On Finding Maximum-Cardinality Symmetric Subsets

Peter Brass

Free University Berlin, Institute of Computer Science,
Takustrasse 9, D-14195 Berlin, Germany,
brass@inf.fu-berlin.de

Abstract. In this paper we study the complexity of the problem of finding a symmetric subset of maximum cardinality among n point in the plane, or in three-dimensional space, and some related problems like the largest repetitive or k-repetitive subsets. For the maximum-cardinality symmetric subset problem in the plane we obtain an algorithm of complexity $O(n^{2+\frac{1}{3}} \log n)$.

1 Results

Finding the symmetries of a set of n points, or more general testing two sets for congruence and finding all congruence mappings between them, is an old and well-studied problem [3, 2, 4, 7–9, 11], which is solved satisfactorily in dimensions two and three ($O(n \log n)$) and remains an interesting problem in higher dimensions. There are at least two ways to make the problem more realistic: allowing for errors in the points (Hausdorff-approximate symmetry) and for errors in the sets (large symmetric subsets). The Hausdorff-approximate symmetry recognition, however, is NP-complete [12], whereas the identification of large symmetric parts in the exact model leads to interesting problems, which are related to combinatorial geometry in a way already apparent in several other exact point pattern matching problems [1, 6].

There are several ways to formalize the notion of 'large symmetric parts' contained in a point set. The most obvious is to determine the largest-cardinality subset with a nontrivial symmetry (Figure 1 shows a set, the largest-cardinality symmetric subset, and another symmetric subset). For this problem Eades [8] gave an $O(n^4 \log n)$-algorithm.

Theorem 1. *The largest-cardinality symmetric subset of a set of n points in the plane can be determined in $O(n^{2+\frac{1}{3}} \log n)$ time.*

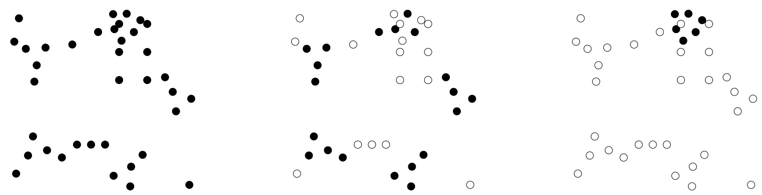

Figure 1

Our algorithm lists as an intermediate result all regular polygons contained in that set. It is remarkable that this can indeed be done in that time, since for each fixed k there are sets of n points containing $c_k n^2$ regular k-gons [10, 14].

A different formalization is to ask for the largest subset Y of the given set X that is repeated: there is a nontrivial motion φ with $Y \subset X$ and $\varphi(Y) \subset X$; or that is r-fold repeated: $Y \subset X$, $\varphi(Y) \subset X$,. . .,$\varphi^r(Y) \subset X$. (Figure 2 shows a set, a 5-fold repeated subset, and a once repeated subset.) This notion captures parts of some bigger symmetric structure, e.g. some finite part of an infinite frieze group symmetry. The special case of equidistant collinear rows of points (ϕ a translation, Y only one point) was also studied previously [5, 13, 16].

Theorem 2. *The largest r-fold repeated subset of a set of n points in the plane can be determined in $O(n^{3+\frac{1}{5}} \log n)$ for $r = 1$ and $O(n^{2+\frac{1}{3}} \log n)$ for $r \geq 2$.*

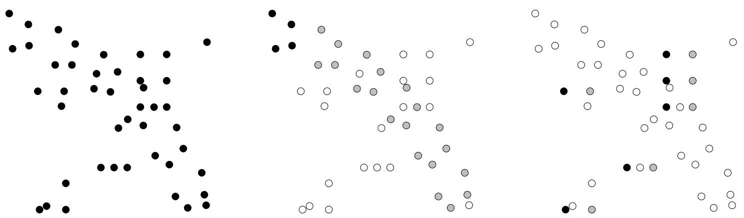

Figure 2

Essentially the same algorithm works for both problems also in three-dimensional space (but not in higher dimensions) where we get a time bound $O(n^3 \log n)$.

2 The basic algorithm

In all the above cases symmetries by reflections are simple, and can be enumerated trivially in $O(n^2 \log n)$ time, since we have to look only at the $\binom{n}{2}$ possible pairs of points that can be exchanged by a reflection, and see which reflection line occurs most frequently. So in the following we will only look for rotation symmetries. Also, the algorithms for the different problems are almost the same (with an important difference only in the case of finding one time repeated sets), so we will give only the first, and state the necessary modifications later.

Given a set X of n points in the plane, the algorithm maintains two search tree data structures, one ($\mathcal{T}$) for isosceles triangles in X (point triples (a, b, c) with $d(a, b) = d(b, c)$) and the other ($\mathcal{S}$) for possible symmetry operations (pairs (p, k) of a centerpoint and a rotation order, with the current number of points $\#(p, k)$ in that symmetric subset and their list appended. The algorithm does the following:

1. Determine for each pair $x_1, x_2 \in X$ the distance $d(x_1, x_2)$, and collect the point pairs having the same distance d to get a partition of $\binom{X}{2}$ into distance graphs G_d (d a distance occuring in X).

2. For each distance d and each point $x \in X$, take each pair of neighbours y_1, y_2 of x in G_d, and insert the tripel (y_1, x, y_2) in a search tree $\mathcal{T}$.

3. As long as $\mathcal{T}$ is not empty, repeat

3.1 Choose any tripel (a, b, c) from $\mathcal{T}$, delete it from $\mathcal{T}$.

3.2 Determine the rotation ψ that maps $a \mapsto b$, $b \mapsto c$.
This rotation determines a polygonal arc $p_0 p_1 p_2 \dots$ with $p_0 = a$, $p_1 = \psi(a) = b$, $p_2 = \psi^2(a) = c$, and generally $p_i = \psi^i(a)$. This may be a regular polygon, an orbit under ψ.

3.3 Construct the sequence of isosceles triangles (p_i, p_{i+1}, p_{i+2}), checking for each of these triangles whether it is contained in $\mathcal{T}$, and deleting it from $\mathcal{T}$, until we either find a triangle that is not contained in $\mathcal{T}$, or arrive at the starting triangle $(p_k, p_{k+1}, p_{k+2}) = (a, b, c)$.

3.3.1 In the first case (the polygonal arc remained incomplete), discard (a, b, c) and the polygonal arc and return to step 3.

3.3.2 In the second case (the polygonal arc closed to a regular k-gon in X), determine the center p of this regular polygon.

3.3.2.1 Insert (p, k) in $\mathcal{S}$, if it does not already exist, increase $\#(p, k)$ by k, and append the regular polygon to the list.

3.3.2.2 For each $j \in \{2, \dots, k\}$ that is coprime to k and each $i = 1, \dots, k$ delete (p_i, p_{i+j}, p_{i+2j}) (all indices $\bmod k$) from $\mathcal{T}$.

4. Traverse $\mathcal{S}$ and determine the element (p, i) for which $\#(p, i)$ is maximal. Output the list of all regular i-gons with center p.

3 Correctness

To determine the maximum cardinality of a subset that has a nontrivial rotation symmetry, we use that each set with a k-fold rotation symmetry is the union of concentric regular k-gons, the orbits of the points under the symmetry. Thus we have only to find all regular polygons, collect those polygons which have the same center and the same order, and determine the point which occurs as common center of the largest group.

To find all regular polygons $r_1 \dots r_k$ contained in the set, we use that any three vertices r_i, r_{i+a}, r_{i+2a} (e.g. three consecutive vertices r_i, r_{i+1}, r_{i+2}) form an isosceles triangle, and for each isosceles triangle in our set there is at most one minimal regular polygon containing them in this way. And this polygon can be found by just following the polygonal arc defined by the rotation around the intersection point of the midperpendiculars of that triangle that maps the first leg of the isosceles triangle on the second. In each path-following step we remove the isosceles triangle we just used from the set of all isosceles triangles, so in the end we either find a regular polygon, or we have removed some partial polygonal arc which does not extend to any regular polygon in our set, and which therefore can be removed.

It remains to avoid that we find the same polygon several times, since each isosceles triangle completes to at most one minimal regular polygon containing that triangle, but the same polygon will be obtained with different numberings of vertices from different isosceles triangles. E.g. a regular pentagon $p_1 \dots p_5$

will be found by following $p_1p_2p_3$ and completing that arc, but will again be found by following $p_1p_3p_5$ and completing that arc $(p_1p_3p_5p_2p_4)$. So after we found the regular polygon $p_1 \dots p_r$ we have to remove all other isosceles triangles $p_ip_{i+a}p_{i+2a}$ which generate the same polygon, which is done in step 3.3.2.2. The same polygon is found exactly for those a which are coprime to the vertex number k of the regular polygon; if k has a nontrivial divisor κ, then the regular k-gon can also be interpreted as union of $\frac{k}{\kappa}$ regular κ-gons, and will be found and stored in $\mathcal{S}$ as set with κ-fold rotational symmetry again. Thus in the structure $\mathcal{S}$ the same set is stored for each symmetry order exactly once, and a simple traversal of $\mathcal{S}$ gives the largest subset with a nontrivial rotational symmetry.

4 Analysis

The construction of the distance graphs in step 1. can be trivially done in $O(n^2 \log n)$ time. If the distance graphs are given, the construction of all isosceles triangles (step 2) can be done in time $O(n^2 + I \log n)$ where I denotes the number of isosceles triangles that are constructed (trivially $I < n^3$). In each of the following steps 3.* one of the isosceles triangles is removed from $\mathcal{T}$, which takes $O(\log n)$, and some further operation of complexity at most $O(\log n)$ is done. Thus the total complexity of step 3 is $O(I \log n)$. Step 4 finally also takes at most $O(I \log n)$ time, since we touch each regular polygon at most once, and there are at most $O(I)$ regular polygons. Thus the total complexity is $O((n^2 + I) \log n)$. Since it is known that $I = O(n^{2+\frac{1}{3}})$ [15] we obtain the upper bound $O(n^{2+\frac{1}{3}} \log n)$ claimed in Theorem 1. Indeed bounding the number of isosceles triangles is simple: the points completing a given base pq to an isosceles triangle are on the mid-perpendicular of pq, so the isosceles triangles containing a fixed point p as base vertex can be counted by constructing the mid-perpendiculars of the segments pq for all other points q, and finding the incidences between all these $n-1$ lines and the $n-1$ points, which is known to be $O(n^{\frac{4}{3}})$; taking the sum over all p gives the $O(n^{\frac{7}{3}})$ bound for I.

A further speedup (perhaps to $O(n^{2+\varepsilon})$) would be possible if one could avoid inserting *all* isosceles triangles into $\mathcal{T}$. Only those triangles are really needed that can occur in a regular polygon; so one really needs only those isosceles triangles with an angle of form $(1-\frac{2}{k})\pi$ at the apex, all others can never lead to a regular polygon. For a fixed vertex x, having a given list of $y_1, \dots, y_m$ neighbours at a fixed distance, it seems probable that there are much less than $\binom{m}{2}$ pairs y_i, y_j which determine an angle of that form $(|\angle y_i x y_j| \in \{\frac{1}{3}\pi, \frac{1}{2}\pi, \frac{3}{5}\pi, \dots, \frac{k-2}{k}\pi, \dots\})$. I have a construction of points $y_1, \dots, y_m$ with $cm \log m$ such pairs, and believe this to be near the correct order. If those pairs could be determined in $O(m^{1+\varepsilon})$, it would allow a speedup of the whole algorithm to $O(n^{2+\varepsilon})$.

The same algorithm works also in three-dimensional space, since the possible symmetries there are also reflections (determined by one point pair, so can be checked in $O(n^2 \log n)$ time) and rotations around a line; so the nontrivial orbits are regular polygons in space. Unfortunately, the bound for the number of isosceles triangles in three-dimensional space is only $O(n^3)$, and that order can be reached (take half of the points on a circle and the other half on the mid-perpendicular of that circle, than any triangle of two points on the circle and one point on the mid-perpendicular is isosceles).

5 Variants for repeated sets

If we are looking for repeated sets it does make a big difference whether we are also interested in sets Y which are once repeated ($Y \subset X$ and $\varphi(Y) \subset X$ for a nontrivial motion φ), or accept only those Y that occur at least r further times, $r \geq 2$. If $Y \subset X$, $\varphi(Y) \subset X$, $\varphi^2(Y) \subset X, \dots \varphi^r(Y) \subset X$, then for each $y \in Y$ the triangle $y, \varphi(y), \varphi^2(y)$ is isosceles, and we can again just follow the paths determined by the motion φ, where we determine the motion from the isosceles triangle. There are, however, two important differences:

- *Two motions are possible*
  If we are looking for complete orbits of isometries, then the only possibe isometries are rotations (and reflections, which have only two-point orbits). If we are also interested in pieces of infinite orbits, then additionally translations and glide-reflections become possible. By this the continuation of an isosceles triangle as an orbit is not anymore unique, but can be a regular polygon (by a rotation) or a zig-zag path (by a glide-reflection).
  To overcome this, we have to insert two copies of each isosceles triangle in $\mathcal{T}$, marked as 'rotation' and 'glide-reflection', and remove the right copy when extending a path.
- *The paths do not close*
  As long as we were looking only for complete regular polygons, we found the whole polygon by just going around. If we also take polygonal arcs, we have to make sure that we remove *maximal* polygonal arcs from $\mathcal{T}$. So we have to follow the path generated by φ from the starting triangle in both directions, forward (φ) as well as backward (φ^{-1}).

In this way we obtain all subsets which are partial orbits of at least three points of some isometry. Each regular polygon found this way should be inserted in $\mathcal{S}$ under the appropriate isometry, and with its full number of points, and all copies of that regular polygon should be deleted as in step 3.3.2.2. If the partial orbit is only a path of length k, and we are looking for an r-fold repeated subset, then it should be discarded if $k \leq r$, otherwise the first $k - r$ points of the path should be inserted in $\mathcal{S}$ under the appropriate isometry. Then in the end again a simple traversal of $\mathcal{S}$ is sufficient to find the maximum-cardinality r-fold repeated subset.

If we are looking for sets $Y \subset X$ which are only once repeated ($Y \subset X$, $\varphi(Y) \subset X$), then the partial orbit of a point consists only of two points, and does not anymore determine the motion. Instead we have to look at the possible images of pairs of points, and count how often which motion φ is determined by them, obtaining essentially the same algorithm as Akutsu, Tamaki and Tokuyama [1] for the 'maximum congruent subsets' problem.

For a pair $y_1, y_2 \in X$ the possible image pairs $z_1 = \varphi(y_1), z_2 = \varphi(y_2)$ are the pairs that have the same distance, so we construct the distance graphs of X, take each pair of edges $(y_1, y_2), (z_1, z_2)$ of the same length, determine the motion that maps the first pair on the second, and increase the count of this motion and its reflected counterpart in a search structure $\mathcal{S}$ for isometries. Then the nontrivial isometry with the largest count gives the maximum-cardinality subset $Y \subset X$ that is once repeated.

The analysis of that algorithm is the same as in [1], giving an $O(n^{3.2} \log n)$ complexity.

References

1. Akutsu, T., Tamaki, H., Tokuyama, T.: Distribution of distances and triangles in a point set and algorithms for computing the largest common point set. Discrete Comput. Geom. **20** (1998) 307–331
2. Alt, H., Mehlhorn, K., Wagener, H., Welzl, E.: Congruence, similarity, and symmetries of geometric objects. Discrete Comput. Geom. **3** (1988) 237–256
3. Alt, H., Guibas, L.: Resemblance of geometric objects. In J.-R. Sack, J. Urrutia (Eds.): Handbook of Computational Geometry Elsevier 1999, 121–153
4. Atallah, M.J.: On symmetry detection. IEEE Trans. Comput. **34** (1985) 663–666
5. Boxer, L., Miller, R.: Parallel algorithms for all maximal equally-spaced collinear sets and all maximal regular coplanar lattices. Pattern Recog. Letters **14** (1993) 14–20
6. Braß, P.: Exact point pattern matching and the number of congruent triangles. Algorithms — ESA 2000 (Mike Paterson, Ed.) LNCS 1879, 112–119
7. Braß, P., Knauer, C.: Testing the congruence of d-dimensional point sets. ACM Symposium on Comput. Geom. 2000, 310–314
8. Eades, P.: Symmetry finding algorithms. In: Computational Morphology. A computational geometric approach to the analysis of form. (Toussaint, Godfried T., Ed.) Machine Intelligence and Pattern Recognition Series 6 (1988) North-Holland, 41–51
9. Eades, P., Ng, H.C.: An algorithm for detecting symmetries in drawings. Ars Comb. **23A** (1987) 95–104
10. Elekes, G., Erdős, P.: Similar configurations and pseudo grids. In: Intuitive geometry 1991 (K. Böröczky et al, Eds.) North-Holland, Colloq. Math. Soc. Janos Bolyai. **63** (1994) 85–104
11. Highnam, P.T.: Optimal algorithms for finding the symmetries of a planar point set. Inf. Proc. Letters **22** (1986) 219–222
12. Iwanowski, S.: Testing approximate symmetry in the plane is NP-hard. Theor. Comput. Sci. **80** (1991) 227–262
13. Kahng, A.B., Robins, G.: Optimal algorithms for extracting spatial regularity in images. Inf. Proc. Letters **12** (1991) 757–764
14. Laczkovich, M., Ruzsa, I.Z.: The number of homothetic subsets. In: The mathematics of Paul Erdős. Vol. II. (R.L. Graham, et al, Eds.) Springer, Algorithms Comb. 14 (1997) 294–302
15. Pach, J., Agarwal, P.K.: Combinatorial Geometry. Wiley, New York 1995
16. Robins, G., Robinson, B.L., Sethi, B.S.: On detecting spatial regularity in noisy images. Inf. Proc. Letters **69** (1999) 189–195