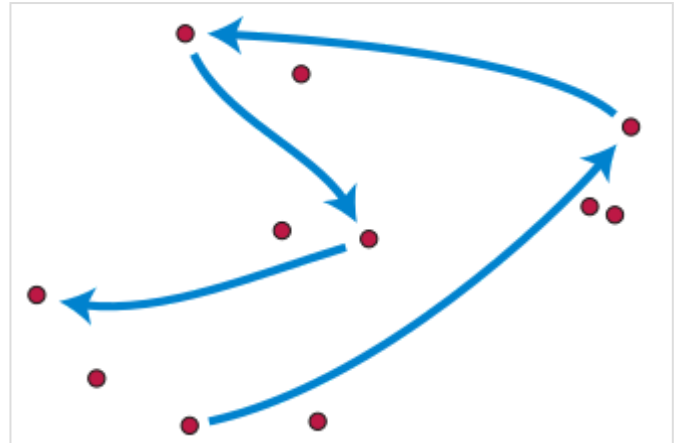# Farthest-first traversal

In computational geometry, the **farthest-first traversal** of a compact metric space is a sequence of points in the space, where the first point is selected arbitrarily and each successive point is as far as possible from the set of previously-selected points. The same concept can also be applied to a finite set of geometric points, by restricting the selected points to belong to the set or equivalently by considering the finite metric space generated by these points.[1] For a finite metric space or finite set of geometric points, the resulting sequence forms a permutation of the points, also known as the **greedy permutation**.[2]

Every prefix of a farthest-first traversal provides a set of points that is widely spaced and close to all remaining points. More precisely, no other set of equally many points can be spaced more than twice as widely, and no other set of equally many points can be less than half as far to its farthest remaining point. In part because of these properties, farthest-point traversals have many applications, including the approximation of the traveling salesman problem and the metric k-center problem. They may be constructed in polynomial time, or (for low-dimensional Euclidean spaces) approximated in near-linear time.

The first five points in a farthest-first traversal of a planar point set. The first point is chosen arbitrarily and each successive point is as far as possible from all previously chosen points.

Definition and properties

A farthest-first traversal is a sequence of points in a compact metric space, with each point appearing at most once. If the space is finite, each point appears exactly once, and the traversal is a permutation of all of the points in the space. The first point of the sequence may be any point in the space. Each point p after the first must have the maximum possible distance to the set of points earlier than p in the sequence, where the distance from a point to a set is defined as the minimum of the pairwise distances to points in the set. A given space may have many different farthest-first traversals, depending both on the choice of the first point in the sequence (which may be any point in the space) and on ties for the maximum distance among later choices.[2]

Farthest-point traversals may be characterized by the following properties. Fix a number k, and consider the prefix formed by the first k points of the farthest-first traversal of any metric space. Let r be the distance between the final point of the prefix and the other points in the prefix. Then this subset has the following two properties:

- All pairs of the selected points are at distance at least r from each other, and
- All points of the metric space are at distance at most r from the subset.

Conversely any sequence having these properties, for all choices of k, must be a farthest-first traversal. These are the two defining properties of a <u>Delone set</u>, so each prefix of the farthest-first traversal forms a Delone set.[3]

Applications

<u>Rosenkrantz, Stearns & Lewis (1977)</u> used the farthest-first traversal to define the farthest-insertion <u>heuristic</u> for the <u>travelling salesman problem</u>. This heuristic finds approximate solutions to the travelling salesman problem by building up a tour on a subset of points, adding one point at a time to the tour in the ordering given by a farthest-first traversal. To add each point to the tour, one edge of the previous tour is broken and replaced by a pair of edges through the added point, in the cheapest possible way. Although Rosenkrantz et al. prove only a logarithmic <u>approximation ratio</u> for this method, they show that in practice it often works better than other insertion methods with better provable approximation ratios.[4]

Later, the same sequence of points was popularized by <u>Gonzalez (1985)</u>, who used it as part of <u>greedy approximation algorithms</u> for two problems in clustering, in which the goal is to partition a set of points into k clusters. One of the two problems that Gonzalez solve in this way seeks to minimize the maximum <u>diameter</u> of a cluster, while the other, known as the <u>metric k-center</u> problem, seeks to minimize the maximum radius, the distance from a chosen central point of a cluster to the farthest point from it in the same cluster. For instance, the k-center problem can be used to model the placement of fire stations within a city, in order to ensure that every address within the city can be reached quickly by a fire truck. For both clustering problems, Gonzalez chooses a set of k cluster centers by selecting the first k points of a farthest-first traversal, and then creates clusters by assigning each input point to the nearest cluster center. If r is the distance from the set of k selected centers to the next point at position $k + 1$ in the traversal, then with this clustering every point is within distance r of its center and every cluster has diameter at most $2r$. However, the subset of k centers together with the next point are all at distance at least r from each other, and any k-clustering would put some two of these points into a single cluster, with one of them at distance at least $r/2$ from its center and with diameter at least r. Thus, Gonzalez's heuristic gives an <u>approximation ratio</u> of 2 for both clustering problems.[3]

Gonzalez's heuristic was independently rediscovered for the metric k-center problem by <u>Dyer & Frieze (1985)</u>, who applied it more generally to weighted k-center problems.[5] Another paper on the k-center problem from the same time, <u>Hochbaum & Shmoys (1985)</u>, achieves the same approximation ratio of 2,[6] but its techniques are different.[5] Nevertheless, Gonzalez's heuristic, and the name "farthest-first traversal", are often incorrectly attributed to Hochbaum and Shmoys.[7] For both the min-max diameter clustering problem and the metric k-center problem, these approximations are optimal: the existence of a polynomial-time heuristic with any constant approximation ratio less than 2 would imply that <u>P = NP</u>.[3][6]

As well as for clustering, the farthest-first traversal can also be used in another type of facility location problem, the max-min facility dispersion problem, in which the goal is to choose the locations of k different facilities so that they are as far apart from each other as possible. More precisely, the goal in this problem is to choose k points from a given metric space or a given set of candidate points, in such a way as to maximize the minimum pairwise distance between the selected points. Again, this can be

approximated by choosing the first k points of a farthest-first traversal. If r denotes the distance of the kth point from all previous points, then every point of the metric space or the candidate set is within distance r of the first $k - 1$ points. By the pigeonhole principle, some two points of the optimal solution (whatever it is) must both be within distance r of the same point among these first $k - 1$ chosen points, and (by the triangle inequality) within distance $2r$ of each other. Therefore, the heuristic solution given by the farthest-first traversal is within a factor of two of optimal.[8][9][10]

Other applications of the farthest-first traversal include color quantization (clustering the colors in an image to a smaller set of representative colors),[11] progressive scanning of images (choosing an order to display the pixels of an image so that prefixes of the ordering produce good lower-resolution versions of the whole image rather than filling in the image from top to bottom),[12] point selection in the probabilistic roadmap method for motion planning,[13] simplification of point clouds,[14] generating masks for halftone images,[15][16] hierarchical clustering,[1] finding the similarities between polygon meshes of similar surfaces,[17] choosing diverse and high-value observation targets for underwater robot exploration,[18] fault detection in sensor networks,[19] modeling phylogenetic diversity,[20] matching vehicles in a heterogenous fleet to customer delivery requests,[21] uniform distribution of geodetic observatories on the Earth's surface[22] or of other types of sensor network,[23] generation of virtual point lights in the instant radiosity computer graphics rendering method,[24] and geometric range searching data structures.[25]

Algorithms

Greedy exact algorithm

The farthest-first traversal of a finite point set may be computed by a greedy algorithm that maintains the distance of each point from the previously selected points, performing the following steps:[3]

- Initialize the sequence of selected points to the empty sequence, and the distances of each point to the selected points to infinity.
- While not all points have been selected, repeat the following steps:
    - Scan the list of not-yet-selected points to find a point p that has the maximum distance from the selected points.
    - Remove p from the not-yet-selected points and add it to the end of the sequence of selected points.
    - For each remaining not-yet-selected point q, replace the distance stored for q by the minimum of its old value and the distance from p to q.

For a set of n points, this algorithm takes $O(n^2)$ steps and $O(n^2)$ distance computations.[3]

Approximations

A faster approximation algorithm, given by Har-Peled & Mendel (2006), applies to any subset of points in a metric space with bounded doubling dimension, a class of spaces that include the Euclidean spaces of bounded dimension. Their algorithm finds a sequence of points in which each successive point has distance within a $1 - \varepsilon$ factor of the farthest distance from the previously-selected point, where ε can be chosen to be any positive number. It takes time $O(n \log n)$.[2]

The results for bounded doubling dimension do not apply to high-dimensional Euclidean spaces, because the constant factor in the big O notation for these algorithms depends on the dimension. Instead, a different approximation method based on the Johnson–Lindenstrauss lemma and locality-sensitive hashing has running time $O\left(\varepsilon^{-2}n^{1+1/(1+\varepsilon)^2+o(1)}\right)$. For metrics defined by shortest paths on weighted undirected graphs, a randomized incremental construction based on Dijkstra's algorithm achieves time $O\left(\varepsilon^{-1}m\log n\log\frac{n}{\varepsilon}\right)$, where n and m are the numbers of vertices and edges of the input graph, respectively.[26]

Incremental Voronoi insertion

For selecting points from a continuous space such as the Euclidean plane, rather than from a finite set of candidate points, these methods will not work directly, because there would be an infinite number of distances to maintain. Instead, each new point should be selected as the center of the largest empty circle defined by the previously-selected point set.[12] This center will always lie on a vertex of the Voronoi diagram of the already selected points, or at a point where an edge of the Voronoi diagram crosses the domain boundary. In this formulation the method for constructing farthest-first traversals has also been called **incremental Voronoi insertion**.[27] It is similar to Delaunay refinement for finite element mesh generation, but differs in the choice of which Voronoi vertex to insert at each step.[28]

See also

- Lloyd's algorithm, a different method for generating evenly spaced points in geometric spaces

References

1. Dasgupta, S.; Long, P. M. (2005), "Performance guarantees for hierarchical clustering", *Journal of Computer and System Sciences*, **70** (4): 555–569, doi:10.1016/j.jcss.2004.10.006 (https://doi.org/10.1016%2Fj.jcss.2004.10.006), MR 2136964 (https://mathscinet.ams.org/mathscinet-getitem?mr=2136964)
2. Har-Peled, S.; Mendel, M. (2006), "Fast construction of nets in low-dimensional metrics, and their applications", *SIAM Journal on Computing*, **35** (5): 1148–1184, arXiv:cs/0409057 (https://arxiv.org/abs/cs/0409057), doi:10.1137/S0097539704446281 (https://doi.org/10.1137%2FS0097539704446281), MR 2217141 (https://mathscinet.ams.org/mathscinet-getitem?mr=2217141), S2CID 37346335 (https://api.semanticscholar.org/CorpusID:37346335)
3. Gonzalez, T. F. (1985), "Clustering to minimize the maximum intercluster distance", *Theoretical Computer Science*, **38** (2–3): 293–306, doi:10.1016/0304-3975(85)90224-5 (https://doi.org/10.1016%2F0304-3975%2885%2990224-5), MR 0807927 (https://mathscinet.ams.org/mathscinet-getitem?mr=0807927)
4. Rosenkrantz, D. J.; Stearns, R. E.; Lewis, P. M. II (1977), "An analysis of several heuristics for the traveling salesman problem", *SIAM Journal on Computing*, **6** (3): 563–581, doi:10.1137/0206041 (https://doi.org/10.1137%2F0206041), MR 0459617 (https://mathscinet.ams.org/mathscinet-getitem?mr=0459617), S2CID 14764079 (https://api.semanticscholar.org/CorpusID:14764079)

5. Dyer, M. E.; Frieze, A. M. (1985), "A simple heuristic for the p-centre problem" (https://www.math.cmu.edu/~af1p/Texfiles/simpheur.pdf) (PDF), *Operations Research Letters*, **3** (6): 285–288, doi:10.1016/0167-6377(85)90002-1 (https://doi.org/10.1016%2F0167-6377%2885%2990002-1), MR 0797340 (https://mathscinet.ams.org/mathscinet-getitem?mr=0797340)

6. Hochbaum, Dorit S.; Shmoys, David B. (1985), "A best possible heuristic for the k-center problem", *Mathematics of Operations Research*, **10** (2): 180–184, doi:10.1287/moor.10.2.180 (https://doi.org/10.1287%2Fmoor.10.2.180), MR 0793876 (https://mathscinet.ams.org/mathscinet-getitem?mr=0793876)

7. For prominent examples of incorrect attribution of the farthest-first heuristic to Hochbaum & Shmoys (1985), see, e.g.,

    - Dasgupta, Sanjoy (2002), "Performance guarantees for hierarchical clustering", in Kivinen, Jyrki; Sloan, Robert H. (eds.), *Computational Learning Theory, 15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 8-10, 2002, Proceedings*, Lecture Notes in Computer Science, vol. 2375, Springer, pp. 351–363, doi:10.1007/3-540-45435-7_24 (https://doi.org/10.1007%2F3-540-45435-7_24), ISBN 978-3-540-43836-6 (corrected in the 2005 journal version of the same paper)

    - Agarwal, Sameer; Ramamoorthi, Ravi; Belongie, Serge J.; Jensen, Henrik Wann (2003), "Structured importance sampling of environment maps", *ACM Trans. Graph.*, **22** (3): 605–612, doi:10.1145/882262.882314 (https://doi.org/10.1145%2F882262.882314)

    - Baram, Yoram; El-Yaniv, Ran; Luz, Kobi (2004), "Online choice of active learning algorithms" (https://jmlr.org/papers/volume5/baram04a/baram04a.pdf) (PDF), *J. Mach. Learn. Res.*, **5**: 255–291

    - Basu, Sugato; Bilenko, Mikhail; Banerjee, Arindam; Mooney, Raymond J. (2006), "Probabilistic semi-supervised clustering with constraints", in Chapelle, Olivier; Schölkopf, Bernhard; Zien, Alexander (eds.), *Semi-Supervised Learning*, The MIT Press, pp. 73–102, doi:10.7551/mitpress/9780262033589.003.0005 (https://doi.org/10.7551%2Fmitpress%2F9780262033589.003.0005), ISBN 978-0-262-03358-9

    - Lima, Christiane Ferreira Lemos; Assis, Francisco M.; de Souza, Cleonilson Protásio (2011), "A comparative study of use of Shannon, Rényi and Tsallis entropy for attribute selecting in network intrusion detection", *IEEE International Workshop on Measurement and Networking, M&N 2011, Anacapri, Italy, October 10-11, 2011*, IEEE, pp. 77–82, doi:10.1109/IWMN.2011.6088496 (https://doi.org/10.1109%2FIWMN.2011.6088496), S2CID 7510040 (https://api.semanticscholar.org/CorpusID:7510040)

    - "Class FarthestFirst" (https://weka.sourceforge.io/doc.dev/weka/clusterers/FarthestFirst.html), *Weka, version 3.9.5*, University of Waikato, December 21, 2020, retrieved 2021-11-06 – via SourceForge

8. White, Douglas J. (1991), "The maximal-dispersion problem", *IMA Journal of Mathematics Applied in Business and Industry*, **3** (2): 131–140 (1992), doi:10.1093/imaman/3.2.131 (https://doi.org/10.1093%2Fimaman%2F3.2.131), MR 1154657 (https://mathscinet.ams.org/mathscinet-getitem?mr=1154657); White credits the use of the farthest-first traversal as a heuristic for this problem to Steuer, R. E. (1986), *Multiple-Criteria Optimization: Theory, Computation, and Applications*, New York: Wiley

9. Tamir, Arie (1991), "Obnoxious facility location on graphs", *SIAM Journal on Discrete Mathematics*, **4** (4): 550–567, doi:10.1137/0404048 (https://doi.org/10.1137%2F0404048), MR 1129392 (https://mathscinet.ams.org/mathscinet-getitem?mr=1129392)

10. Ravi, S. S.; Rosenkrantz, D. J.; Tayi, G. K. (1994), "Heuristic and special case algorithms for dispersion problems", *Operations Research*, **42** (2): 299–310, doi:10.1287/opre.42.2.299 (https://doi.org/10.1287%2Fopre.42.2.299), JSTOR 171673 (https://www.jstor.org/stable/171673), S2CID 16489402 (https://api.semanticscholar.org/CorpusID:16489402)

11. Xiang, Z. (1997), "Color image quantization by minimizing the maximum intercluster distance", *ACM Transactions on Graphics*, **16** (3): 260–276, doi:10.1145/256157.256159 (https://doi.org/10.1145%2F256157.256159), S2CID 17713417 (https://api.semanticscholar.org/CorpusID:17713417)

12. Eldar, Y.; Lindenbaum, M.; Porat, M.; Zeevi, Y. Y. (1997), "The farthest point strategy for progressive image sampling", *IEEE Transactions on Image Processing*, **6** (9): 1305–1315, Bibcode:1997ITIP....6.1305E (https://ui.adsabs.harvard.edu/abs/1997ITIP....6.1305E), doi:10.1109/83.623193 (https://doi.org/10.1109%2F83.623193), PMID 18283019 (https://pubmed.ncbi.nlm.nih.gov/18283019)

13. Mazer, E.; Ahuactzin, J. M.; Bessiere, P. (1998), "The Ariadne's clew algorithm", *Journal of Artificial Intelligence Research*, **9**: 295–316, arXiv:1105.5440 (https://arxiv.org/abs/1105.5440), doi:10.1613/jair.468 (https://doi.org/10.1613%2Fjair.468)

14. Moenning, C.; Dodgson, N. A. (2003), "A new point cloud simplification algorithm", *3rd IASTED International Conference on Visualization, Imaging, and Image Processing*

15. Gotsman, Craig; Allebach, Jan P. (1996), "Bounds and algorithms for dither screens" (https://www.cs.technion.ac.il/~gotsman/AmendedPubl/BoundsAndAlgorithms/BoundsAndAlg-scanned.pdf) (PDF), in Rogowitz, Bernice E.; Allebach, Jan P. (eds.), *Human Vision and Electronic Imaging*, Proc. SPIE, vol. 2657, pp. 483–492, doi:10.1117/12.238746 (https://doi.org/10.1117%2F12.238746), S2CID 10608234 (https://api.semanticscholar.org/CorpusID:10608234)

16. Shahidi, R.; Moloney, C.; Ramponi, G. (2004), "Design of farthest-point masks for image halftoning", *EURASIP Journal on Applied Signal Processing*, **2004** (12): 1886–1898, Bibcode:2004EJASP2004...45S (https://ui.adsabs.harvard.edu/abs/2004EJASP2004...45S), doi:10.1155/S1110865704403217 (https://doi.org/10.1155%2FS1110865704403217)

17. Lipman, Y.; Funkhouser, T. (2009), "Möbius voting for surface correspondence", *Proc. ACM SIGGRAPH*, pp. 72:1–72:12, doi:10.1145/1576246.1531378 (https://doi.org/10.1145%2F1576246.1531378), ISBN 978-1-60558-726-4, S2CID 6001942 (https://api.semanticscholar.org/CorpusID:6001942)

18. Girdhar, Y.; Giguère, P.; Dudek, G. (2012), "Autonomous adaptive underwater exploration using online topic modelling" (http://cim.mcgill.ca/~yogesh/publications/iser2012.pdf) (PDF), *Proc. Int. Symp. Experimental Robotics*

19. Altinisik, U.; Yildirim, M.; Erkan, K. (2012), "Isolating non-predefined sensor faults by using farthest first traversal algorithm", *Ind. Eng. Chem. Res.*, **51** (32): 10641–10648, doi:10.1021/ie201850k (https://doi.org/10.1021%2Fie201850k)

20. Bordewich, Magnus; Rodrigo, Allen; Semple, Charles (2008), "Selecting taxa to save or sequence: Desirable criteria and a greedy solution", *Systematic Biology*, **57** (6): 825–834, doi:10.1080/10635150802552831 (https://doi.org/10.1080%2F10635150802552831), PMID 19085326 (https://pubmed.ncbi.nlm.nih.gov/19085326)

21. Fisher, Marshall L.; Jaikumar, Ramchandran (1981), "A generalized assignment heuristic for vehicle routing", *Networks*, **11** (2): 109–124, doi:10.1002/net.3230110205 (https://doi.org/10.1002%2Fnet.3230110205), MR 0618209 (https://mathscinet.ams.org/mathscinet-getitem?mr=0618209); as cited by Gheysens, Filip; Golden, Bruce; Assad, Arjang (1986), "A new heuristic for determining fleet size and composition", in Gallo, Giorgio; Sandi, Claudio (eds.), *Netflow at Pisa*, Mathematical Programming Studies, vol. 26, Springer, pp. 233–236, doi:10.1007/bfb0121103 (https://doi.org/10.1007%2Fbfb0121103), ISBN 978-3-642-00922-8

22. Hase, Hayo (2000), "New method for the selection of additional sites for the homogenisation of an inhomogeneous cospherical point distribution", in Rummel, Reinhard; Drewes, Hermann; Bosch, Wolfgang; et al. (eds.), *Towards an Integrated Global Geodetic Observing System (IGGOS): IAG Section II Symposium Munich, October 5-9, 1998, Posters – Session B*, International Association of Geodesy Symposia, vol. 120, Springer, pp. 180–183, doi:10.1007/978-3-642-59745-9_35 (https://doi.org/10.1007%2F978-3-642-59745-9_35), ISBN 978-3-642-64107-7

23. Vieira, Luiz Filipe M.; Vieira, Marcos Augusto M.; Ruiz, Linnyer Beatrys; Loureiro, Antonio A. F.; Silva, Diógenes Cecílio; Fernandes, Antônio Otávio (2004), "Efficient Incremental Sensor Network Deployment Algorithm" (https://web.archive.org/web/20150716075437/http://www.lbd.dcc.ufmg.br/colecoes/sbrc/2004/001.pdf) (PDF), *Proc. Brazilian Symp. Computer Networks*, pp. 3–14, archived from the original (http://www.lbd.dcc.ufmg.br/colecoes/sbrc/2004/001.pdf) (PDF) on 2015-07-16, retrieved 2015-07-16

24. Laine, Samuli; Saransaari, Hannu; Kontkanen, Janne; Lehtinen, Jaakko; Aila, Timo (2007), "Incremental instant radiosity for real-time indirect illumination", *Proceedings of the 18th Eurographics Conference on Rendering Techniques (EGSR'07)*, Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, pp. 277–286, doi:10.2312/EGWR/EGSR07/277-286 (https://doi.org/10.2312%2FEGWR%2FEGSR07%2F277-286), ISBN 978-3-905673-52-4, S2CID 18626929 (https://api.semanticscholar.org/CorpusID:18626929)

25. Abbar, S.; Amer-Yahia, S.; Indyk, P.; Mahabadi, S.; Varadarajan, K. R. (2013), "Diverse near neighbor problem", *Proceedings of the 29th Annual Symposium on Computational Geometry*, pp. 207–214, doi:10.1145/2462356.2462401 (https://doi.org/10.1145%2F2462356.2462401), hdl:1721.1/87000 (https://hdl.handle.net/1721.1%2F87000), ISBN 978-1-4503-2031-3, S2CID 6286186 (https://api.semanticscholar.org/CorpusID:6286186)

26. Eppstein, David; Har-Peled, Sariel; Sidiropoulos, Anastasios (2020), "Approximate greedy clustering and distance selection for graph metrics", *Journal of Computational Geometry*, **11** (1): 629–652, doi:10.20382/jocg.v11i1a25 (https://doi.org/10.20382%2Fjocg.v11i1a25), MR 4194877 (https://mathscinet.ams.org/mathscinet-getitem?mr=4194877), S2CID 18316279 (https://api.semanticscholar.org/CorpusID:18316279)

27. Teramoto, Sachio; Asano, Tetsuo; Katoh, Naoki; Doerr, Benjamin (2006), "Inserting points uniformly at every instance" (https://search.ieice.org/bin/summary.php?id=e89-d_8_2348), *IEICE Transactions on Information and Systems*, E89-D (8): 2348–2356, Bibcode:2006IEITI..89.2348T (https://ui.adsabs.harvard.edu/abs/2006IEITI..89.2348T), doi:10.1093/ietisy/e89-d.8.2348 (https://doi.org/10.1093%2Fietisy%2Fe89-d.8.2348), hdl:2433/84849 (https://hdl.handle.net/2433%2F84849)

28. Ruppert, Jim (1995), "A Delaunay refinement algorithm for quality 2-dimensional mesh generation", *Journal of Algorithms*, **18** (3): 548–585, doi:10.1006/jagm.1995.1021 (https://doi.org/10.1006%2Fjagm.1995.1021)