



## Linear Programming Solutions to the Linear Ordering Problem in Major American Sports

*Coleman, B.J.*

*Coggin College of Business, University of North Florida*

### Abstract

We examine the efficacy of two linear programming models for optimizing the linear ordering problem (LOP) of ranking teams in major American sports, whether measured by head-to-head win counts or by collective victory margins. Using standard solution software, both models are tested over 42 problems involving full seasons from the National Football League, the National Basketball Association, the National Hockey League, and Major League Baseball, as well as major college football, baseball, and men's basketball. We find that optimal solutions can be achieved rapidly for all four professional sports from either model, regardless of whether victory margins or win counts are used. However, solution speeds were much slower for the collegiate sports, particularly for baseball and men's basketball where the number of teams were the largest, and especially when win counts were the performance measure. Such problems have a large and sparse pairwise comparison matrix with low values and variation in its non-zero elements, characteristics that have been found to be challenging in the general LOP literature. However, a modified minimum violations model demonstrated dramatic efficiency advantages when solving the collegiate cases, and illustrated that optimal solutions are largely achievable in reasonable time frames even for those sports.

**KEYWORDS:** LINEAR ORDERING, RANKING, OPTIMIZATION, LINEAR PROGRAMMING, MINIMUM VIOLATIONS

## Introduction

The linear order problem (LOP) is a classical combinatorial optimization problem in which there exists a square pairwise comparison matrix (PCM) of order  $N$ , containing all possible comparisons of  $N$  items to every other item. Such comparisons can be binary – i.e., item  $i$  is preferred over item  $j$  – or can represent the degree by which  $i$  is preferred over  $j$ . The objective of the LOP is to rank order the items – i.e., the rows and columns in the PCM – such that the sum of the values in the PCM's super-diagonal is maximized, or equivalently, that the sum of the values in its sub-diagonal is minimized. Common non-sports applications of the LOP include triangulating input-output matrices in economics and aggregating individual preferences, but also include problems in task/machine scheduling, archaeology, anthropology, psychology, and graph theory (see citations provided in Grötschel et al. (1984) and Santucci and Ceberio (2020) for examples of each).

The LOP is ubiquitous in sports: entries in the PCM represent the head-to-head game or match results among a sport's participants. Those entries can be the number of head-to-head wins by team  $i$  over team  $j$  in the respective season being examined, or they can represent the collective victory margin of team  $i$  over team  $j$  during the season. Determining an optimal solution to the LOP in sports is equivalent to generating a ranking of the participants. When the PCM entries represent the number of head-to-head wins, and the objective is to minimize the sum of the elements in the PCM's sub-diagonal, solving the LOP is also equivalent to identifying a minimum violations ranking. A game result is violated if the winning team in the game is ranked behind the team it defeated.

Solving the LOP optimally has been determined to be NP-hard (Garey & Johnson, 1979). Thus, while optimization of the LOP using linear programming (LP) or other exact methods is computationally tractable for relatively small problems, most of the LOP literature has focused on developing heuristic solutions. These have included but are not necessarily limited to Tabu search, simulated annealing, variable neighborhood search, iterated local search, scatter search, genetic algorithms, memetic algorithms, ant colony optimization, differential evolution algorithms, and estimation of distribution algorithms (Baiocchi, Milani, and Santucci, 2015; Ceberio, Mendiburu, and Lozano, 2015).

However, optimal solutions are obviously preferred in practice when tractable and would certainly be preferred in sports, particularly if the solution could impact participation in postseason play or the identification of an eventual champion. Moreover, optimal solutions that can be generated using conventional methods such as LP and using standard solution software are likely favored by practitioners as well as researchers due to ease of implementation. Numerous exact methods for identifying optimal solutions to the LOP have been offered (e.g., branch and bound, branch and cut, and cutting planes (Ceberio et al., 2015), as well as constraint relaxation algorithms (Pedings et al., 2012)), but these require specialized coding, algorithms, and/or software that goes beyond that which is commonly available and quickly applicable for potential users. LP models are also likely much more directly modifiable to address secondary, tertiary, or even more objectives when multiple optimality exists (as done in Coleman (2014)), a condition that is common in LOPs (Anderson et al., 2022), including those in sports. Thus, the assessment of LP as a feasible solution methodology for the LOP in sports represents a substantial potential contribution to both academic and practitioner communities.

## ***Solving the LOP in Major American Sports***

Despite the pervasiveness of the LOP in sports, very few published works have presented exact solutions for a full season of games across all participating teams in any of the major American sports. The only known exceptions are Coleman (2005, 2014), Miles, Fowks, and Coulter

(2010), and Chartier et al. (2010). Coleman (2005, 2014) presented mixed binary integer linear programming (BILP) models to determine a minimum violations ranking of major college football teams. The first paper presented results for the 1994 through 2003 seasons, and the second presented results for the 2004 through 2011 seasons for a multi-objective BILP that used the Coleman (2005) model as the basis. Miles et al. (2010) improved the efficiency of the BILP model in Coleman (2005) by adding a set of constraints that reduced the size of the solution space, and presented results for the 2002 and 2008 seasons of major college football. Chartier et al. (2010) presented results from a BILP solved using relaxation techniques for the 2009 season of all divisions of college football, and the 2009-10 seasons of NCAA major college basketball and all-division college basketball. Whereas the PCMs in Coleman (2005, 2014) and Miles et al. (2010) represented win counts, the PCMs addressed by Chartier et al. (2010) generally represented various measures of points scored and/or total or average point differentials.

Coleman (2005) reported his original model to be relatively efficient for win-count LOPs in major college football involving between 107 and 117 teams. Nine of his 10 problems averaged 106 seconds to reach optimality, using LINGO v8.0. However, one case ran for over 10 hours before being terminated. Miles et al. (2010) found solution times to be substantially faster by adding their additional constraint set; they reported that the solution time for the 2002 season (also using LINGO v8.0) was reduced from 36 seconds to 6 seconds. Coleman (2014) reported that a five-level multi-objective mixed BILP solved in an average of 74 seconds using CPLEX v12.2. Chartier et al. (2010) found solutions in under a minute using their BILP and relaxation technique when solving one of the LOPs for major college basketball in 2009-10 (347 teams), and solved a college basketball LOP involving 1041 teams in 1344 minutes.

Occasionally, problems in other sports have been used as test cases for various LOP solution algorithms. For example, a library of 207 problems (Martí, 2010) that has been widely employed in the LOP literature (Lugo et al., 2022) as a test bed for evaluating various solution algorithms includes eight problems comparing varying numbers of players in the Association of Tennis Professionals (ATP) in 1993-94. For the three largest of those cases, involving PCMs with 134, 163, and 452 players, there is still no known optimal solution reported in the literature (Lugo et al., 2022).

That finding is illustrative of the issue of solving the LOP in many sports: the dearth of models and solutions is at least in part due to the size of the problems encountered. Whereas professional sports in the U.S. involve only 30 to 32 teams, problems in American college sports such as basketball can involve 350 or more, with an associated exponential increase in the size of their PCM matrices. Moreover, characteristics other than size that have been found to be particularly challenging in studies of the general LOP are common in college sports. In professional sports, every team plays every other team at least once (in the case of Major League Baseball) or twice (in the NBA and NHL), or plays nearly half of the other teams in the league (i.e., 14 of 31 in the NFL), resulting in PCMs with relatively few zeroes. In contrast, the college sports have much sparser PCMs: teams generally play no more than about 10% of the other available teams. When PCM entries represent win counts, the matrices also have low values and low coefficients of variation.

Those characteristics are problematic. Lugo et al. (2022) have highlighted PCM sparsity as a contributor to LOP complexity, and Schiavinotto and Stützle (2004) identified high sparsity as well as a low coefficient of variation as indicators of problem hardness. Martí, Reinelt, and Duarte (2012) also noted that more narrow ranges of values in the PCM increase solution difficulty. The PCMs in the ATP problems noted above represented net win totals for each player versus every other player, with values ranging from 0 to 4. There were many cases in which two players never played or their win total against the other was identical (thereby yielding a net win

total of zero for each player), resulting in a very sparse PCM. In addition to size, these characteristics are part of the reason why the three largest versions of those problems have never yet been solved. They also contribute to the challenge of solving the LOP in American major college sports, which contain three contributors to solution difficulty – problem size, PCM sparsity, and low variation in PCM entries if/when they represent win counts – that make such problems particularly complex to solve.

Despite the challenges, there is interest in the sports community for LOP solutions, as well as evidence for the lack of LOP solutions for major sports. Kenneth Massey, one of the contributors to the amalgamated ranking that was used by the former Bowl Championship Series during 1998 through 2013, regularly compiles a composite of numerous rankings in college football, basketball, and baseball (Massey, 2025a, 2025b, 2025c). As part of those composites, he also computes each ranking's violation percentage, essentially representing how well that ranking solves the LOP in that sport. The Massey composite for major college football has frequently included a minimum-violations ranking (optimal LOP solution) from Coleman (Massey, 2025c). The computed violation percentage from that ranking thereby represents the optimal violations comparative for every other ranking in the composite. That the composites for sports other than football have not similarly included a minimum violations ranking is indicative of the lack of LOP solutions being computed in other sports, and/or the perceived onerous challenges associated with deriving one. It is that condition that partly prompts the current research.

### **Research Questions**

The current research seeks to evaluate the efficacy of using LP and standard solution software to optimize the LOP in seven major American sports: the National Football League (NFL), the National Basketball Association (NBA), Major League Baseball (MLB), the National Hockey League (NHL), the Football Bowl Subdivision of NCAA college football (CFB), NCAA Division I men's college basketball (MBB), and NCAA Division I college baseball (CBB). Other than the prior studies cited above, no known prior literature has addressed the LOP for any of these sports, and no existing studies have assessed the efficacy of achieving optimal solutions using LP and standard software. Moreover, even the prior studies that have addressed some of these problems were performed more than a decade ago, meaning that solution efficacy could now be quite different using current hardware and software.

In addition, no prior literature has identified problem characteristics that differentiate LOP solution efficiency across the same set of sports. The current research seeks to address that gap as well. In doing so, it also partially addresses a call in the general LOP literature (Lugo et al., 2022) for further study of the impact of sparsity on problem complexity and solution efficiency.

Finally, the current research seeks to assess the performance differences of two different LP formulations of the LOP. One formulation is the classical formulation that has been widely used and/or referenced in prior studies of LOP solution methods. The other is a modified version of the minimum violations formulation used in Coleman (2005, 2014) and Miles et al. (2010), which – other than in those studies – has not been applied to any other LOP within or outside of sports. It has also not been compared to the classical model over any set of problems, sports or otherwise. In doing so, the current research not only represents a contribution to quantitative analysis in sports, but also potentially to the body of knowledge for the general LOP.

## Linear Programming Formulations

### Classical Formulation

The classical linear programming (CLP) model for generating optimal solutions to the LOP is a pure binary integer program (Martí et al., 2012) where:

$$\begin{aligned} X_{ij} &= 1 \text{ if item } i \text{ is ranked above item } j, 0 \text{ otherwise,} \\ C_{ij} &= \text{value of paired comparison between item } i \text{ and } j. \end{aligned}$$

CLP is then formulated as

$$\begin{aligned} \text{MAX } & \sum C_{ij} X_{ij} \\ & \forall i \neq j \end{aligned} \quad (1)$$

$$\begin{aligned} \text{ST } & \\ & X_{ij} + X_{ji} = 1 \quad \forall i, \forall j; i < j \quad (2) \\ & X_{ij} + X_{jk} + X_{ki} \leq 2 \quad \forall i, \forall j, \forall k; i < j, i < k, j \neq k \quad (3) \\ & X_{ij} \in \{0,1\} \quad \forall i \neq j \quad (4) \end{aligned}$$

Very few alternate LP models have been offered in the literature. A recent exception is Dupin (2022), who presents six alternate formulations based on modifications of models for the asymmetric travelling salesman problem. Dupin examines their solution times using CPLEX v12.4 over 14 sets of 30 randomly generated problems of size 20, 30, 40, 50, and 100, and finds that none of the formulations represent more than a slight improvement over the CLP.

### Minimum Violations Formulation

However, the literature also includes the formulation from Coleman (2005, 2014) for minimizing violations of past game results, denoted as MinV. MinV is a mixed binary linear program where:

$$\begin{aligned} V_{ij} &= 1 \text{ if preferred item } i \text{ is ranked below non-preferred item } j, 0 \text{ otherwise;} \\ T_i &= \text{value of item } i; \\ M &= \text{arbitrarily large value;} \\ D &= \text{minimum distance between the values of two items necessary to avoid a violation.} \end{aligned}$$

MinV is then formulated as

$$\begin{aligned} \text{MIN } & \sum C_{ij} V_{ij} \\ & \forall C_{ij} > 0 \end{aligned} \quad (5)$$

$$\begin{aligned} \text{ST } & \\ & T_i - T_j + M V_{ij} \geq D \quad \forall C_{ij} > 0 \quad (6) \\ & 0 \leq T_i \leq ND \quad \forall i \quad (7) \\ & V_{ij} \in \{0,1\} \quad \forall C_{ij} > 0 \quad (8) \\ & T_i \in [0, N-1] \text{ and integer} \quad \forall i \quad (9) \end{aligned}$$

The initial version of MinV in Coleman (2005) included only (5), (6), and (8) and used  $M=1000$ , with  $D$  being somewhat arbitrarily chosen. Values of  $T_i$  were allowed to be continuous. The later version in Coleman (2014) also used  $M=1000$  and set  $D=1$ , and added (7) as an adoption of the approach presented in Miles et al. (2010). Miles et al. reported that adding (7) with an upper bound equal to  $ND$  substantially reduced solution times by essentially creating “slots” or ranking positions for each of the  $N$  items to be ranked.

## Modifications to Minimum Violations Formulation

Two additional modifications to the MinV in Coleman (2014) were added in the research reported herein. Both were consistent with the concept from Miles et al. (2010) to create a limited number of slots or ranking positions and thereby reduce the solution space. Moreover, both were consistent with one of the models from Dupin (2022) that is based on the Miller, Tucker, and Zemlin (1960) formulation of the asymmetric traveling salesman problem. In addition to the binary variables  $X_{ij}$  used in the CLP, Dupin employs a variable representing the position of the item in the ranking. By nature of the constraints in Dupin's model, this variable takes on integer values between 0 and  $N-1$ . In the modified version of MinV assessed below,  $T_i$  was adjusted to be analogous to Dupin's ranking position variable using the following two adjustments.  $M$  in constraint (6) was set equal to  $N$  as opposed to an arbitrarily large number, with (as in Coleman (2014))  $D$  set to 1 in constraints (6) and (7). Also, constraint (9) was added to restrict values of  $T_i$  in (9) to be general integer between 0 and  $N-1$ .

An advantage of MinV is that it requires a constraint only for the non-zero paired comparisons in the PCM. Any empty (zero) cells in the PCM are ignored. In contrast, CLP requires constraints to account for all pairwise comparisons in (2) and all transitives in (3), regardless of whether those comparisons are zero in the PCM.

## Methods

Three recent full seasons of game results for the NFL, the NBA, MLB, the NHL, CFB, MBB and CBB were acquired from Massey (2024). Across the three seasons examined, the number of teams was 30 in the NBA and MLB, 32 in the NFL and NHL, between 130 and 133 in CFB, between 301 and 305 in CBB, and between 358 and 363 in MBB.

Two versions of each problem were generated: one in which the PCM represented the number of head-to-head wins by team  $i$  over team  $j$  in the respective season, and one in which the PCM represented the collective victory margin of team  $i$  over team  $j$  during the season. This allowed for an examination of the impact of the size and variation in the non-zero PCM values as mitigating factors in performance. The first, using win counts of each team versus every other team, resulted in PCMs with low non-zero values with small variation, whereas the second, using victory margins, resulted in PCMs with nearly identical sparsity but with larger non-zero values with larger variations.

This approach generated a total of 42 example problems for the analysis. In all cases, each PCM was converted to its normal form (Martí et al., 2012) where all entries were not only integral but also nonnegative with  $\min\{C_{ij}, C_{ji}\} = 0$ .

All problems were solved in CPLEX Version 22.1.1 using CLP as well as MinV. CPLEX default parameters were applied. Each model was solved to optimality or to the best available solution after five minutes of processing on a 13<sup>th</sup> generation Intel Core i9-13900K system operating at 3000Mhz with 24 cores, 32 logical processors, and 64GB of RAM. To make solutions directly comparable across the two models, the objective function of CLP was switched to a minimization.

Three performance metrics were computed: (1) whether a problem was solved to optimality within five minutes of run time, (2) the mean number of seconds of run time necessary to reach optimality if the optimal solution was reached in under five minutes, and (3) the mean gap between the model solution and the optimal solution.

In order to identify the impact of key problem characteristics on model performance, also computed for each problem were the degree of sparsity (i.e., percentage of non-zero elements

among the non-diagonals) and the mean and standard deviation of the non-zero elements in the PCM.

## Results

Results are shown in Tables 1 and 2. Optimal solutions were identified by MinV within five minutes for 38 of the 42 test cases. For the remaining four problems, MinV was run without a time limit to determine the optimal solutions (or in the case of MBB in 2024 in Table 2, the best solution once an out-of-memory condition was reached) that were used as comparatives. The MinV solution times required in those cases are shown in bold italics in the table, as is the solution at the point of the out-of-memory condition. By comparison, CLP solved only 27 of the 42 cases within five minutes.

Table 1. Solution results for MinV and CLP for seven major American sports, using collective victory margins; Opt = optimal objective function value; Secs = seconds of solution time; Gap = gap between solution and optimal solution.

| Sport    | Teams | Opt  | Zeros | Non-Zeros |       | MinV         |       | CLP  |      |
|----------|-------|------|-------|-----------|-------|--------------|-------|------|------|
|          |       |      |       | Mean      | SD    | Secs         | Gap   | Secs | Gap  |
| CBB 2022 | 301   | 1977 | 96.9% | 9.91      | 9.00  | 36.8         | -     |      | 568% |
| CBB 2023 | 305   | 2153 | 96.8% | 9.85      | 8.26  | 140.2        | -     |      | 549% |
| CBB 2024 | 305   | 2165 | 96.8% | 9.68      | 8.29  | 29.2         | -     |      | 510% |
| CFB 2021 | 130   | 483  | 96.2% | 16.75     | 13.12 | 0.6          | -     | 14.7 | -    |
| CFB 2022 | 131   | 584  | 96.2% | 15.65     | 12.99 | 0.8          | -     | 13.6 | -    |
| CFB 2023 | 133   | 526  | 96.3% | 16.23     | 12.71 | 0.9          | -     | 15.2 | -    |
| MBB 2022 | 358   | 3161 | 97.3% | 15.24     | 11.95 | 141.9        | -     |      | 661% |
| MBB 2023 | 363   | 3584 | 97.2% | 14.94     | 11.88 | <b>770.7</b> | 0.22% |      | 608% |
| MBB 2024 | 362   | 3666 | 97.2% | 15.16     | 12.06 | 203.0        | -     |      | 619% |
| MLB 2021 | 30    | 366  | 71.7% | 12.95     | 12.88 | 1.0          | -     | 0.1  | -    |
| MLB 2022 | 30    | 304  | 70.9% | 12.22     | 11.21 | 0.7          | -     | 0.1  | -    |
| MLB 2023 | 30    | 561  | 55.7% | 9.61      | 8.54  | 1.3          | -     | 0.1  | -    |
| NBA 2022 | 30    | 1070 | 54.4% | 24.53     | 19.97 | 1.0          | -     | 0.1  | -    |
| NBA 2023 | 30    | 1212 | 54.6% | 21.15     | 17.88 | 1.4          | -     | 0.1  | -    |
| NBA 2024 | 30    | 904  | 54.1% | 25.28     | 20.63 | 1.2          | -     | 0.1  | -    |
| NFL 2021 | 32    | 349  | 80.7% | 14.38     | 12.82 | 0.8          | -     | 0.1  | -    |
| NFL 2022 | 32    | 229  | 80.5% | 10.81     | 9.68  | 0.6          | -     | 0.1  | -    |
| NFL 2023 | 32    | 279  | 80.0% | 12.49     | 11.46 | 0.6          | -     | 0.1  | -    |
| NHL 2022 | 32    | 260  | 57.9% | 4.26      | 2.88  | 1.2          | -     | 0.1  | -    |
| NHL 2023 | 32    | 259  | 58.2% | 4.24      | 2.75  | 1.8          | -     | 0.2  | -    |
| NHL 2024 | 32    | 290  | 59.5% | 4.16      | 2.81  | 1.6          | -     | 0.1  | -    |

Table 2. Solution results for MinV and CLP for seven major American sports, using win counts; Opt = optimal objective function value; Secs = seconds of solution time; Gap = gap between solution and optimal solution.

| Sport    | Teams | Opt        | Zeros | Non-Zeros |      | MinV           |       | CLP  |      |
|----------|-------|------------|-------|-----------|------|----------------|-------|------|------|
|          |       |            |       | Mean      | SD   | Secs           | Gap   | Secs | Gap  |
| CBB 2022 | 301   | 465        | 97.2% | 1.73      | 1.01 | 102.2          | -     |      | 356% |
| CBB 2023 | 305   | 488        | 97.0% | 1.70      | 0.98 | 252.2          | -     |      | 345% |
| CBB 2024 | 305   | 499        | 97.0% | 1.65      | 0.95 | 67.3           | -     |      | 324% |
| CFB 2021 | 130   | 72         | 96.2% | 1.01      | 0.07 | 0.7            | -     |      | 226% |
| CFB 2022 | 131   | 84         | 96.3% | 1.00      | 0.05 | 2.2            | -     |      | 220% |
| CFB 2023 | 133   | 74         | 96.3% | 1.00      | 0.05 | 2.0            | -     |      | 238% |
| MBB 2022 | 358   | 347        | 97.7% | 1.28      | 0.51 | <b>2,166</b>   | -     |      | 421% |
| MBB 2023 | 363   | 417        | 97.6% | 1.27      | 0.50 | <b>31,453</b>  | 5.76% |      | 346% |
| MBB 2024 | 362   | <b>451</b> | 97.6% | 1.25      | 0.47 | <b>230,815</b> | 7.78% |      | 330% |
| MLB 2021 | 30    | 91         | 74.3% | 3.00      | 2.48 | 1.0            | -     | 0.1  | -    |
| MLB 2022 | 30    | 95         | 71.0% | 2.84      | 2.34 | 0.8            | -     | 0.1  | -    |
| MLB 2023 | 30    | 149        | 57.6% | 2.17      | 1.60 | 1.1            | -     | 0.1  | -    |
| NBA 2022 | 30    | 63         | 70.1% | 2.20      | 0.91 | 0.6            | -     | 0.1  | -    |
| NBA 2023 | 30    | 78         | 71.7% | 2.20      | 0.92 | 0.8            | -     | 0.1  | -    |
| NBA 2024 | 30    | 61         | 71.7% | 2.25      | 0.99 | 0.6            | -     | 0.1  | -    |
| NFL 2021 | 32    | 37         | 82.1% | 1.17      | 0.38 | 0.7            | -     | 0.1  | -    |
| NFL 2022 | 32    | 28         | 82.9% | 1.13      | 0.36 | 0.5            | -     | 0.1  | -    |
| NFL 2023 | 32    | 38         | 82.6% | 1.14      | 0.34 | 0.6            | -     | 0.1  | -    |
| NHL 2022 | 32    | 105        | 65.9% | 1.97      | 0.82 | 1.0            | -     | 0.1  | -    |
| NHL 2023 | 32    | 99         | 67.5% | 1.93      | 0.83 | 1.1            | -     | 0.1  | -    |
| NHL 2024 | 32    | 114        | 69.0% | 1.97      | 0.84 | 1.2            | -     | 0.2  | -    |

Both models solved all 12 professional sports problems very rapidly, regardless of whether win count or victory margin was the outcome measure in the PCM. CLP solved the win count cases within 0.1 seconds on average, with MinV taking an average of 0.8 seconds. Performance across the victory margin cases was similar to the win count cases: CLP required 0.1 seconds and MinV required 1.1 on average. In sum, CLP was the better model for smaller PCMs with relatively low sparsity and higher non-zero values and variation, although the differences in solution efficiency were *de minimis*.

However, for the much larger and sparser college examples, the performance advantage for MinV was dramatic. For CFB, MinV reached optimality within an average of 1.6 seconds (for win counts) and 0.8 seconds (for victory margins). In contrast, CLP failed to reach optimality on any of the win-count PCMs within five minutes, with solutions at the five-minute mark that were 228% higher than optimal on average. However, CLP performance was much better when victory margins were used, where optimality was reached within 15 seconds. Thus, the larger and more variable non-zero values in the PCM greatly mitigated the sparsity issue for CLP.

However, the larger number of teams in CBB and MBB – more than double the size of CFB – created an even wider performance gap. Using victory margin, MinV optimally solved all three CBB problems in an average of 69 seconds. CLP failed to reach optimality for any of the three

victory margin cases, with solutions at the five-minute mark that were over 542% larger than optimal. Similar results were found for MBB: MinV solved two of the three cases within five minutes, taking an average of 173 seconds on those two, and the solution for the third was within 0.22% of optimal at the five-minute mark. CLP failed to solve any of the three, with solutions after five minutes that averaged over 629% larger than optimal.

Performance differences were also wide when using win counts for CBB and MBB. The lower values and variation in the PCM associated with win counts also challenged both models more substantially. Whereas MinV solved all three CBB win-count problems in an average of 141 seconds, it failed to solve any of the three MBB examples in five minutes. However, its MBB solutions after five minutes averaged within 4.5% of optimal. CLP failed to identify optimal solutions for any of the six CBB and MBB cases, and its solutions after five minutes were 353% larger than optimal.

In sum, MinV performed substantially better than CLP on the large and sparse PCMs found in college baseball and basketball, and particularly so when the PCM entries represented small counts with tight variation.

## Discussion

LP was found to be efficient at determining optimal solutions to the LOP for the large majority of the seven major American sports examined, even when using standard solution software. Problems in the four major American professional sports were rapidly solved using either the classical model of the LOP or the modified minimum violations formulation MinV. For problems in college sports, with much larger numbers of items to be ranked and sparse pairwise comparison matrices, solution speeds were much slower and/or the performances of the two models were much more differentiated. This was particularly true when the non-zero entries in the PCM had low values and low variation, which is the case when entries represent win counts. However, even for most of those problems, MinV was able to reach optimality in well under five minutes of run time.

The MinV model solved problems in college football within about two seconds regardless of whether victory margins or win counts were used in the PCM. In college baseball, where the number of teams was more than double that of college football, MinV identified an optimal solution within an average of about 105 seconds. In men's college basketball, where the number of teams was the largest, MinV reached optimality within five minutes for two of the three victory-margin problems and was within 0.22% of optimal on the third. For the three win-count cases, its solution after five minutes averaged within 4.5% of optimal.

In contrast, the classical formulation failed to reach optimality within five minutes for 15 of the 18 college sports cases, with solutions at the five-minute mark that averaged 411% over optimal.

The MinV performance was also dramatically improved over the results reported in prior literature for college football. The mean solution time for the win-count cases – for problems averaging 131 teams – was more than 98% lower than that reported in Coleman (2005) for problems averaging 113 teams, even if the one 10-hour solution time in Coleman (2005) is ignored. As a direct comparison to both Coleman (2005) and Miles et al. (2010), the 2002 season – the only one with reported solution times in both articles – was also solved using MinV. MinV required 0.69 seconds to solve the 117-team CFB problem in 2002, as compared to 38 seconds in Coleman (2005) and six seconds in Miles et al. (2010). The one 10-hour case in Coleman (2005), for the 2000 season, was also solved using MinV, and required only 0.81 seconds.

## Conclusion

We find that optimal solutions via linear programming and standard solution software are largely achievable in reasonable time frames for the linear ordering problem in major American sports. Of the two formulations examined, MinV appears to be the strongly preferred option, particularly when addressing problems in collegiate sports. Its superiority versus the classical formulation in addressing the types of large and sparse paired comparison matrices found in college sports has not been illustrated by prior literature, and suggests that it might also be a preferred option in non-sports applications with similar characteristics.

## References

- Anderson, P.E., Chartier, T.P., Langville, A.N., & Pedings-Behling, K.E. (2022). Fairness and the set of optimal rankings for the linear ordering problem. *Optimization and Engineering* 23, 1289-1317. <https://doi.org/10.1007/s11081-021-09650-y>
- Baiocchi, M., Milani, A., & Santucci, V. (2015). Linear ordering optimization with a combinatorial differential evolution. In *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2135-2140). Hong Kong. <https://doi.org/10.1109/SMC.2015.373>
- Ceberio, J., Mendiburu, A., & Lozano, J.A. (2015). The linear ordering problem revisited. *European Journal of Operational Research* 241, 686-696. <https://doi.org/10.1016/j.ejor.2014.09.041>
- Chartier, T.P., Kreutzer, E., Langville, A.N., Pedings, K., & Yamamoto, Y. (2010). Minimum violations sports ranking using evolutionary optimization and binary integer linear program approaches. In A. Bedford & M. Owens (eds.), *Proceedings of the Tenth Australian Conference on Mathematics and Computers in Sport* (pp. 13-20). New South Wales, Australia. [https://www.academia.edu/75423491/Minimum\\_Violations\\_Sports\\_Ranking\\_Using\\_Evolutionary\\_Optimization\\_and\\_Binary\\_Integer\\_Linear\\_Program\\_Approaches#outer\\_page\\_7](https://www.academia.edu/75423491/Minimum_Violations_Sports_Ranking_Using_Evolutionary_Optimization_and_Binary_Integer_Linear_Program_Approaches#outer_page_7)
- Coleman, B.J. (2005). Minimizing game score violations in college football rankings. *Interfaces* 35(6), 483-496. <https://doi.org/10.1287/inte.1050.0172>
- Coleman, B.J. (2014). Minimum violations and predictive meta-rankings for college football. *Naval Research Logistics* 61(1), 17-33. <https://doi.org/10.1002/nav.21563>
- Dupin, N. (2022). *Integer linear programming reformulations for the linear ordering problem*. HAL open science. hal-03607145, <https://hal.science/hal-03607145v1/document>
- Garey, M.R., & Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W.H. Freeman & Co. <https://perso.limos.fr/~palafour/PAPERS/PDF/Garey-Johnson79.pdf>
- Grötschel, M., Jünger, M., & Reinelt, G. (1984). A cutting plane algorithm for the linear ordering problem. *Operations Research* 32(6), 1195-1220, <https://doi.org/10.1287/opre.32.6.1195>
- Lugo, L., Segura, C., & Miranda, G. (2022). A diversity-aware memetic algorithm for the linear ordering problem. *Memetic Computing* 14, 395-409, <https://doi.org/10.1007/s12293-022-00378-5>
- Martí, R. (2010). lolib\_2010. Retrieved August 27, 2024, from [https://www.dropbox.com/s/fk105g63jmi3i1d/lolib\\_2010.zip?dl=0](https://www.dropbox.com/s/fk105g63jmi3i1d/lolib_2010.zip?dl=0)
- Martí, R., Reinelt, G., & Duarte, A. (2012). A benchmark library and a comparison of heuristic methods for the linear ordering problem. *Computational Optimization and Applications* 51, 1297-1317, <https://doi.org/10.1007/s10589-010-9384-9>

- Massey, K. (2024) Massey ratings - scores and schedule data archive. Retrieved August 27, 2024, from <https://masseyratings.com/data>
- Massey, K. (2025a) Massey ratings – college baseball rankings. Retrieved July 1, 2025, from <https://masseyratings.com/ranks?s=cbase&top=-1>
- Massey, K. (2025b) Massey ratings – college basketball rankings. Retrieved July 1, 2025, from <https://masseyratings.com/ranks?s=cb&top=-1>
- Massey, K. (2025c) Massey ratings – college football rankings. Retrieved July 1, 2025, from <https://masseyratings.com/ranks?s=cf&d=20250121&top=-1>
- Miles, W.W., Fowks, G.T., & Coulter, L.O. (2010). AccuV college football ranking model. *Journal of Quantitative Analysis in Sports* 6(3), <https://doi.org/10.2202/1559-0410.1189>
- Miller, C., Tucker, A., & Zemlin, R (1960). Integer programming formulation of traveling salesman problems, *Journal of the ACM* 7(4), 326-329, <https://doi.org/10.1145/321043.321046>
- Pedings, K.E., Langville, A.N., & Yamamoto, Y.A. (2012). A minimum violations ranking method. *Optimization and Engineering* 13, 349-370, <https://doi.org/10.1007/s11081-011-9135-5>
- Santucci, V., & Ceberio, J. (2020). Using pairwise precedences for solving the linear ordering problem, *Applied Soft Computing* 87. <https://doi.org/10.1016/j.asoc.2019.105998>
- Schiavinotto, T., & Stützle, T. (2004). The linear ordering problem: instances, search space analysis and algorithms. *Journal of Mathematical Modeling and Algorithms* 3, 367-402. <https://doi.org/10.1023/B:JMMA.0000049426.06305.d8>