The Linear Distance Traveling Tournament Problem

Richard Hoshino and Ken-ichi Kawarabayashi
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

Abstract

We introduce a linear distance relaxation of the n-team Traveling Tournament Problem (TTP), a simple yet powerful heuristic that temporarily "assumes" the n teams are located on a straight line, thereby reducing the $\binom{n}{2}$ pairwise distance parameters to just $n-1$ variables. The modified problem then becomes easier to analyze, from which we determine an approximate solution for the actual instance on n teams. We present combinatorial techniques to solve the Linear Distance TTP (LD-TTP) for $n = 4$ and $n = 6$, without any use of computing, generating the complete set of optimal distances regardless of where the n teams are located. We show that there are only 295 non-isomorphic schedules that can be a solution to the 6-team LD-TTP, and demonstrate that in *all* previously-solved benchmark TTP instances on 6 teams, the distance-optimal schedule appears in this list of 295, even when the six teams are arranged in a circle or located in three-dimensional space. We then extend the LD-TTP to multiple rounds, and apply our theory to produce a nearly-optimal regular-season schedule for the Nippon Pro Baseball league in Japan. We conclude the paper by generalizing our theory to the n-team LD-TTP, producing a feasible schedule whose total distance is guaranteed to be no worse than $\frac{4}{3}$ times the optimal solution.

Introduction

The Traveling Tournament Problem (TTP) is a well-known problem in the area of sports scheduling that has attracted much research activity in recent years (Kendall et al. 2010). Inspired by the real-life problem of optimizing the regular-season schedule for Major League Baseball (MLB), the goal of the TTP is to determine the optimal double round-robin tournament schedule for an n-team sports league that minimizes the sum total of distances traveled by all n teams (Easton, Nemhauser, and Trick 2001). The proposers of the TTP serve as consultants to MLB, and have created the league's regular-season schedules for seven of the past eight years.

There is an online set of benchmark n-team TTP data sets (Trick 2012). For example, NLn are the instances for MLB's National League on n teams, and CIRCn are the instances where the n teams correspond to vertices of a circle graph, with a distance of 1 unit between neighbouring vertices.

Several techniques have been applied to solve TTP instances, including local search techniques as well as integer and constraint programming. Solutions to TTP instances are often found after weeks of computation on high-performance machines using parallel computing; the first solution to NL6 required over fifteen minutes of computation time on twenty parallel machines (Easton, Nemhauser, and Trick 2002). A recently-developed branch-and-price heuristic (Irnich 2010) solved NL6 in one minute, CIRC6 in three hours, and NL8 in twelve hours, all on a single processor.

In many ways, the TTP is a variant of the well-known Traveling Salesman Problem (TSP), asking for a distance-optimal schedule linking venues that are close to one another. The computational complexity of the TSP is NP-hard; recently, it was shown that solving the TTP is strongly NP-hard (Thielen and Westphal 2011).

The purpose of this paper is to introduce the *Linear Distance* Traveling Tournament Problem (LD-TTP), where we assume the n teams are located on a straight line, thereby reducing its complexity. This straight line relaxation is a natural heuristic when the n teams are located in cities connected by a common train line running in one direction, modelling the actual context of domestic sports leagues in countries such as Chile, Sweden, Italy, and Japan. As we will demonstrate in the paper, solving the LD-TTP is considerably easier, and for the cases $n = 4$ and $n = 6$, we can determine the complete set of possible solutions through elementary combinatorial techniques without *any* use of computing.

The LD-TTP contributes a simple yet powerful idea to the field of tournament scheduling, where the straight-line relaxation enables us to generate approximate solutions to large n-team TTP benchmark sets by "pretending" the n teams lie on a straight line, solving the modified problem to find an "optimal" tournament schedule, and then applying the actual distance matrix on this schedule to find a feasible solution to the TTP. We find that this technique surprisingly generates the distance-optimal schedule for *all* benchmark sets on 6 teams. We then extend the LD-TTP to multiple rounds in order to generate a close-to-optimal solution for Japan's pro baseball league, and determine a general solution to the LD-TTP for any $n \equiv 4 \pmod 6$ that is guaranteed to be a $\frac{4}{3}$-approximation of the distance-optimal schedule.

The Traveling Tournament Problem

Let $\{t_1, t_2, \ldots, t_n\}$ be the n teams in a sports league, where n is even. Let D be the $n \times n$ *distance matrix*, where entry $D_{i,j}$ is the distance between the home stadiums of teams t_i and t_j. By definition, $D_{i,j} = D_{j,i}$ for all $1 \leq i, j \leq n$, and all diagonal entries $D_{i,i}$ are zero. We assume the distances form a *metric*, i.e., $D_{i,j} \leq D_{i,k} + D_{k,j}$ for all i, j, k.

The TTP requires a tournament lasting $2(n-1)$ days, where every team has exactly one game scheduled each day with no byes or days off (this explains why n must be even.) The objective is to minimize the total distance traveled by the n teams, subject to the following conditions:

(a) *each-venue*: Each pair of teams plays twice, once in each other's home venue.

(b) *at-most-three*: No team may have a home stand or road trip lasting more than three games.

(c) *no-repeat*: A team cannot play against the same opponent in two consecutive games.

When calculating the total distance, we assume that every team begins the tournament at home and returns home after playing its last away game. Furthermore, whenever a team has a road trip consisting of multiple away games, the team doesn't return to their home city but rather proceeds directly to their next away venue.

To illustrate with a specific example, Table 1 lists the distance-optimal schedule (Easton, Nemhauser, and Trick 2001) for the NL6 benchmark set. In this schedule, as with all subsequent schedules presented in this paper, home games are marked in bold.

Team	1	2	3	4	5	6	7	8	9	10
Florida (F)	A	Ph	**N**	**Pi**	N	M	Pi	**Ph**	**M**	**A**
Atlanta (A)	**F**	**N**	**Pi**	Ph	M	Pi	**Ph**	**M**	N	F
Pittsburgh (Pi)	**N**	**M**	A	F	**Ph**	**A**	**F**	N	Ph	M
Philadelphia (Ph)	M	**F**	**M**	**A**	Pi	**N**	A	F	**Pi**	N
New York (N)	Pi	A	F	**M**	**F**	Ph	M	**Pi**	**A**	**Ph**
Montreal (M)	**Ph**	Pi	Ph	N	**A**	**F**	**N**	A	F	**Pi**

Table 1: An optimal TTP solution for NL6.

For example, the total distance traveled by Florida is $D_{F,A} + D_{A,Ph} + D_{Ph,F} + D_{F,N} + D_{N,M} + D_{M,Pi} + D_{Pi,F}$. Based on the NL6 distance matrix (Trick 2012), the tournament schedule in Table 1 requires 23916 miles of total team travel, which is the minimum distance possible.

The 4-Team LD-TTP

In the Linear Distance TTP, we assume the n home stadiums lie on a straight line, with t_1 at one end and t_n at the other. Thus, $D_{i,j} = D_{i,k} + D_{k,j}$ for all triplets (i, j, k) with $1 \leq i < k < j \leq n$. Since the Triangle Inequality is replaced by the Triangle Equality, we no longer need to consider all $\binom{n}{2}$ entries in the distance matrix D; each tournament's total travel distance is a function of $n-1$ variables, namely the set $\{D_{i,i+1} : 1 \leq i \leq n-1\}$. For notational convenience, denote $d_i := D_{i,i+1}$ for all $1 \leq i \leq n-1$.

Table 2 gives a feasible solution to the 4-team LD-TTP. We claim that this solution is optimal, for all possible 3-tuples (d_1, d_2, d_3). To see why this is so, define ILB_{t_i} to be the *independent lower bound* for team t_i, the minimum possible distance that can be traveled by t_i in order to complete its games, independent of the other teams' schedules. Then a trivial lower bound for the total travel distance is $TLB \geq \sum_{i=1}^{n} ILB_{t_i}$.

Team	1	2	3	4	5	6
t_1	$\boldsymbol{t_4}$	$\boldsymbol{t_3}$	t_2	t_4	t_3	$\boldsymbol{t_2}$
t_2	$\boldsymbol{t_3}$	$\boldsymbol{t_4}$	$\boldsymbol{t_1}$	t_3	t_4	t_1
t_3	t_2	t_1	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_4
t_4	t_1	t_2	t_3	$\boldsymbol{t_1}$	$\boldsymbol{t_2}$	$\boldsymbol{t_3}$

Table 2: An optimal LD-TTP solution for $n = 4$.

Since t_i must play a road game against each of the other three teams, $ILB_{t_i} = 2(d_1 + d_2 + d_3)$ for $1 \leq i \leq 4$. This implies that $TLB \geq 8(d_1 + d_2 + d_3)$. Since Table 2 is a tournament schedule whose total distance is the trivial lower bound, this completes the proof.

We remark that Table 2 is not the unique solution - for example, we can generate another optimal schedule by simply reading Table 2 from right to left. Assuming the first match between t_1 and t_2 occurs in the home city of t_2, a straightforward computer search finds 18 non-isomorphic schedules with total distance $8(d_1+d_2+d_3)$. Thus, by symmetry, there are 36 optimal schedules for the 4-team LD-TTP.

The 6-Team LD-TTP

Unlike the previous section, the analysis for the 6-team LD-TTP requires more work.

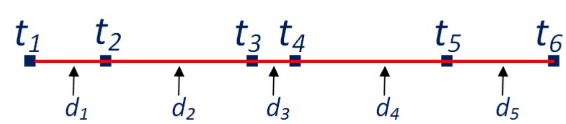

Figure 1: The general instance of the LD-TTP for $n = 6$.

Any 6-team instance of the LD-TTP can be represented by the five-tuple $(d_1, d_2, d_3, d_4, d_5)$. We define $S = 14d_1 + 16d_2 + 20d_3 + 16d_4 + 14d_5$. We claim the following:

Theorem 1 *Let Γ be a 6-team instance of the LD-TTP. The optimal solution to Γ is a schedule with total distance*

$$S+2\min\{d_2+d_4, d_1+d_4, d_3+d_4, 3d_4, d_2+d_5, d_2+d_3, 3d_2\}.$$

We will prove Theorem 1 through elementary combinatorial arguments with no computing, thus demonstrating the utility of this linear distance relaxation and presenting new techniques to attack the general TTP in ways that differ from integer/constraint programming. Our proof will follow from several lemmas, which we now prove one by one.

Lemma 1 *Any feasible schedule of Γ must have total distance at least S.*

Proof For each $1 \leq k \leq 5$, define c_k to be the total number of times a team crosses the "bridge" of length d_k, connecting the home stadiums of teams t_k and t_{k+1}. Let Z be the total travel distance of this schedule. Since Γ is linear, $Z = \sum_{k=1}^{5} c_k d_k$. Since each team crosses every bridge an even number of times, c_k is always even.

Let L_k be the home venues of $\{t_1, t_2, \ldots, t_k\}$ and R_k be the home venues of $\{t_{k+1}, \ldots, t_6\}$. By the *each-venue* condition, every team in L_k plays a road game against every team in R_k. By the *at-most-three* condition, every team in L_k must make at least $2\lceil \frac{6-k}{3} \rceil$ trips across the bridge, with half the trips in each direction. Similarly, every team in R_k must make at least $2\lceil \frac{k}{3} \rceil$ trips across the bridge, implying that $c_k \geq 2k\lceil \frac{6-k}{3} \rceil + 2(6-k)\lceil \frac{k}{3} \rceil$.

Thus, $c_1 \geq 14$, $c_2 \geq 16$, $c_4 \geq 16$, and $c_5 \geq 14$. We now show that $c_3 \geq 20$, which will complete the proof that $Z = \sum c_k d_k \geq 14d_1 + 16d_2 + 20d_3 + 16d_4 + 14d_5 = S$.

Since there are $n = 6$ teams, there are $2(n-1) = 10$ days of games. For each $1 \leq i \leq 9$, let $X_{i,i+1}$ be the total number of times the d_3-length bridge is crossed as the teams move from their games on the i^{th} day to their games on the $(i+1)^{\text{th}}$ day. Let $X_{start,1}$ and $X_{10,end}$ respectively be the number of times the teams cross this bridge to play their first game, and return home after having played their last game. Then $c_3 = X_{start,1} + \sum_{i=1}^{9} X_{i,i+1} + X_{10,end}$.

For each $1 \leq i \leq 9$, let $f(i)$ denote the number of games played in L_3 on day i. Thus, on day i, exactly $2f(i)$ teams are to the left of this bridge and $6 - 2f(i)$ teams are to the right. So $f(i) \in \{0, 1, 2, 3\}$ for all i. Since $|L_3|$ and $|R_3|$ are odd, we have $X_{start,1} \geq 1$ and $X_{10,end} \geq 1$.

If $f(i) < f(i+1)$, then $X_{i,i+1} \geq 2$, as at least two teams who played in R_3 on day i must cross over to play their next game in L_3. Similarly, if $f(i) > f(i+1)$, then $X_{i,i+1} \geq 2$.

If $f(i) = f(i+1) = 1$, then on day i, two teams p and q play in L_3 while the other four teams play in R_3. If $X_{i,i+1} = 0$ then no team crosses the bridge after day i, forcing p and q to play against each other on day $i+1$, thus violating the *no-repeat* condition. Thus, at least one of p or q must cross the bridge, exchanging positions with at least one other team who must cross to play in L_3. Thus, $X_{i,i+1} \geq 2$. Similarly, if $f(i) = f(i+1) = 2$, then $X_{i,i+1} \geq 2$.

If $f(i) = f(i+1) = 0$, then all teams play in R_3 on days i and $i+1$. Then $X_{start,1} = 3$ if $i = 1$ and $X_{10,end} = 3$ if $i = 9$. If $2 \leq i \leq 8$, then each of $\{t_1, t_2, t_3\}$ must play a home game on either day $i-1$ or day $i+2$, in order to satisfy the *at-most-three* condition. Thus, on one of these two days, at least two teams in $\{t_1, t_2, t_3\}$ play at home, implying at least four teams are in L_3. Therefore, we must have $X_{i-1,i} \geq 4$ or $X_{i+1,i+2} \geq 4$.

We derive the same results if $f(i) = f(i+1) = 3$. We have $X_{start,1} = 3$ if $i = 1$, $X_{10,end} = 3$ if $i = 9$, and either $X_{i-1,i} \geq 4$ or $X_{i+1,i+2} \geq 4$ if $2 \leq i \leq 8$.

So in our double round-robin schedule, if the sequence $\{f(1), \ldots, f(10)\}$ has no pair of consecutive 0s or consecutive 3s, then $c_3 = X_{start,1} + \sum_{i=1}^{9} X_{i,i+1} + X_{10,end} \geq 1 + 9 \cdot 2 + 1 = 20$. And if this is not the case, we still have $c_3 \geq 20$ from the results of the previous two paragraphs. We have therefore proven that $Z \geq S$. ∎

Lemma 2 *Consider a feasible schedule of Γ with total distance $Z = \sum c_k d_k$. If $c_2 = 16$, then teams t_1 and t_2 must play against each other on Days 1 and 10.*

Proof Like we did in Lemma 1, for each $1 \leq i \leq 9$ define $X^*_{i,i+1}$ be the total number of times the d_2-length bridge is crossed as the teams move from their games on the i^{th} day to their games on the $(i+1)^{\text{th}}$ day. Similarly define $X^*_{start,1}$ and $X^*_{10,end}$ so that $c_2 = X^*_{start,1} + \sum_{i=1}^{9} X^*_{i,i+1} + X^*_{10,end}$.

By a nearly-identical case-analysis argument as in the previous proof, we can show that $\sum_{i=1}^{9} X^*_{i,i+1} \geq 16$. Therefore, if $c_2 = 16$, then we must have $X^*_{start,1} = X^*_{10,end} = 0$, implying that on Days 1 and 10, t_1 and t_2 stay in L_2 while the other four teams stay in R_2. Since t_1 and t_2 are the only teams in L_2, clearly this forces these two teams to play against each other, to begin and end the tournament. ∎

Lemma 3 *Let S_1 be the set of tournament schedules with distance $S + 2(d_2 + d_4)$, S_2 with distance $S + 2(d_1 + d_4)$, S_3 with distance $S + 2(d_3 + d_4)$, S_4 with distance $S + 6d_4$, S_5 with distance $S + 2(d_2 + d_5)$, S_6 with distance $S + 2(d_2 + d_3)$, and S_7 with distance $S + 6d_2$. Then each set in $\{S_1, S_2, \ldots, S_7\}$ is non-empty.*

Proof For each of these seven sets, it suffices to find just one feasible schedule with the desired total distance. For each of $\{S_1, S_2, S_3, S_4\}$, at least one such set has appeared previously in the literature, as the solution to a 6-team benchmark set or in some other context. (As we will see in the following section, we can label the six teams of NL6 so that Table 1 is an element of S_4.) The solution to CIRC6 (Trick 2012), where $D_{i,j} = \min\{j-i, 6-(j-i)\}$ for all $1 \leq i < j \leq 6$, is an element of S_1. Table 3 provides this schedule. For each $1 \leq k \leq 5$, we list the number of times the d_k bridge is crossed by each of the six teams.

	1	2	3	4	5	6	7	8	9	10	d_1	d_2	d_3	d_4	d_5
t_1	t_2	t_3	t_4	$\boldsymbol{t_6}$	$\boldsymbol{t_3}$	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_2}$	4	4	4	2	2
t_2	$\boldsymbol{t_1}$	t_6	t_5	t_4	$\boldsymbol{t_6}$	t_3	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_3}$	t_1	2	4	2	2	2
t_3	t_4	$\boldsymbol{t_1}$	t_6	t_5	t_1	$\boldsymbol{t_2}$	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_2	$\boldsymbol{t_4}$	2	4	4	2	2
t_4	$\boldsymbol{t_3}$	t_5	$\boldsymbol{t_1}$	$\boldsymbol{t_2}$	$\boldsymbol{t_5}$	t_6	t_2	t_1	$\boldsymbol{t_6}$	t_3	2	2	4	4	2
t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$	$\boldsymbol{t_3}$	t_4	$\boldsymbol{t_1}$	t_3	t_2	t_1	$\boldsymbol{t_6}$	2	2	2	4	2
t_6	$\boldsymbol{t_5}$	$\boldsymbol{t_2}$	$\boldsymbol{t_3}$	t_1	t_2	$\boldsymbol{t_4}$	$\boldsymbol{t_1}$	t_3	t_4	t_5	2	2	2	4	4

Table 3: An optimal TTP solution for CIRC6, with total distance $S + 2(d_2 + d_4) = 14d_1 + 18d_2 + 20d_3 + 18d_4 + 14d_5$.

We conclude the proof by noting that $|S_{i+3}| = |S_i|$ for $2 \leq i \leq 4$, as we can label the teams backward from t_6 to t_1 to create a feasible schedule where each distance d_k is replaced by d_{6-k}. Therefore, we have shown that each S_i is non-empty. ∎

We are now ready to prove Theorem 1, that the optimal solution to any 6-team instance Γ is a schedule that appears in $S_1 \cup S_2 \cup \ldots \cup S_7$. We note that any of these seven optimal distances can be the minimum, depending on the 5-tuple $(d_1, d_2, d_3, d_4, d_5)$.

Proof Suppose the optimal solution to Γ has total distance $Z = \sum c_k d_k$. By Lemma 1, $c_1, c_5 \geq 14$, $c_2, c_4 \geq 16$, and $c_3 \geq 20$. Recall that each coefficient c_k is even.

By Lemma 3, S_1 is non-empty, and so a schedule cannot be optimal if $Z > S + 2(d_2 + d_4)$. Thus, if $c_2, c_4 \geq 18$, then we must have $(c_1, c_2, c_3, c_4, c_5) = (14, 18, 20, 18, 14)$ so that $Z = S + 2(d_2 + d_4)$, forcing the schedule to be in set S_1.

Suppose that $c_2 \leq c_4$, so that it suffices to check the possibility $c_2 = 16$. By Lemma 2, t_1 and t_2 must play against each other on Days 1 and 10. There are three cases:

Case 1: $c_2 = 16, c_1 = 14$.

Case 2: $c_2 = 16, c_1 \geq 16, c_4 = 16$.

Case 3: $c_2 = 16, c_1 \geq 16, c_4 \geq 18$.

In Case 1, every team must travel the minimum number of times across the d_1- and d_2-bridges, i.e., t_1 must take exactly two road trips, and each of $\{t_3, t_4, t_5, t_6\}$ must play their road games against t_1 and t_2 on consecutive days. By symmetry, we may assume that the first match between t_1 and t_2 occurs in the home city of t_2. Then a simple case analysis shows that for some permutation $\{p, q, r, s\}$ of $\{3, 4, 5, 6\}$, the schedule for teams t_1 and t_2 must be

	1	2	3	4	5	6	7	8	9	10
t_1	t_2	$t_?$	$\boldsymbol{t_p}$	$\boldsymbol{t_q}$	$t_?$	$t_?$	$t_?$	$\boldsymbol{t_r}$	$\boldsymbol{t_s}$	$\boldsymbol{t_2}$
t_2	$\boldsymbol{t_1}$	$\boldsymbol{t_p}$	$\boldsymbol{t_q}$	$t_?$	$t_?$	$t_?$	$\boldsymbol{t_r}$	$\boldsymbol{t_s}$	$t_?$	t_1

This structural characterization reduces the search space considerably, and from this we show that either $c_4 \geq 22$, or $c_3 \geq 22$ and $c_4 \geq 18$. By Lemma 3, the latter implies $Z = S + 2(d_3 + d_4)$ and the former implies $Z = S + 6d_4$. Therefore, this optimal schedule must be in S_3 or S_4.

In Case 2, we demonstrate that no structural characterization exists if $c_2 = c_4 = 16$. To do this, we use Lemma 2 (for $c_2 = 16$) and its symmetric analogue (for $c_4 = 16$) to show that in order not to violate the *at-most-three* or *no-repeat* conditions, t_3 and t_4 must play each other on Days 1 and 10, as well as on some other Day i with $2 \leq i \leq 9$. But then this violates the *each-venue* condition. Hence, we may eliminate this case.

In Case 3, if $c_1 \geq 16$ and $c_4 \geq 18$, then Z is at least $S + 2(d_1 + d_4)$. By Lemma 3, we must have $Z = S + 2(d_1 + d_4)$ and this optimal schedule must be in S_2.

So we have shown that if $c_2 = 16$, then the schedule appears in $S_2 \cup S_3 \cup S_4$. By symmetry, if $c_4 = 16$, then the schedule appears in $S_5 \cup S_6 \cup S_7$. Finally, if $c_2, c_4 \geq 18$, the schedule appears in S_1. This concludes the proof. ∎

By Theorem 1, there are only seven possible optimal distances. For each optimal distance, we can enumerate the set of tournament schedules with that distance, thus producing the complete set of possible LD-TTP solutions, over all instances, for the case $n = 6$.

Theorem 2 *Consider the set of all feasible tournaments for which the first game between t_1 and t_2 occurs in the home city of t_2. Then there are 295 non-isomorphic schedules whose total distance appears in $S_1 \cup S_2 \cup \ldots \cup S_7$, grouped as follows:*

Total Distance	$\in S_1$	$\in S_2$	$\in S_3$	$\in S_4$	$\in S_5$	$\in S_6$	$\in S_7$
# of Schedules	223	4	8	24	4	8	24

We derive Theorem 2 by a computer search. For each of $\{S_1, S_2, S_3, S_4\}$, we develop a structural characterization theorem, similar to Case 1 above, that shows that a feasible schedule in that set must have a certain form. This characterization reduces the search space, from which a brute-force search (using Maplesoft) enumerates all possible schedules. While it took several hours to enumerate the 223 schedules in S_1, Maplesoft took less than 100 seconds to enumerate the set of schedules in each of S_2, S_3, and S_4. As noted earlier, once we have the set of schedules in S_i (for $2 \leq i \leq 4$), we immediately have the set of schedules in S_{i+3} by symmetry. Complete details appear in our journal paper (Hoshino and Kawarabayashi 2012).

Application to Benchmark Sets

We now apply Theorems 1 and 2 to all benchmark TTP sets on 6 teams. In addition to NL6, we examine a six-team set from the Super Rugby League (SUPER6), six galaxy stars whose coordinates appear in three-dimensional space (GALAXY6), our earlier six-team circular distance instance (CIRC6), and the trivial constant distance instance (CON6) where each pair of teams has a distance of one unit.

For all our benchmark sets, we first order the six teams so that they approximate a straight line, either through a formal "line of best fit" or an informal check by inspection. Having ordered our six teams, we determine the five-tuple $(d_1, d_2, d_3, d_4, d_5)$ from the distance matrix and ignore the other $\binom{6}{2} - 5 = 10$ entries. Modifying our benchmark set and assuming the six teams lie on a straight line, we solve the LD-TTP via Theorem 1. Using Theorem 2, we take the set of tournament schedules achieving this optimal distance and apply the actual distance matrix of the benchmark set (with all $\binom{6}{2}$ entries) to each of these optimal schedules and output the tournament with the minimum total distance.

This simple process, each taking approximately 0.3 seconds of computation time per benchmark set, generates a feasible solution to the 6-team TTP. To our surprise, this algorithm outputs the distance-optimal schedule in all five of our benchmark sets. This was an unexpected result, given the non-linearity of our data sets: for example, CIRC6 has the teams arranged in a circle, while GALAXY6 uses three-dimensional distances. To illustrate our theory, let us begin with NL6, ordering the six teams from south to north:

Figure 2: Location of the six NL6 teams.

Thus, Florida is t_1, Atlanta is t_2, Pittsburgh is t_3, Philadelphia is t_4, New York is t_5, and Montreal is t_6. From the NL6 distance matrix (Trick 2012), we have $(d_1, d_2, d_3, d_4, d_5) = (605, 521, 257, 80, 337)$.

Since $2\min\{d_2 + d_4, d_1 + d_4, d_3 + d_4, 3d_4, d_2 + d_5, d_2 + d_3, 3d_2\} = 6d_4 = 480$, Theorem 1 tells us that the optimal LD-TTP solution has total distance $S+6d_4 = 14d_1+16d_2+20d_3 + 22d_4 + 14d_5 = 28424$. By Theorem 2, there are 24 schedules in set S_4, all with total distance $S + 6d_4$. Two of these 24 schedules are presented in Table 4.

	1	2	3	4	5	6	7	8	9	10
t_1	t_2	t_4	$\boldsymbol{t_5}$	$\boldsymbol{t_3}$	t_5	t_6	t_3	$\boldsymbol{t_4}$	$\boldsymbol{t_6}$	$\boldsymbol{t_2}$
t_2	$\boldsymbol{t_1}$	$\boldsymbol{t_5}$	$\boldsymbol{t_3}$	t_4	t_6	t_3	$\boldsymbol{t_4}$	$\boldsymbol{t_6}$	t_5	t_1
t_3	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_2	t_1	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_5	t_4	t_6
t_4	t_6	$\boldsymbol{t_1}$	$\boldsymbol{t_6}$	$\boldsymbol{t_2}$	t_3	$\boldsymbol{t_5}$	t_2	t_1	$\boldsymbol{t_3}$	t_5
t_5	t_3	t_2	t_1	$\boldsymbol{t_6}$	$\boldsymbol{t_1}$	t_4	t_6	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_4}$
t_6	$\boldsymbol{t_4}$	t_3	t_4	t_5	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	$\boldsymbol{t_5}$	t_2	t_1	$\boldsymbol{t_3}$

	1	2	3	4	5	6	7	8	9	10
t_1	t_2	t_5	$\boldsymbol{t_6}$	$\boldsymbol{t_3}$	t_6	t_4	t_3	$\boldsymbol{t_5}$	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$
t_2	$\boldsymbol{t_1}$	$\boldsymbol{t_6}$	$\boldsymbol{t_3}$	t_5	t_4	t_3	$\boldsymbol{t_5}$	$\boldsymbol{t_4}$	t_6	t_1
t_3	$\boldsymbol{t_6}$	t_4	t_2	t_1	$\boldsymbol{t_5}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_6	t_5	$\boldsymbol{t_4}$
t_4	$\boldsymbol{t_5}$	$\boldsymbol{t_3}$	t_5	t_6	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	$\boldsymbol{t_6}$	t_2	t_1	t_3
t_5	t_4	$\boldsymbol{t_1}$	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$	t_3	$\boldsymbol{t_6}$	t_2	t_1	$\boldsymbol{t_3}$	t_6
t_6	t_3	t_2	t_1	$\boldsymbol{t_4}$	$\boldsymbol{t_1}$	t_5	t_4	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_5}$

Table 4: Two LD-TTP solutions with total distance $S + 6d_4$.

Removing this straight line assumption, we now apply the actual NL6 distance matrix to determine the total distance traveled for each of these 24 schedules from set S_4, which will naturally produce different sums. The top schedule in Table 4 is best among the 24 schedules, with total distance 23916, while the bottom schedule is the worst, with total distance 24530. We note that the top schedule, achieving the optimal distance of 23916 miles, is identical to Table 1.

We repeat the same analysis with the other four benchmark sets. In each, we mark which of the sets $\{S_1, S_2, \ldots, S_7\}$ produced the optimal schedule.

Benchmark Data Set	Optimal Solution	LD-TTP Solution	Optimal Schedule
NL6	23916	23916	$\in S_4$
SUPER6	130365	130365	$\in S_3$
GALAXY6	1365	1365	$\in S_1$
CIRC6	64	64	$\in S_1$
CON6	43	43	$\in S_1$

Table 5: Comparing LD-TTP to TTP on benchmark data sets.

A sophisticated branch-and-price heuristic (Irnich 2010) solved NL6 in just over one minute, yet required three hours to solve CIRC6. The latter problem was considerably more difficult due to the inherent symmetry of the data set, which required more branching. However, through our LD-TTP approach, both problems can be solved to optimality in the same amount of time – approximately 0.3 seconds.

Based on the results of Table 5, we ask whether there exists a 6-team instance Γ where the optimal TTP solution is different from the optimal LD-TTP solution. This will be discussed at the conclusion of the paper.

Application to Japanese Baseball

The Multi-Round Balanced Traveling Tournament Problem (Hoshino and Kawarabayashi 2011b) was motivated by the actual regular-season structure of Nippon Professional Baseball (NPB), Japan's largest and most well-known professional sports league. The mb-TTP extends the TTP to $r = 2k$ rounds, for any arbitrary $k \geq 1$, so that k double round-robin tournaments are concatenated together.

In the case of the NPB, we have $n = 6$ and $k = 4$, as the six teams play $k(2n-2) = 40$ *sets* of three games against the other five teams. Our analysis for the LD-TTP is particularly suitable for the NPB Central League, as the home stadiums of the six teams lie on the same bullet train line:

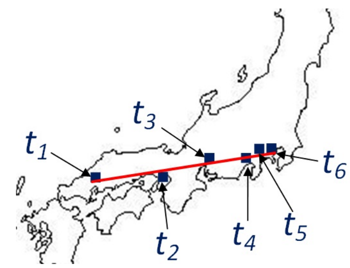

Figure 3: Location of the six NPB Central League teams.

In addition to the *each-venue*, *at-most-three*, and *no-repeat* conditions, the NPB schedule requires two further "balancing" constraints to ensure competitive fairness:

(d) *each-round*: Each pair of teams must play exactly one (three-game) set in each 5-set round.

(e) *diff-two*: $|H_{i,s} - R_{i,s}| \leq 2$ for all $1 \leq i \leq n$ and $1 \leq s \leq 2k(n-1)$, where $H_{i,s}$ and $R_{i,s}$ are the number of home and road sets played by team i in the first s sets.

We found the distance-optimal regular-season schedule for both the NPB Central and Pacific Leagues (Hoshino and Kawarabayashi 2011c), achieving a total reduction of 25% as compared to the actual distance traveled by the teams during the 2010 season. To do this, we applied the theory of perfect matchings to enumerate all $169,728$ double round-robin tournaments satisfying these five conditions, requiring 67 hours of computation time. We then took these pre-computed tournaments and turned the mb-TTP into a shortest-path problem with vertices corresponding to ten-set "blocks" and edge weights corresponding to travel distances, a process requiring a further five hours of computation. Then Dijkstra's Algorithm generated the distance-optimal tournaments: $114,169$ kilometres for the Pacific League and $57,836$ kilometres for the Central League.

By the theories developed in this paper, we can develop a close-to-optimal tournament for both leagues, at a fraction of the computational cost. We do this by taking the 295 schedules in Theorem 2, and noting that only four satisfy these additional balancing constraints (all belonging to set

S_1). The remaining 291 must be thrown away as they either have some pair of teams meeting twice within the first half of the schedule (five sets of games), or have a team begin or end the season with three consecutive home/road sets. Including the schedules where we play the games from right to left, there are only eight schedules that satisfy the five mb-TTP conditions, including Table 6 below.

	1	2	3	4	5	6	7	8	9	10
t_1	t_4	t_2	$\boldsymbol{t_3}$	$\boldsymbol{t_6}$	$\boldsymbol{t_5}$	t_3	t_6	t_5	$\boldsymbol{t_4}$	$\boldsymbol{t_2}$
t_2	t_3	$\boldsymbol{t_1}$	$\boldsymbol{t_6}$	$\boldsymbol{t_5}$	t_4	t_6	t_5	$\boldsymbol{t_4}$	$\boldsymbol{t_3}$	t_1
t_3	$\boldsymbol{t_2}$	t_5	t_1	$\boldsymbol{t_4}$	$\boldsymbol{t_6}$	$\boldsymbol{t_1}$	t_4	t_6	t_2	$\boldsymbol{t_5}$
t_4	$\boldsymbol{t_1}$	t_6	t_5	t_3	$\boldsymbol{t_2}$	$\boldsymbol{t_5}$	$\boldsymbol{t_3}$	t_2	t_1	$\boldsymbol{t_6}$
t_5	t_6	$\boldsymbol{t_3}$	$\boldsymbol{t_4}$	t_2	t_1	t_4	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	$\boldsymbol{t_6}$	t_3
t_6	$\boldsymbol{t_5}$	$\boldsymbol{t_4}$	t_2	t_1	t_3	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	$\boldsymbol{t_3}$	t_5	t_4

Table 6: A schedule in S_1 satisfying the mb-TTP conditions.

By restricting our attention to 8 candidate blocks, rather than the full set of 169, 728, our Dijkstra-based shortest-path algorithm takes just 0.5 seconds to output a feasible tournament satisfying the five conditions of the mb-TTP, whose total travel distance is just slightly worse than the provably-optimal solutions. The results are shown in Table 7, with the solution for the Central League given in Table 8.

NPB League	Optimal Solution	LD-TTP Solution	Percentage Difference
Central	57836	59079	2.1%
Pacific	114169	118782	4.0%

Table 7: Comparing LD-TTP to TTP for the NPB League.

	1-5	6-10	11-15	16-20
t_1	$t_3\ t_2\ \boldsymbol{t_4}\ \boldsymbol{t_6}\ \boldsymbol{t_5}$	$t_4\ t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_2}$	$t_3\ t_2\ \boldsymbol{t_4}\ \boldsymbol{t_6}\ \boldsymbol{t_5}$	$t_4\ t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_2}$
t_2	$t_4\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ \boldsymbol{t_5}\ t_3$	$t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_4}\ t_1$	$t_4\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ \boldsymbol{t_5}\ t_3$	$t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_4}\ t_1$
t_3	$\boldsymbol{t_1}\ t_6\ t_5\ \boldsymbol{t_4}\ \boldsymbol{t_2}$	$\boldsymbol{t_5}\ t_4\ t_2\ t_1\ \boldsymbol{t_6}$	$\boldsymbol{t_1}\ t_6\ t_5\ \boldsymbol{t_4}\ \boldsymbol{t_2}$	$\boldsymbol{t_5}\ t_4\ t_2\ t_1\ \boldsymbol{t_6}$
t_4	$\boldsymbol{t_2}\ t_5\ t_1\ t_3\ \boldsymbol{t_6}$	$\boldsymbol{t_1}\ \boldsymbol{t_3}\ t_6\ t_2\ \boldsymbol{t_5}$	$\boldsymbol{t_2}\ t_5\ t_1\ t_3\ \boldsymbol{t_6}$	$\boldsymbol{t_1}\ \boldsymbol{t_3}\ t_6\ t_2\ \boldsymbol{t_5}$
t_5	$t_6\ \boldsymbol{t_4}\ \boldsymbol{t_3}\ t_2\ t_1$	$t_3\ \boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ t_4$	$t_6\ \boldsymbol{t_4}\ \boldsymbol{t_3}\ t_2\ t_1$	$t_3\ \boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ t_4$
t_6	$\boldsymbol{t_5}\ \boldsymbol{t_3}\ t_2\ t_1\ t_4$	$\boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_4}\ t_5\ t_3$	$\boldsymbol{t_5}\ \boldsymbol{t_3}\ t_2\ t_1\ t_4$	$\boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_4}\ t_5\ t_3$

	21-25	26-30	31-35	36-40
t_1	$t_3\ t_2\ \boldsymbol{t_4}\ \boldsymbol{t_6}\ \boldsymbol{t_5}$	$t_4\ t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_2}$	$t_4\ t_2\ \boldsymbol{t_3}\ \boldsymbol{t_6}\ \boldsymbol{t_5}$	$t_3\ t_6\ t_5\ \boldsymbol{t_4}\ \boldsymbol{t_2}$
t_2	$t_4\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ \boldsymbol{t_5}\ t_3$	$t_6\ t_5\ \boldsymbol{t_3}\ \boldsymbol{t_4}\ t_1$	$t_3\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ \boldsymbol{t_5}\ t_4$	$t_6\ t_5\ \boldsymbol{t_4}\ \boldsymbol{t_3}\ t_1$
t_3	$\boldsymbol{t_1}\ t_6\ t_5\ \boldsymbol{t_4}\ \boldsymbol{t_2}$	$\boldsymbol{t_5}\ t_4\ t_2\ t_1\ \boldsymbol{t_6}$	$\boldsymbol{t_2}\ t_5\ t_1\ \boldsymbol{t_4}\ \boldsymbol{t_6}$	$\boldsymbol{t_1}\ t_4\ t_6\ t_2\ \boldsymbol{t_5}$
t_4	$\boldsymbol{t_2}\ t_5\ t_1\ t_3\ \boldsymbol{t_6}$	$\boldsymbol{t_1}\ \boldsymbol{t_3}\ t_6\ t_2\ \boldsymbol{t_5}$	$\boldsymbol{t_1}\ t_6\ t_5\ t_3\ \boldsymbol{t_2}$	$\boldsymbol{t_5}\ \boldsymbol{t_3}\ t_2\ t_1\ \boldsymbol{t_6}$
t_5	$t_6\ \boldsymbol{t_4}\ \boldsymbol{t_3}\ t_2\ t_1$	$t_3\ \boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ t_4$	$t_6\ \boldsymbol{t_3}\ \boldsymbol{t_4}\ t_2\ t_1$	$t_4\ \boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_6}\ t_3$
t_6	$\boldsymbol{t_5}\ \boldsymbol{t_3}\ t_2\ t_1\ t_4$	$\boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_4}\ t_5\ t_3$	$\boldsymbol{t_5}\ \boldsymbol{t_4}\ t_2\ t_1\ t_3$	$\boldsymbol{t_2}\ \boldsymbol{t_1}\ \boldsymbol{t_3}\ t_5\ t_4$

Table 8: A nearly-optimal schedule for the NPB Central League.

As expected, the gap for the Pacific League is worse than that of the Central League, as the six stadiums in the former do not have the nice "straight line" property of the latter (see Figure 3). Nonetheless, a 4.0% difference is surprisingly small, given that our multi-round schedule was generated in just half a second, as compared to the three days it took to generate the optimal schedule.

An Approximation Algorithm

We have solved the LD-TTP for $n = 4$ and $n = 6$, and developed a multi-round generalization of the 6-team LD-TTP. A natural follow-up question is whether our techniques scale for larger values of n. To answer this question, we show that for all $n \equiv 4 \pmod 6$, we can develop a solution to the n-team LD-TTP whose total distance is at most 33% higher than that of the optimal solution, although in practice this optimality gap is actually much lower.

While our construction is *only* a $\frac{4}{3}$-approximation, we note that this ratio is stronger than the currently best-known $(\frac{5}{3} + \epsilon)$-approximation for the general TTP (Yamaguchi et al. 2011). Our schedule is based on an "expander construction", and is completely different from previous approaches that generate approximate TTP solutions. We now describe this construction.

Let m be a positive integer. We first create a single round-robin tournament U on $2m$ teams, and then expand this to a double round-robin tournament T on $n = 6m - 2$ teams.

We use a variation of the Modified Circle Method (Fujiwara et al. 2007) to build U, our single round-robin schedule. Let $\{u_1, u_2, \ldots, u_{2m-1}, x\}$ be the $2m$ teams. Then each team plays $2m - 1$ games, according to this three-part construction:

(a) For $1 \leq k \leq m$, team k plays the other teams in the following order: $\{2m - k + 1, 2m - k + 2, \ldots, 2m - 1, 1, 2, \ldots, k - 1, x, k + 1, k + 2, \ldots, 2m - k\}$.

(b) For $m + 1 \leq k \leq 2m - 1$, team k plays the other teams in the following order: $\{2m - k + 1, 2m - k + 2, \ldots, k - 1, x, k + 1, k + 2, \ldots, 2m - 1, 1, 2, \ldots, 2m - k\}$.

(c) Team x plays the other teams in the following order: $\{1, m + 1, 2, m + 2, \ldots, m - 1, 2m - 1, m\}$.

	1	2	3	4	5	6	7
u_1	$\textcircled{x}$	$\boldsymbol{u_2}$	$\boldsymbol{u_3}$	$\boldsymbol{u_4}$	$\boldsymbol{u_5}$	$\boldsymbol{u_6}$	$\boldsymbol{u_7}$
u_2	u_7	u_1	$\textcircled{x}$	$\boldsymbol{u_3}$	$\boldsymbol{u_4}$	$\boldsymbol{u_5}$	$\boldsymbol{u_6}$
u_3	u_6	u_7	u_1	u_2	$\textcircled{x}$	$\boldsymbol{u_4}$	$\boldsymbol{u_5}$
u_4	u_5	u_6	u_7	u_1	u_2	u_3	$\textcircled{x}$
u_5	$\boldsymbol{u_4}$	$\textcircled{x}$	u_6	u_7	u_1	u_2	u_3
u_6	$\boldsymbol{u_3}$	$\boldsymbol{u_4}$	$\boldsymbol{u_5}$	$\textcircled{x}$	u_7	u_1	u_2
u_7	$\boldsymbol{u_2}$	$\boldsymbol{u_3}$	$\boldsymbol{u_4}$	$\boldsymbol{u_5}$	$\boldsymbol{u_6}$	$\textcircled{x}$	u_1
x	u_1	u_5	u_2	u_6	u_3	u_7	u_4

Table 9: The single round-robin construction for $2m = 8$ teams.

For all games not involving team x, we designate one home team and one road team as follows: for $1 \leq k \leq m$, u_k plays only road games until it meets team x, before finishing the remaining games at home. And for $m + 1 \leq k \leq 2m - 1$, we have the opposite scenario, where u_k plays only home games until it meets team x, before finishing the remaining games on the road. As an example, Table 9 provides this single round-robin schedule for the case $m = 4$.

This construction ensures that for any $1 \leq i, j \leq 2m - 1$, the match between u_i and u_j has exactly one home team and one road team. To verify this, note that u_i is the home team and u_j is the road team iff i occurs before j in the set $\{1, 2m - 1, 2, 2m - 2, \ldots, m - 1, m + 1, m\}$.

Now we "expand" this single round-robin tournament U on $2m$ teams to a double round-robin tournament T on $n = 6m-2$ teams. To accomplish this, we keep x and transform u_k into three teams, $\{t_{3k-2}, t_{3k-1}, t_{3k}\}$, so that the set of teams in T is $\{t_1, t_2, t_3, \ldots, t_{6m-5}, t_{6m-4}, t_{6m-3}, x\}$.

Suppose u_i is the home team in its game against u_j, played in time slot r. Then we expand that time slot in U into *six* time slots in T, namely the slots $6r-5$ to $6r$. We describe the match assignments in Table 10.

	$6r-5$	$6r-4$	$6r-3$	$6r-2$	$6r-1$	$6r$
t_{3i-2}	$\boldsymbol{t_{3j-1}}$	$\boldsymbol{t_{3j}}$	t_{3j-2}	t_{3j-1}	t_{3j}	$\boldsymbol{t_{3j-2}}$
t_{3i-1}	$\boldsymbol{t_{3j}}$	t_{3j-2}	t_{3j-1}	t_{3j}	$\boldsymbol{t_{3j-2}}$	$\boldsymbol{t_{3j-1}}$
t_{3i}	t_{3j-2}	t_{3j-1}	t_{3j}	$\boldsymbol{t_{3j-2}}$	$\boldsymbol{t_{3j-1}}$	$\boldsymbol{t_{3j}}$
t_{3j-2}	$\boldsymbol{t_{3i}}$	$\boldsymbol{t_{3i-1}}$	$\boldsymbol{t_{3i-2}}$	t_{3i}	t_{3i-1}	t_{3i-2}
t_{3j-1}	t_{3i-2}	$\boldsymbol{t_{3i}}$	$\boldsymbol{t_{3i-1}}$	$\boldsymbol{t_{3i-2}}$	t_{3i}	t_{3i-1}
t_{3j}	t_{3i-1}	t_{3i-2}	$\boldsymbol{t_{3i}}$	$\boldsymbol{t_{3i-1}}$	$\boldsymbol{t_{3i-2}}$	t_{3i}

Table 10: Expanding one time slot in U to six time slots in T.

Before proceeding further, let us explain the idea behind this construction. Recall that by the *each-venue* condition, each team in T must visit every opponent's home stadium exactly once, and by the *at-most-three* condition, road trips are at most three games. We will build a tournament that maximizes the number of three-game road trips, and ensure that the majority of these road trips involve three venues closely situated to one another, to minimize total travel. In Table 10 above, if $\{t_{3j-2}, t_{3j-1}, t_{3j}\}$ are located in the same region, then each of the teams in $\{t_{3i-2}, t_{3i-1}, t_{3i}\}$ can play their three road games against these teams in a highly-efficient manner.

We now explain how to expand the time slots in games involving team x. For each $1 \le k \le m$, consider the game between u_k and x. We expand that time slot in U into six time slots in T, as described in Table 11.

	$6r-5$	$6r-4$	$6r-3$	$6r-2$	$6r-1$	$6r$
t_{3k-2}	$\boldsymbol{x}$	$\boldsymbol{t_{3k}}$	t_{3k-1}	x	t_{3k}	$\boldsymbol{t_{3k-1}}$
t_{3k-1}	t_{3k}	$\boldsymbol{x}$	$\boldsymbol{t_{3k-2}}$	$\boldsymbol{t_{3k}}$	x	t_{3k-2}
t_{3k}	$\boldsymbol{t_{3k-1}}$	t_{3k-2}	x	t_{3k-1}	$\boldsymbol{t_{3k-2}}$	$\boldsymbol{x}$
x	t_{3k-2}	t_{3k-1}	$\boldsymbol{t_{3k}}$	$\boldsymbol{t_{3k-2}}$	$\boldsymbol{t_{3k-1}}$	t_{3k}

Table 11: The six time slot expansion for $1 \le k \le m$.

And for each $m+1 \le k \le 2m-1$, consider the game between u_k and x. We expand that time slot in U into six time slots in T, as described in Table 12.

	$6r-5$	$6r-4$	$6r-3$	$6r-2$	$6r-1$	$6r$
t_{3k-2}	x	$\boldsymbol{t_{3k}}$	$\boldsymbol{t_{3k-1}}$	$\boldsymbol{x}$	t_{3k}	t_{3k-1}
t_{3k-1}	$\boldsymbol{t_{3k}}$	x	t_{3k-2}	t_{3k}	$\boldsymbol{x}$	$\boldsymbol{t_{3k-2}}$
t_{3k}	t_{3k-1}	t_{3k-2}	$\boldsymbol{x}$	$\boldsymbol{t_{3k-1}}$	$\boldsymbol{t_{3k-2}}$	x
x	$\boldsymbol{t_{3k-2}}$	$\boldsymbol{t_{3k-1}}$	t_{3k}	t_{3k-2}	t_{3k-1}	$\boldsymbol{t_{3k}}$

Table 12: The six time slot expansion for $m+1 \le k \le 2m-1$.

This construction builds a double round-robin tournament T with $n = 6m-2$ teams and $2n-2 = 12m-6$ time slots. To give an example, Table 13 provides T for the case $m = 2$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
t_1	$\boldsymbol{x}$	$\boldsymbol{t_3}$	t_2	x	t_3	$\boldsymbol{t_2}$	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_4	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_8}$	$\boldsymbol{t_9}$	t_7	t_8	t_9	$\boldsymbol{t_7}$
t_2	t_3	$\boldsymbol{x}$	$\boldsymbol{t_1}$	$\boldsymbol{t_3}$	x	t_1	$\boldsymbol{t_6}$	t_4	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_9}$	t_7	t_8	t_9	$\boldsymbol{t_7}$	$\boldsymbol{t_8}$
t_3	$\boldsymbol{t_2}$	t_1	x	t_2	$\boldsymbol{t_1}$	$\boldsymbol{x}$	t_4	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_7	t_8	t_9	$\boldsymbol{t_7}$	$\boldsymbol{t_8}$	$\boldsymbol{t_9}$
t_4	$\boldsymbol{t_9}$	$\boldsymbol{t_8}$	$\boldsymbol{t_7}$	t_9	t_8	t_7	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3	t_2	t_1	$\boldsymbol{x}$	$\boldsymbol{t_6}$	t_5	x	t_6	$\boldsymbol{t_5}$
t_5	t_7	$\boldsymbol{t_9}$	$\boldsymbol{t_8}$	$\boldsymbol{t_7}$	t_9	t_8	t_1	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3	t_2	t_6	$\boldsymbol{x}$	$\boldsymbol{t_4}$	$\boldsymbol{t_6}$	x	t_4
t_6	t_8	t_7	$\boldsymbol{t_9}$	$\boldsymbol{t_8}$	$\boldsymbol{t_7}$	t_9	t_2	t_1	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3	$\boldsymbol{t_5}$	t_4	x	t_5	$\boldsymbol{t_4}$	$\boldsymbol{x}$
t_7	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_4	t_5	t_6	$\boldsymbol{t_4}$	x	$\boldsymbol{t_9}$	$\boldsymbol{t_8}$	$\boldsymbol{x}$	t_9	t_8	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3	t_2	t_1
t_8	$\boldsymbol{t_6}$	t_4	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_9}$	x	t_7	t_9	$\boldsymbol{x}$	$\boldsymbol{t_7}$	t_1	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3	t_2
t_9	t_4	t_5	t_6	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	$\boldsymbol{t_6}$	t_8	t_7	$\boldsymbol{x}$	$\boldsymbol{t_8}$	$\boldsymbol{t_7}$	x	t_2	t_1	$\boldsymbol{t_3}$	$\boldsymbol{t_2}$	$\boldsymbol{t_1}$	t_3
x	t_1	t_2	$\boldsymbol{t_3}$	$\boldsymbol{t_1}$	$\boldsymbol{t_2}$	t_3	$\boldsymbol{t_7}$	$\boldsymbol{t_8}$	t_9	t_7	t_8	$\boldsymbol{t_9}$	t_4	t_5	$\boldsymbol{t_6}$	$\boldsymbol{t_4}$	$\boldsymbol{t_5}$	t_6

Table 13: The case $m = 2$, producing a 10-team tournament.

It is straightforward to verify that this tournament schedule on $n = 6m-2$ teams is feasible for all $m \ge 1$, i.e., it satisfies the *each-venue*, *at-most-three*, and *no-repeat* conditions. We now show that this expander construction gives a $\frac{4}{3}$-approximation to the LD-TTP, regardless of the values of the distance parameters $d_1, d_2, \ldots, d_{n-1}$.

Let Γ be an n-team instance of the LD-TTP, with $n = 6m-2$ for some $m \ge 1$. Let S be the total distance of the optimal solution of Γ. Using our expander construction, we generate a feasible tournament with total distance less than $\frac{4}{3}S$. This gives a $\frac{4}{3}$-approximation to the LD-TTP.

Let $y_1, y_2, \ldots, y_n$ be the $n = 6m-2$ teams of Γ, in that order, with d_k being the distance from y_k to y_{k+1} for all $1 \le k \le n-1$. Now we map the set $\{t_1, t_2, \ldots, t_{n-1}, x\}$ to $\{y_1, y_2, \ldots, y_n\}$ as follows: $t_i = y_i$ for $1 \le i \le 3m-3$, $x = y_{3m-2}$, and $t_i = y_{i+1}$ for $3m-2 \le i \le 6m-3$. In Figure 4 below, we illustrate this mapping for the case $m = 2$, where the $n = 6m-2$ teams are divided into three triplets and a singleton x:

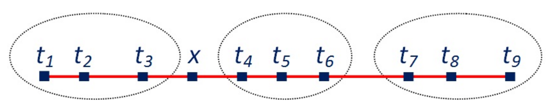

Figure 4: The labeling of the $n = 6m-2$ teams, for $m = 2$.

We then apply this labeling to our expander construction to create a feasible n-team tournament T. For each $1 \le k \le n-1$, let f_k be the total number of times the d_k "bridge" is crossed, so that the total distance of T is $\sum_{k=1}^{n-1} f_k d_k$. We now provide an exact formula for f_k, where we separate the analysis into six cases, depending on the position of x (left or right of the d_k bridge), and the value of k modulo 3.

Position of x	k mod 3	f_k, the value of the d_k coefficient
Right	0	$4k(n-k)/3 + (4n+6k-16)/3$
Right	1	$4k(n-k)/3 + (6n+8k-20)/3$
Right	2	$4k(n-k)/3 + (4n+12k-20)/3$
Left	0	$4k(n-k)/3 + (4n-2k-4)$
Left	1	$4k(n-k)/3 + (8n-4k-22)/3$
Left	2	$4k(n-k)/3 + (14n-10k-16)/3$

Table 14: Formulas for d_k coefficient, for each of the six cases.

We can show (Hoshino and Kawarabayashi 2012) that there are five exceptions to Table 14, as follows:

(a) If $k = 1$ then $f_k := f_k - 2(n-4)/3$.
(b) If $k = \frac{n}{2} - 1$, then $f_k := f_k + 2$.
(c) If $k = \frac{n}{2}$, then $f_k := f_k - 2$.
(d) If $k = \frac{n}{2} + 1$, then $f_k := f_k - 4$.
(e) If $k = n - 1$ then $f_k := f_k - 2(n-4)/3$.

For example, for the case $m = 2$ (see Table 13), the total travel distance of T is $24d_1 + 36d_2 + 42d_3 + 48d_4 + 56d_5 + 52d_6 + 38d_7 + 36d_8 + 26d_9$. Let us prove the formula $f_k = 4k(n-k)/3 + (4n + 6k - 16)/3$ for the first case in Table 14; the remaining cases follow by the same reasoning.

There are k teams to the left of the d_k bridge. By our expander construction, $(k+6)/3$ of these teams cross the bridge $2(n-k+2)/3$ times, and the remaining $(2k-6)/3$ teams cross the bridge $2(n-k+5)/3$ times. And of the $n-k-1$ teams to the right of the bridge (not including team x), $(n-k-1)/3$ of these teams cross the bridge $2k/3$ times and the remaining $2(n-k-1)/3$ teams cross the bridge $(2k+6)/3$ times. Finally, team x crosses the bridge $4k/3$ times. From there, we sum up the cases and determine that $f_k = 4k(n-k)/3 + (4n + 6k - 16)/3$.

Let $S = \sum_{k=1}^{n-1} c_k d_k$ be the total distance of the optimal solution of Γ. Then as we described in the proof of Lemma 1, we have $c_k \geq 2k\lceil\frac{n-k}{3}\rceil + 2(n-k)\lceil\frac{k}{3}\rceil$ because each of the k teams to the left of the d_k bridge must make at least $2\lceil\frac{n-k}{3}\rceil$ trips across the bridge, and the $n-k$ teams to the right of this bridge must make at least $2\lceil\frac{k}{3}\rceil$ trips across.

For $m \geq 3$, it is straightforward to verify that $\frac{f_k}{c_k} < \frac{4}{3}$ for all $1 \leq k \leq n-1$, thus establishing our $\frac{4}{3}$-approximation for the LD-TTP. This ratio of $\frac{4}{3}$ is the best possible due to the case $k = 3$, which has $f_3 = \frac{16n-34}{3}$ and $c_3 = 4n - 8$, implying $\frac{f_3}{c_3} \to \frac{4}{3}$ as $n \to \infty$. This worst-case scenario is achieved when $d_k = 0$ for all $k \neq 3$, i.e., when teams $\{t_1, t_2, t_3\}$ are located at one vertex, and the remaining $n-3$ teams are located at another vertex.

Therefore, 33.3% is the worst possible gap between the optimal solution and the solution produced by our expander construction. In practice, this ratio is much lower, which we demonstrate by applying our construction to five instances: CONSn for $n = 10, 16, 22$ and CIRCn for $n = 10, 16$. The optimal solutions to the first four instances are known (Trick 2012). As we see in Table 15, this percentage gap is extremely small for the constant instances, and is quite reasonable even for the (obviously non-linear) circular instances.

Instance	Optimal	Our Solution	Percentage Gap
CONS10	124	128	3.2%
CONS16	327	334	2.1%
CONS22	626	636	1.6%
CIRC10	242	276	14.0%
CIRC16	$[846, 916]$	994	$[8.5\%, 17.5\%]$

Table 15: Comparing our construction to the optimal solution.

A natural question is whether there exist similar constructions for $n \equiv 0$ and $n \equiv 2 \pmod 6$. In these cases, we ask whether a $\frac{4}{3}$-approximation is best possible. This is just one of many open questions arising from this work. We now conclude the paper with several other avenues for further research.

Conclusion

In many professional sports leagues, teams are divided into two conferences, where teams have *intra*-league games within their own conference as well as *inter*-league games against teams from other conference. The TTP models intra-league tournament play. The NP-complete Bipartite Traveling Tournament Problem (Hoshino and Kawarabayashi 2011a) models inter-league play, and it would be interesting to see whether our linear distance relaxation can also be applied to bipartite instances to help formulate new ideas in inter-league tournament scheduling.

We conclude the paper by proposing two new benchmark instances for the Traveling Tournament Problem, as well as three additional open problems on the Linear Distance TTP. We first begin with the benchmark instances.

For each $n \geq 4$, define LINEn to be the instance where the n teams are located on a straight line, with a distance of one unit separating each pair of adjacent teams, i.e., $d_k = 1$ for all $1 \leq k \leq n-1$. And define INCRn to be the increasing-distance scenario where the n teams are arranged so that $d_k = k$ for all $1 \leq k \leq n-1$. Figure 5 illustrates the location of each team in INCR6.

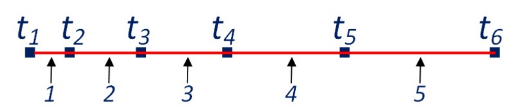

Figure 5: The instance INCR6.

By definition, the TTP solution matches the LD-TTP solution for each of these two instances. By Theorem 1, the optimal solutions for LINE6 and INCR6 have total distance 84 and 250, respectively. This naturally motivates the following problem:

Problem 1 *Solve the TTP for the instances LINEn and INCRn, for $n \geq 8$.*

Theorem 2 listed all seven possible optimal distances for the 6-team LD-TTP, which leads us to ask the following:

Problem 2 *Let PD_n denote the number of possible distances that can be a solution to the n-team LD-TTP. Determine PD_n for $n \geq 8$.*

For example, $PD_4 = 1$ and $PD_6 = 7$. If we can show PD_n is exponential in n, an immediate corollary is the non-existence of a polynomial-time algorithm to solve the n-team LD-TTP.

Finally, for any instance Γ on n teams, define X_Γ to be the total distance of an optimal TTP solution, and X^*_Γ to be the total distance of an optimal LD-TTP solution. Define OG_n to be the *maximum optimality gap*, the largest value of $\frac{X^*-X}{X}$ taken over all instances Γ.

A brute-force enumeration of all 1920 feasible 4-team tournaments, combined with several applications of the Triangle Inequality, shows that $OG_4 = 0\%$. We conjecture that $OG_6 > 0\%$ but have yet to find a 6-team instance with a positive optimality gap. This motivates our final question.

Problem 3 *Determine the value of OG_n for $n \geq 6$.*

Suppose that $OG_6 = 5\%$. Then Theorem 2 guarantees a tournament schedule that is at most 5% higher than the optimal TTP solution, at a fraction of the computational cost. Of course, this is not necessary for the case $n = 6$ as we can use integer and constraint programming to output the TTP solution in a reasonable amount of time. However, for larger values of n, this linear distance relaxation technique would allow us to quickly generate close-to-optimal solutions when the exact optimal total distance is unknown or too difficult computationally. We are hopeful that this approach will help us develop better upper bounds for large unsolved benchmark instances.

Acknowledgements

This research has been partially supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research), the C & C Foundation, the Kayamori Foundation, and the Inoue Research Award for Young Scientists. The authors thank Brett Stevens from Carleton University for suggesting the idea of the Linear Distance TTP during the 2010 Winter Meeting of the Canadian Math Society. Finally, the authors thank the reviewers for their insightful comments that significantly improved the presentation of this paper.

References

Easton, K.; Nemhauser, G.; and Trick, M. 2001. The traveling tournament problem: description and benchmarks. *Proceedings of the 7th International Conference on Principles and Practice of Constraint Programming* 580–584.

Easton, K.; Nemhauser, G.; and Trick, M. 2002. Solving the travelling tournament problem: A combined integer programming and constraint programming approach. *Proceedings of the 4th International Conference on the Practice and Theory of Automated Timetabling* 319–330.

Fujiwara, N.; Imahori, S.; Matsui, T.; and Miyashiro, R. 2007. Constructive algorithms for the constant distance traveling tournament problem. *Lecture Notes in Computer Science* 3867:135–146.

Hoshino, R., and Kawarabayashi, K. 2011a. The inter-league extension of the traveling tournament problem and its application to sports scheduling. *Proceedings of the 25th AAAI Conference on Artificial Intelligence* 977–984.

Hoshino, R., and Kawarabayashi, K. 2011b. The multi-round balanced traveling tournament problem. *Proceedings of the 21st International Conference on Automated Planning and Scheduling (ICAPS)* 106–113.

Hoshino, R., and Kawarabayashi, K. 2011c. A multi-round generalization of the traveling tournament problem and its application to Japanese baseball. *European Journal of Operational Research* 215:481–497.

Hoshino, R., and Kawarabayashi, K. 2012. Generating approximate solutions to the traveling tournament problem using a linear distance relaxation. *Manuscript*.

Irnich, S. 2010. A new branch-and-price algorithm for the traveling tournament problem. *European Journal of Operational Research* 204:218–228.

Kendall, G.; Knust, S.; Ribeiro, C.; and Urrutia, S. 2010. Scheduling in sports: An annotated bibliography. *Computers and Operations Research* 37:1–19.

Thielen, C., and Westphal, S. 2011. Complexity of the traveling tournament problem. *Theoretical Computer Science* 412:345–351.

Trick, M. 2012. Challenge traveling tournament problems. [Online; accessed 2-April-2012].

Yamaguchi, D.; Imahori, S.; Miyashiro, R.; and Matsui, T. 2011. An improved approximation algorithm for the traveling tournament problem. *Annals of Operations Research* 61(4):1077–1091.