 + .7, \tag{2}$$

then P_1P_2 is the first segment of the path. If (2) is not satisfied, and if $\nu > 2$, we consider the paths $P_1P_2P_3$, $P_1P_3P_2$. If one of the inequalities

$$P_1P_2P_3 < x_3 - x_1 + 1.4, \qquad P_1P_3P_2 < x_2 - x_1 + 1.4 \tag{3}$$

is satisfied, we choose such a one, and the corresponding pair of segments is the first pair of the path. If none of the inequalities (2), (3) are satisfied, and if $\nu > 3$, we consider the four paths $P_1P_2P_3P_4$, $P_1P_2P_4P_3$, $P_1P_3P_2P_4$, $P_1P_3P_4P_2$. If one of them, say $P_1 \cdots P_r$, satisfies

$$P_1 \cdots P_r < x_r - x_1 + 2.1, \tag{4}$$

we choose such a one, and the corresponding triad of segments is the first triad of the path. If none of the inequalities (2), (3), or (4) are satisfied, and if $\nu > 4$, we consider the four paths $P_1P_2P_3P_4P_5$, $P_1P_2P_4P_3P_5$, $P_1P_2P_3P_5P_4$, $P_1P_2P_4P_5P_3$. It turns out that at least one of these, say $P_1 \cdots P_r$, satisfies

$$P_1 \cdots P_r < x_r - x_1 + 2.8. \tag{5}$$

We choose such a one, and the corresponding quadruple of segments is the first quadruple of the path. When none of the inequalities (2), (3), or (4) are satisfied, we have

$$P_1P_2P_3P_4 < x_4 - x_1 + 2.8. \tag{6}$$

For if

$$P_1P_2P_3P_4 \geqq x_4 - x_1 + 2.8, \tag{7}$$

then since

$$P_iP_{i+1} \leqq x_{i+1} - x_i + | y_{i+1} - y_i | \qquad (i = 1, 2, 3)$$

we have

$$\sum_{i=1}^{3} | y_{i+1} - y_i | \geqq 2.8.$$

But each term does not exceed 1. Hence each term is at least .8. We may suppose, by a reflection in $y=1/2$ if necessary, that $y_1>y_2$. Then

$$y_1 > .8, \quad y_2 < .2, \quad y_3 > .8, \quad y_4 < .2.$$

This is the case $\alpha\delta\alpha\delta$ in the notation of §7. It is shown in §8 that either $P_1P_3P_2P_4$ is an admissible path through P_1, P_2, P_3, P_4 which satisfies (4) (a possibility which is here excluded by hypothesis), or else

$$P_1P_2P_3 < x_3 - x_1 + 1.5.$$

A fortiori

$$P_1P_2P_3P_4 < x_4 - x_1 + 2.5,$$

which contradicts (7).

6. **Proof of lemma continued.** The result so far is: if $\nu>4$, then we can construct a portion $P_1 \cdots P_r$ $(r>1)$ of the required path, consisting of k segments, such that

$$(8) \qquad P_1 \cdots P_r < x_r - x_1 + .7k.$$

If $\nu=2$, 3, 4, then either the same conclusion holds, or else the construction fails.

When the construction does not fail, and the portion constructed does not contain all the ν points, we attempt to repeat the argument starting from P_r and ignoring the points which precede it in the symbol $P_1 \cdots P_r$. It is important to notice that with the above meaning of k, the letters in the symbol $P_1 \cdots P_r$ are a permutation of $P_1, P_2, \cdots P_{k+1}$. Thus, if we ignore the points which precede P_r in the symbol $P_1 \cdots P_r$, we are left with P_r and certain P_ρ where $\rho>r$.

By the statements of §5, the construction can always be repeated if there are at least four points not on the portion already constructed. If there are only one, two, or three such points, the repetition may or may not be possible. We repeat the construction if possible. We then obtain a second portion $P_r \cdots P_s$ $(s>r)$ of the required path, consisting of κ segments, such that

$$(9) \qquad P_r \cdots P_s < x_s - x_1 + .7\kappa.$$

If $P_1 \cdots P_r \cdots P_s$ does not contain all the ν points, we attempt to repeat the argument starting from P_s and ignoring the points which precede it in the symbol $P_1 \cdots P_r \cdots P_s$; and so on. The process can come to an end only in one of the following four cases.

(i) *All the ν points are on the path constructed.* By adding the inequalities (8), (9), $\cdots$, we find that the path $P_1 \cdots P_j$, say, satisfies

$$P_1 \cdots P_j < x_j - x_1 + .7(\nu - 1).$$

(ii) *One point is not on the path constructed.* Then since, at any stage, the points not on the portion so far constructed have suffixes greater than those on the portions constructed, the point not on the path is P_ν. On adding (8), (9), $\cdots$ we obtain for the path $P_1 \cdots P_t$ through the $\nu-1$ points $P_1, P_2, \cdots, P_{\nu-1}$,

$$P_1 \cdots P_t < x_t - x_1 + .7(\nu - 2).$$

But

$$P_tP_\nu < x_\nu - x_t + 1.$$

Hence

$$P_1 \cdots P_tP_\nu < x_\nu - x_1 + .7\nu.$$

(iii) *Two points are not on the path constructed.* For the reason mentioned in (ii), the two points are $P_{\nu-1}$, P_ν. The path constructed, $P_1 \cdots P_t$, satisfies

$$P_1 \cdots P_t < x_t - x_1 + .7(\nu - 3).$$

But

$$P_tP_{\nu-1} < x_{\nu-1} - x_t + 1, \quad P_{\nu-1} P_\nu < x_\nu - x_{\nu-1} + 1.$$

Hence

$$P_1 \cdots P_tP_{\nu-1}P_\nu < x_\nu - x_1 + .7\nu.$$

(iv) *Three points are not on the path constructed.* For the reason mentioned in (ii), the three points are $P_{\nu-2}$, $P_{\nu-1}$, P_ν. The path constructed, $P_1 \cdots P_t$, satisfies

$$P_1 \cdots P_t < x_t - x_1 + .7(\nu - 4).$$

By §5 (replacing P_1, P_2, P_3, P_4 by P_t, $P_{\nu-2}$, $P_{\nu-1}$, P_ν respectively), either there is a path through P_t, $P_{\nu-2}$, $P_{\nu-1}$, P_ν, beginning at P_t, ending at P_j, and satisfying

$$P_t \cdots P_j < x_j - x_t + 2.1,$$

or else

$$P_tP_{\nu-2}P_{\nu-1}P_\nu < x_\nu - x_t + 2.8.$$

In the first case $P_1 \cdots P_t \cdots P_j$ is a path as required; in the second case $P_1 \cdots P_t P_{\nu-2} P_{\nu-1} P_\nu$ is such a path.

We thus see that if we justify the assertions of §5, the lemma will follow.

7. **Proof of lemma completed.** Divide the strip $0 \leqq y \leqq 1$ into four strips

α: $1 \geqq y \geqq 4/5$, β: $4/5 \geqq y \geqq 1/2$, γ: $1/2 \geqq y \geqq 1/5$, δ: $1/5 \geqq y \geqq 0$.

The possible sets $P_1, P_2, \cdots, P_\nu$ can be divided into four classes (or cases) $\alpha, \beta, \gamma, \delta$ according as P_1 is in α, β, γ, or δ. Each class can be divided into four subclasses according as P_2 is in α, β, γ, or δ; and so on. Thus, the class $\alpha\delta\alpha\delta$ is the class in which P_1 is in α, P_2 in δ, P_3 in α, and P_4 in δ. There is some overlapping among the classes, but this is irrelevant.

We need only consider the cases α and β, since γ and δ are obtained from them by a reflection in the line $y=1/2$. The case β is easier than α. In case α we consider the subcases $\alpha\alpha, \cdots, \alpha\delta$ and see if (2) is satisfied. For any subcase in which this is not so, we consider its subcases, and see if one of (3) is satisfied. For any subcase in which this is not so, we consider its subcases, and see if (4) is satisfied. The reader will have an adequate idea of the method, if we consider in detail the most difficult of the subcases, and the only one which requires the consideration of P_5, namely, $\alpha\delta\alpha\delta$.

We use the elementary fact that if AB is a segment which makes an acute angle ϕ with a given line, then

$$AB = AB \cos \phi + AB \sin \phi \tan \frac{\phi}{2} \cdot$$

In the applications, we know upper bounds h, v for $AB \cos \phi$, $AB \sin \phi$ respectively, and an upper bound $\alpha < \pi/2$ for ϕ. We can then infer that

$$AB \leqq h + v \tan \frac{\alpha}{2} \cdot$$

An example of this is when two of the points P_i, say P_1, P_2, both belong to α or to δ. Since, by hypothesis,

$$P_1P_2 > .35, \qquad |y_1 - y_2| \leqq .2,$$

we have $\sin \phi < 4/7$, where ϕ is the inclination of P_1P_2 to the x-axis. Then $\tan (\phi/2) < 1/3$, and

$$P_1P_2 \leqq x_2 - x_1 + |y_2 - y_1|/3 < x_2 - x_1 + .07.$$

The case $\alpha\delta\alpha\delta$. We use the abbreviations

$$d_2 = x_3 - x_2, \qquad d_3 = x_4 - x_3, \qquad d_4 = x_5 - x_4.$$

We have

$$\begin{aligned} P_1P_3P_2P_4 &\leqq x_3 - x_1 + .07 + (1 + d_2^2)^{1/2} + x_4 - x_2 + .07 \\ &\leqq x_4 - x_1 + d_2 + (1 + d_2^2)^{1/2} + .14. \end{aligned}$$

Hence

$$P_1P_3P_2P_4 < x_4 - x_1 + 2.1 \tag{10}$$

if $d_2 \leqq .72$. We may therefore suppose that $d_2 > .72$. Then if ϵ is the inclination of P_1P_3 to the x-axis,

$$\tan \epsilon < 5/18, \qquad \tan \frac{\epsilon}{2} < .14,$$

and so

$$P_1P_3 \leqq x_3 - x_1 + .14 | y_1 - y_3 | < x_3 - x_1 + .03,$$

and similarly,

$$P_2P_4 < x_4 - x_2 + .03.$$

We can therefore repeat the above argument with .03 in place of .07. We then find that (10) holds if $d_2 \leqq .77$. We may therefore suppose that $d_2 > .77$. Then if θ is the inclination of P_2P_3 to the x-axis,

$$\tan \theta < \frac{1}{.77}, \qquad \tan \frac{\theta}{2} < .5, \qquad P_2P_3 < d_2 + .5,$$

and so

$$P_1P_2P_3 < x_3 - x_1 + 1.5. \tag{11}$$

We now consider the four subcases.

Case $\alpha\delta\alpha\delta\alpha$. We have

$$\begin{aligned} P_1P_2P_4P_3P_5 &< x_2 - x_1 + 1 + x_4 - x_2 + .03 + (1 + d_3^2)^{1/2} \\ &\quad + x_5 - x_3 + .07 \\ &< x_5 - x_1 + d_3 + (1 + d_3^2)^{1/2} + 1.1 \\ &< x_5 - x_1 + 2.8 \end{aligned}$$

if $d_3 \leqq .55$. We may therefore suppose that $d_3 > .55$. Then if ϕ is the inclination of P_3P_4 to the x-axis,

$$\tan\phi < \frac{1}{.55}, \qquad \tan\frac{\phi}{2} < .6, \qquad P_3P_4 < d_3 + .6,$$

and, by (11), $P_1P_2P_3P_4$ is a path as required.

Case $\alpha\delta\alpha\delta\beta$. Clearly,

$$P_1P_2P_4P_5P_3 < x_3 - x_1 + 2(x_5 - x_3) + 2 < x_3 - x_1 + 2.8$$

if $x_5 - x_3 \leqq .4$. We may therefore suppose that $x_5 - x_3 > .4$. Then if ψ is the inclination of P_3P_5 to the x-axis,

$$\tan\psi < \frac{5}{4}, \qquad \tan\frac{\psi}{2} < \frac{1}{2},$$

and

$$P_3P_5 \leqq x_5 - x_3 + (1 - y_5)/2 \leqq x_5 - x_3 + .25.$$

Now

$$P_3P_5P_4 \leqq x_5 - x_3 + (1 - y_5)/2 + (d_4^2 + y_5^2)^{1/2} \leqq d_3 + 1.3$$

if

$$d_4 + (1 - y_5)/2 + (d_4^2 + y_5^2)^{1/2} \leqq 1.3. \tag{12}$$

If (12) is satisfied, then, by (11), $P_1P_2P_3P_5P_4$ is a path as required. But (12) is satisfied if $d_4 \leqq 1/3$. It suffices to prove this when $d_4 = 1/3$. For this value of d_4, the first member of (12) is an increasing function of y_5 in the range of y_5, namely, $1/2 \leqq y_5 \leqq 4/5$. It therefore suffices to prove (12) when $d_4 = 1/3$, $y_5 = 4/5$, in which case (12) becomes an equality.

We may therefore suppose that $d_4 > 1/3$. If ω is the inclination of P_4P_5 to the x-axis, then

$$\tan\omega \leqq \frac{y_5}{d_4} < \frac{12}{5}, \qquad \tan\frac{\omega}{2} < \frac{2}{3},$$

and so

$$P_4P_5 < d_4 + 2(y_5 - y_4)/3 < d_4 + .6. \tag{13}$$

Now

$$\begin{aligned} P_1P_2P_4P_3P_5 &< x_2 - x_1 + 1 + x_4 - x_2 + .03 + (1 + d_3^2)^{1/2} \\ &\quad + x_5 - x_3 + .25 \\ &< x_5 - x_1 + 2.8 \end{aligned}$$

if $d_3 \leqq .43$. We may therefore suppose that $d_3 > .43$. Then

$$\tan\phi < \frac{1}{.43}, \qquad \tan\frac{\phi}{2} < .66, \qquad P_3P_4 < d_3 + .66,$$

and, by (11) and (13), $P_1P_2P_3P_4P_5$ is a path as required.

Case $\alpha\delta\alpha\delta\gamma$. As in the preceding case, we may suppose that $x_5 - x_3 > .4$. Then

$$\tan\psi < 2, \qquad \tan\frac{\psi}{2} < .62,$$

$$P_3P_5 \leqq x_5 - x_3 + .62(y_3 - y_5) < x_5 - x_3 + .5.$$

Now

$$\begin{aligned} P_1P_2P_4P_3P_5 &< x_2 - x_1 + 1 + x_4 - x_2 + .03 + (1 + d_3^2)^{1/2} \\ &\quad + x_5 - x_3 + .5 \\ &< x_5 - x_1 + 2.8 \end{aligned}$$

if $d_3 \leqq .23$. We may therefore suppose that $d_3 > .23$. Then

$$\tan\phi < \frac{1}{.23}, \qquad \tan\frac{\phi}{2} < .8, \qquad P_3P_4 < d_3 + .8.$$

Further,

$$P_4P_5 < d_4 + .5.$$

By (11), $P_1P_2P_3P_4P_5$ is a path as required.

Case $\alpha\delta\alpha\delta\delta$. We have

$$P_3P_4P_5 < 1 + d_3 + d_4 + .07,$$

so that, by (11), $P_1P_2P_3P_4P_5$ is a path as required.

QUEEN'S UNIVERSITY, BELFAST