THE MEAN DISTANCE IN 2-SPACE

G. Yuval

June 1976

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa., 15213

An $O(N^2)$ lower bound is proven for the mean distance between N points in 2-space, using methods from complex function theory; but if any finite error is allowed, an O(NlogN) algorithm is shown for computing the mean distance to within that finite error.

This research was supported in part by the National Science Foundation under grant MCS75-222-55 and the Office of Naval Research under contract N00014-76-C-0370, NR 044-422.

Introduction

In this paper, we prove a lower bound for a problem -- finding the mean Euclidean distance between N points in 2-space -- by methods drawn from complex function theory. The crucial fact we shall use is that two complex functions (under the appropriate caveats) can either be identical everywhere, or else they can be identical (almost) nowhere; there is no in-between possibility.

Because of this, a well-behaved function can be continued uniquely everywhere, given its values in any interval, no matter how small. If it has singularities, we can continue it along any path that circumvents them; this may cause the function to become multivalued (depending on whether we circumvent a singularity from the left or from the right), in which case we say it has a many-sheeted Riemann surface†.

The square-root function, used in the definition of the mean distance, has a 2-sheeted Riemann surface; that is, a square root can have 2 values. This causes the mean distance to have very many possible values. Using analytic continuation, we show that <u>any</u> algorithm that computes it on the reals has to compute it everywhere, and must therefore give equally many different results on the different sheets. A simple argument of accounting for all this ambiguity shows that this implies a large complexity.

The methods used here should be applicable to other algorithms that use functions with nontrivial Riemann surfaces (e.g. the <u>harmonic</u> mean distance in 2-space, or the gravitational potential energy of N stars).

I wish to thank D.Jefferson for some most valuable discussions.

†Readers unfamiliar with Riemann surfaces might find [1] useful.

Straight-line algorithms

Let N points (X_i,Y_i) be given in 2-space. The mean distance D between them is defined as

$$D = \frac{1}{M}\sum_{i<j}\sqrt{(X_i-X_j)^2+(Y_i-Y_j)^2}$$

where M is N(N-1)/2.

We can consider D as a function of 2N complex variables. For each of the M square roots, there are (complex) points where that square root alone is singular, and all the others are analytic†. Continuing D on a contour around such a point will reverse the sign of the relevant square root, leaving the others unchanged. Therefore, the Riemann hypersurface of D has 2^M sheets, and each can be continued to any of the others.

Consider now any straight-line algorithm A to compute D, using the five arithmetic operations +,-,*,/ and square-root.

A can be considered to apply to complex inputs as well as real ones, and computes some function f, meromorphic (=analytic, except for poles) on its Riemann surface. On the reals, f=D. We can therefore apply, from ch.II of [2], their

"Theorem 4: If f_1 is analytic in a domain D_1, and f_2 in a domain D_2, if the intersection of D_1 and D_2 is a nonempty domain, and if $f_1(z),f_2(z)$ have equal values in a real environment of a point of $D_1.D_2$, then $f_1(z),f_2(z)$ are analytic continuations of each other; i.e. there exists a unique function f(z) analytic in D_1+D_2 which coincides with f_1 in D_1 and with f_2 in D_2." (N.B. z is a complex vector; the extension to meromorphic functions is trivial, by removing small neighbourhoods of the singularities from the domains D_1,D_2)

We therefore find that f and D are the same complex function. By choosing the appropriate continuation paths, it is seen that they have the same minimal Riemann surface. Therefore, the full analytic continuation of f is 2^M-valued almost everywhere.

This high degree of ambiguity can only result from square-root steps, since all our other operations are unambiguous. Let us try to account for all the ambiguity:

†for the square root of $(X_1-X_2)^2+(Y_1-Y_2)^2$, such a point is given by $X_k=Y_k=k, k>1$; $X_1=1$; $Y_1=2+i$. This does not hold in 1-space.

If, at each square-root operation of A, we choose a branch of the square-root function, f becomes single-valued. Since these choices must give us 2^M different results, there must be at least M of them; thus, A must include M square-root operations. We have thus proved:

<u>Theorem 1</u>: If we are limited to +,-,*,/ and square-root operations, the mean distance in 2-space has complexity $O(N^2)$.

The extension to higher dimension is trivial, since K-space includes 2-space.

In the following sections, we shall remove some of the restrictions on A.

Decisions

If A is not a straight-line algorithm, but includes decision steps, we cannot use the previous section's proof of theorem 1. Let A have some bounded number of decision steps in it. Each decision can go one of two ways, and therefore, depending on the input data, A behaves like one out of a finite family of straight-line algorithms $A_1, A_2, \ldots$. For every real input set, at least one of the A_i gives the right answer D. Let A_i compute the function f_i, meromorphic on its Riemann hypersurface.

We shall now prove, by <u>reductio</u> <u>ad</u> <u>absurdum</u>, that at least one of the f_i is equal to D everywhere. For suppose not. Then none of the functions f_i-D is idntically zero. By theorem II.4 of [2], they cannot vanish on any real neighbourhood. Since each f_i-D is continuous almost everywhere†, every real neighbourhood (i.e. non-empty open set) contains a neighbourhood in which it nowhere vanishes (since it contains a point at which f_i-D is continuous and nonzero).

Let us now choose some neighbourhood N_0. It must contain a neighbourhood N_1 in which f_1-D nowhere vanishes. This N_1 must contain a neighbourhood N_2 in which f_2-D nowhere vanishes. Continuing, we find a neighbourhood in which none of the f_i-D ever vanishes, but their product is identically zero; which is absurd. Q.E.D.

Therefore, A must reduce, for some inputs, to a straight-line algorithm that gives D everywhere, and theorem 1 applies even if decisions are allowed.

In particular, since the absolute value function (on the reals) can be computed by decisions and arithmetic steps, it does not help in speeding up the computation of the mean distance.

†technically, it is continuous in an open set, dense in real 2N-space.

Higher-order roots

It might seem that theorem 1 is limited to a computation model that allows +,-,*,/, and square roots only, and that higher-order roots may get round the lower bound of M operations. But permutation-group considerations forbid this:

Consider the set of 2^M values of (1) in some singularity-free region. The various possible closed paths, from this region to itself, along which analytic continuation is carried out, generate a group of permutations between these 2^M values. All elements of this group are of characteristic 2; that is, if we go twice along any closed path, every sheet of the Riemann hypersurface returns to itself.

We can do the analytic continuation, along a path, not only of the final values given by algorithms A' and T, but also of every intermediate value. If, at some stage of A', we take (e.g) a 4-th root, the result of this step has a fourfold multiplicity (=number of branches), but the structure of that multiplicity is all wrong: the 4 values of this step divide into two pairs, and there is a continuation path such that going twice around it will lead from one member of the pair to the other†.

In the final result, such a double path leaves everything invariant; the extra multiplicity is therefore "wasted", and this step only contributes a factor of 2 to the multiplicity of the final result, since that result can only depend on whether we choose an odd, or an even, branch of the root, but not on which of the even branches, (or which of the odd ones), we choose.

Therefore, high-order roots (and, for similar reasons, logarithms) have no effect on theorem 1. They can be used to reduce the number of square roots, but not of total operations.

†If, for instance, we take the 4-th root of 1, the 2 pairs are ± 1 and $\pm i$.

References

[1] Bieberbach, Conformal Mapping, New York 1953.

[2] Bochner and Martin, Several Complex Variables, Princeton 1948.

G.Yuval,comp.sci.,CMU,15213

Fast approximations to the mean distance in 2-space

Keywords: mean distance, approximation

Introduction

It has recently [1] been shown that the mean euclidean distance D between N points in 2-space requires $O(N^2)$ steps to compute. In this paper, we shall show that, if any finite relative error ϵ is permitted, there is an O(NlogN) algorithm that computes D to within $+\epsilon D$.

To do this, we define a series of metrics for 2-space. The mean distance for each metric takes only O(NlogN) steps, and these metrics approximate the Euclidean metric to within closer and closer tolerances.

The results of this paper have been derived independently by D. Jefferson (private communication).

A foolish metric

The crudest metric we shall consider defines the length |V| of the vector V=(X,Y) to be the absolute value of X. The rule |aV| = a|V| (a positive scalar, V vector) does hold for this metric, but the unit "circle" is an infinite strip, parallel to the Y axis, of width 2. Let us call this metric F.

Fancier metrics

If we rotate this foolish metric F (i.e. use the projection on some line different from the X-axis), we get a different, equally foolish, metric. The length of a vector, under all such metrics, is a convex, piecewise linear, function of the vector. This property (and also |aV|=a|V|), remains true if we add two or more such metrics, defining the new length of a vector to be the sum of its lengths under the different metrics we add. Therefore, the new unit "circle" is a convex polygon; in general, it will have twice as many edges as there are metrics added up. This can be verified by considering which combinations of the linear parts of the various foolish metrics are possible.

If we rotate F by $K*180^0/L$ for K=0,...,L-1, and add up the resulting metrics, the unit "circle" will be a regular 2L-gon†. As L gets bigger, it will look more and more like a circle‡. Since |aV|=a|V|, this is enough to ensure that these metrics (scaled by a suitable multiplicative constant) converge to the Euclidean metric.

†For L=2, we get the L^1 norm |V|=|X|+|Y|.

‡This follows from its convexity and 2L-fold rotational symmetry

Finding the mean distance

Since, under all our metrics, the length of a vector is the sum of its F measure and of similar measures, it is enough to compute the mean F-distance, which is

$$\frac{2}{N(N-1)} \sum_{i<j} |X_i - X_j| \quad .$$

If we sort the N points by their X coordinate, this number equals

$$\frac{2}{N(N-1)} \sum_{i<j} (X_j - X_i) \quad .$$

If we collect terms, we find that each X_i appears as many times as the number of j-s which are less than i, minus the number of j-s which are greater than i. The sum therefore equals

$$\frac{2}{N(N-1)} \sum_{i} (2i-N-1)X_i \quad .$$

Thus, the mean F distance (and therefore, the mean distances under all our metrics) can be found in O(NlogN) steps.

References

[1] G.Yuval, The complexity of the mean distance in 2-space, submitted to Inf.Proc.Letts.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

THE MEAN DISTANCE IN 2-SPACE

5. TYPE OF REPORT & PERIOD COVERED

Interim

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

G. Yuval

8. CONTRACT OR GRANT NUMBER(s)

N00014-76-C-0370, NR 044-422

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Carnegie-Mellon University
Computer Science Dept.
Pittsburgh, PA 15213

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Office of Naval Research
Arlington, VA 22217

12. REPORT DATE

June 1976

13. NUMBER OF PAGES

12

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An $O(N^2)$ lower bound is proven for the mean distance between N points in 2-space, using methods from complex function theory; but if any finite error is allowed, an O(N log N) algorithm is shown for computing the mean distance to within that finite error.