**Two snowball algorithms in metric space**

By A.R.J. Marchand

CNRS UMR 5287, Bordeaux

I introduce here two novel "snowball" algorithms. One of them very efficiently solves the problem of the minimum enclosing disk over a set of points, the other algorithm addresses the question of the maximal distance between two points in a set in Euclidean space of any dimension. Both algorithms rely on distance only (not necessarily Euclidean distance) and tend to run in linear time.

Metric space is a set of elements equipped with a notion of distance. The most familiar example is Euclidean space, with distance defined as the square root of the sum of squares of differences between point coordinates. Many algorithms take advantage of the topology of metric space based on one of the properties of a distance of any type, namely the triangular inequality. In Euclidean space, it is furthermore possible to use angles and distances along abstract lines to partition the explored space into several parts. This allows algorithms of the “divide and conquer” type to reduce the combinatorial complexity of problems by solving them on a smaller number of elements in each part.

I introduce here two novel algorithms of a class I dubbed "snowball". Snowball algorithms rely heavily on the notion of distance, but this distance need not be Euclidean. Snowball algorithms search cyclically through the set of elements to be explored, keeping only a small subset, while maximizing some function of distance on this subset. The search is limited to the first element that satisfies a particular condition, and the algorithm terminates when no such element can be found. These characteristics tend to yield algorithms that run in approximately linear time.

1. **Minimal enclosing disk**

Over several decades, progress has been made in reducing the complexity of algorithms for the minimal enclosing disk (Elzinga & Hearn, 1972; Megiddo, 1983; Skyum, 1991; Welzl, 1991; Efrat, Sharir & Ziv, 1994; Har-Peled & Mazumdar, 2005; Yildirim, 2008; Gao & Wang, 2018; Smolik & Skala, 2022). A currently standard solution is that proposed by Welzl (1991). Welzl’s algorithm runs in *O*(*n*), meaning that the number of steps required to solve the problem increases linearly with the number of points. Its performance is however hampered by its recursive nature, which stems from the requirement to include at each step all the points that have been examined so far. The snowball algorithm presented here avoids this requirement.

The algorithm relies on the well-known properties of the minimal circle:

* It is unique for a given set of points in the plane;
* It is determined by three points on its circumference if these three points form a triangle that is not obtuse. Otherwise, two points suffice, determining a diameter of the minimum circle.

Therefore, a minimal circle enclosing any sample of four points can always be defined by some subset of two or three of them (forming a diameter or a non-obtuse triangle). One and maybe two points are unnecessary. Adding new points to the necessary subset provides a new sample of four points. This is the heart of the snowball algorithm, as well as of the earlier algorithm of Elzinga & Hearn (1972). Note that in general, the unnecessary points should not be eliminated, as they may be needed at a later step.

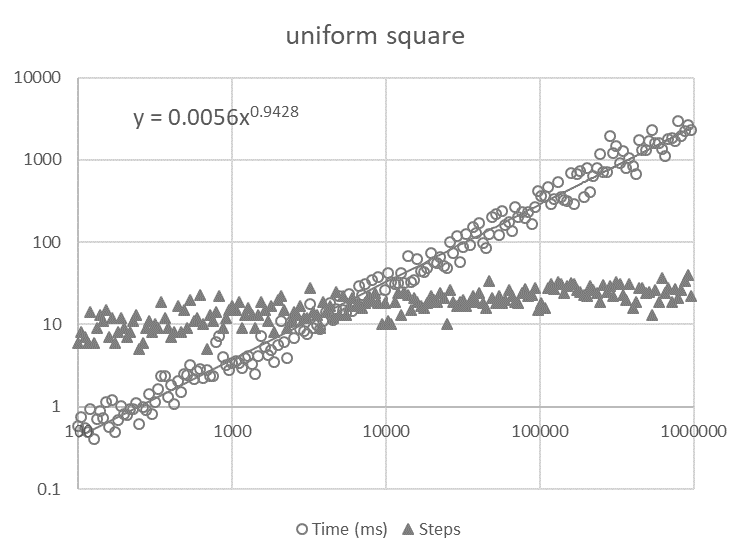
At each step of the snowball algorithm, after a minimal circle has been determined for by two or three points, the sample is complemented to four by the *first point(s) found outside* the circle in the whole set. The set is *always searched forward* and when it is exhausted, the search starts again at the first point. Unlike in Elzinga & Hearn (1972), no point is tested and eliminated. Unlike in Welzl (1991), no attempt is made to recursively ensure that the current circle includes all previously explored points.

**Correctness**

The correctness of the algorithm is proved by the following four properties:

1. At each step, the circle defined by the necessary subset is the minimal circle for the subset as well as the minimal circle for the sample four points. So *if* a circle obtained in this way ultimately encloses all the points in the analyzed set, then it must necessarily be minimal enclosing circle for the whole set;
2. At each step which includes a new point outside the circle, the radius of the new enclosing circle strictly increases;
3. The algorithm only stops when no point can be found outside the circle;
4. The set of possible circles is finite and includes the unique solution.

**Performance**



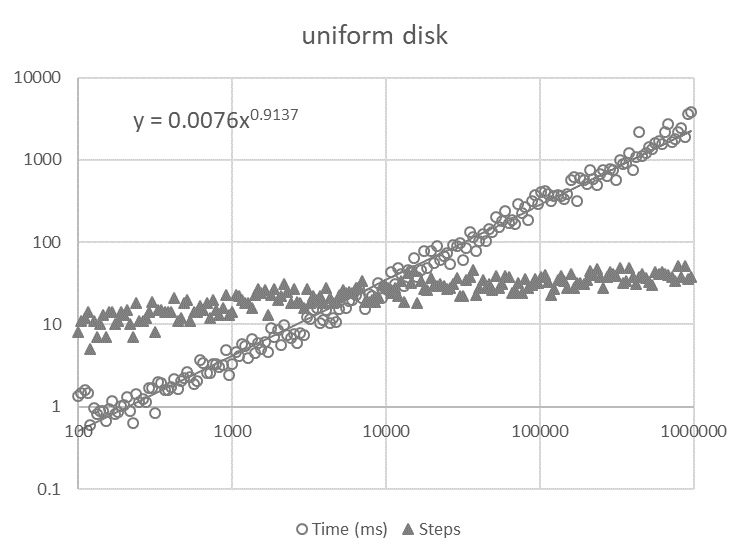


Figure 1. Time complexity and number of steps evaluated for n points uniformly distributed in a square (upper panel) or a disk (lower panel). Each symbol represents a simulation for increasingly larger sets. Execution time increased as N0.94 and N0.91 respectively. The number of steps increased much more slowly to approximately 30 and 45 respectively for N=106 points.

The performance of the snowball algorithm was simulated with the Python script shown below (Annex 1). The efficiency of the method was remarkable. With n points uniformly distributed in a square or a disk (Fig. 1), execution time increased in an approximately linear way in the range N=102 to 106. This is empirical evidence that the algorithm belongs to class *O*(*n*).

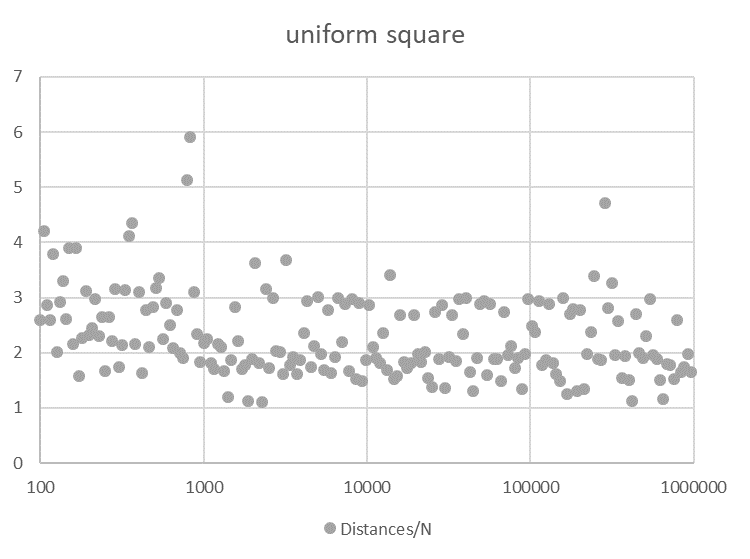


Figure 2. Total number of distances computed relative to the size N of the set. For large sets, this number comes close to 2N.

Moreover, the radius of the circle increased rapidly at each step, so that the number of steps was quite small for large sets. Time performance was thus dominated by the search for points outside the current circle, i.e. computing the distance (squared) from the center to the next points.

The total number of distances computed was close to 2N (Figure 2). This value (2.21 and 2.08 for n=5000) compares favorably with Welzl’s minidisk algorithm, which reports an average of 4.9 and 5.6 distance computations per point (MTFBALL algorithm).

Other trials with Gaussian radial or annular distributions and with elongated distributions yielded essentially the same results.

**Discussion**

It is not immediately obvious that the snowball algorithm is correct or efficient. Some of the unnecessary points at step n may be necessary at a later step. This may have led earlier researchers to disregard this relatively straightforward method. However, it can be proved that all points get eventually enclosed and that the circle found is the minimum one.

Not only is the algorithm correct, but its performance was found to be exceptional. Its time complexity appears to be *O*(*n*) and it requires fewer distance computations than the widely recognized Welzl (1972) algorithm.

The recursive nature of Welzl’s algorithm precludes its use for very large sets because of the depth of recursion and the necessity to recursively store subsets. The snowball algorithm does not suffer from this limitation and can therefore accommodate larger sets. Points may be accessed sequentially which is another simplification.

The algorithm requires little working memory and few operations because only the current subset of two or three points and the center and radius of their enclosing disk need to be kept in memory at each step. The computation of the minimal circle of four points requires more calculations than a simple distance check, but since the radius increases in a small number of steps relative to the total number of points, most of the computing time is taken by distance computations to verify the inclusion of single points (essentially two multiplications and four additions).

Of course, performance will necessarily degrade if the set of points exhibits a sequential spatial trend. This issue was already raised by Welzl who indicated that it is sufficient to randomize the set at the start. Provided that the trend is not periodic, it is even possible to avoid the cost of randomizing by browsing the set with a large step p so that p and N are relative primes and p is not too close to a divisor of N. Combined with the cyclical search, it guarantees that all points will be explored in N steps with no additional cost compared to a unit step.

It should be noted that attempts to speed up the search by eliminating points inside the convex hull, as proposed for instance by Skyum (1991) or Smolik & Skala (2022), were unsuccessful. The snowball algorithm tests each point approximately twice over the whole process. This leaves little room for improvement. Indeed, just to verify that the final circle encloses all points already requires at least one distance measurement per point. So, it is not clear that a point can be eliminated with a cost of less than two distance computations.

Finally, the snowball algorithm holds promises for the same generalizations as Welzl’s to higher dimensions and to ellipsoids.

1. **Maximal distance within a set**

A brute force search for the maximal distance in a set of size M requires M(M-1)/2 distance calculations, which correspond to class *O*(*n²*). Faster algorithms have been described, such as that of Skala & Majdisova (2015) which runs in in Euclidean space of dimension two.

Based on the performance of the previous snowball algorithm, the first objective was here to obtain an algorithm running in *O*(*n*). The second objective was to have a simple algorithm applicable to higher dimensions.

We start with any points A and B considered as the diameter of a hypersphere. If the hypersphere encloses all points, the distance AB is the solution, because no pair of points enclosed in the hypersphere can be further apart than the diameter AB. The first point C found not to be included in the hypersphere is used to look for the first point D such that the distance CD is strictly larger than the distance AB. If no point D is found, C is eliminated and replaced by the next point outside the hypersphere. If D is found, CD is substituted to AB and the process is iterated. The set is *always searched forward* and when it is exhausted, the search cyclically starts again at the first point.

The largest distance is unique, but the two points found as a solution may not be unique. Other points may be found by removing one or the other of these two points and starting the search again.

**Correctness**

The snowball algorithm relies firstly on the fact that for a distance CD to be larger than AB, at least one of the points C must lie outside the hypersphere of diameter AB. Secondly, because CD>AB, the radius of the hypersphere is strictly increasing at the next step. Thirdly, the algorithm only stops when no point can be found outside the hypersphere, which is why unsuitable points C must be eliminated. Together, these conditions guarantee the correctness of the algorithm, but not its performance.

**Performance**

The performance of the snowball algorithm was simulated with the Python script shown below (Annex 2). With n points uniformly distributed in a square or a disk (Fig. 1), execution time increased in an approximately linear way in dimensions 2 to 4 with 102 to 105 points, and slightly more than linearly in dimensions 5 to 7. This is empirical evidence that the algorithm belongs to class *O*(*n*).

The number of distances computed per point started around 6.5 in 2D and increased roughly as the square of the number of dimensions.

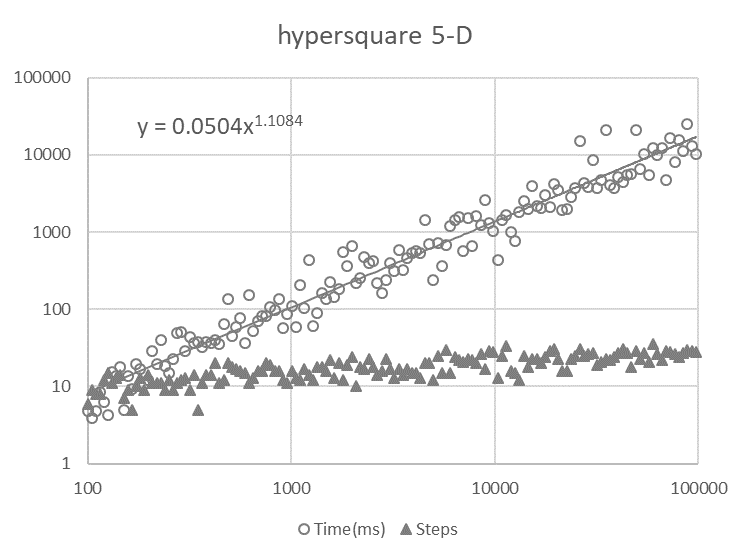


Figure 3. Time complexity and number of steps evaluated for n points uniformly distributed in a hypersquare in 5 dimensions. Each symbol represents a simulation for increasingly larger sets. Execution time increased as N1.1084 i.e. slightly more than linearly. The number of steps increased much more slowly to approximately 27 for N=105 points.

Figure 4. Total number of distances computed relative to the size N of the set. For large sets, this number averaged 6.3, 11.7, 21.1, 40.3, 61.5 and 111 times N in dimensions 2 to 7 resp.

**Discussion**

The algorithm presents analogies with the algorithm for the minimum circle, in particular because it relies on a strictly increasing radius, it uses a forward-only search, and because the search stops as soon as a point fulfilling a condition is found.

In all the cases examined, the snowball algorithm largely outperforms the brute force one in terms of the number of distances computed, with the exception of very small sets in high dimension.

The algorithm requires little working memory and few operations because only the current subset of two points and the center and radius of their enclosing disk need to be kept in memory at each step. It is subject to the same randomization constraints as other algorithms if the ordered set presents a spatial trend.

An advantage of snowball algorithm is that they may be applied to a variety of metric spaces, not necessarily Euclidean.

**References**

Efrat, A., Sharir, M., Ziv, A. (1994). Computing the smallest k-enclosing circle and related problems. Computational Geometry, 4(3), 119–136.

Elzinga, J., Hearn, D. W. (1972). "The minimum covering sphere problem", Management Science, 19: 96–104.

Gao, S.,Wang, C. (2018). A new algorithm for the smallest enclosing circle. In: 2018 8th International Conference on Management, Education and Information (MEICI 2018), Atlantis Press.

Har-Peled, S., Mazumdar, S. (2005). Fast algorithms for computing the smallest k-enclosing circle. Algorithmica, 41(3), 147–157.

Megiddo, N. (1983). Linear-time algorithms for linear programming in R3 and related problems. SIAM Journal on Computing, 12(4), 759–776.

Skyum, S. (1991). A simple algorithm for computing the smallest enclosing circle. Information Processing Letters, 37(3), 121–125.

Skala, V., Majdisova, Z. (2015). Fast Algorithm for Finding Maximum Distance with Space Subdivision in E2. In: Zhang, YJ. (eds) Image and Graphics. ICIG 2015. Lecture Notes in Computer Science, vol 9218. Springer, Cham.

Smolik, M., Skala, V. (2022). Efficient Speed-Up of the Smallest Enclosing Circle Algorithm. Informatica, 33(3): 623-633.

Welzl, E. (1991). Smallest enclosing disks (balls and ellipsoids). In: Maurer, H. (Ed.), New Results and New Trends in Computer Science, Lecture Notes in Computer Science, Vol. 555. Springer, Berlin, Heidelberg.

Yildirim, E. A. (2008). Two algorithms for the minimum enclosing ball problem, SIAM J. Optim., 19: 1368–1391.

**Annex 1: Python procedure for the minimum disk algorithm**

def minimum\_disk (points, subset):

"""

compute minimum circle enclosing points

Parameters: points, a list of points, tuples of 2 float

subset, a list of four points to start with

Returns: subset, a list of two or three points

O, a point, tuple of 2 float, center of the circumscribed circle

r\_squared, a float

"""

while True:

# try solving with 2 points

P, s, [A, B], ok = min\_diameter(subset)

if ok:

O, r\_squared = P, s

subset = [A, B]

# solve with 3 points

else:

O, r\_squared, [A, B, C] = min\_circumscribed(subset)

subset = [A, B, C]

# try adding one outside point

D, position = point\_outside(points, position, O, r\_squared)

if not D:

return subset, O, r\_squared # all points are inside: return

else:

subset += [D]

# try adding another outside point

if len(subset) == 3:

D, position = point\_outside(points, position, O, r\_squared)

if D:

subset += [D]

**Annex 2: Python procedure for the maximum distance algorithm**

def maximum\_distance (points, diameter):

"""

compute minimum circle enclosing points

Parameters: points, a list of points, tuples of 2 float

diameter, a list of two points to start with

Returns: diameter, a list of two points

r\_squared, a float, one fourth the square of diameter

"""

while True:

O,r\_squared = circle(diameter)

C, position = point\_outside(points, size, position, O, r\_squared)

if not C:

return diameter, r\_squared # AB is solution

D, position = point\_outside(points, size, position, C, 4\*r\_squared)

if not D:

position = (position-1)%size

points, size = eliminate(points, position, size-1) # eliminate C

if position == size: position = 0

continue

diameter = [C, D]