Linear and $O(n \log n)$ Time Minimum-Cost Matching Algorithms for Quasi-convex Tours (Extended Abstract)

Samuel R. Buss* Peter N. Yianilos†

Abstract

Let G be a complete, weighted, undirected, bipartite graph with n red nodes, n' blue nodes, and symmetric cost function $c(x, y)$. A maximum matching for G consists of $\min\{n, n'\}$ edges from distinct red nodes to distinct blue nodes. Our objective is to find a minimum-cost maximum matching, i.e. one for which the sum of the edge costs has minimal value. This is the weighted bipartite matching problem; or as it is sometimes called, the assignment problem.

We report a new and very fast algorithm for an abstract special case of this problem. Our first requirement is that the nodes of the graph are given as a 'quasi-convex tour'. This means that they are provided circularly ordered as $x_1, \ldots, x_N$ where $N = n + n'$, and that for any x_i, x_j, x_k, x_ℓ, not necessarily adjacent but in tour order, with x_i, x_j of one color and x_k, x_ℓ of the opposite color, the following inequality holds:

$$c(x_i, x_\ell) + c(x_j, x_k) \leq c(x_i, x_k) + c(x_j, x_\ell)$$

If $n = n'$, our algorithm then finds a minimum-cost matching in $O(N \log N)$ time. Given an additional condition of 'weak analyticity', the time complexity is reduced to $O(N)$. In both cases only linear space is required. In certain cases these results apply even if $n \neq n'$.

Our algorithm is conceptually elegant, straightforward to implement, and free of large hidden constants. As such we expect that it may be of practical value in several problem areas.

Many natural graphs satisfy the quasi-convexity condition. These include graphs which lie on a line or circle with the canonical tour ordering, and costs given by any concave-down function of arclength — or graphs whose nodes lie on an arbitrary convex planar figure with costs provided by Euclidean distance.

The weak-analyticity condition applies to points lying on a circle with costs given by Euclidian distance, and we thus obtain the first linear-time algorithm for the minimum-cost matching problem in this setting (and also where costs are given by the L_1 or L_∞ metrics).

Given two symbol strings over the same alphabet, we may imagine one to be red and the other blue, and use our algorithms to compute string distances. In this formulation, the strings are embedded in the real line and multiple independent assignment problems are solved; one for each distinct alphabet symbol.

While these examples are somewhat geometrical, it is important to remember that our conditions are purely abstract; so that our algorithms may find application to problems in which no direct connection to geometry is evident.

Keywords — Assignment problem, bipartite weighted matching, computational geometry, concave penalty function, convexity, linear time, Monge property, quadrangle inequality, string comparison.

1. Introduction

The abstract above gives a short overview of the contents of the paper and we shall give an in-depth discussion of our definitions, results and algorithm below. However, we first give a quick review of prior related work on matching. We shall consider graphs G which have N nodes, the nodes are partitioned into a set of n red nodes and n' blue nodes with $N = n + n'$. G is *balanced* if it has equal numbers of red and blue nodes. There is a symmetric cost function $c(x, y)$, which gives the cost of an edge from node x to node y, with x and y of distinct colors. A matching is a set of edges with no endpoints in common that match all the nodes of one color with the same number of nodes of the opposite color. The cost of a matching is the sum of the costs of its edges. The problem of finding a minimal-cost matching for a general bipartite graph is known to have an $O(N^3)$ time algorithm (see Lawler [15] for this and other background on matching), and for graphs with nodes in the plane with the Euclidean distance as cost function, there is a $O(N^{2.5} \log N)$ time algorithm, due to Vaidya [19].

The minimum-cost matching problem is substantially easier in the case where the nodes are linearly or circularly ordered. The simplest version of linear/circular ordering is where the points lie on a line or on a curve homeomorphic to a circle, and the cost $c(x, y)$ of an edge between x and y is equal to the shortest arclength distance between the nodes. The matching problem for this arclength cost function has been studied by Karp-Li [11], Aggarwal et al. [1], Werman et al. [20] and others, and is the 'Skis and Skiers' problem of Lawler [15]. Karp-Li have given linear time algorithms for this matching problem; Aggarwal et al. have generalized the linear time algorithm to the transportation problem.

A more general version of the matching problem for linearly ordered graphs has been studied by Gilmore-Gomory [8] (see [15]). In this version, the cost of an edge from a red node x forward to a blue node y is defined to equal $\int_x^y f$ and from a blue node x forward to a red node y to equal $\int_x^y g$, for some functions f and g. This matching problem has a linear time algorithm provided $f + g \geq 0$.

Another version of the matching problem for linearly ordered graphs is considered by Aggarwal et al.[1]: they use graphs which satisfy a "Monge" property which states that the inequality (1) below holds except with the inequality sign's direction reversed. They give a linear time algorithm for the matching problem for (unbalanced) Monge graphs.

In the prior work most closely related to this paper, Marcotte and Suri [17] consider the matching problem for a circularly ordered, balanced tour in which the nodes are the vertices of a convex polygon and the cost function is equal to Euclidean distance. This matching problem is substantially more complicated than the comparatively simple 'Skis and Skiers' type problems; nonetheless, Marcotte and Suri give an $O(N \log N)$ time algorithm which solves this minimum-cost matching problem. For the case where the nodes are the vertices of a simple polygon and the cost function is equal to the shortest Euclidean distance *inside* the polygon, they give an $O(N \log^2 N)$ time algorithm.

The main results of this paper apply to all the above matching problems on circularly or linearly ordered tours, except for the sole exception of the unbalanced, Monge graphs. For the 'Skis and Skiers' and the Gilmore-Gomory

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92092-0112. Email: sbuss@ucsd.edu. Supported in part by NSF grant DMS-9205181.

†NEC Research Institute, 4 Independence Way, Princeton, NJ 08540. Email: pny@research.nj.nec.com.

problems, Theorem 2 gives new linear time algorithms which find minimum-cost matchings which are different than the traditional minimum-cost matchings (and our algorithms are more complicated than is necessary for these simple problems). Our algorithms subsume those of Marcotte and Suri and give some substantial improvements: First, with the weak analyticity condition, we have linear time algorithms for many important cases, whereas Marcotte and Suri's algorithm takes $O(N \log N)$ time. Second, our assumption of quasi-convexity is considerably more general than their planar geometrical setting and allows diverse applications. Third, our algorithms are conceptually simpler than the divide-and-conquer methods used by Marcotte and Suri, and we expect that our algorithms are easier to implement.

We list some sample applications of our algorithms in Examples 1-8 below.

Our quasi-convex property is equivalent to the "inverse quadrangle inequality" used, for instance, by [6], but is weaker than the similar "inverse Monge property" of [4].[2] Dynamic programming problems based on cost functions which satisfy the (inverse) quadrangle inequality and some closely related matrix-search problems have been studied by many authors, including [2, 3, 4, 5, 6, 7, 10, 12, 13, 14, 16, 21, 22]. However, there seems to be no direct connection between our quasi-convex matching problem and the problems solved by these authors.

We now give the definitions necessary to state the main results of this paper. We think of the nodes of the graph G as being either a linear or circular tour of the graph; in the case of a circular tour, we think of the node x_1 as following again after x_N.

Definition A sequence of nodes $x_{i_1}, x_{i_2}, \ldots, x_{i_\ell}$ are *in input order* if and only if $i_1 < i_2 < \cdots < i_\ell$. The nodes are defined to be *in tour order* if and only if there exists a k such that the sequence $x_{i_k}, \ldots, x_{i_\ell}, x_{i_1}, \ldots, x_{i_{k-1}}$ is in input order.

Definition The nodes $x_1, \ldots, x_N$ form a *quasi-convex* tour if and only if, whenever x_i, x_j, x_k, x_ℓ are in tour order, with x_i and x_j of one color and x_k and x_ℓ of the other color, then

$$c(x_i, x_\ell) - c(x_i, x_k) \quad \leq \quad c(x_j, x_\ell) - c(x_j, x_k). \qquad (1)$$

Reordering terms in (1) gives

$$c(x_i, x_\ell) + c(x_j, x_k) \ \leq \ c(x_i, x_k) + c(x_j, x_\ell).$$

To give a geometric intuition to quasi-convexity, note that when x_i, x_j, x_k, x_ℓ are the vertices of a quadrilateral, the inequality states that the sum of the lengths of diagonals is greater than or equal to the sum of the lengths of two of the sides.

Definition The tour $x_1, \ldots, x_N$ of G is *linear* if and only if, the following holds: For all $i < j < k$, we have

$$c(x_i, x_j) \leq c(x_i, x_k)$$

if x_i is of opposite color from x_j and x_k; and we have

$$c(x_i, x_k) \geq c(x_j, x_k)$$

if x_k is of opposite color from x_i and x_j.

[2]Any Monge matching problem may be trivially transformed into a quasi-convex matching problem, but not vice-versa.

The property of quasi-convexity is defined independently of the starting point of the tour; i.e., the nodes of the tour can be 'rotated' without affecting quasi-convexity. Obviously, the definition of linear tours is sensitive to the choice of starting point of the tour.

Examples: Our main theorems give either $O(N \log N)$ or $O(N)$ time algorithms for all of the following examples, with the exception of (7):

(1) Let the nodes $x_1, \ldots, x_N$ be sequentially ordered points on a line (e.g., they are real numbers indicating points on the x-axis), and let $||x_j - x_i||$ be the Euclidean distance from x_i to x_j. Let f be any concave down function, so $f''(x) \leq 0$ for all x. If the cost function is defined by

$$c(x_i, x_j) \ = \ f(||x_j - x_i||), \qquad (2)$$

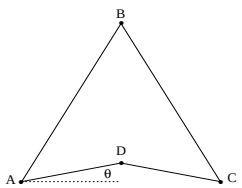
then $x_1, \ldots, x_N$ are a quasi-convex tour. Prior work for examples (1) and (2) gave linear time matching algorithms *only* for the case where $f(x)$ is a linear function [11, 1].

(2) Now let the points $x_1, \ldots, x_N$ lie on a smooth curve C which is homeomorphic to a circle, with the points listed in, say, counterclockwise order. And let $||x_j - x_i||$ equal the shortest arclength along C from x_i to x_j. Again let $f(x)$ be any concave down function. With the cost function given by equation (2), the nodes $x_1, \ldots, x_N$ form a quasi-convex tour.

(3) Suppose $x_1, \ldots, x_N$ lie, in that order, on a circle. Let $c(x_i, x_j)$ equal the Euclidean distance from x_i to x_j. Since Euclidean distance is a concave down function of the circular arclength, this is a special case of Example (2) and the nodes form a quasi-convex tour. In this case, the weak analyticity condition always holds and Main Theorem 2 gives an $O(N)$ time algorithm. The best prior algorithm was $O(N \log N)$ time [17].

(4) More generally, if $x_1, \ldots, x_N$ are the vertices of a convex polygon listed in, say, counterclockwise order, and if the cost function is equal to Euclidean distance, then the nodes form a quasi-convex tour. The prior algorithm for this case was $O(N \log N)$ time [17] and our algorithms are either $O(N)$ or $O(N \log N)$ time depending on whether the weak analyticity condition holds.

(5) Some non-convex polygons also have vertices which form a quasi-convex tour. For example, in a polygon shaped as below, the vertices A, B, C, D will form a quasi-convex tour, provided the angle θ not too large. (This is reason we use 'quasi-convex' instead of 'convex' to describe tours which satisfy equation (1).)



(6) Examples (4) and (5) are also quasi-convex under other distance metrics such as the L_1 and L_∞ metrics.

(7) Marcotte and Suri consider graphs where the nodes are the vertices of a simple polygon and the cost function is equal to the length of the shortest connecting path inside the polygon. The nodes of such a polygon form a quasi-convex tour. The prior algorithm and the algorithm of this paper are $O(N \log^2 N)$ time for this example, since the cost function requires $O(\log N)$ time to compute.

(8) In string matching algorithms, the cost of shifting a character's position is specified as a function of the distance

shifted. The authors have worked in the past on string matching algorithms [23, 24] in which the cost function is a linear function of distance. These prior algorithms have been quite successfully used in commercial applications, especially natural language search; and, we expect that the use of a concave down distance function will significantly improve the matching quality. As we discuss in section 5, the setting of example (1) above is precisely what is needed to allow (near) linear time string matching algorithms with concave-down cost functions. A number of authors, including [5, 6], have studied concave down cost functions for string matching; their string matching algorithms are based on least-edit-distance and, in this regard, are quite different from ours. Least-edit-distance string matching algorithms are widely used because they provide rich and flexible string comparison functions; on the other hand, the best general algorithms for computing least-edit-distance require $O(N^2)$ time (see [18]). Our string matching algorithms are not as flexible, but can be tailored to work well for many applications: they have the advantage of being linear time computable.

Main Theorem 1 (a) *There is an $O(N \log N)$ time algorithm for the minimum-cost matching problem for linear quasi-convex tours.*

(b) *There is an $O(N \log N)$ time algorithm for the minimum-cost matching problem for balanced quasi-convex tours.*

Remark The running times of the algorithms are given in terms of the number N of nodes, even though the input size may in some cases need to be $\Omega(N^2)$ to fully specify the values of the cost function. However, in all the examples above, the input size is $O(N)$ since the cost function is specified by the nodes' positions on a line, on a curve, or in the plane. In any event, our runtime analysis assumes that any value $c(x_i, x_j)$ of the cost function can be computed in constant time. If this is not the case, then the runtimes are to be multiplied by the time needed to compute a value of the cost function (this is the situation in Example (7) above).

We next define an "weak analyticity" condition which will allow yet faster algorithms.

Definition Suppose that x_i and x_j are red (blue) nodes, that $\delta \geq 0$, and that there is a blue (resp, red) node x_k such that

$$c(x_i, x_k) - c(x_j, x_k) \quad < \quad \delta,$$

The *δ-crossover point* of x_i and x_j is defined to be the first such x_k, where 'first' means in tour order starting from x_j and ending at x_i. If no such x_k exists, then the δ-crossover point does not exist.

It is not hard to see that the property of quasi-convexity implies that, if the δ-crossover point x_k exists, then $c(x_i, x_\ell) - c(x_j, x_\ell) \geq \delta$ whenever x_i, x_j, x_ℓ, x_k are in tour order and $c(x_i, x_\ell) - c(x_j, x_\ell) < \delta$ whenever x_i, x_j, x_k, x_ℓ are in tour order. Thus binary search provides an $O(\log N)$-time procedure which, given x_i, x_j and δ, will determine if x_k exists and, if so, which node x_k is. This is the approach taken in the algorithms of Theorem 1, and is the source of the $\log N$ factor in the runtime. However, in some cases, x_k can be found in constant time and we define:

Definition A quasi-convex tour satisfies the *strong analyticity condition* provided there is a constant time algorithm which can determine if the δ-crossover point of x_i and x_j exists and, if so, can determine which node it is.

A quasi-convex tour satisfies the *analyticity condition* provided there is a constant time algorithm which can answer the following question (as a function of similarly colored nodes x_i, x_j, x_k in tour order and of $\epsilon, \delta > 0$, where the δ-crossover of x_i and x_j is known to exist):

> "Do x_j and x_k have an ϵ-crossover point which either equals or precedes in tour order the δ-crossover point of x_i and x_j?"

Even the analyticity condition is too strong to be satisfied in many situations, so we also define a 'weak analyticity condition' as follows.

Definition Let x be a node and y and z be denotations of nodes. We write $y \prec_x z$ to denote that either (1) y and z exist and are distinct and y precedes z in the tour order beginning at x, or (2) y exists and z does not.

A *relative crossover procedure* is a procedure Ω such that, given ϵ, δ, x_i, x_j and x_k as input, and letting y be the δ-crossover of x_i and x_j, and z be the ϵ-crossover of x_j and x_k, then

(1) If $y \prec_{x_j} z$, then Ω outputs "Yes".

(2) If $z \prec_{x_j} y$, then Ω outputs "No".

(3) Otherwise Ω may output either answer.

Note that Ω is not required to determine y and z. The difference between weak analyticity and ordinary analyticity is that when condition (3) holds, Ω may output either answer.

Definition The *weak analyticity condition* holds provided there is a constant-time relative crossover procedure.

Clearly the strong analyticity condition implies the analyticity condition, which in turn implies the weak analyticity condition. In most applications, we do not have the analyticity or strong analyticity conditions, but the weak analyticity condition does hold in many natural situations. In particular, examples (1), (2), (3) and (4) do satisfy the weak analyticity condition, provided that the concave-down function is sufficiently natural. Consider, for instance, example (1) with the concave-down function $f(x) = x$, $f(x) = \sqrt{x}$, or $f(x) = \log x$, etc. For example (1), the input nodes $x_1, \ldots, x_N$ are given with a sequence of real numbers $r_1 \leq r_2 \leq \cdots \leq r_N$ which are the positions of the nodes on the real line. Given nodes x_i, x_j and $\delta > 0$, the first possible position for the δ-crossover of x_i and x_j can be found by solving the equation $f(y - r_i) = \delta + f(y - r_j)$ for y; since we assume that arithmetic operations take constant time, the solution y can be found in constant time. Note that y is only the *theoretical* crossover point; the actual crossover is the first node x_k such that $y \leq r_k$. Unfortunately, even after y is known, it will not be possible to determine x_k in constant time, unless some additional information is given about the distribution of the nodes on the real line. Thus, the analyticity condition and strong analyticity conditions do not hold in general for example (1). The reason the analyticity condition does not hold is that, if the theoretical ϵ-crossover point occurs after the theoretical δ-crossover point, then the analyticity algorithm must output 'No' if there is a node after the theoretical δ-crossover point and before or at the theoretical ϵ-crossover point, and must output 'Yes' otherwise (because in the latter case the two actual crossover points coincide). Unfortunately, there is no general way to decide this in constant time, so the analyticity condition is false. However, the weak analyticity condition does hold, since the function Ω may operate by

computing the theoretical δ-crossover of x_i and x_j and the theoretical ϵ-crossover of x_j and x_k and outputting "Yes" if the former is less than the latter.

For similar reasons, example (3) satisfies the weak analyticity condition: in this case, since the nodes lie on a circle and the cost function is Euclidean distance, the theoretical crossover position is computed (in constant time) as the intersection of a hyperbola and the circle. Likewise, the weak analyticity condition also holds for Example (2) if the concave-down function is sufficiently nice, and it holds for Example (6), where nodes lie on a circle under the L_1 and L_∞ metrics. Example (4), where the nodes form the vertices of a convex polygon, does not seem to satisfy the weak analyticity condition in general; however, some important special cases do. For example, if the vertices of the convex polygon are known to lie on a polygon with a bounded number of sides, on an oval, or an a branch of a hyperbola, then the weak analyticity condition does hold.

The analyticity condition has been implicitly used by Hirschberg-Larmore [10] who defined a *Bridge* function which is similar to our Ω function: they give a special case in which *Bridge* is constant-time computable and thus the analyticity condition holds. Later, Galil-Giancarlo [6] defined a "closest zero property" which is equivalent to our strong analyticity condition.[3] As we illustrated above, the analyticity and strong analyticity conditions rarely hold. Thus it is interesting to note that the algorithms of Hirschberg-Larmore and of Galil-Giancarlo will still work, with only minor modifications, if only the weak analyticity condition holds.

Our second main theorem implies that these examples which satisfy the weak analyticity condition have linear time algorithms for minimum-cost matching:

Main Theorem 2 (a) *There is an $O(N)$ time algorithm for the minimum-cost matching problem for linear quasi-convex tours which satisfy the weak analyticity condition.*

(b) *There is an $O(N)$ time algorithm for the minimum-cost matching problem for balanced quasi-convex tours which satisfy the weak analyticity condition.*

Remark In order to achieve the linear time algorithms, it is necessary that nodes of the graph be input in their tour order. This assumption is necessary, since without it, is possible to give a linear time reduction of sorting to the matching problem for linear tours.

Our main theorems also apply to minimum-cost matchings for some non-bipartite quasi-convex tours. If a non-bipartite graph G has N nodes and has cost function c, then a matching for G is a set of $\lfloor \frac{1}{2}N \rfloor$ edges with all endpoints distinct. Parts (a) of Main Theorems 1 and 2 hold also for non-bipartite graphs which are linear quasi-convex tours. And parts (b) of Main Theorems 1 and 2 hold also for non-bipartite graphs which are quasi-convex tours with an even number of nodes. The non-bipartite cases are discussed in section 4; the algorithms are simple modifications of the algorithms for the bipartite tours.

It is apparent that our algorithms can be parallelized but we have not investigated the precise runtime and processor count that is needed for a parallel implementation. He [9] has given a PRAM implementation of Marcotte and Suri's algorithm which uses N processors and $O(\log^2 N)$ time and it is clear that our algorithm can be computed with the same number of processors with the same time bounds using He's methods.

We wish to thank Dina Kravets, Dave Robinson and Warren Smith for helpful discussions —and Dave Robinson for implementing the algorithms described below.

2. Reductions and Lemmas

2.1 Reduction to tours of alternating colors The first step to giving our minimum-cost matching algorithms is to reduce to the special case of tours in which the colors of the nodes alternate. In other words, we will be able to assume w.l.o.g. that $x_1, x_3, x_5, \ldots$ are red and that $x_2, x_4, x_6, \ldots$ are blue.

Definition Let x_i and x_j be nodes. We write $[x_i, x_j]$ to denote the sequence of nodes obtained by starting with x_i and advancing in tour order to x_j. We write $(x_i, x_j]$, $[x_i, x_j)$ and (x_i, x_j) for this sequence minus the starting node, the ending node, or both.

If x is a node, let $d(x)$ denote the number of red nodes in $[x_1, x)$ minus the number of blue nodes in $[x_1, x)$. The *level* of x, $level(x)$, is equal to $d(x)$ if x is red and is equal to $d(x)+1$ if x is blue. We write $x \sim y$ to mean that $d(x) = d(y)$; obviously, $\sim$ is an equivalence relation. It is easy to see that if y is the first node after x in input order such that $x \sim y$, then x and y are of opposite colors. Also, if $x \sim y$ and x, y are in input order and are of opposite colors, then (x, y) contains equal number of red and blue nodes. For balanced tours, the $\sim$-equivalence relation is invariant under circular rotation of the nodes in the tour.

Given a matching on the nodes of a graph, we write $x_i \leftrightarrow x_j$ to indicate the presence of an edge between x_i and x_j in the matching. We say that x_j *immediately follows* x_i in tour order if $j = i+1$ or if $i = N$ and $j = 1$. Two nodes x_i and x_j are *adjacent* iff one of them immediately follows the other. An edge $x_i \leftrightarrow x_j$ is called a *jumper* if x_i and x_j are not adjacent. Two jumpers are said to *cross* if they are of the form $x_i \leftrightarrow x_k$ and $x_j \leftrightarrow x_\ell$ with x_i, x_j, x_k, x_ℓ in tour order.

Lemma 3 *Let G be either a linear quasi-convex tour or a balanced quasi-convex tour. Then G has a minimum-cost matching in which every edge $x_i \leftrightarrow x_j$ satisfies $x_i \sim x_j$. In other words, some minimum-cost matching for G can be obtained as a union of minimum-cost matchings on the $\sim$-equivalence classes of G.*

To prove Lemma 3 we use:

Lemma 4 *G has a minimum-cost matching in which no jumpers cross.*

Proof (Sketch) If a minimum-cost matching does have a pair of jumpers which cross, the quasi-convexity property allows them to be 'uncrossed' without increasing the total cost. Repeatedly uncrossing jumpers will eventually yield a minimum-cost matching with no crossing jumpers. (See Lemma 1 of [1] for a detailed proof of this.)

Lemma 3 is proved by noting that a minimum-cost matching with no crossing jumpers must respect the $\sim$-equivalence classes. This is because, if a jumper $x_i \leftrightarrow x_j$ is in a crossing-free matching with $i < j$, then the nodes in the interval (x_i, x_j) must be matched which each other and thus (x_i, x_j) must have equal numbers of red and blue nodes. In the unbalanced, linear case, this also depends on the fact that, w.l.o.g., there is no jumper which crosses an unmatched node (this is an immediate consequence of the linearity).

[3] The definition of the "closest zero property" is misstated in [6]: it should be defined as saying that it is possible to find the first r such that $w(l,r) - w(k,r) - a \leq 0$ (note their w corresponds to our cost function c, and a is a real). However, their algorithm explicitly uses the correct definition of "closest zero property" (see their Fact 2).

By Lemma 3, in order to find a minimum-cost matching, it suffices to extract the $\sim$-equivalence classes, and find minimum-cost matchings for each equivalence class independently. It is an easy matter to extract the $\sim$-equivalent classes in linear time by using straightforward counting. Each equivalence class consists of an alternating color subtour: in the balanced case, there are an even number of nodes in each equivalence class, and in the linear case, there may be an even or odd number of nodes. Thus, to give (near) linear time algorithms for finding matchings, it will suffice to restrict our attention to tours in which the nodes are of alternating colors.

In view of Lemma 4, we may restrict our attention to matchings which contain no crossing jumpers. Such a matching will be called *crossing-free*.

Finally, we can assume w.l.o.g. that the tour is balanced. To see why we can assume this, suppose that $x_1, \ldots, x_N$ is an unbalanced, linear tour of alternating colors. This means that x_1 and x_N are the same color, say red. We can add a new node x_{N+1} to the end of the tour, label it blue, and let $c(x_i, x_{N+1}) = 0$ for all red x_i. These $N+1$ nodes no longer form a linear tour; however, they do form a balanced quasi-convex tour. Solving the matching problem for the $N+1$ nodes immediately gives a solution to the matching problem on the original N nodes.

2.2 Some important lemmas Since we are now working only with balanced quasi-convex tours of alternating colors, we shall often change the names of the nodes to $R_1, B_1, \ldots, R_M, B_M$; so R_i and B_j refer to the i-th red node and the j-th blue node in the tour, respectively. (So x_{2i-1} is the same as R_i and x_{2i} is the same as B_i.) Note that this means $N = 2M$. To simplify notation, we define

$$c_i \;=\; c(R_i, B_i) \qquad \text{and} \qquad c'_i \;=\; c(B_i, R_{i+1}).$$

A *greedy* matching is a matching which contains no jumpers, i.e., every node is matched to an adjacent node. There are two greedy matchings, namely, the one containing all edges $R_i \leftrightarrow B_i$ and the one containing all edges $B_{i-1} \leftrightarrow R_i$ and the edge $B_N \leftrightarrow R_1$. For x_i and x_j nodes of opposite color, a matching σ is said to be *greedy on* (x_i, x_j) provided it contains as a submatching the unique matching of adjacent nodes contained in the interval (x_i, x_j). We similarly define the notion of σ being greedy on a balanced interval I, for I one of the intervals $[x_i, x_j)$, $[x_i, x_j]$ or $(x_i, x_j]$, but with the additional provisos that $x_i \leftrightarrow x_{i+1}$ is in σ in the first two cases and that $x_{j-1} \leftrightarrow x_j$ is in σ in the second two cases.[4]

The notation $[R_i, B_j]$ has already been defined. In addition, the notation $[i, j]$ denotes the interval of integers $i, i+1, \ldots, j$ if $i < j$, or the (circular) interval $i, i+1, \ldots, M, 1, 2, \ldots, j$ if $j < i \leq M$. We also use the notations $(i, j]$, $[i, j)$ and (i, j) for the intervals with one or both of the endpoints omitted.

Definition Let R_i and B_j be nodes; we write $R_i \rightarrow B_j$ to denote a *directed* edge going from R_i forward (in tour order) to B_j. That is, we think of $R_i \rightarrow B_j$ jumping over the nodes $R_i, B_i, R_{i+1}, \ldots, R_j, B_j$. We say that $R_i \rightarrow B_j$ is a *candidate* (meaning, a candidate for a jumper), if

$$c(R_i, B_j) + \sum_{\ell \in [i,j)} c'_\ell \;<\; \sum_{\ell \in [i,j]} c_\ell.$$

The intuitive meaning $R_i \rightarrow B_j$ being a candidate is that it would be of lower cost to use the jumper $R_i \leftrightarrow B_j$ plus the greedy matching of adjacent nodes in (R_i, B_j) in place of just the greedy matching of adjacent nodes in $[R_i, B_j]$.

[4]Note that, of the two greedy matchings for G, one is greedy on $[x_1, x_N]$ and the other is greedy on $[x_2, x_1]$.

A similar definition is used to define what it means for an edge $B_i \rightarrow R_j$ to be a candidate; namely, $B_i \rightarrow R_j$ is a candidate iff

$$c(B_i, R_j) + \sum_{\ell \in (i,j)} c_i \;<\; \sum_{\ell \in [i,j)} c'_i.$$

Candidates always have endpoints of opposite colors and are directed. It is possible to have both $R_i \rightarrow B_j$ and $B_j \rightarrow R_i$ be (distinct) candidates, or to have one or neither of them candidates.

It is an easy observation that if there are no candidates, then the greedy assignment(s) are minimum-cost matchings. To prove this, suppose σ is a minimum-cost matching which contains a jumper: by Lemma 4, σ may be picked to contain no crossing jumpers. Since there are no crossing jumpers, σ must contain a jumper $x_i \leftrightarrow x_j$ such that σ is greedy on (x_i, x_j) (namely, pick the jumper so as to minimize the tour-order distance from x_i to x_j). Let σ' be the matching which is the same as σ, except greedy on $[x_i, x_j]$. Clearly σ' has one fewer jumper than σ, and since $x_i \rightarrow x_j$ is not a candidate, σ' has cost no greater than σ. Iterating this construction shows that at least one of the jumper-less greedy matchings must be minimum-cost. To show they are both minimum-cost, let σ_0 and σ_1 be the greedy matchings which contain the edges $x_1 \leftrightarrow x_2$ and $x_1 \leftrightarrow x_N$, respectively. Then σ_0 can not have cost lower than (respectively, higher than) the cost of σ_1 since otherwise, $x_2 \rightarrow x_1$ ($x_1 \rightarrow x_N$, respectively) would be a candidate.

Definition A candidate $x_i \rightarrow x_j$ is a *minimal* candidate iff there is no other candidate $x_k \rightarrow x_\ell$ in its interior; that is to say, there is no candidate $x_k \rightarrow x_\ell$ with $[x_k, x_\ell]$ a proper subset of $[x_i, x_j]$.

Lemma 5 *Consider a balanced quasi-convex tour of alternating colors.*

(a) *Suppose $R_a \rightarrow B_b$ is a minimal candidate. Then every minimum-cost, crossing-free matching is greedy on the interval (R_a, B_b). That is to say, every minimum-cost, crossing-free matching contains the edges $B_{\ell-1} \leftrightarrow R_\ell$ for all $\ell \in (a, b]$.*

(b) *Suppose $B_a \rightarrow R_b$ is a minimal candidate. Then every minimum-cost, crossing-free matching is greedy on the interval (B_a, R_b). That is to say, every minimum-cost, crossing-free matching contains the edges $R_\ell \leftrightarrow B_\ell$ for all $\ell \in (a, b)$.*

Note that Lemma 5 says only that the edges connecting adjacent nodes in the *interior* of the minimal candidate are in every minimum-cost matching; it does not say that the minimal candidate itself is a jumper in any minimum-cost matching. The proof of Lemma 5 is fairly involved and we postpone it until section 2.3. Lemma 5 also holds for linear tours with alternating colors for candidates $x \rightarrow y$ with x, y in input order.

Lemma 5 suggests an algorithm for finding a minimum-cost matching. Namely, if there is a minimal candidate, greedily assign edges in its interior according to Lemma 5. This induces a matching problem on the remaining unassigned nodes, and it is clear that any minimum-cost matching on this smaller problem will lead to a minimum-cost matching for the original problem. Iterating this, one can continue removing nodes in the interiors of minimal candidates and reducing the problem size. Eventually a matching problem with no candidates will

be reached; in this case, it suffices to greedily match the remaining nodes.

Unfortunately, this algorithm suggested by Lemma 5 is not linear time (yet); thus we need to refine Lemma 5 somewhat:

Definition We define:

$$\mathsf{Bnft}[R_a, B_b] = \Big(\sum_{i\in[a,b]} c_i - \sum_{i\in[a,b)} c'_i\Big) - c(R_a, B_b),$$
$$\mathsf{Bnft}[B_a, R_b] = \Big(\sum_{i\in[a,b)} c'_i - \sum_{i\in(a,b)} c_i\Big) - c(B_a, R_b),$$

and, for x and y the same color, $\mathsf{Bnft}[x, y] = -\infty$.

It is immediate that $\mathsf{Bnft}[x, y] > 0$ iff $x \to y$ is a candidate; in fact, $\mathsf{Bnft}[x, y]$ measures the benefit (i.e., the reduction in cost), of using $x \leftrightarrow y$ as a minimal jumper instead of the greedy matching on $[x, y]$.

The next lemma forms the basis for the correctness of the algorithm given in section 3 for the serial transitive closure problem. The general idea is that the algorithm will scan the nodes in tour order until at least one candidate is found and then, according to Lemma 6, the algorithm will choose an interval (x_ℓ, x_k) to greedily match. Once the interval (ℓ, k) has been greedily matched, the algorithm need only solve the induced matching problem on the remaining nodes.

Lemma 6 *Let G be a balanced quasi-convex tour matching problem. Let $1 < k \le N$ and suppose $\mathsf{Bnft}[x_i, x_j] \le 0$ for all $1 \le i < j < k$. Suppose $m \stackrel{\text{def}}{=} \max\{\mathsf{Bnft}[x_i, x_k] : i < k\} > 0$ and let $\ell \stackrel{\text{def}}{=} \max\{i < k : \mathsf{Bnft}[x_i, x_k] = m\}$. Then every minimum-cost, crossing-free matching is greedy on (x_ℓ, x_k).*

Proof The proof is, in essence, an iteration of Lemma 5. We argue by induction on k. Let G, k, m and ℓ satisfy the hypothesis of the lemma. Let $s = \max\{i < k : \mathsf{Bnft}[x_i, x_k] > 0\}$, so $x_s \to x_k$ is a minimal candidate. By Lemma 5, any minimum-cost, crossing-free solution for G is greedy on the interval (x_s, x_k). Hence, it will suffice to let G' be the matching problem obtained from G by discarding the nodes $x_{s+1}, \ldots, x_{k-1}$ and prove that any minimum-cost, crossing-free solution for G' is greedy on $(x_\ell, x_s]$. If $\ell = s$, there is nothing to prove, so we assume $\ell < s$. Note that x_k is now the $(s+1)$-st node in the G' tour order. We use Bnft' to denote the Bnft function for G'.

Claim:
(a) If $1 \le i < j \le s$, $\mathsf{Bnft}'[x_i, x_j] = \mathsf{Bnft}[x_i, x_j]$.
(b) If $1 \le i \le s$, $\mathsf{Bnft}'[x_i, x_k] = \mathsf{Bnft}[x_i, x_k] - \mathsf{Bnft}[x_s, x_k]$.

Claim (a) is immediate from the definition of Bnft. The intuitive meaning of (b) is that the benefit of using the jumper $x_i \leftrightarrow x_k$ is reduced by the benefit already obtained from the jumper $x_s \leftrightarrow x_k$. We formally prove (b) for the case that x_i and x_s are red and x_k is blue, the opposite colored case has a similar proof. Assume $x_i = R_a$, $x_s = R_b$ and $x_k = B_c$. Then

$$\mathsf{Bnft}'[R_a, B_c] = \sum_{\ell\in[a,b)} c_i + c(R_b, B_c) - \sum_{\ell\in[a,b)} c'_i - c(R_a, B_c)$$
$$\mathsf{Bnft}[R_a, B_c] = \sum_{\ell\in[a,c]} c_i - \sum_{\ell\in[a,c)} c'_i - c(R_a, B_c)$$
$$\mathsf{Bnft}[R_b, B_c] = \sum_{\ell\in[b,c]} c_i - \sum_{\ell\in[b,c)} c'_i - c(R_b, B_c)$$

From these three equations, Claim (b) follows immediately.

Now let $m' = \max\{\mathsf{Bnft}'[x_i, x_k] : i < s\}$. By Claim (b), $m' = m - \mathsf{Bnft}[x_s, x_k]$; since $\ell < s$, $m' > 0$. Likewise, $\ell = \max\{i < s : \mathsf{Bnft}'[x_i, x_k] = m'\}$. Thus, by the induction hypothesis, any minimum-cost solution for G' is greedy on $(x_\ell, x_s]$ and Lemma 6 is proved. □

Definition The Δ function is defined by:

$$\Delta[R_a, R_b] = \sum_{\ell\in[a,b)} c_\ell - \sum_{\ell\in[a,b)} c'_\ell$$
$$\Delta[B_a, B_b] = \sum_{\ell\in[a,b)} c'_\ell - \sum_{\ell\in(a,b]} c_\ell$$

Lemma 7

(a) $\mathsf{Bnft}[R_a, B_c] > \mathsf{Bnft}[R_b, B_c]$ *if and only if* $c(R_a, B_c) - c(R_b, B_c) < \Delta[R_a, R_b]$.

(b) $\mathsf{Bnft}[B_a, R_c] > \mathsf{Bnft}[B_b, R_c]$ *if and only if* $c(B_a, R_c) - c(B_b, R_c) < \Delta[B_a, B_b]$.

Lemma 7 follows immediately from the definitions.

Lemma 8 *Let u, v, x, y be in tour order with nodes u and v of one color and x and y of the other color. Then*

$$\mathsf{Bnft}[u, x] > \mathsf{Bnft}[v, x] \quad \Rightarrow \quad \mathsf{Bnft}[u, y] > \mathsf{Bnft}[v, y]$$

Proof By Lemma 7, $\mathsf{Bnft}[u, x] > \mathsf{Bnft}[v, x]$ is equivalent to $c(u, x) - c(v, x) < \Delta[u, v]$, and $\mathsf{Bnft}[u, y] > \mathsf{Bnft}[v, y]$ is equivalent to $c(u, y) - c(v, y) < \Delta[u, v]$. Now, by quasi-convexity, $c(u, x) - c(v, x) \ge c(u, y) - c(v, y)$, which suffices to prove the lemma. □

Let R_a and R_b be distinct red nodes. The previous two lemmas show that if there is any node R_c (with R_a, R_b and R_c in tour order) such that $\mathsf{Bnft}[R_a, R_c]$ is greater than $\mathsf{Bnft}[R_b, R_c]$, then the first such R_c is the $\Delta[R_a, R_b]$-crossover point of R_a and R_b. We shall denote this first R_c, if it exists, by $\chi[R_a, R_b]$; if it does not exist, then $\chi[R_a, R_b]$ is said to be undefined. Similarly, $\chi[B_a, B_b]$ is defined to the be the $\Delta[B_a, B_b]$-crossover point of B_a and B_b, and, if defined, is the first R_c where $\mathsf{Bnft}[B_a, R_c]$ is greater than $\mathsf{Bnft}[B_b, R_c]$.

We now assume that we have a procedure $\Omega(x, y, z)$, which given nodes x, y, z in tour order returns "True" if $\chi[x, y] \prec_y \chi[y, z]$ and returns "False" if $\chi[y, z] \prec_y \chi[x, y]$. (If neither condition holds, then $\Omega(x, y, z)$ may return an arbitrary truth value.) If the weak analyticity condition holds, then Ω is constant time computable. Without this assumption, Ω is $O(\log N)$ time computable since Lemma 8 allows $\chi[-, -]$ to be computable by binary search.

The general idea of the algorithm given in section 3 below is that it will scan the nodes in tour order searching for candidates. Whenever a node is reached that is the head of candidate, the algorithm will take the candidate specified in Lemma 6 (the one that was denoted $x_\ell \to x_k$) and greedily match the nodes in its interior. The greedily matched nodes are then dropped from consideration and the algorithm resumes its search for a candidate. Suppose the u and v are two nodes already scanned in this process that are being remembered as potential endpoints of candidates. Lemma 7 tells us that if a node x is found where $\mathsf{Bnft}[u, x] > \mathsf{Bnft}[v, x]$, then at all succeeding nodes y, $\mathsf{Bnft}[u, y] > \mathsf{Bnft}[v, y]$. By the criterion of Lemma 6, this means that after the node x is found, there is no further reason to consider candidates that begin at node v, since any candidate $v \to y$ would be subsumed by the better candidate $u \to y$.

To conclude this section we describe the algorithm in very general terms; in section 3 we give the precise specification of the algorithm. The algorithm scans nodes (starting with node x_1, say) and maintains three lists. The first list, $\mathcal{M}$, contains the nodes in tour order which have been examined so far. The second list, $\mathcal{L}^{-1}$, contains all the red nodes that need to be considered as potential endpoints of candidates (so $\mathcal{L}^{-1}$ is guaranteed to contain all the nodes satisfying the criterion of Lemma 6). The third list, $\mathcal{L}^1$, similarly contains all the blue nodes that need to be considered as potential endpoints of candidates. At any point during the scan, the lists will be of the form:

$$\mathcal{M} = x_1, \ldots, x_{r-1}$$

$$\mathcal{L}^{-1} = R_{a_1}, \ldots, R_{a_p}$$
$$\mathcal{L}^{1} = B_{b_1}, \ldots, B_{b_q}$$

with $\mathcal{L}^{-1}$ and $\mathcal{L}^{1}$ subsequences of $\mathcal{M}$. The following four conditions will be maintained during execution:

(1) $x_1, \ldots, x_{r-1}$ are the nodes scanned but not matched, are in tour order, and there are no candidates $x_i \to x_j$ with $1 \le i < j < r$.

(2) x_{r-1} precedes $\chi[R_{a_{p-1}}, R_{a_p}]$ in tour order.

(3) For all $1 \le i \le p-2$, $\Omega(R_{a_i}, R_{a_{i+1}}, R_{a_{i+2}})$ is false.

(4) For all $1 \le i \le q-2$, $\Omega(B_{b_i}, B_{b_{i+1}}, B_{b_{i+2}})$ is false.

(5) At any possible future node x_k following x_{r-1} such that x_k is the first point where a candidate is discovered; if the x_ℓ which satisfies Lemma 6 is among $x_1, \ldots, x_{r-1}$ then it is already on the list $\mathcal{L}^{-1}$ or $\mathcal{L}^{1}$ (depending on which color it is).

When scanning the next node x_r, the algorithm must do the following (we assume x_r is blue, similar actions are taken for red nodes):

(β) While $p \ge 2$ and $\mathsf{Bnft}[R_{a_{p-1}}, x_r] > \mathsf{Bnft}[R_{a_p}, x_r]$, pop R_{a_p} from $\mathcal{L}^{-1}$ and decrement p.

(γ) If $\mathsf{Bnft}[R_{a_p}, x_k] > 0$, greedily match nodes in the interval (R_{a_p}, x_k). The matched nodes are discarded from the lists $\mathcal{M}$, $\mathcal{L}^{-1}$ and $\mathcal{L}^{1}$ (the remaining nodes are to be implicitly renumbered at this point).

(δ) While $q \ge 2$ and $\Omega(B_{a_{q-1}}, B_{a_q}, x_r)$, pop B_{a_q} from $\mathcal{L}^{1}$ and decrement q.
Then push x_r onto the end of $\mathcal{L}^{1}$ (and increment q).

Step (β) is justified by recalling that if x_r is past $\chi[R_{a_{p-1}}, R_{a_p}]$, then R_{a_p} may be removed from consideration as an endpoint of a candidate (by Lemma 6).

Step (δ) is justified as follows: suppose $R_i = \chi[B_{a_{q-1}}, B_{a_q}]$ equals or precedes $R_j = \chi[B_{a_q}, x_r]$ (using tour order, beginning at B_{a_q}). Then at any future candidate endpoint x_k, either x_k follows or equals R_i, in which case $\mathsf{Bnft}[B_{a_{q-1}}, x_k]$ is greater than $\mathsf{Bnft}[B_{a_q}, x_k]$, or x_k precedes R_j, in which case, $\mathsf{Bnft}[x_r, x_k]$ is greater than $\mathsf{Bnft}[B_{a_q}, x_k]$. Thus B_{a_q} will never be the starting endpoint of a candidate satisfying the criteria of Lemma 6, and we may drop it from consideration.

To justify Step (γ) we must show that the candidate $R_{a_p} \to x_r$ satisfies the criteria from Lemma 6: in view of the correctness of the rest of the algorithm, for this it will suffice to show that $\mathsf{Bnft}[R_{a_i}, x_r] \le \mathsf{Bnft}[R_{a_p}, x_r]$ for all $1 \le i < p$. For this, note that Step (β) and condition (3) above ensure that x_r precedes $\chi[R_{a_i}, R_{a_{i+1}}]$ for all $1 \le i < p$. This, in turn, implies $\mathsf{Bnft}[R_{a_i}, x_r] \le \mathsf{Bnft}[R_{a_{i+1}}, x_r]$ for all i, which proves the desired inequality.

After the algorithm has scanned all the nodes once, it will have found and processed all candidates $x_i \to x_j$ where $i < j$. However, since the tour is circular, it is necessary to process candidates $x_i \to x_j$ with $i > j$. At the end of the first scan, the list $\mathcal{M}$ consists of all nodes, $x_1, \ldots, x_n$ which have not been matched yet and $\mathcal{L}^{-1}$ and $\mathcal{L}^{1}$ contain nodes $R_{a_1}, \ldots, R_{a_p}$ and $B_{b_1}, \ldots, B_{b_q}$, as usual. During the second scan, the algorithm is searching for any candidates of the form $R_{a_i} \to B_j$ with $j < a_i$ or of the form $B_{a_i} \to R_j$ with $j \le a_i$ (and only for such candidates). To process a node during the second scan, the algorithm pops x_1 off the left end of $\mathcal{M}$, implicitly renames x_1 to x_n and the rest of the nodes x_i to x_{i-1}, sets $r = n$, does Step (α): (still assuming x_r is blue)

(α) If x_r equals B_{b_1}, then pop B_{b_1} from the list $\mathcal{L}^{1}$ and implicitly renumber $\mathcal{L}^{1}$, decrementing q,

and then does steps (β)-(δ), except that in Step (δ), the node x_r is not added to the end of $\mathcal{L}^{1}$. The reason for Step (α) is that once a node B_{b_i} is encountered on the second scan, B_{b_i} is no longer a possible starting endpoint for a candidate. The reason for not adding x_r to the end of $\mathcal{L}^{1}$ in Step (δ) is that can not be the starting endpoint of a candidate, because any such candidate would have already been found earlier.

The second scan will stop as soon as both $\mathcal{L}$ lists become empty. At this point no candidates remain and a greedy matching may be used for the remaining nodes in the $\mathcal{M}$ list.

The actual description of the algorithm with an efficient implementation is given in section 3, and it is there proved that the algorithm is linear time with the weak analyticity condition and $O(N \log N)$ time otherwise. Although we described steps (α)-(δ) only for blue x_r above, the algorithm in section 3 uses a toggle ψ to handle both colors with the same code. Finally, one more important feature of the algorithm is the way in which it computes the values of the Bnft function and of the $\Delta[x, y]$ function: it uses intermediate values $I[x]$ which are defined as follows.

Definition The $I[x]$ function is defined by

$$I[R_a] = \Delta[R_1, R_a]$$
$$I[B_a] = I[R_a] + c(R_a, B_a).$$

Note that $I[R_{a+1}] = I[B_a] - c(B_a, R_{a+1})$.

It is immediate from the definitions that, if x, y are tour order (starting from x_1), then

$$\begin{array}{rl} \Delta[x,y]{=}I[y] - I[x] & \text{for } x \text{ and } y \text{ red.} \\ \Delta[x,y]{=}I[x] - I[y] & \text{for } x \text{ and } y \text{ blue.} \\ \mathsf{Bnft}[x,y]{=}I[y] - I[x] - c(x,y) & \text{for } x \text{ red, } y \text{ blue.} \\ \mathsf{Bnft}[x,y]{=}I[x] - I[y] - c(x,y) & \text{for } x \text{ blue, } y \text{ red.} \end{array}$$

These equalities permit the values of Δ and Bnft to be computed in constant time from the values of $I[-]$. Also, it is important to note that only the relative $I[-]$ values are needed; in other words, it is OK if the $I[-]$ values are shifted by a constant additive constant, since we always use the difference between two $I[-]$ values.

The $I[-]$ function is not only easy to compute, but also provides an intuitive graphical means of understanding the above lemmas and algorithm description. For example, in Figure 1, $R_1 \to B_3$ is a (minimal) candidate whereas $R_1 \to B_1$ and $R_1 \to B_2$ are not candidates. In Figure 2(a), the node B_4 is the relative crossover, $\chi[R_1, R_2]$, of R_1 and R_2; on the other hand, in Figure 2(b), the relative crossover does not exist. Figure 3(a) shows an example where $\Omega(R_1, R_2, R_3)$ is true and Figure 3(b) shows an example where $\Omega(R_1, R_2, R_3)$ is false.

2.3 Proof of Lemma 5 We now prove Lemma 5 — by symmetry, it will suffice to prove part (a). Since the lemma is trivial in case $a = b$, we assume $a \ne b$. Let σ be a crossing-free matching: we must prove that σ is greedy on (R_a, B_b). By the crossing-freeness of σ and by the fact that $R_a \to B_b$ is a *minimal* candidate,

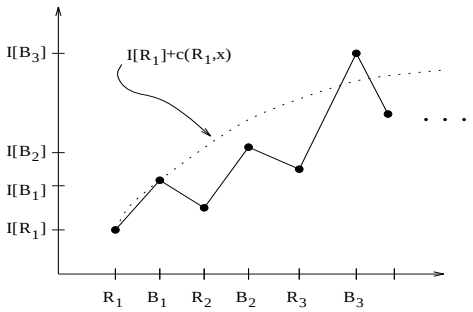


Figure 1: $R_1 \rightarrow B_3$ is a candidate as $I[R_1] + c(R_1, B_3) < I[B_3]$, which is equivalent to $\mathsf{Bnft}[R_1, B_4] > 0$.

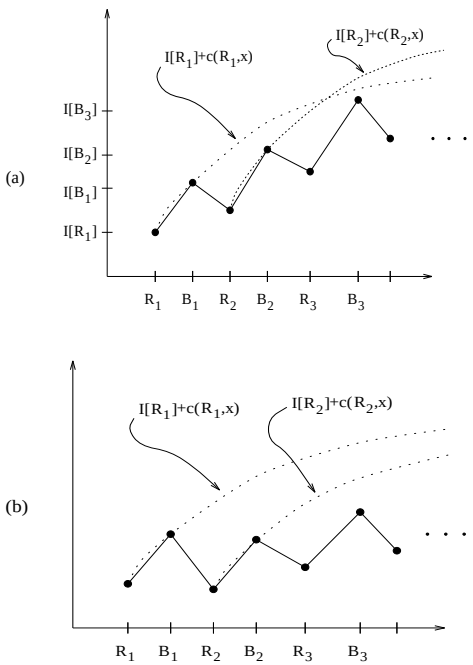


Figure 2: Illustrations of the relative crossover, $\chi[R_1.R_2]$, of R_1 and R_2. In (a), B_3 is $\chi[R_1, R_2]$, since it is the first node x to satisfy $I[R_2] + c(R_2, x) > I[R_1] + c(R_1, x)$. In (b), the relative crossover does not exist.

σ does not contain any jumper with both endpoints in $[R_a, R_b]$, except possibly $R_a \leftrightarrow B_b$ itself. If $R_a \leftrightarrow B_b$ is in σ, then the same reasoning shows that σ is greedy on (R_a, B_b); so we suppose that $R_a \leftrightarrow B_b$ is not in σ. Since we are dealing (with no loss of generality) with balanced tours, we may assume that $b = N$, by renumbering nodes if necessary.

Claim (1): $R_a \leftrightarrow B_a$ is not in σ.
Suppose, for a contradiction, that $R_a \leftrightarrow B_a$ is in σ. Let v be the least value such that $R_v \leftrightarrow B_q$ is in σ for some $q < a < v$. Note that such a v, $a < v \leq N$, must exist since there are no jumpers in $[R_a, B_N]$ and since σ is not greedy on $[R_a, B_N]$ (it can not be greedy on $[R_a, B_N]$, since $R_a \rightarrow B_N$ is a candidate). By choice of v, σ is greedy on $[R_a, R_v)$. Since $R_a \rightarrow B_N$ is a minimal candidate, $\mathsf{Bnft}[R_a, B_N] > \mathsf{Bnft}[R_v, B_N]$, so Lemma 7 implies

$$\sum_{i\in[a,v)} c_i - \sum_{i\in[a,v)} c'_i > c(R_a, B_N) - c(R_v, B_N).$$

Since R_a, R_v, B_N and B_q are in tour order, quasi-convexity

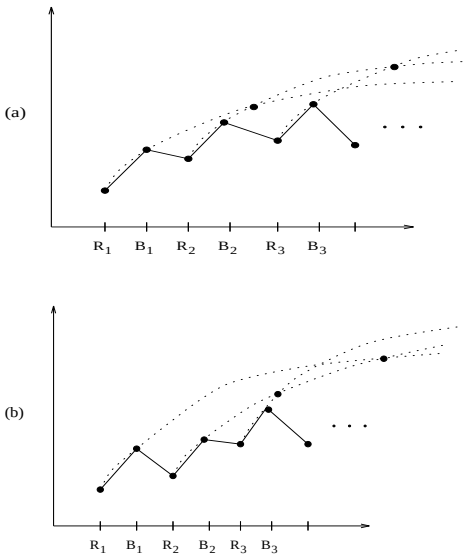


Figure 3: $\Omega(R_1, R_2, R_3)$ is true in (a) and false in (b).

implies

$$c(R_a, B_N) - c(R_v, B_N) \geq c(R_a, B_q) - c(R_v, B_q).$$

Combining these inequalities yields

$$\sum_{i\in[a,v)} c_i + c(R_v, B_q) > \sum_{i\in[a,v)} c'_i + c(R_a, B_q). \qquad (3)$$

Let σ' be the matching obtained from σ by replacing the jumper $B_q \leftrightarrow R_v$ and the greedy matching on $[R_a, R_v)$ with the edge $B_q \leftrightarrow R_a$ and the greedy matching on $(R_a, R_v]$. By (3), σ' has cost strictly less than the cost of σ, which is a contradiction.

Claim (2): $R_N \leftrightarrow B_N$ is not in σ.
Claim (2) is proved by an argument similar to Claim (1). Alternatively, reverse the colors and the tour order and Claim (2) is a version of Claim (1).

Claim (3): The matching σ is greedy on (R_a, B_b).
Suppose, for a contradiction that σ is not greedy on (R_a, B_b). In view of Claims (1) and (2) and since σ has no jumpers in $[R_a, B_N]$, this means that there exist u and v such that u is the least value such that σ contains $B_u \leftrightarrow R_r$ with $r < a \leq u$ and v is the least value such that σ contains $R_v \leftrightarrow B_q$ with $q < a \leq v$. Namely, let u be the least value $\geq a$ such that $B_u \leftrightarrow R_{u+1}$ is not in σ and v be the least value $> u$ such that $R_v \leftrightarrow B_v$ is not in σ. For these choices of u and v, it must be that $q < r < a \leq u \leq v \leq N$ and that σ is greedy on $[B_a, R_u]$ and on $[R_{u+1}, B_{v-1}]$.

Since $R_a \rightarrow B_N$ is a candidate,

$$\sum_{i\in[a,N]} c_i > c(R_a, B_N) + \sum_{i\in[a,N)} c'_i.$$

And since it is minimal, neither $R_a \rightarrow B_u$ nor $R_v \rightarrow B_N$ are candidates; i.e.,

$$\sum_{i\in[a,u]} c_i \leq c(R_a, B_u) + \sum_{i\in[a,u)} c'_i$$

$$\sum_{i\in[v,N]} c_i \leq c(R_v, B_N) + \sum_{i\in[v,N)} c'_i.$$

Combining these three inequalities gives

$$\sum_{i\in(u,v)} c_i + c(R_a, B_u) + c(R_v, B_N) > \sum_{i\in[u,v)} c'_i + c(R_a, B_N). \qquad (4)$$

Since R_a, R_v, B_N, B_q and R_r, R_a, B_u, B_q are in tour order, quasi-convexity implies the two inequalities

$$c(R_a, B_q) + c(R_v, B_N) \leq c(R_a, B_N) + c(R_v, B_q)$$

$$c(R_r, B_q) + c(R_a, B_u) \leq c(R_r, B_u) + c(R_a, B_q)$$

which combine to yield:

$$c(R_a, B_N) - c(R_a, B_u) - c(R_v, B_N) \geq \quad (5)$$
$$c(R_r, B_q) - c(R_r, B_u) - c(R_v, B_q).$$

Using (4) and (5) gives the inequality

$$\sum_{i\in(u,v)} c_i + c(R_r, B_u) + c(R_v, B_q) > \sum_{i\in[u,v)} c'_i + c(R_r, B_q).$$

Let σ' be the matching obtained from σ by replacing the jumpers $R_r \leftrightarrow B_u$ and $R_v \leftrightarrow B_q$ and the greedy matching on (B_u, R_v) with the edge $R_r \leftrightarrow B_q$ and the greedy matching on $[B_u, R_v]$. The final inequality says that σ' has cost strictly less than the cost of σ, which is a contradiction. □

3. The Algorithm

In this section, we give the actual algorithm for the Main Theorems. The correctness of the algorithm follows from the development in section 2.2.

3.1 Preliminaries As mentioned above, the algorithm maintains three lists of nodes called deques (for "double ended queues", since we will have to access both ends of the lists). The three deques are the "main" deque $\mathcal{M}$, and two "left" deques $\mathcal{L}^1$ and $\mathcal{L}^{-1}$. The latter two are called "left deques" since they contain possible left endpoints for candidates. The deques will be updated by *push-right* operations which add a new node to the right end, by *pop-right* operations which pop the rightmost node off the deque, and by *pop-left* operations. However, *push-left* operations are never required. Deque operations can be efficiently implemented by using contiguous memory locations to store the deque elements and maintaining pointers to the left and right endpoints; each deque operation can then be performed in constant time. For our algorithm, it will suffice to reserve enough space for $2N$ deque elements (with no possibility that a deque will grow leftward since push-left's are not used).

Subscripts R, L, and $R-1$ are used to select the rightmost item, leftmost item, and the item preceding the rightmost, respectively. So $\mathcal{L}^{-1}_L$ refers to the leftmost element of $\mathcal{L}^{-1}$, $\mathcal{M}_{R-1}$ refers to the item just before the rightmost member of $\mathcal{M}$, etc. Each deque element is actually a pair, for example, $\mathcal{M}_R = (X, I)$; the first entry X of the pair is a node and the second entry I is a numerical value, namely $I = I[X]$ as defined in section 2.2. To simplify notation, we shall use the same notation for a deque element as for the node which is its first component. Thus, $\mathcal{M}_L$ also denotes the node which is its first component. We write $I[\mathcal{M}_L]$ to denote its second (numerical) component. Similar conventions apply to the $\mathcal{L}^{\pm 1}$ deques. To simplify our presentation of the algorithm we deal with boundary effects by augmenting the definition of primitive operations as necessary. For example, accessing a non-existent deque element will return an *undefined* indicator $\emptyset$ and, in general, functions of undefined operands are false or zero (in particular, the cost function $c(-,-)$ and the $I[-]$ functions return zero if they have $\emptyset$ as an argument).

Function Input() returns the next vertex from an imagined input tape which moves in the forward direction only; and is assumed to hold a balanced, alternating color tour. When the tape's end is reached, '*undefined*' is returned. Procedure Output() is used to write an individual matching to an imagined output tape. They are written as discovered; but can easily be output in tour order (with only an extra $O(N)$-time computation).

To use the same code for red nodes and blue nodes, a variable ψ tracks vertex color by toggling between -1 and 1. Our convention is that $\psi = 1$ corresponds to blue, and $\psi = -1$ to red.

3.2 Narrative Description of the Algorithm. Initialization consists of setting the three deques to be empty and setting the color toggle $\psi := -1$.

The algorithm first reads nodes from the input and pushes them onto the right end of the $\mathcal{M}$-deque, and then twice scans the nodes in tour order. During the two scans, nodes are popped from the left end of $\mathcal{M}$ and then pushed onto its right end.[5] In addition, while processing a node, some nodes may be popped off the right end of $\mathcal{M}$ to be matched. It will always be the case that $\mathcal{M}$ contains a sequence of contiguous nodes in tour order and that the node currently being scanned immediately follows the (formerly) rightmost element of $\mathcal{M}$.

The variable ψ will be maintained as a color toggle, so that ψ is equal to -1 if the node currently being processed is red and to 1 if the current node is blue. The algorithm used for pushing an element onto the right end of $\mathcal{M}$ is:

Algorithm 1 *This procedure pushes a vertex X onto the right of the $\mathcal{M}$ deque, and computes the corresponding $I[X]$ value which is pushed along with X.*

```
procedure Push_Main(X)
    I := I[M_R] + ψ · c(M_R, X)
    push-right (X,I) onto M
    return()
```

Algorithm 1 merely computes the $I[-]$ value for a node X and pushes the node and its $I[-]$ value on the right end of M. To justify the computation of the value of $I[X]$, note that if X is blue, then $\psi = 1$ and $I[X]$ was defined to equal $I[\mathcal{M}_R] - c(\mathcal{M}_R, X)$; whereas, if X is red then $\varphi = -1$ and $I[X]$ equals $I[\mathcal{M}_R] + c(\mathcal{M}_R, X)$. (Unless $\mathcal{M}$ is empty, in which case, $I[X] = 0$.)

Once the current node has been pushed onto the right end of $\mathcal{M}$, the following code implements Step (β) from section 2.2:

```
while c(L^{-ψ}_{R-1}, M_R) − c(L^{-ψ}_R, M_R) < ψ · (I[L^{-ψ}_R] − I[L^{-ψ}_{R-1}])
    pop-right L^{-ψ}
```

To justify the correctness of the while condition, suppose that the currently scanned node is red, so $\varphi = -1$. By Lemma 7, $\mathsf{Bnft}[\mathcal{L}^{-\psi}_{R-1}, \mathcal{M}_R] > \mathsf{Bnft}[\mathcal{L}^{-\psi}_R, \mathcal{M}_R]$ iff $c(\mathcal{L}^{-\psi}_{R-1}, \mathcal{M}_R) - c(\mathcal{L}^{-\psi}_R, \mathcal{M}_R) < \Delta[\mathcal{L}^{-\psi}_{R-1}, \mathcal{L}^{-\psi}_R]$. Furthermore, $\Delta[\mathcal{L}^{-\psi}_{R-1}, \mathcal{L}^{-\psi}_R]$ is equal to $\psi \cdot (I[\mathcal{L}^{-\psi}_R] - I[\mathcal{L}^{-\psi}_{R-1}])$ since $\mathcal{L}^{-\psi}$ contains blue nodes and $\psi = -1$ (by the equalities at the end of section 2.2). In this case, $\mathcal{M}_R$ is past the crossover point of $\mathcal{L}^{-\psi}_{R-1}$ and $\mathcal{L}^{-\psi}_R$, so $\mathcal{L}^{-\psi}_R$ may be discarded from consideration as a left endpoint of a candidate. A similar calculation justifies the case when the current node is blue.

To implement Step (γ), the following code is used:

```
if c(M_R, L^{-ψ}_R) < ψ · (I[M_R] − I[L^{-ψ}_R])
    X := pop-right M
    while M_R ≠ L^{-ψ}_R
        Match_Pair()
    Push_Main(X)
```

where *Match_Pair* is defined below. The above **if** statement checks whether $\mathcal{L}^{-\psi}_R \to \mathcal{M}_R$ is a candidate; if so, the algorithm greedily assigns edges to node in the interior of the candidate

[5] For linear tours, only the first scan is needed; however, we treat only the more general (circular) case.

(where 'greedily' means with respect to the nodes that have not already been assigned). Before the greedy assignment is started, the rightmost entry is popped from $\mathcal{M}$ and is saved as X to pushed back on the right end afterwards. There are two reasons for this: firstly, this gets the current node X out of the way of Match_Pair's operation, and secondly and more importantly, when X is pushed back onto $\mathcal{M}$, the $I[-]$ value for the current node is recomputed so as to be correct for the reduced matching problem in which the greedily matched nodes are no longer present. *Match_Pair* is the following procedure:

procedure Match_Pair()
    Output("$\mathcal{M}_{R-1} \leftrightarrow \mathcal{M}_R$")
    pop-right $\mathcal{M}$
    if $\mathcal{M}_R = \mathcal{L}_R^{\psi}$
        pop-right $\mathcal{L}^{\psi}$
    pop-right $\mathcal{M}$
    return()

The procedure Match_Pair assigns a jumper $\mathcal{M}_{R-1} \leftrightarrow \mathcal{M}_R$ and discards a matched node from the deque $\mathcal{L}^{\psi}$ if appears there. Because of the **while** condition controlling calls to *Match_Pair*, it is not possible for a matched node to occur in $\mathcal{L}^{-\psi}$, so we do not check for this condition.

To implement Step (δ), the following code is used:

while $\Omega(\mathcal{L}_{R-1}^{\psi}, \mathcal{L}_R^{\psi}, \mathcal{M}_R)$ = "Yes"
    pop-right $\mathcal{L}^{\psi}$
push-right $\mathcal{M}_R$ onto $\mathcal{L}^{\psi}$ (without popping $\mathcal{M}_R$)

That completes the description of the how nodes are processed during the first scan. As mentioned earlier, the last instruction (the *push-right*) is omitted from Step (δ) during the second scan. Other than this, the processing for Steps (β)-(δ) is identical in the two scans.

One potentially confusing aspect of the second scan is that the $I[-]$ values are no longer actually the correct $I[-]$ values: for example, it is no longer the case that $I[\mathcal{M}_L]$ is necessarily equal to zero. Strictly speaking, the $I[-]$ values all shift by an additive constant when an entry is popped from the left end of $\mathcal{M}$; however, it is not necessary to implement this shift, since the algorithm only uses differences between $I[-]$ values. The end result is that nothing special needs to be done to the I values when we pop-left $\mathcal{M}$.

After both scans are completed, any remaining nodes may be greedily matched. As discussed above, there are two possible greedy matchings and both have the same (optimal) cost. Thus either one may be used: the algorithm below just calls Match_Pair repeatedly to assign one of these greedy matchings.

The complete matching algorithm is shown as Algorithm 2. We claim that its runtime is either $O(N)$ or $O(N \log N)$ depending on whether the weak analyticity condition holds. To see this, note that the initialization and the windup processing both take $O(N)$ time. The loops for the each of the two scans are executed $\leq N$ times. Except for the **while** loops, each call to Process_Node takes constant time. The second **while** loop (which calls Match_Pair) is executed more than once only when edges are being output. If the first or third **while** loop is executed more than once, then vertices are being popped from the $\mathcal{L}$ stacks. Since $\lfloor \frac{1}{2} N \rfloor$ edges are output and since $O(N)$ vertices are pushed onto the $\mathcal{L}$ stacks, each of these **while** loops are executed only $O(N)$ times during the *entire* execution of the algorithm. An iteration of the first or second

Algorithm 2 *This is the matching algorithm for balanced quasi-convex tours. All variables are global.*

"Initialization"
$\mathcal{M}, \mathcal{L}^{-1}, \mathcal{L}^{1} := \emptyset$
ψ := -1
"Read Input into the $\mathcal{M}$ deque"
while [X := Input()] $\neq \emptyset$
    Push_Main (X)
    $\psi := -\psi$
"The First Scan"
while $\mathcal{L}^{\psi}$ is empty or $\mathcal{M}_L \neq \mathcal{L}_L^{\psi}$
    X := pop-left $\mathcal{M}$
    Process_Node()
    push-right $\mathcal{M}_R$ onto $\mathcal{L}^{\psi}$
    $\psi := -\psi$
"The Second Scan"
while $\mathcal{L}^{-1}$ and $\mathcal{L}^{1}$ are not both empty
    X := pop-left $\mathcal{M}$
    if X = $\mathcal{L}_L^{\psi}$
        pop-left $\mathcal{L}^{\psi}$
    Process_Node()
    $\psi := -\psi$
"Windup Processing"
while $\mathcal{M}$ is not empty
    Match_Pair()
Exit.
procedure Process_Node()
    Push_Main(X)
    while $c(\mathcal{L}_{R-1}^{-\psi}, \mathcal{M}_R) - c(\mathcal{L}_R^{-\psi}, \mathcal{M}_R) < \psi \cdot (I[\mathcal{L}_R^{-\psi}]] - I[\mathcal{L}_{R-1}^{-\psi}])$
        pop-right $\mathcal{L}^{-\psi}$
    if $c(\mathcal{M}_R, \mathcal{L}_R^{-\psi}) < \psi \cdot (I[\mathcal{M}_R] - I[\mathcal{L}_R^{-\psi}])$
        X := pop-right $\mathcal{M}$
        while $\mathcal{M}_R \neq \mathcal{L}_R^{-\psi}$
            Match_Pair()
        Push_Main(X)
    while $\Omega(\mathcal{L}_{R-1}^{\psi}, \mathcal{L}_R^{\psi}, \mathcal{M}_R)$ = "Yes"
        pop-right $\mathcal{L}^{\psi}$
    return

while loop takes constant time, while an iteration of the third **while** loop takes either constant time or $O(\log N)$ time, depending on whether the weak analyticity property holds.

Because of space limitations, this extended abstract does not include code for the Ω predicate. When the weak analyticity condition holds, the Ω predicate typically operates by computing two theoretical relative crossovers and comparing their positions. Without the weak analyticity condition, the Ω-predicate runs in logarithmic time, by using a binary search of the $\mathcal{M}$-deque.

There are a couple of improvements that can be made to the algorithm which will increase execution speed by a constant factor. Firstly, the calls to Match_Pair made during the "Windup Processing" do not need to check if $\mathcal{M}_R = \mathcal{L}_R^{\psi}$, since $\mathcal{L}^{\psi}$ is empty at this time. Secondly, if computing the cost function $c(-,-)$ is more costly than simple addition, then it is possible for Push_Main() to use an alternative method during the two scans to compute the cost $c(\mathcal{M}_R, X)$ for nodes X which have just been popped from the left of $\mathcal{M}$ (except for the first one popped from the left in the first scan). Namely, the algorithm can save the old $I[X]$ value for the node X as it is left-popped off the deque $\mathcal{M}$. Then the cost function can

be computed by computing the difference between the $I[-]$ value of X and the $I[-]$ of the previous node left-popped from $\mathcal{M}$. This second improvement applies only to the first Push_Main call in Process_Node.

4. Non-bipartite, quasi-convex tours

In this section we show how the earlier algorithms can be applied to non-bipartite, quasi-convex tours. The principal observation is that non-bipartite tours may be made bipartite by the simple construction of making the nodes alternate in color. This is already observed by Marcotte-Suri [17] in a more restrictive setting; we repeat the construction here for the sake of completeness.

First, it is apparent that the proof of Lemma 4 still works in the non-bipartite case, and thus any non-bipartite, quasi-convex tour has a minimum-cost matching in which no jumpers cross. This fact implies the following two lemmas:

Lemma 9 *Let $x_1, \ldots, x_N$ be a non-bipartite, quasi-convex tour with N even. Then there exists a minimum-cost matching such that every edge in the tour is of the form $x_i \leftrightarrow x_j$ with i even and j odd.*

Proof It will suffice to show that any crossing-free matching has this property. Suppose $x_i \leftrightarrow x_j$ is a jumper in a crossing-free matching, with $i < j$. Since N is even, the matching is complete in that every node is matched. The crossing free property thus implies that the nodes in (x_i, x_j) are matched with each other; so there are an even number of such nodes, i.e., one of i and j is even and the other is odd. □

Lemma 10 *Let $x_1, \ldots, x_N$ be a non-bipartite linear quasiconvex tour. Then there exists a minimum-cost matching such that every edge in the tour is of the form $x_i \leftrightarrow x_j$ with i even and j odd.*

Proof If N is even then this lemma is just a special case of the former lemma. If N is odd, then add an additional node x_{N+1} to the end of the tour, with $c(x_i, x_{N+1}) = 0$ for all i. The resulting tour is again quasi-convex and of even length; so the lemma again follows immediately from the former lemma.

When Lemmas 9 and 10 apply, we may color the even nodes red and the odd nodes blue and reduce the non-bipartite matching problem to a bipartite matching problem. As an immediate corollary, we have that the two Main Theorems also apply in the non-bipartite setting; namely, for non-bipartite, quasi-convex tours of even length and for non-bipartite, linear, quasi-convex tours, the matching problem can always be solved in $O(N \log N)$ time and it can be solved in $O(N)$ time if the weak analyticity condition holds.

We do not know whether similar algorithms exist for the case of general (i.e., non-linear) quasi-convex tours of odd length. Similarly, we do not know any linear or near-linear time algorithms for bipartite, quasi-convex tours which are neither balanced nor linear.

We conclude this section by mentioning a tantalizing connection between our work and the work of F. F. Yao[22]. Yao gave a quadratic runtime algorithm for solving the dynamic programming problem

$$d(i,j) \;=\; c(i,j) + \min\{d(i,k-1) + d(k,j) : i < k \leq j\}$$

for linear quasi-convex tours with cost function c (improving on the obvious cubic-time algorithm). Our non-bipartite matching problem can be stated as a similar dynamic programming problem; namely, the minimum-cost, $MC(i,j)$, of a complete matching on the nodes in $[x_i, x_j]$ can be recursively defined to equal

$$\min\{c(i,k) + MC(i+1,k-1) + MC(k+1,j) : i < k \leq j\}.$$

(A similar dynamic programming algorithm can be given for the bipartite matching problem.) The obvious naive algorithm for computing $MC(-,-)$ is cubic-time, however, our main results give (near) linear time algorithms for linear quasi-convex tours. This raises the possibility that the dynamic programming problem considered by Yao may also have a near-linear time solution.

5. Applications to String Matching

As a final topic we briefly discuss the application of our matching results to string comparison – but a full treatment is beyond the scope of this paper. Given two symbol strings $v = a_1 a_2 \cdots a_n$ and $w = b_1 b_2 \cdots b_n$, our goal is to measure a particular notion of *distance* between them. Intuitively, distance acts as a measure of similarity, i.e. strings that are highly similar (highly dissimilar) are to have a small (large) distance between them. The purpose of such formulations is usually to approximate human similarity judgements within a pattern classification or information retrieval system.

Suppose $f(x)$ is a monotonely increasing, concave-down function with $f(0) = 0$. Let symbols $a_1, \ldots, a_n$ in v be a graph's red nodes, $b_1, \ldots, b_n$ in w be its blue nodes, and consider bipartite matchings of these $2n$ symbols. In the simplest formulation we define the cost of an edge $a_i \leftrightarrow b_j$ as $f(|j-i|)$ if a_i and b_j are the same symbol, and as $f(n)$ if a_i and b_j are distinct symbols. The cost of matching unequal characters can also be set to be any other fixed value instead of $f(n)$. Our *distance*, $\sigma(v,w)$, between strings v and w is then the minimum cost of any such bipartite matching.

As an example, consider the two strings "delve" and "level" and let $f(x) = \sqrt{x}$. Then the distance between these two strings is $\sqrt{5} + \sqrt{0} + \sqrt{2} + \sqrt{1} + \sqrt{1} \approx 5.65$.

As we have set up our problem above, the computation of $\sigma(v,w)$ is not directly an instance of the quasi-convex matching problem. However we can compute the σ function by considering each alphabet symbol α separately, and solving the quasi-convex matching problem σ_α which results from restricting attention to occurrences of a single alphabet symbol at a time. To make this clear, we introduce a special symbol "-" which indicates the absence of an alphabet symbol. The value of σ("delve", "level") can be expressed as the sum

$$\begin{aligned}\sigma_d(\text{“}d\text{----”},\text{“-----”}) + \sigma_e(\text{“-}e\text{--}e\text{”},\text{“-}e\text{-}e\text{-”}) \\ +\sigma_l(\text{“--}l\text{--”},\text{“}l\text{---}l\text{”}) + \sigma_v(\text{“---}v\text{-”},\text{“--}v\text{--”})\end{aligned}$$

To make the summed σ_α terms equal σ as originally defined, each σ_α is defined to be the subproblem's minimum matching cost, plus $f(n)/2$ times the number of unmatched symbols.

We will loosely refer to distance functions that result from this kind of formulation as σ-distances. Assuming that $f(x)$ satisfies the weak analyticity condition, it is not too difficult to show that it is possible to compute $\sigma(v,w)$ in linear time. If the weak analyticity condition does not hold, then our results give an $O(n \log n)$ time algorithm.

A novel feature of our σ-distances is that distinct alphabet symbols are treated independently. This is in contrast to most prior work which has used 'least edit distance' for string comparison (see [18] for a survey). As an illustration of the difference between our distance measure and the 'edit distance' approach, consider comparing the word "abcde" with its mirror image "edcba". Our approach recognizes some

similarity between these two forms, while the most standard 'edit distance' approach sees only that the two strings have "c" in common — in essence substituting the first two and last two symbols of the string without noticing the additional occurrences of the same symbols at the other end of the other string.

A special form of our σ-distance measure in which $f(x) = x$ was introduced earlier by the authors and shown to have a simple linear time algorithm [23, 24]. This earlier algorithm has been successfully used in commercial applications, especially for spelling correction in word processing software, typewriters, and hand-held dictionary devices (we estimate that that over 15,000,000 such software/hardware units have been sold by Proximity Technology, Franklin Electronic Publishers and their licensees). Other less prominent commercial applications include database field search (e.g. looking up a name or address), and the analysis of multi-field records such as mailing addresses, in order to eliminate near-duplicates. In both of these applications, the strict global left-right ordering imposed by $O(n^2)$ time 'edit distance' methods, can be problematic. On the other hand, very local left-right order preservation seems to be an important part of similarity perception in humans. One simple adaptation of our σ-distance methods which goes a long way towards capturing this characteristic, consists of extending the alphabet beyond single symbols to include digraphs or multi-graphs. The result is increased sensitivity to local permutation. Another effective alphabet extension technique involves the addition of *feature symbols* to the alphabet to mark events such as likely phonetic transitions. We expect that the use of general concave-down distance functions (as opposed to $f(x) = x$) will improve the quality of the similarity judgements possible within the σ-distance framework.

The development above considers strings of equal length only. The unequal length case is not a difficult generalization; but considering it does highlight the issue of *embedding*. By this we mean that it is implicit in our formulation that the two strings are in a sense embedded into the real line. The particular rather natural embedding we've assumed so far, maps a_i and b_i to value i on the real line – but others are possible.

A detailed comparison of our methods with 'edit distance' approaches is beyond the scope of this paper. But we must point out that the 'edit distance' formulation is in several senses richer than ours. First, the cost of matching different alphabet members need not be fixed. Also, our distance formulation depends on a designated embedding while the 'edit distance' method requires no such specification. Finally, for some problems, left-right order preservation may be desirable. On the other hand, even the simplest 'edit distance' approach is $O(n^2)$; compared with the $O(n)$ or $O(n \log n)$ complexity of our method. We therefore feel that additional work is needed to better understand the applications of — and perhaps extend our approach.

References

[1] A. Aggarwal, A. Bar-Noy, S. Khuller, D. Kravets, and B. Schieber, *Efficient minimum cost matching using quadrangle inequality*, in Proceedings of the 33th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, 1992, pp. 583–592.

[2] A. Aggarwal and M. Klawe, *Applications of generalized matrix searching to geometric algorithms*, Discrete Applied Mathematics, (1990), pp. 3–23.

[3] A. Aggarwal, M. Klawe, S. Moran, P. Shor, and R. Wilber, *Geometric applications of a matrix-searching algorithm*, Algorithmica, 2 (1987), pp. 195–208.

[4] A. Aggarwal and J. Park, *Notes on searching in multidimensional monotone arrays*, in Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science, IEEE Computer Society, 1988, pp. 497–512.

[5] D. Eppstein, *Sequence comparison with mixed convex and concave costs*, Journal of Algorithms, 11 (1990), pp. 85–101.

[6] Z. Galil and R. Giancarlo, *Speeding up dynamic programming with applications to molecular biology*, Theoretical Computer Science, 64 (1989), pp. 107–118.

[7] Z. Galil and K. Park, *A linear-time algorithm for concave one-dimensional dynamic programming*, Inf. Process. Lett., 33 (1990), pp. 309–311.

[8] P. Gilmore and R. Gomory, *Sequencing a one state-variable machine: A solvable case of the traveling salesman problem*, Operations Research, 12 (1964), pp. 655–679.

[9] X. He, *An efficient parallel algorithm for finding minimum weight matching for points on a convex polygon*, Inf. Process. Lett., 37 (1991), pp. 111–116.

[10] D. S. Hirschberg and L. L. Larmore, *The least weight subsequence problem*, SIAM J. Comput., 16 (1987) 628–638.

[11] R. M. Karp and S.-Y. R. Li, *Two special cases of the assignments problem*, Discrete Mathematics, 13 (1975), pp. 129–142.

[12] M. M. Klawe and D. J. Kleitman, *An almost linear time algorithm for generalized matrix searching*, SIAM J. Disc. Math., 3 (1990), pp. 81–97.

[13] D. Kravets and J. K. Park, *Selection and sorting in totally monotone arrays*, Math. Systems Theory, 24 (1991) 201–220.

[14] L. L. Larmore and B. Schieber, *On-line dynamic programming with applications to the prediction of RNA secondary structure*, Journal of Algorithms, (1991), pp. 490–515.

[15] E. Lawler, *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehart and Winston, 1976.

[16] Y. Mansour, J. K. Park, B. Schieber, and S. Sen, *Improved selection in totally monotone arrays*, in Proceedings, 11th Conf. on Foundations of Software Technology and Theoretical Computer Science, 1991, pp. 347–359.

[17] O. Marcotte and S. Suri, *Fast matching algorithms for points on a polygon*, SIAM J. Comput., 20 (1991), pp. 405–422.

[18] D. Sankoff and J. B. Kruskal, *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, Addison-Wesley, 1983.

[19] P. M. Vaidya, *Geometry helps in matching*, SIAM J. Comput., 18 (1989), pp. 1201–1225.

[20] M. Werman, S. Peleg, R. Melter, and T. Kong, *Bipartite graph matching for points on a line or a circle*, Journal of Algorithms, 7 (1986), pp. 277–284.

[21] R. Wilber, *The concave least-weight subsequence problem revisited*, Journal of Algorithms, 9 (1988), pp. 418–425.

[22] F. F. Yao, *Speed-up in dynamic programming*, SIAM J. Alg. Disc. Meth., 3 (1982), pp. 523–540.

[23] P. N. Yianilos, *The definition, computation and application of symbol string similarity functions*, Master's thesis, Emory University, 1978.

[24] P. N. Yianilos and S. R. Buss, *Associative memory circuit system and method, continuation-in-part.* U.S. Patent #4490811, December 1984.