A Ptolemaic Partitioning Mechanism

Richard Connor

University of St Andrews, St Andrews, Scotland
`rchc@st-andrews.ac.uk`

Abstract. For many years, exact metric search relied upon the property of triangle inequality to give a lower bound on uncalculated distances. Two exclusion mechanisms derive from this property, generally known as pivot exclusion and hyperplane exclusion. These mechanisms work in any proper metric space and are the basis of many metric indexing mechanisms. More recently, the Ptolemaic and four-point lower bound properties have been shown to give tighter bounds in some subclasses of metric space.

Both triangle inequality and the four-point lower bound directly imply straightforward *partitioning* mechanisms: that is, a method of dividing a finite space according to a fixed partition, in order that one or more classes of the partition can be eliminated from a search at query time. However, up to now, no partitioning principle has been identified for the Ptolemaic inequality, which has been used only as a filtering mechanism. Here, a novel partitioning mechanism for the Ptolemaic lower bound is presented. It is always better than either pivot or hyperplane partitioning. While the exclusion condition itself is weaker than Hilbert (four-point) exclusion, its calculation is cheaper. Furthermore, it can be combined with Hilbert exclusion to give a new maximum for exclusion power with respect to the number of distances measured per query.

Keywords: Metric Search · Partitioning · Ptolemaic inequality · Supermetric space

1 Background and Related Work

The context of interest is querying a large finite space (S, d) which is a subset of an infinite metric space (U, d).

In most general terms, querying the space (S, d) with query $q \in U$ is the task of finding a subset $\{s \leftarrow S \mid d(q, s) \leq t\}$, for some value t which gives a suitable size of solution set. It is generally assumed that $|S|$ is large or the cost of applying the function d is high, and so the simple solution of applying $d(q, s)$ to all $s \in S$ is intractable [1, 10].

Table 1 gives a summary of these and other notations used throughout the article.

Table 1: Notation used throughout

Symbols	Meaning
(U, d)	An infinite metric space with domain U and distance d
(S, d)	A large finite space $S \subset U$ over which search is performed
P	A small reference set $P \subset U$, usually $P \subset S$
$u, u_0, u_1, \ldots$	Elements of the infinite domain U
$s, s_0, s_1, \ldots$	Elements of the finite domain S
$p, p_0, p_1, \ldots$	Elements of the reference set P
M	A fixed radius used to define a partition with a given $p \in P$
q, t	A query $q \in U$ associated with a numeric query threshold t
$\mathcal{P}$	A partition of S defined according to distances to P
$\mathcal{S}$	A class of $\mathcal{P}$ which may be excluded given a particular q, t
$\mathcal{Q}$	A subset of U defined according to a particular q, t
$\mathbb{R}^n$	An n-dimensional real domain
ℓ_2	The Euclidean distance metric
τ	A numeric parameter of the Ptolemaic partitioning mechanism ($\tau \geq 0.5$, typically $\tau \approx 1$)

1.1   Filtering and Partitioning

All metric search solutions rely upon algebraic properties of (U, d). A relatively small set of distinguished reference points $P = \{p_0, \ldots, p_m\}$ (typically, $P \subset S$) is used to avoid direct calculation of $d(q, s)$, after the distances $d(s, P)$ and $d(q, P)$ have been calculated. $d(s, P)$ is calculated ahead of query time, during a pre-processing phase. Two types of usage are distinguished as follows:

filtering: given a query $q \in U$, a specific datum $s \in S$, and the distances $d(q, P)$ and $d(s, P)$, it may be possible to determine that $d(q, s) > t$ for some t without having to calculate $d(q, s)$.

partitioning: given a partition $\mathcal{P}$ of S determined at pre-processing time with respect to $d(S, P)$, and the distances $d(q, P)$, it may be possible to determine that some classes of $\mathcal{P}$ do not contain any elements s such that $d(q, s) \leq t$.

Both types of mechanism have their place in metric search, see [1, 10] for many examples. Filtering approaches however imply linear-time solutions, whereas partitioning can be used to construct an indexing mechanism, typically where a very large data set is recursively partitioned, in order to achieve a sub-linear search time.

For filtering, the algebraic properties are required to give a lower-bound on the distance $d(q, s)$ with reference to the sets of distances $d(q, P)$ and $d(s, P)$. For partitioning, a further requirement is to identify a partition that can be determined at pre-processing time, of which one or more classes may be excluded at query time according to $d(q, P)$.

Table 2 shows partitioning mechanisms which derive from various known lower-bound properties. The contribution of this paper is a novel partitioning

mechanism for Ptolemaic inequality, shown in bold type in the table. Until now, such a partitioning mechanism has been missing from the literature.

Table 2: Five different partition functions and their corresponding exclusion conditions. In all cases the partition criterion is used to form a distinguished subset of S at pre-processing time, for all $s \in S$. The exclusion condition is evaluated with respect to the query q and a query radius t. Row 5 summarises the novel contribution of this paper.

	Underlying property	Partition criterion	Exclusion condition
1	triangle inequality	$d(s,p) \leq M$	$d(q,p) > M + t$
2	triangle inequality	$d(s,p) \geq M$	$d(q,p) < M - t$
3	triangle inequality	$d(s,p_0) \leq d(s,p_1)$	$d(q,p_0) - d(q,p_1) > 2t$
4	four-point lower bound	$d(s,p_0) \leq d(s,p_1)$	$\frac{d(q,p_0)^2 - d(q,p_1)^2}{d(p_0,p_1)} > 2t$
5	**Ptolemaic inequality**	$\boldsymbol{d(s,p_0) \leq d(s,p_1)}$ $\boldsymbol{\wedge\ d(s,p_1) \geq \tau d(p_0,p_1)}$	$\boldsymbol{d(q,p_0) - d(q,p_1) > t/\tau}$

The remainder of this section introduces some necessary preliminaries. In Section 2 the underlying geometry of the partition mechanism is given, and Section 3 gives its formal definition. Section 4 gives a quantitative analysis of its value.

1.2   Subclasses of Metric Space

Properties (1-3) listed in Table 2 are possessed by all proper metric spaces. Property (4) is found only in *supermetric* spaces [2], which include all spaces which are isometrically embeddable in Hilbert space[1], while property (5) is found in any Hadamard space[2].

Any Hilbert-embeddable space is also a Hadamard space; although Hadamard spaces are a little more general, it is not clear that any practical non-Hilbert spaces fall in this category. Details of Hilbert spaces are elaborated in [4]; in this context it is sufficient to know that the following classes of metric space are members of both classes: Euclidean, Cosine, Jensen-Shannon, Quadratic Form, Triangular, and Mahalanobis spaces[3]. Furthermore, the square root of any proper metric gives a space in both classes. The partition mechanism described here is thus applicable to any of these spaces.

---

[1] see eg `https://en.wikipedia.org/wiki/Hilbert_space`
[2] see eg `https://en.wikipedia.org/wiki/Hadamard_space`
[3] For appropriate formulations; see [4].

1.3   Ptolemaic and Four-point Lower Bounds

The Ptolemaic inequality was identified for use as a distance lower-bound for certain metric spaces in [6], and used further in a number of studies for example [7, 8]. For any four objects $u_0, u_1, u_2, u_3 \in U$, the Ptolemaic inequality states:

$$d(u_0, u_2) \cdot d(u_1, u_3) \leq d(u_0, u_1) \cdot d(u_2, u_3) + d(u_1, u_2) \cdot d(u_3, u_0)$$

In $(\mathbb{R}^n, \ell_2)$ this is more simply stated as the product of the lengths of the diagonals of any quadrilateral being no greater than the sum of the products of the pairs of opposing sides. Given this property, a lower bound on the distance $d(q, s)$ can be determined whenever, for two reference values p_0, p_1, all the distances $d(s, p_0), d(s, p_1), d(q, p_0), d(q, p_1)$ and $d(p_0, p_1)$ are known. This lower bound is much tighter than those available via simple triangle inequality, and has been used to great effect for filtering objects during search, particularly in the context of a very expensive distance function [8]. The mechanisms used to incorporate this lower bound into metric search techniques include the Ptolemaic pivot table, the Ptolemaic PM-Tree, and the Ptolemaic M-Index [7]. In all cases, the inequality is used as an extra filtering mechanism superimposed onto an existing filtering or partitioning structure.

The four-point lower bound property, and the Hilbert exclusion mechanism, were first identified in [2], and investigated further in [4, 5]. Any supermetric space (U, d) has the four-point property: for any four objects $u_0, u_1, u_2, u_3 \in U$, there exists a tetrahedron with vertices $u_0', u_1', u_2', u_3' \in \mathbb{R}^3$ where the distances between pairs of points are preserved, i.e. $d(u_i, u_j) = \ell_2(u_i', u_j')$.

The four-point property thus implies the Ptolemaic property, but not vice-versa.

The four-point lower-bound property applies to the case where five of the six edge lengths of a tetrahedron are known. In this case, two adjacent faces of the tetrahedron can be constructed. A lower bound of the unknown distance is obtained by notionally rotating these faces around their common edge to minimise the final edge length, which occurs when a planar tetrahedron is formed.

1.4   Projection into 2-dimensional Space

Together these properties imply that if the Ptolemaic inequality is applied to a quadrilateral in two dimensions, when that quadrilateral has been formed according to five distances measured among four objects in a supermetric space, then the inequality applies also to the original space. Figure 1 shows an example of this.

The figure shows a projection in $(\mathbb{R}^2, \ell_2)$ of four objects p_0, p_1, s_0 and s_1 selected from a supermetric space (U, d). All distances other than $d(s_0, s_1)$ have been calculated in (U, d). The projections of objects p_0 and p_1 are plotted at the points $(0, 0)$ and $(0, d(p_0, p_1))$ respectively[4]. The projections of objects s_0 and s_1 are plotted at the unique points above the X-axis which preserve their

---

[4] This choice is arbitrary, any two points which preserve $d(p_0, p_1)$ could be used
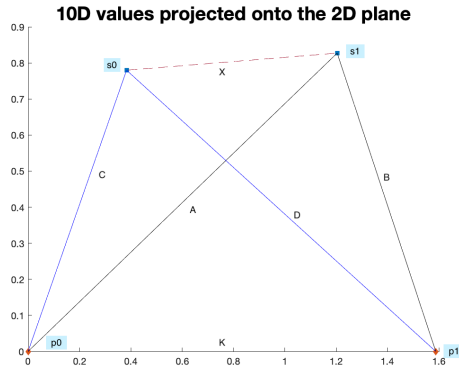
Fig. 1: Four objects p_0, p_1, s_0, s_1 selected from a supermetric space are projected onto a 2D plane according to the known distances K, A, B, C, D. Although $d(s_0, s_1)$ is not known, it is known that a tetrahedron with these four vertices exists in 3 dimensions. By the four-point lower bound, $d(s_0, s_1) \geq X$. By the Ptolemaic lower bound in 2 dimensions, $X \geq \frac{AD - BC}{K}$

distances from p_0 and p_1. The supermetric properties imply that the tetrahedron p_0, p_1, s_0, s_1 must exist in $(\mathbb{R}^3, \ell_2)$, therefore the unknown distance $d(s_0, s_1)$ is lower-bounded by the sixth edge of the planar tetrahedron plotted in $(\mathbb{R}^2, \ell_2)$.

By the four-point lower bound property, $d(s_0, s_1) \geq X$. By the Ptolemaic lower-bound property, $X \geq \frac{AD - BC}{K}$. Therefore, in the original supermetric space, $d(s_0, s_1) \geq \frac{AD - BC}{K}$. For the rest of this article, only 2D projections like these are considered, where two distinguished reference objects p_0, p_1 are used to form a planar projection of the rest of the data set, and rely on the Ptolemaic property with the context of planar quadrilaterals. This restricts the outcome to Hilbert-embeddable spaces, although as noted this is not a significant practical restriction.

It is worth noting that while the derivation and correctness of the mechanism rely upon the existence of the 2D projection, the projection itself does not require to be calculated. As shown in Table 2, the calculations required are restricted to simple calculations over distances measured in the original space.

2 The Underlying Geometry

Partitioning mechanisms differ from filtering in that, for each possibility of exclusion, it is necessary to identify two subsets of the universal space:

1. a *static* subset $\mathcal{S}$, which can be identified and indexed during the pre-processing of the finite data set, and
2. a *dynamic* subset $\mathcal{Q}$, which is identified only after the query and (typically) its associated search radius become apparent.

Exclusion of $\mathcal{S}$ can be performed when every element of $\mathcal{Q}$ is separated by at least the search radius from every element of $\mathcal{S}$. In the following section objects denoted by s and q are referred to, representing elements of $\mathcal{S}$ and $\mathcal{Q}$ respectively.
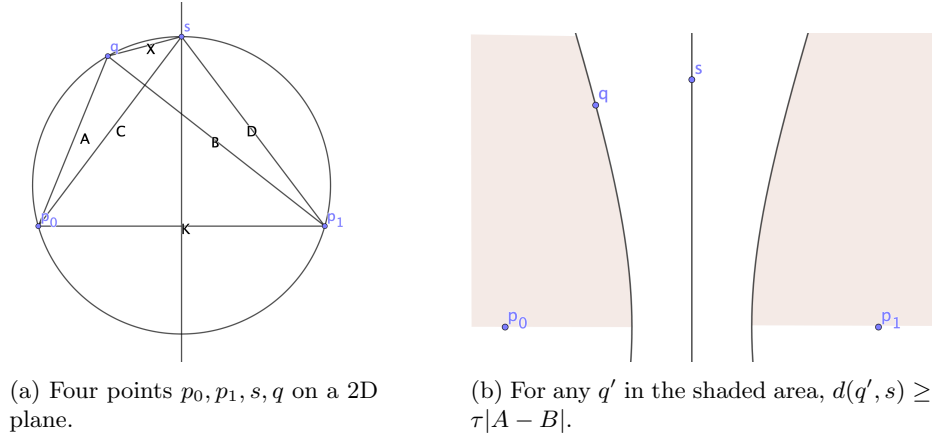
2.1   2D Geometry

(a) Four points p_0, p_1, s, q on a 2D plane.

(b) For any q' in the shaded area, $d(q', s) \geq \tau |A - B|$.

Fig. 2: In 2a, line segments among the points are annotated by their lengths. s is chosen according to the parameter $\tau = C/K$, and q is a point on the circle defined by p_0, p_1 and s. q is the unique point with $X = \tau(B - A)$. The regions shown in the figure represent regions of the original, potentially non-Euclidean, space; the inequalities established are generally applicable to the original space.

Figure 2a shows four points p_0, p_1, q and s drawn on a plane. These points represent the 2D projections of two reference (or pivot) values p_0, p_1, a query value q and a potential solution value s. The figure is annotated with line segments labelled $A - D$, K and X, where the labels represent the lengths of the respective lines. K is the inter-pivot distance, and X is a lower bound of the unknown distance $d(q, s)$.

For the moment, values have been chosen such that

- $C = D$
- the parameter τ defines the ratio C/K
- point q lies on the same circle as p_0, p_1 and s

The Ptolemaic inequality states

$$BC \leq AD + KX$$

so in this case:

$$X \geq \tau(B - A)$$

The boundary of this region defines a hyperbola with foci p_0, p_1 and semi-major axis $X/2\tau$, as shown in Figure 2b. It follows that any point within the shaded region is at least distance X from the point s.

As q, p_0, p_1 and s are co-circular, q is the unique closest point on the (left-hand) hyperbola to s and $X = \tau(B - A)$. The line segment sq is therefore perpendicular to the tangent of the hyperbola at q. As the gradient of the tangent is negative, it follows that any point above and to the right of s is further than X from any point to the left of the hyperbola, as illustrated in Figure 3a.
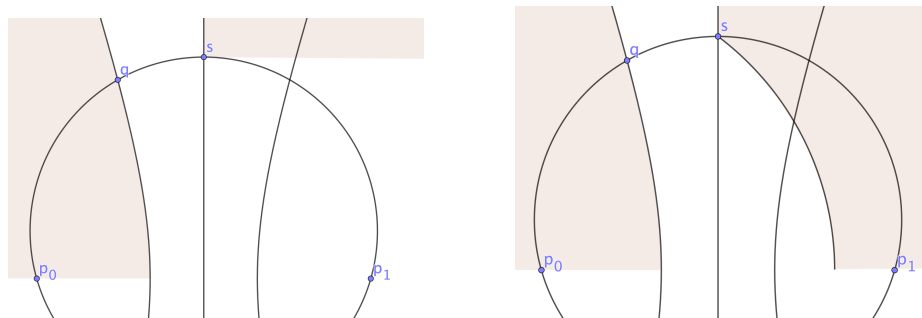
(a) The shaded area to the right of the central axis contains points which are at least $d(q, s)$ from that on the left.

(b) The shaded area here also con-tains contains points which are at least $d(q, s)$ from that on the left.

Fig. 3: The shaded area to the left of each central axis denotes the locus defined by $d(q', p_0) - d(q', p_1) > \tau t$ for any $q' \in U$. In both cases, for any $s' \in S$ in the shaded area to the right of the central axis, $d(q', s') > t$.

However, the static partition can be extended to include more of the finite search space, by including any value $s' \in S$ to the right of the central axis where also $d(s', p_0) \geq d(s, p_0)$, as illustrated in Figure 3b. This not only increases the cardinality of the potentially excluded subset, but also avoids the requirement to calculate the 2D projection.

The static and dynamic classes represented in Figure 3b are now formally defined as

$$
\begin{aligned}
\mathcal{S} &= \{s' \leftarrow S \mid d(s', p_0) \geq d(s', p_1) \quad \wedge \quad d(s', p_0) \geq \tau d(p_0, p_1)\} \\
\mathcal{Q} &= \{q' \leftarrow U \mid d(q', p_1) - d(q', p_0) > t/\tau\}
\end{aligned}
$$

with the property that it is impossible for any element $q' \in \mathcal{Q}$ to be within distance t of any element $s' \in \mathcal{S}$.

The validity of this extension seems evident from the illustration in Figure 3b, but this needs to be demonstrated for the general case. A full justification of its correctness is included in an extended version of this paper available at [3].

3   The Partition Mechanism

The addition of the criterion $d(s', p_0) \geq \tau d(p_0, p_1)$ to the static partition allows further exclusion potential relying on the normal triangle inequality method, i.e. if $d(q, p_0) < \tau d(p_0, p_1) - t$. This extension to the exclusion criterion is illustrated in Figure 4a.
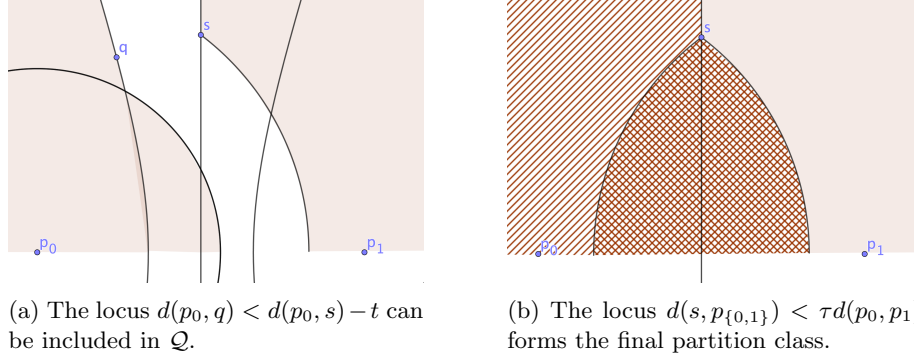

(a) The locus $d(p_0, q) < d(p_0, s) - t$ can be included in $\mathcal{Q}$.

(b) The locus $d(s, p_{\{0,1\}}) < \tau d(p_0, p_1)$ forms the final partition class.

Fig. 4

Furthermore, when also including the symmetric opposite criteria, the static partition now defines three subclasses as shown in Figure 4b. It may further be noted that the third of these subclasses may also independently excluded if $d(q, p_0)$ or $d(q, p_1) \geq \tau K + t$, again relying only on triangle inequality.

So finally, according to the geometry established in Section 2.1, a static partition $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$ of S can be established for a pair of reference points p_0, p_1 with $K = d(p_0, p_1)$ and a given value of τ as follows:

$$
\begin{aligned}
\mathcal{S}_1 &= \{s \leftarrow S \mid d(s, p_0) \geq d(s, p_1) \quad \wedge \quad d(s, p_0) \geq \tau K\} \\
\mathcal{S}_2 &= \{s \leftarrow S \mid d(s, p_0) < d(s, p_1) \quad \wedge \quad d(s, p_1) \geq \tau K\} \\
\mathcal{S}_3 &= \{s \leftarrow S \mid d(s, p_0) < \tau K \quad \wedge \quad d(s, p_1) < \tau K\}
\end{aligned}
$$

These static regions are illustrated on the 2D plane in Figure 4b.

For a given query object q with threshold t, where $A = d(q, p_0)$ and $B = d(q, p_1)$, these classes can be excluded from a search as follows:

$$
\begin{aligned}
\mathcal{S}_1 &: \quad B - A > t/\tau \quad \vee \quad A < \tau K - t \\
\mathcal{S}_2 &: \quad A - B \geq t/\tau \quad \vee \quad B < \tau K - t \\
\mathcal{S}_3 &: \quad A \geq \tau K + t \quad \vee \quad B \geq \tau K + t
\end{aligned}
$$

Note that it is possible for the exclusion of region $\mathcal{S}_3$ to occur in conjunction with that of $\mathcal{S}_1$ or $\mathcal{S}_2$. The mechanism resulting from these definitions is now evaluated in Section 4.

4   Evaluation

Before proceeding with a full quantitative evaluation, it is interesting to view graphical representations based on a sample from a particular data set, in order to give a more pragmatic view of the Ptolemaic partition mechanism in comparison with hyperplane (hyperbolic) and Hilbert (four-point) partition mechanisms.
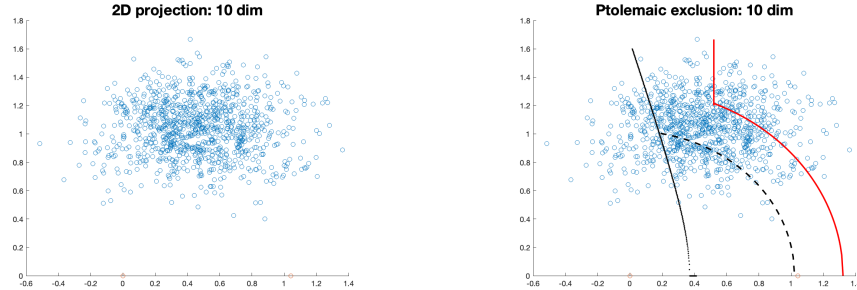
(a) A scatter plot based on randomly selected reference points. p_0 is plotted at $(0,0)$ and p_1 at $(0, d(p_0, p_1))$. The data set is plotted according to the distance of each value from p_0 and p_1.

(b) The boundary of the class $\mathcal{S}_1$ is plotted in red. Values lying to the right of this boundary can be excluded for queries lying to the left of either black boundary.

Fig. 5: Graphical view of the Ptolemaic partition mechanism based on a 2D projection. $\tau = 1.3$ and the query threshold is 0.3.

Figure 5a shows a scatter plot of 1,000 values randomly generated in a 10-dimensional Euclidean space, each projected onto a 2D plane according to their distances from two randomly generated pivot values.

Figure 5b shows the same projection superimposed with one of the partition boundaries of the Ptolemaic partition mechanism with a τ value of 1.3, and a query threshold of 0.3, which is the mean nearest-neighbour distance. The boundary of the static region $\mathcal{S}_1$ is shown in red; those points lying to the right of the red boundary are thus subset to exclusion when either $d(q, p_1) - d(q, p_0) > 1.3 \cdot 0.3$, or if $d(q, p_1) < 1.3 \cdot d(p_0, p_1) - 0.3$. The boundaries of these regions are shown by solid and dotted black lines respectively; every point to the left of either boundary represents a value for $q' \in \mathcal{Q}$ which allows exclusion of the static class.

Figure 6 shows the same plot with boundaries for standard hyperplane and Hilbert exclusion, in 6a and 6b respectively.

It is clear, at least in this case, that the Ptolemaic mechanism always excludes a smaller subset, while the probability of the exclusion being possible is higher. From Figure 6a it is evident that the dimensionality of the data set challenges hyperplane exclusion, while both Ptolemaic and Hilbert mechanisms continue

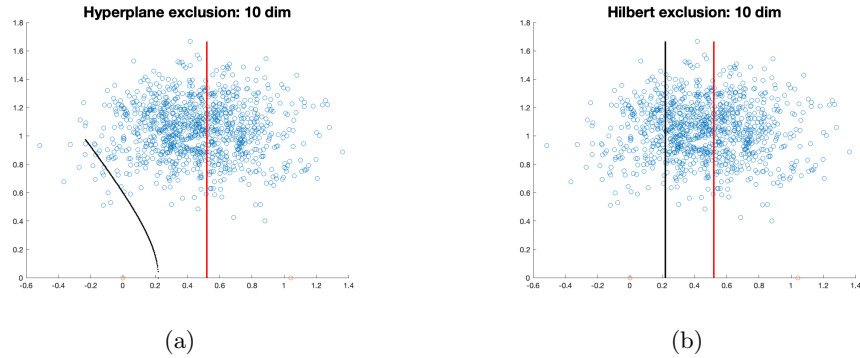(a)   (b)

Fig. 6: Equivalent graphical views of hyperplane and Hilbert partition mechanisms. In each case, queries falling to the left of the black line can be used to exclude the subset of data falling to the right of the red line. Again a query threshold of 0.3 has been used.

to remain effective. Finally, while it is not possible to judge the relative efficacy of Ptolemaic vs. Hilbert from these diagrams, it can be observed that neither is a proper subset of the other, and it is therefore possible to use both Ptolemaic and Hilbert with respect to the same pair of reference points. This would allow a hybrid mechanism, more effective that either in isolation, based on the same dynamic measurements of $d(q, p_0)$ and $d(q, p_1)$.

4.1   Quantitative Evaluation

Quantitative evaluation is performed over sets of uniformly generated Euclidean data, from between 8 and 20 dimensions. 50k data objects are used and 1k non-intersecting queries are evaluated. The threshold used for each query corresponds to the 5nn distance as pre-calculated over the data.

Experiments were performed over Ptolemaic, Hilbert, and hyperplane mechanisms. For each experiment, a fixed number of reference points was used, and each of the $\binom{n}{2}$ pairs of reference points was used to construct a partition over the space. The single outcome is the mean proportion per query of values that were successfully excluded, this value being between 0 and 1. For the majority of the experiments 10 reference points used, this giving 45 different partitions. Thus all results given correspond to the proportion of the data that can be successfully excluded at cost of only 10 distance calculations per query.

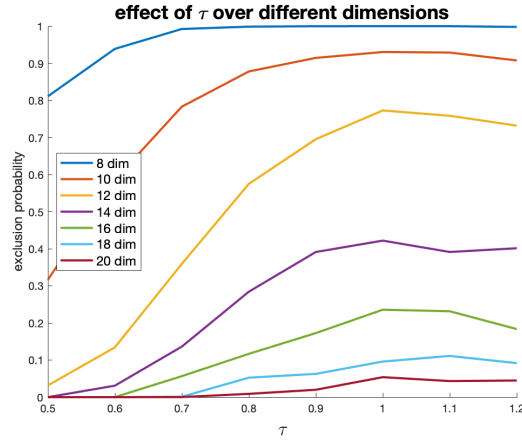All experiments were performed using MatLab and the code is available[5].

Fig. 7: Probability of successful exclusion for differing values of τ in different dimensions. Note that when $\tau = 0.5$ the mechanism is identical to traditional hyperplane (Hyperbolic) exclusion.

4.2 Choosing τ

First, different values for τ are examined. As mentioned, when $\tau = 0.5$ the mechanism reverts to simple hyperplane exclusion; while a value of less than 0.5 is technically possible, there is no value in such a choice. As τ gets large, then ever fewer data will be present in the partition which may be excluded, and again the mechanism will become useless. Early tests showed that a value somewhere around 1 is usually close to optimal, although for specific reference point pairs an optimum values of between around 0.8 and 1.2 were observed.

It would in fact be possible to optimise τ based on each particular pair of reference points, which we have not yet investigated thoroughly. In this experiment a fixed value of τ is applied to all partitions, which is possibly more realistic for many scenarios.

Figure 7 shows the results of various values of τ when applied to data of between 8 and 20 dimensions. As can be seen there is a general trend of larger values being better as dimensions increase, but only within quite a small margin; while there is clearly an element of noise in this experiment, the best value in each case is either 1.0 or 1.1, although further investigation is warranted. For further experiments described over the data of different dimensions, the best value of τ found in this experiment was used.
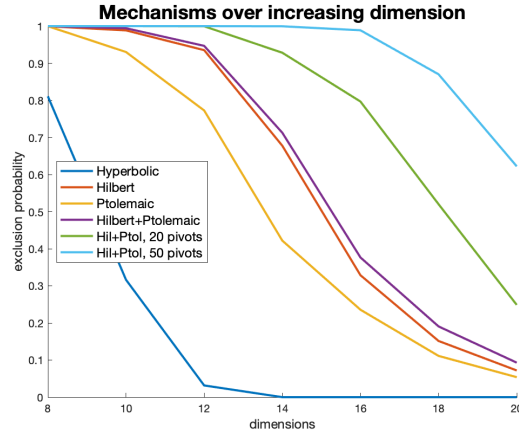
Fig. 8: Performance of different exclusion mechanisms as dimensionality increases, measured as the probability of a non-solution being excluded based on the distances among query and reference values.

4.3 Evaluation over High Dimensional Data

Having picked a value for τ, outcomes for some different mechanisms over data ranging from 8 to 20 dimensions are given. Four mechanisms are used: Hyperplane (Hyperbolic) exclusion; Hilbert exclusion; Ptolemaic exclusion, and finally a combination of Hilbert and Ptolemaic exclusion.

For this combination, each pair of reference points was used to construct both Ptolemaic and Hilbert partitions, and the union of non-excluded data was calculated. The observation here is that it is possible for different data to be excluded by each mechanism. As the essential query-time cost of performing the exclusions is the cost of the two distance operations and some relatively cheap arithmetic, taking the union of all possible exclusions makes practical sense as the distance calculations are amortised. As Hilbert exclusion always allows exclusion from a superset of queries identified by Hyperbolic exclusion, there is no point in combining that mechanism also.

Figure 8 shows the outcome. As can be seen, while the performance of Hyperbolic exclusion falls rapidly away after around 8-10 dimensions, both Ptolemaic and Hilbert perform much better into the higher dimensional range. Hilbert always performs better than Ptolemaic, which is not very surprising as the four-point lower bound property is stronger then the Ptolemaic inequality, and technically applies to a smaller subset of metric spaces. What is more interesting, however, is that the combination of Ptolemaic and Hilbert gives a strictly better result than Hilbert alone; that is, the data sets identified for exclusion by the two mechanisms are not in a strict subset relation. Again it is noted that the

---

[5] https://bitbucket.org/richardconnor/partitions

inherent query-time cost of the joint mechanism is very similar to the cost of just one, as in all cases the query to pivot distances calculated are reused in both mechanisms.

The final plots in the graph show the use of 20 and 50 pivot values for the combination mechanism. Although only doubling the number of query-to-pivot distances required, 20 pivots gives $\binom{20}{2}$ i.e. 190 partitions to apply, and as can be seen the increase makes for a much higher exclusion ratio. Similarly, 50 pivots gives 1,225 partitions. The important observation however is that there is a clear degree of orthogonality in the randomly selected partitions, allowing almost perfect exclusion in 12 and 16 dimensions respectively.

5   Conclusions and Future Work

This paper fills a significant gap in the literature, that is a set partition that can be used as an exclusion mechanism for the Ptolemaic inequality; for some years, other distance lower-bounds have had known mechanisms and in this sense the Ptolemaic inequality has been an outlier.

In its simplest form, the mechanism is quantitatively much better than traditional hyperplane partitioning, and not quite as good as Hilbert partitioning. This is almost inevitable, as the class of spaces to which the inequalities can be applied are in a strict subset relation. Should this mechanism have been identified before Hilbert exclusion it would have been deserving of significant excitement, but this is nowadays tempered by the existence of the more effective Hilbert exclusion over essentially the same subclass of metric spaces.

However, it is the case that the individual data objects which the new mechanism excludes are not a proper subset of those identified by Hilbert exclusion, and as shown the two mechanisms may operate in conjunction to give a unified mechanism which, for the same cost of distance calculations against reference points, gives a better exclusion outcome than either in isolation. Particularly in high-dimensional spaces, this therefore gives a further increment in the limit of dimensionality for which exact search can be effective. While the "rule of thumb" used to be that 8-10 dimensions was the effective limit for exact search [9], with the combined mechanism 16 dimensions can be effectively searched while avoiding almost all explicit distance calculations.

Some further avenues are worth exploring. First, it is feasible to calculate individual τ values customised to each particular pair of pivot points, rather than to choose a single value for the whole set. This would be expected to give significant, if incremental, improvement in performance.

Finally, there are many other contexts beyond a simple recursive decomposition of a large data set where such mechanisms can be used. It is therefore of potential value in its own right for this previous gap in knowledge to be filled.

6   Acknowledgements

References

1. Chávez, E., Navarro, G., Baeza-Yates, R., Marroquín, J.L.: Searching in metric spaces **33**(3), 273–321. https://doi.org/10.1145/502807.502808, `https://dl.acm.org/doi/10.1145/502807.502808`
2. Connor, R., Vadicamo, L., Cardillo, F.A., Rabitti, F.: Supermetric search with the four-point property. In: Proceedings of 9th International Conference on Similarity Search and Applications (SISAP 2016). pp. 51–64. LNCS, Springer (2016)
3. Connor, R.: A ptolemaic partitioning mechanism (2022). https://doi.org/10.48550/ARXIV.2208.09324, `https://arxiv.org/abs/2208.09324`
4. Connor, R., Cardillo, F.A., Vadicamo, L., Rabitti, F.: Hilbert exclusion: improved metric search through finite isometric embeddings. ACM Transactions on Information Systems (TOIS) **35**(3), 1–27 (2016)
5. Connor, R., Vadicamo, L., Cardillo, F.A., Rabitti, F.: Supermetric search. Information Systems **80**, 108–123 (2019)
6. Hetland, M.L.: Ptolemaic indexing. Journal of Computational Geometry **6**, 165–184–165–184 (2015)
7. Hetland, M.L., Skopal, T., Lokoč, J., Beecks, C.: Ptolemaic access methods: Challenging the reign of the metric space model. Information Systems **38**(7), 989–1006 (2013). https://doi.org/https://doi.org/10.1016/j.is.2012.05.011, `https://www.sciencedirect.com/science/article/pii/S0306437912000786`
8. Lokoč, J., Hetland, M.L., Skopal, T., Beecks, C.: Ptolemaic indexing of the signature quadratic form distance. In: Proceedings of the Fourth International Conference on SImilarity Search and APplications. p. 9–16. SISAP '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/1995412.1995417, `https://doi.org/10.1145/1995412.1995417`
9. Weber, R., Schek, H.J., Blott, S.: A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In: Proc. of 24th VLDB. vol. 98, pp. 194–205. Morgan Kaufmann (1998)
10. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity search: the metric space approach, vol. 32. Springer Science & Business Media (2006)