On the complexity of the partial least-squares matching Voronoi diagram

Matthias Henze* Rafel Jaume† Balázs Keszegh‡

Abstract
Given two point sets of sizes \( n \) and \( m \), we study the partial matching problem of translating the smaller point set to a position where it is best resembled by an equally sized subset of the larger point set. Measuring the similarity is done by the sum of squares of the Euclidean distances between the matched points in either set. A Voronoi-type diagram can be associated with this matching problem, we are interested in finding an optimal injective match of \( \{ a_1, \ldots, a_n \} \) and \( \{ b_1, \ldots, b_m \} \) be finite point sets in \( \mathbb{R}^d \) with \( m \leq n \). In the partial point matching problem, we are interested in finding an optimal injective match of \( B \) within \( A \) with respect to a given measure of similarity and a set of available transformations. In this paper, we study the sum of squared Euclidean distances as a measure and will allow the point set \( B \) to be translated to any desired position, that is, we want to solve

\[
\begin{align*}
\text{minimize} & \quad c_\pi(t) = \sum_{i=1}^m \| b_i + t - a_{\pi(i)} \|^2 \\
\text{subject to} & \quad \pi : [m] \rightarrow [n] \text{ an injection,} \\
& \quad t \in \mathbb{R}^d.
\end{align*}
\]

The injectivity requirement on the matching is crucial. Dropping it reduces the problem to traversing regions of an overlay of the \( m \) Voronoi diagrams of the point sets \( A - b_i, i \in [m] \), which can be done efficiently.

As an approach to decide whether problem (1) can be solved in polynomial time in \( n \) and \( m \), Rote [8] associated a Voronoi-type diagram with this matching problem, that may be called the partial (least-squares) matching Voronoi diagram \( V(A, B) \) of \( A \) and \( B \). In this diagram each injection \( \pi : [m] \rightarrow [n] \) defines a (possibly empty) region \( P_\pi \) that consists of all translations \( t \) for which \( \pi \) is the optimal assignment for \( B + t \). More precisely,

\[
P_\pi = \{ t \in \mathbb{R}^d : c_\pi(t) \leq c_\sigma(t) \text{ for all } \sigma : [m] \rightarrow [n] \}.
\]

Rote [8] observed that every \( m \)-element subset of \( A \) is the image of at most one injection that contributes a region to the diagram. This is due to the translation invariance of the optimal matching between two equally sized sets. Further, by

\[
c_\pi(t) = \sum_{i=1}^m \| b_i - a_{\pi(i)} \|^2 + 2 \left( t, \sum_{i=1}^m (b_i - a_{\pi(i)}) \right) + m\| t \|^2,
\]

the regions \( P_\pi \) are intersections of finitely many affine half-spaces and hence the partial matching Voronoi diagram is a polyhedral subdivision of \( \mathbb{R}^d \). Allowing multiplicities in the point set \( B \), we get as a special case the \( m \)th order Voronoi diagram \( V_m(A) \) of \( A \) by setting \( b_1 = \ldots = b_m = 0 \).

A polynomial bound on the number of cells, i.e., full-dimensional regions, of \( V(A, B) \) would yield an efficient algorithm to solve the matching problem (1). Such an algorithm would traverse all cells and solve the minimization problem inside a fixed cell by means of quadratic programming. For more details on this approach we refer to [8]. In the special case \( V_m(A) \) the number of cells is known to be polynomial in \( n \) and \( m \). A rough bound of \( O(m^{2d}) \) follows from the complexity of hyperplane arrangements (see [3]), and the best bound that is sensitive to \( m \) is due to Clarkson & Shor [2] and reads \( O(m^{2d}/m! \frac{d}{2}) \) for \( n/m \rightarrow \infty \).

The following result implies that, for \( d = 1 \), the number of cells in \( V(A, B) \) is at most quadratic.

**Theorem 1 (Rote [8])** Every line intersects the interior of at most \( n(n-m)+1 \) cells of \( V(A, B) \).

This bound is best possible which can be seen by taking \( A \) and \( B \) to be equally spaced points on a line, and where the spacing between the points in \( A \) is very small and between the points in \( B \) very large.

We may generalize this example to arbitrary dimension by putting such an instance on \( \frac{n}{2} \) and \( \frac{m}{2} \) points in each of the \( d \) coordinate axes. Doing this carefully leads to a diagram with \( \Omega \left( \frac{m! (n-m)!}{d^{2d}} \right) \) cells and motivates the subsequent conjecture.
Observation 1 The partial matching Voronoi diagram of point sets $A, B \subset \mathbb{R}^d$ of sizes $n = |A| \geq |B| = m$ consists of $O(m^d(n - m)^d)$ cells.

2 Counting stable matchings

In this section, we present a combinatorial result that will be used later on in order to derive first bounds on the complexity of $\mathcal{V}(A, B)$. The combinatorial problem we are considering is a stable marriage type question.

The stable marriage problem was first introduced by Gale & Shapley [4] in 1962. It is usually stated as trying to marry $n$ men with $n$ women, each of them with a ranking of the people of the opposite gender, in a way that no non-married pair would bilaterally want to have an affair. In the classical paper [4], it is shown that such a stable marriage exists for any given set of rankings. In fact, in general there are a lot of stable marriages for a single instance. Knuth [7] and Irving & Leather [5] gave some bounds on this number.

A lot of variants of the problem have been studied (see [6] for a survey), e.g., stable marriages with incomplete rankings or with ties, the stable roommates problem or the hospital/residents problem. The variant we are interested in was first introduced by Shapley & Scarf [9] in 1974, and is called the “House Allocation Problem.” The literature on this problem focuses on algorithmic questions, while we are interested in a bound on the number of stable marriages.

Throughout this section, we let $\mathcal{D} = (d_1, \ldots, d_m) \in S^n_m$ be a list of permutations on $n$ elements. These may be regarded as $m$ linear orderings on $[n]$ in the sense that $a \leq b$ if and only if $d_i^{-1}(a) \leq d_j^{-1}(b)$. The elements of $[m]$ are usually denoted by $i, j, \ldots$ and the elements of $[n]$ by $a, b, \ldots$.

Definition 1 We say that an injection $\mu : [m] \to [n]$ is a matching and its image $M(\mu) = \mu([m])$ is called the matched set of $\mu$.

- A matching $\mu$ is said to be better than $\nu$ (with respect to $\mathcal{D}$), if
  $\mu(i) \leq \nu(i)$ for all $i \in [m]$.

- A matching $\mu$ is called stable (with respect to $\mathcal{D}$), if there is no better matching $\nu \neq \mu$.

The following observations about this kind of matchings can be easily proved.

Observation 1

i) Stable matchings are non-parallel, i.e., given two stable matchings $\mu$ and $\nu$, there are $i, j \in [m]$ such that $\mu(i) < \nu(i)$ and $\nu(j) < \mu(j)$.

ii) If $\mu$ is stable, then
$\{a \in [n] : \exists i \in [m] \text{ s.t. } a \leq \mu(i) \} \subseteq M(\mu)$.

In particular, only the $m$ smallest values in each of the orderings $\leq_1, \ldots, \leq_m$ can appear in a stable matching. Hence, at most $m^2$ different elements $a \in [n]$ are used.

iii) If $\mu$ is stable, there is no sequence $i_1, \ldots, i_m$ of length $m' \leq m$ satisfying
$\mu(i_{k-1}) <_{i_k} \mu(i_k)$ for all $k \in [m']$,
where $i_0 := i_{m'}$.

Definition 2 Let $\rho \in S_m$ be a permutation on $[m]$. The $\rho$-greedy matching $\mu$ for $\mathcal{D}$ is the unique injection that satisfies $\mu(i) \leq \rho(i) b$, for all $i \in [m]$ and for all $b \in [n] \setminus \{\mu(1), \ldots, \mu(i - 1)\}$. A matching is called greedy if it is $\rho$-greedy for some $\rho \in S_m$.

The following characterization of stable matchings already appears in [1, Lem. 1] for the case $m = n$. Therein, stable matchings are called Pareto-efficient and greedy matchings serial-dictatorship mechanisms. We give a more focused and shorter proof of this result that moreover covers the general case $m \leq n$.

Theorem 2 A matching is stable if and only if it is greedy.

Proof. First of all, let $\mu$ be a $\rho$-greedy matching and let us assume on the contrary that there is a better matching $\nu$. Let $i \in [m]$ be the smallest index such that $\nu(i) <_{\rho(i)} \mu(i)$. This contradicts the definition of the greedy matching $\mu$ since it would have chosen the element $\nu(i)$. Hence, $\mu$ is stable.

Conversely, if $\mu$ is a stable matching, then there must be some index $i \in [m]$ such that $\mu(i) = d_i(1)$, i.e., $\mu(i)$ is the smallest in the order $\leq_i$. Assuming the contrary means, by Observation 1 iii), that for all $i \in [m]$ there exists an $i' \in [m] \setminus \{i\}$ such that $\mu(i') = d_i(1)$. In particular, there is a cycle $i_1, \ldots, i_m$ such that $\mu(i_{k-1}) = d_{i_k}(1)$, for all $k \in [m']$, where $i_0 := i_m$. The matching $\nu$ obtained from $\mu$ by exchanging $\mu(i_k)$ for $d_{i_k}(1)$, for all $k \in [m']$, is then better than $\mu$, contradicting stability.

We now construct the permutation $\rho$ as follows. Let $\rho(1) = i_1$, where $\mu(i_1) = d_{i_1}(1)$. Then, we consider the induced matching $\mu_1 : [m] \setminus \{i_1\} \to [n] \setminus \{\mu(i_1)\}$. It is easy to see that this is a stable matching, as a better matching would yield a better matching for $\mu$ as well. Again we find an element $i_2 \in [m] \setminus \{i_1\}$ such that $\mu_1(i_2) = d_{i_2}(1)$ and set $\rho(2) = i_2$. Repeating this process yields the permutation $\rho \in S_m$ and $\mu$ is $\rho$-greedy. $\square$
Corollary 3 The number of stable matchings with respect to a given \( D \) is at most \( mn! \) and, in general, this bound cannot be improved.

Proof. By Theorem 2, stable matchings are greedy matchings. There exists only one greedy matching for each of the \( mn! \) permutations of \([n]\).

For the lower bound construction observe that if all the permutations \( d_i \) are equal, then every \( \rho \in S_m \) induces a different greedy matching. \( \square \)

Motivated by the translation invariance discussed in the introduction, we are actually interested in the number of matched sets of stable matchings. To this end, we define

\[ m^g = \{ M(\mu) : \mu \text{ is a stable/greedy matching} \} \]

and

\[ M^g = \bigcup_{\mu \text{ is a stable/greedy matching}} M(\mu). \]

The latter set contains those elements of \([n]\) that appear in some stable/greedy matching. Clearly, \( |m^g| \leq \binom{|M^g|}{m} \), which motivates us to determine the order of magnitude of \( |M^g| \) in the worst case. We have already seen in Observation 1 ii), that \( |M^g| \leq m^2 \). We are not aware of an upper bound of the type \( o(m^2) \). Our best lower bound construction is quite apart from this upper bound and comes from the following recursive example:

Claim 1 There exists a \( \mathcal{D} \) such that

\[ |M^g| \geq \frac{m}{2} \log 4m \in \Omega(m \log m). \]

Proof. We give a recursive construction. In each step we double the size of \( m \). For \( m = 1 \), evidently \( |M^g| = 1 \). We say that a construction reaches some element \( a \in [n] \) if \( a \in M^g \), i.e., \( a \in M(\mu) \) for some stable/greedy matching \( \mu \). Assuming that we already constructed a \( \mathcal{D} \) for \( m \) that reaches \( \frac{m}{2} \log 4m \) elements, we construct a \( \mathcal{D}' \) for \( 2m \) that reaches \( m \log 8m \) elements. In order to do that, we let \( m \) new elements be the smallest in the preference orders of \( d_1, \ldots, d_m \) and let the same elements in the same order be the smallest in the preference orders for \( d_{m+1}, \ldots, d_{2m} \). Thus \( d_i(1) = d_{i+m}(1) \) for all \( i \in [m] \). The remaining preferences of \( d_1, \ldots, d_m \) are chosen according to the construction for \( \mathcal{D} \). The rest of the preferences of \( d_{m+1}, \ldots, d_{2m} \) are set equivalently, just using a new set of elements of \([n]\) (we consider \( n \) not to be fixed). In this way, any \( \rho \) that starts with the first \( m \) elements of \( \lfloor 2m \rfloor \) in some order and then traverses the second \( m \) elements behaves exactly like a \( \rho' \) preference order on the previous construction \( \mathcal{D} \). The same thing applies to any \( \rho \) that starts on the second \( m \) elements. Thus, the number of elements of \([n]\) that belong to some set \( M(\rho) \) is at least \( m + |M^g| + |M^g| = m + 2 \frac{m}{2} \log 4m = m \log 8m \), as desired. \( \square \)

A curious observation is that it is NP-hard to decide whether an element of \([n]\) belongs to \( M^g \).

We now investigate \( |m^g| \) directly, because from the considerations above we only get \( |m^g| \leq \binom{|M^g|}{m} \leq \binom{m^2}{m} \in O((me)^m) \). The trivial bound \( |m^g| \leq mn \) is much better and basically this is the best upper bound we know up to the order of magnitude.

Our best lower bound comes from the following simple example:

Claim 2 There exists a \( \mathcal{D} \) such that

\[ |m^g| \geq \left( \frac{m}{\sqrt{2}} \right) \in \Omega\left( \frac{m^{m}}{\sqrt{m}} \right). \]

Proof. Choose \( \mathcal{D} \) such that for every \( \leq i \), the \( \left( \frac{m}{2} \right) \) smallest elements are the same, i.e., \( d_i(k) = d_j(k) \) for all \( i, j \in [m], k \in \left[ \left( \frac{m}{2} \right) \right] \), and the next elements are all different, i.e., \( d_i\left( \left( \frac{m}{2} \right) + 1 \right) \neq d_j\left( \left( \frac{m}{2} \right) + 1 \right) \) for \( i \neq j \). Now, when we greedily choose the smallest elements according to the order of some \( \rho \in S_m \), in the first \( \left( \frac{m}{2} \right) \) steps we choose one of the smallest \( \left( \frac{m}{2} \right) \) elements and after exactly \( \left( \frac{m}{2} \right) \) steps all of these have been chosen. Then, in the last \( \left( \frac{m}{2} \right) \) steps we choose by construction the \( \left( \frac{m}{2} \right) + 1 \)st smallest element with respect to the actual \( \leq i \). Thus, any subset of size \( \left( \frac{m}{2} \right) \) appears in the second half of some ordering \( \rho \), and hence we have \( \left( \frac{m}{\sqrt{2}} \right) \) different sets in \( m^g \). \( \square \)

Problem 1 Determine if the two previous claims exhibit worst case examples, i.e., is it true that \( |m^g| \in O(2^m) \) and \( |M^g| \in O(m \log m) \)?

3 A polynomial bound in the size of the bigger set

We are now prepared to discuss the geometric part of the argument that establishes an upper bound for the number of cells of \( V(A, B) \) which is polynomial in the size of the bigger set \( A \). This result is formalized in the following theorem.

Theorem 4 The partial matching Voronoi diagram \( V(A, B) \) of point sets \( A, B \subset \mathbb{R}^d \) of sizes \( n = |A| \geq |B| = m \) consists of \( O(m! m^d n^{2d}) \) cells.

The bisector between two points \( a_i, a_j \in A \) is the hyperplane defined by

\[ B(a_i, a_j) = \{ x \in \mathbb{R}^d : \| x - a_i \| = \| x - a_j \| \}. \]

Consider the hyperplane arrangement \( A \) in \( \mathbb{R}^d \) consisting of the \( m \binom{n}{2} \) hyperplanes \( B(a_i, a_j) - b_k \), where \( i, j \in [n] \) and \( k \in [m] \).

Lemma 5 Every cell \( C \in A \) intersects the interior of at most \( m! \) different cells of \( V(A, B) \).
Proof. Let $C \in \mathcal{A}$ be a cell and let $b_k \in B$. Every $t \in \text{int } C$ defines a linear ordering $\leq_k$ on $[n]$; For $i, j \in [n]$, we write $i \leq_k j$ if and only if $\|b_k + t - a_i\| \leq \|b_k + t - a_j\|$. This ordering induces a permutation $d_k$ on $[n]$ that, by definition of $\mathcal{A}$, depends on $k$ and $C$, but not on the actual $t \in \text{int } C$. In addition, it is easy to see that the optimal matching for any $t$ must be stable in the sense of Definition 1. Otherwise, a better matching would attain a smaller value for the cost function, contradicting the optimality of the first. Hence, by Corollary 3, the optimal matching can then be found in $O(n)$ time.

Proof. [Theorem 4] With the help of Lemma 5, Theorem 4 follows by observing that the number of optimal assignments will then depend on the complexity of the arrangement of bisectors in the parameter space. Note, that the bisectors are defined by the chosen distance function.

A particularly interesting example is the case in which, instead of translating $B$, we look to find optimal assignments when we are allowed to transform $B$ subject to any linear mapping. In this case, the parameter space we want to subdivide is $\mathbb{R}^d$, the space of all $(d \times d)$-matrices. The preference orders are now defined by the relations $\|M \cdot b_k - a_i\| \leq \|M \cdot b_k - a_j\|$, where $M \in \mathbb{R}^d$, which again correspond to linear inequalities in the parameter space. The associated hyperplane arrangement in $\mathbb{R}^d$ consists of $O(m^d n^{2d})$ cells, and hence the number of optimal assignments is $O(m^d n^{2d})$.

The particular case of rotating $B$ around a fixed point was posed as a problem in [8]. Still, our bound is only useful when $m$ is much smaller than $n$ and the polynomiality remains open even for equally sized sets $A$ and $B$.

4 Conclusions and further work

We considered the problem of bounding the number of optimal assignments of a point set $B$ to a subset of a bigger set $A$ with respect to the sum of squared distances and translations. Our main result is a bound that is polynomial in the size of the bigger set, independently of the size of the smaller one. We have also seen that our arguments extend to assignments with respect to more general cost functions and other sets of available transformations.

In view of Conjecture 1, the natural line of further investigations is reducing the factor $m!$ that appears in the bound of Theorem 4. As illustrated by Claim 2, our combinatorial approach cannot yield an upper bound of the type $o\left(\frac{m^n}{n!}\right)$. Further improvements have to rely more on the geometric properties of the problem. For instance, in the least-squares matching Voronoi diagram, each subset $A' \subseteq A$ of size $|A'| = m$ gives rise to at most one cell, and also not every stable matching that is counted in Section 2 necessarily comes from an optimal assignment. Both are properties that we have not been able to exploit so far.

Acknowledgement. This work was carried out during a visit of the third author in Berlin, funded by a EUROCORES short term visit grant in the framework of the EuroGIGA programme.

References