Baseball, Optimization and the World Wide Web

Ilan Adler [1] Alan L. Erera [2] Dorit S. Hochbaum [3] Eli V. Olinick [4]

July 19, 1998

[1]Research supported in part by ONR contract N00014-91-J-1241.
[2]Research supported in part by ONR contract N00014-91-J-1241.
[3]Research supported in part by ONR contract N00014-91-J-1241 and by NSF award No. DMI-9713482
[4]Research supported in part by ONR contract N00014-91-J-1241 and by NSF award No. DMI-9713482

Abstract

The competition for baseball playoff spots—the fabled "pennant race"— is one of the most closely-watched American sports traditions. Baseball fans, famous for their love of statistics, check newspapers (and now, web sites) daily looking for measures on their team's progress (or lack thereof!). While informative, traditionally reported statistics such as *games back* and the "magic number" are overly conservative and ignore the remaining schedule of games. By using simple optimization techniques one can model these schedule effects explicitly and determine when a team has locked up a playoff spot or is truly "mathematically eliminated" from contention. We describe how we used optimization and the Internet to develop a popular web site to provide automatic daily updates of new, improved playoff race statistics during the 1996, 1997 and current baseball seasons, and how this site fits into a larger operations research web project, the Berkeley Remote Interactive Optimization Testbed.

SF CHRON CLIPPING HERE

Figure 1: SF Chronicle declares Giants eliminated from first place on 9/10/96.

1 Introduction

Fans of professional sports teams are known for their insatiable desire for information about the performance of their favorite teams. This phenomenon is particularly true for fans of Major League Baseball (MLB) in the United States. Of primary concern for many is their teams' prospects for reaching the post-season playoffs. The competition for baseball playoff spots—the fabled "pennant race"—is one of the most closely-watched American sports traditions. Fans check newspapers (and now, web sites) daily for updates on their team's progress (or lack thereof!).

The media typically report a number of statistical measures for teams in the MLB playoff hunt to help fans sort out the playoff picture. *Games back* and *magic number* are two common statistics. While informative, these statistics are calculated on a pairwise basis that ignores the remaining schedule of games. Similar myopic calculations are used to determine when a particular team no longer has any chance of finishing first. A team is declared to be "mathematically eliminated" if it does not have enough games remaining to attain the number of wins currently held by the first place team; thus, even if the trailing team were to win the rest of its games and the first place team were not to win again, the trailing team would finish no better than second. For example, this is the situation described in the article shown in Figure 1 from the September 10, 1996 edition of the San Francisco Chronicle. Despite the headline, however, the

San Francisco Giants actually had been eliminated three days earlier: the news of the Giants unfortunate demise was first announced on September 8 on our World Wide Web (WWW) site, shown in Figure 2 below.

Figure 2: INSERT GIANTS ELIMINATION HTML

Figure 2: RIOT Web site declares Giants eliminated from first place on 9/8/96.

The traditional method for determining when a trailing team is eliminated is overly conservative. The deficiency of the method is simple to explain: it ignores the important fact that each time the first place team (or any team for that matter) loses a game, some other team *must* win. If the first place team loses all of its remaining games, for example, it may "give" enough wins to another team (or teams) such that that team must finish with a better record than the trailing team under consideration. In this case, the trailing team cannot finish in first place even if it were to win all the games it has left to play. By using simple optimization techniques, however, one can model these schedule effects explicitly and determine when a team is truly mathematically eliminated. The project described in this paper represents an attempt to provide optimization-based playoff race statistics to the general public.

Playoff race elimination problems can be useful for teaching basic operations research ideas. Elimination problems are covered in textbooks by Schrage [10] and Ahuja, Magnanti and Orlin [1]. Robinson [9] argues that many students can relate to the problem subject intuitively, and find the results interesting. We agree, and furthermore suggest that the Internet can be used to present the problem and broadcast its results to the public to attract the interest of both students of operations research, as well as people who otherwise might never be exposed to optimization concepts. By constructing a web site containing not only daily updates of the improved statistics but also the details behind their calculation, we hope to fulfill an educational goal in addition to simply providing entertaining information.

In this paper we describe how we used optimization and the Internet to develop a web site to provide daily updates of improved playoff race statistics during the 1996 and 1997 baseball seasons, and how this site fits into a larger operations research web project, the Berkeley Remote Interactive Optimization Testbed (RIOT). Up and running during the baseball season, the site can be accessed at `http://riot.ieor.berkeley.edu/~baseball`. After the site was listed in the Yahoo! internet directory late in the 1996 season, computer-savvy baseball fans were immediately attracted to the pages. During September, when the pennant race is most heated, the pages attracted 100-200 hits each day. As further testament to the popularity of the pages, a public radio program called *Beyond Computers* featured the site on a recent program dedicated to the evolving role of technology in sports.

In Section 2, we discuss baseball elimination and clinching problems and how standard optimization techniques can be used to provide a better picture of how the playoff race is shaping up. We introduce in this section the set of new measures of a baseball team's prospects for post-season play that we provide on the web site. In Section 3, we discuss how these statistics are generated automatically each day of the baseball season. In Section 4 we show how a single linear program (LP) can be used to determine which, if any, of the teams in a division have been eliminated from finishing atop the division. This LP is also used to used to specify the minimum number of games that any trailing team must win to have any chance of finishing in first place. We also treat the problem of elimination from the playoffs (in addition to first place), which is more complicated because of the "wildcard" possibility. In Section 5 we consider the problem of determining the minimum number of games a team must win to ensure

itself (clinch) a playoff spot. Finally, we make some brief conclusions about our site specifically and our general approach to using optimization on the Web.

2 Problem description

Realizing that many readers may not be aficionados of America's "national pastime," we begin by reviewing the Major League Baseball playoff structure. The MLB teams are partitioned into two leagues, the American and National. Each league is further subdivided into three divisions. Four teams make the playoffs from each of the two leagues: the three teams which finish with the best records in their respective divisions, and a fourth team, the wildcard, with the best record among all second place teams in the league. Ties in the final standings for either first place in a division or the wildcard spot are settled by special one-game playoffs. Each league then conducts a playoff tournament with its four invited teams to determine its "pennant winner." Finally, the American League pennant winner and the National League pennant winner play in the World Series for the MLB championship.

As the baseball regular season progresses, followers of the sport are naturally interested in their teams' chances of earning a spot in the playoffs. Specifically, fans would like to know when a team has "clinched" a playoff spot or when a team has been eliminated from the playoffs. We first consider the problem of determining when a team has been eliminated.

The Elias Sports Bureau is the official statistician for Major League Baseball and it determines elimination from first place in the division using a simple wins-based comparison. If a team trails the first place team in the wins column by more games than that team has remaining, it is eliminated. On September 10, 1996 the Giants had a record of 59 wins and 83 losses (a 59-83 record) with 20 games left. On the same day, the first place team in the National League (NL) West division, the San Diego Padres, had a record of 80 wins and 65 losses with 17 games left to play. Even if the Giants were to win all of their remaining games, they would finish with at best a 79-83 record while the Padres' worst possible record would be 80-82. Since the Giants were doomed to finish behind the Padres regardless of the outcome of any of the remaining games in the season, they were officially declared to be eliminated from first place. It has long been known in the optimization community, however, that this wins-based criterion is sufficient to eliminate from first place, but not necessary (Schwartz [11]).

The problem with the wins-based measure is that it ignores the fact that other teams may influence the race. For example, consider the standings for the NL West on September 8, 1996 shown in Figure 2 above. Glancing at the standings, it appears that San Francisco had a remote chance of catching the first place Dodgers and Padres. If Los Angeles and San Diego each were to lose enough games while San Francisco won enough times, the Giants conceivably could have moved into first place. Considering the remaining schedule, however, reveals that even a miracle winning streak would not have enabled the Giants to finish in first place. At that point in the season, Los Angeles and San Diego were slated to play each other seven more times. Since there are no ties in baseball, one of these teams would win at least four of these games and finish with a record of 82-80, or better. This means that the Giants were eliminated from first place, since they could achieve at best a .500 record of 81 and 81. Note that the best case scenario for a trailing team may not include the first place team losing the rest of its games!

Determining that a particular team has been eliminated can often be much more difficult than this example. A more complicated elimination scenario involving all the teams in the American League East is shown in Figure 3 below. Although it is not apparent from the

standings, it is possible to show that Detroit has been eliminated by using the information in Figure 4. An inspired reader may try the proof as an exercise; details are given in the Appendix A.

It turns out that a simple maximum flow calculation on a small network can be used to determine precisely when a team has been necessarily eliminated from first place (Schwartz [11], Hoffman and Rivlin [5]). Robinson [9] showed that such an optimization approach would have eliminated teams an average of three days earlier than the wins-based criterion in the 1987 season. Despite these results, sportswriters continue to use the incorrect wins-based method.

As noted in [9], the first place elimination problem has a long history in the optimization literature. Schwartz [11] may have been the first to observe that a team can be eliminated from first place earlier than the traditional method indicates and presented the first network formulation to determine first place elimination at the earliest possible date. Hoffman and Rivlin [5] extended this work to show necessary and sufficient conditions for eliminating a team from k^{th} place. McCormick [6, 7] in turned showed that this more general problem of determining elimination from k^{th} is $\mathcal{NP}$-complete.

Gusfield, Martel and Fernandez-Baca [4] showed that determining when a team is eliminated from first place can be solved as a maximum flow problem on a bipartite network. Gusfield and Martel [3] showed that the minimum number of games a given team must win in order to avoid elimination from first place can found by solving a parametric maximum flow problem. By extending a result of Gallo, Grigoriadis and Tarjan [2] and using a binary search procedure, they

Detroit Elimination HTML

Figure 3: Can the Tigers win the pennant?

give a running time of $O(n^3 + n^2 log(nD))$, where n is the number of teams and D the number of games the team of interest has left to play, for finding this number. Recently, McCormick [7] has improved the time bound for solving this parametric maximum flow problem to $O(n^3)$.

Prior to the 1994 season, there were only two divisions in each league and only the two division winners in each league advanced to the playoffs. The addition of the wildcard playoff position complicates the process of determining whether a team is eliminated from playoff contention. Mathematically, the problem is similar to the k^{th} place elimination problem. Robinson discussed the complications introduced by wildcard berths in the context of applying his baseball elimination model to the National Football League (NFL). For various reasons, however, Robinson did not provide a formulation with wildcard teams. First, the NFL uses a procedure to break ties in the standings that is complex and rules-based. Further, professional football teams play far fewer games than Major League Baseball teams, so the combinatorial nature of the elimination problem in the NFL is relatively trivial when compared to baseball. In fact, many newspaper sports sections publish a complete enumeration of the possible NFL playoff scenarios as the end of the season approaches.

Motivated by the idea in [3], we decided that it would be interesting to create a baseball standings web site that not only allows fans to determine accurately whether or not teams have been eliminated from the playoffs, but also gives them an idea of how *close* a team is to elimination. Clearly, as the minimum number of games that a team needs to win in order to avoid elimination approaches the number of games that team has remaining, elimination

Remaining series in AL East HTML

Figure 4: Number of games remaining between teams in the AL East.

becomes imminent. Following [3], define the **first place elimination number** as the minimum number of remaining games that a team must win in order to have any chance of finishing in first place in its division. In other words, if the team fails to win this number of additional games, it cannot come in first place. Additionally, define the **playoff elimination number** as the minimum number games a team must win to have any shot at making the playoffs, whether as a division winner or as the wildcard team. The optimization techniques used to calculate these numbers will be described in Section 4.

Often, baseball fans are more interested in determining whether a team has clinched first place in its division or the wildcard playoff spot. A clinch is a guarantee; after a team has clinched first place, for example, it could lose the rest of its games without jeopardizing its first place position. Currently, the media use "magic numbers" to determine first place clinches, and typically only report a magic number for the team currently holding first place. Assume teams are ranked in ascending order of losses, and suppose the first-place team has l_1 losses and g_1 games remaining, and the second-place team has l_2 losses. The magic number, μ, is defined as $g_1 - (l_2 - l_1)$. Any combination of wins by the first-place team and losses by the second-place team totaling μ guarantees the first-place team at least a tie for first place. Of course, when the first place team's magic number drops to zero, the team has clinched the top spot in the division.

Unlike the case with elimination problems, the schedule of remaining games has little effect (mathematically) on a team's ability to clinch first place. However, although magic numbers give necessary and sufficient conditions for clinching, they suffer from two drawbacks: (1) they do not specify the minimum number of future wins necessary by a team to clinch first place independent of other teams' performance, and (2) they are typically only reported for teams in first place in their respective divisions. Addressing these concerns, we define the **first place clinch number** for each team as the minimum number of games which, if won, guarantee that the team finishes in at least a tie for first place. In Section 5, we describe how we calculate these numbers for each team without relying on optimization. The final interesting statistic that we provide on the web site is the **playoff clinch number** for each team. This number is defined as the minimum number of games which, if won, guarantee that a team finishes in a playoff position, whether as the division winner or as the wildcard team.

3 The baseball playoff races web site

The Giants elimination example presented in Figure 2 was taken from the baseball standings web site that we created as a component of the Berkeley RIOT Internet project. RIOT is a collection of web sites developed and maintained by the Industrial Engineering and Operations Research Department and Haas School of Business at the University of California, Berkeley. To date, the primary focus of the project has been to provide educational information about industrial engineering and operations research, and promote interest in the field via web pages and easy-to-use, interactive Java applets. The RIOT site showcases applications of methods to solve both general and specific problems, such as the planning of mining operations, scheduling, network design, geometric algorithms for manufacturing purposes, clustering and pattern recognition, and the Traveling Salesman Problem. Each RIOT application is implemented with details of the underlying methods used in the problem solution; once visitors have "played" with the application and discovered its utility, they can learn about the optimization models and algorithms used to produce the results.

One of the RIOT's missions is to provide interesting on-line applications of operations research that can be used by students and researchers in the field. However, another goal of the site is to attract those who may be unfamiliar with optimization methods. To do so, we wanted to include applications that appeal to the general Internet audience. The popularity of sports-related web sites such as the ones maintained by ESPN and the National Basketball Association (NBA) speaks to the fact that sports fans love to "surf" the web in search of information, including up-to-date statistics. Baseball fans in particular are famous for their affinity for numbers, so we decided that a playoff elimination application would be of general interest to many baseball fans and could help attract a large audience to the RIOT web site.

Creating the playoff elimination web pages required a two-phased effort. In order to provide the most informative and appealing site to fans, we decided that it was necessary that we extend earlier theoretical work by defining the new optimization-based statistical measures of a team's playoff prospects, described in the previous section. Once the optimization models needed to generate the new statistics were designed, we implemented an automated software system that uses the models to produce daily updates of the measures. The automated system runs in the early morning hours each day of the baseball season, creating and posting on the web an updated standings report which includes the elimination and clinch numbers for each team, along with information traditionally included in newspaper sports sections such as wins, losses and games back.

The system works as follows. Since the calculations require the current win-loss records of each team, and the remaining number of games between teams, the software system updates its database using the results of the previous day's games. A free internet news service called Infobeat (`www.infobeat.com`) automatically sends us e-mail each night containing the final scores of all MLB games. At 2 a.m. Pacific time (long after the e-mail usually arrives), the update process begins by reading the latest results from the e-mail and updating our MLB standings database. This small database is simply a file which contains each team's win-loss record and the number of games remaining between each pair of teams. Next, the program uses the database to generate the mathematical programs described in Sections 4 and 5, and invokes the CPLEX optimization package (`www.cplex.com`) to solve the simple linear and integer formulations. Finally, the optimization results are used to determine each team's elimination and clinching numbers, and updated standings pages are generated in the hypertext markup language (HTML) format used by WWW browser programs (e.g. Netscape Communicator and Microsoft Internet Explorer).

Occasionally, the daily e-mail arrives later than expected and we have to manually issue the command to execute the update program. Aside from that, however, the process described above requires very little human intervention. The update software is implemented on a Sun Microsystems SPARCstation 20 and typically takes about ten minutes to complete its various tasks.

The bulk of the update software is written in the Perl programming language. Perl is specifically designed for writing Unix script programs and it is also very well suited for string manipulation. The update procedure requires parsing text files, such as our nightly email baseball scores and CPLEX output, to extract the relevant information which is then combined with text and graphical images to produce our baseball standings web pages. These type of operations are much easier to code in Perl than languages like C or C++.

Work began on the site in the middle of the 1996 season and one of the more challenging parts of the project turned out to be determining the number of games remaining between

each pair of teams. Although there is a wealth of MLB data available on the web, we found it difficult to find the information we needed in a readily-usable format. For our purposes, we would have liked to find a table or matrix giving the number of games left between pairs of teams. The most common format for this type of information, however, is an actual schedule of games which lists the games to be played for each day of the season. In order to find out how many games were left between say the Boston Red Sox and New York Yankees, we had to read through the list and count the number of times Boston was to play New York. Although this task was fairly easy to automate, discrepancies and inaccuracies in the schedules posted on the web made this a difficult chore.

Canceled and suspended games complicated this process. It is not uncommon for a baseball game to be suspended before it is completed, and then restarted at a later date. This occurs frequently in the early part of the season due to weather conditions. Some web sites report the partial results of these games when they are suspended and then update the final score once the game is completed. For example, if we start generating the remaining schedule of games from July 17 on, we might miss a game that was originally scheduled for May 18 but postponed until August 15 due to rain. Depending on the policies of the particular web site posting the schedule, that game might be listed on May 18, August 15 or both dates and the fact that the game was not completed may or may not be reported. Another wrinkle we encountered was the way "doubleheaders" (pairs of games between the two teams played back-to-back on the same day) were listed on various sites. We initially generated schedule information from multiple sources to ensure accuracy; these inconsistencies in data presentation made it difficult to reconcile the data created from different web pages. In fairness to these web sites, we should point out that this information is provided free of charge and keeping track of the 2,268 MLB games played each season is by no means a trivial job.

For the 1997 season, the task was much simpler since all of our software was ready at the start of the season. We did, however, have to contend with the introduction of interleague play—games between teams in different leagues. When we originally designed the system, we anticipated the possibility of interleague play in the future; therefore, the database and update program required only minor modifications. In 1998, two new teams, the Arizona Diamondbacks and Tampa Bay Devil Rays, joined Major League Baseball. To accommodate these new teams, the Milwaukee Brewers switched from the American to the National League and the Detroit Tigers moved from the American League East to the American League Central Division. As with interleague play, we easily adapted to these changes. As Major League Baseball continues to expand, however, we may have to make more significant changes to keep up. One possibility is that two additional teams will join MLB in the next two years and there will be a so called "realignment" of the teams into two leagues with two eight team divisions. If this happens it is likely that a second wild card team from each league will be added to the playoffs and new mathematical programming formulations for the wildcard elimination and clinch numbers will be needed.

4 Mathematics of playoff elimination

In this section, we describe how first place and playoff elimination statistics for each team in a division are calculated. The formulations described allow first place elimination numbers to be calculated using a single linear program per division. Further, we extend the model to handle the wildcard possibility when determining the playoff elimination number for each team. We

show that these statistics can be generated by solving at most two small mixed-integer programs per division. Importantly for the formulations which follow, one should note that professional baseball games cannot end in ties—each game is played until one team is victorious. Also, by league rules each team is required to play its entire schedule of games each season—there are no canceled games (at least for teams entering the post-season).

4.1 First place elimination

Consider a single division of teams, say the American League East. The first place elimination problem (FPE) can be posed as follows: given the current win-loss records of each team and the remaining schedule of games, determine the first place elimination number for each team, as previously defined. Note that we consider a team to be not eliminated if it can finish in a tie for the best record in the division.

Using the parametric maximum flow formulation of Gusfield and Martel [3] or the linear program proposed by Robinson [9] could have allowed us to determine a full set of first place elimination numbers by solving a new formulation for each team. To reduce computation, however, we formulate a single linear program (of the approximate size of Robinson's) to determine the numbers for each team in a division.

Let N be the set of teams in the division of concern. For each team i in this division , let w_i represent its total number of current wins, g_{ij} the number of games remaining to be played against team j in the division, and t_i the number of games remaining against non-division opponents. Let x_{ij} be the decision variable representing the number of future games team i wins against team j in the division; let x represent an entire scenario of future wins, $x = \{x_{ij} \mid i, j \in N\}$.

Consider first the following integer linear program:

(P1) First Place Elimination Integer Formulation

$$
\begin{array}{lll}
\min v & & \\
\text{subject to} & & \\
(1) \quad & x_{ij} + x_{ji} = g_{ij} & \forall\, i, j \in N,\ i < j \\
(2) \quad & v \geq w_i + \sum_{j \in N} x_{ij} & \forall\, i \in N \\
(3) \quad & x_{ij} \geq 0 & \forall\, i, j \in N \\
(4) \quad & v \text{ integer} & \\
(5) \quad & x_{ij} \text{ integer} & \forall\, i, j \in N
\end{array}
$$

The problem (P1) is guaranteed to have an optimal solution since it is feasible and bounded. Let us denote the optimal value of v as v^*. We now claim that v^* represents the minimum number of wins that any team finishing in first place must have at season's end, and that any team that can attain v^* wins has a chance of finishing in first:

Theorem 4.1 *Suppose v^* is optimal for (P1). Then, team $i \in N$ is eliminated from first place contention if, and only if:*

$$w_i + t_i + \sum_{j \in N} g_{ij} \ < \ v^* \tag{1}$$

The proof is straightforward. It follows directly from the formulation that v^* is the minimum number of games *any* team must win to finish atop the division. Since the left-hand side of the above is the maximum number of possible wins for team i given that it wins the rest of its games, the sufficiency of the condition is clear. To establish necessity, it can be shown that if team ℓ can attain $v^\ell \geq v^*$ wins by season's end, there is always a scenario in which it finishes

atop the division. See Appendix B for details. ∎

Using Theorem 4.1, we can determine whether or not any team is eliminated at any given point during the season. Since Major League Baseball consists of six divisions, we need to solve six problems of type (P1) to determine the appropriate v^* for each division. We can use these optimal values v^* further, however, to determine first place elimination numbers for each team that is not eliminated. This result is summarized in the following corollary to Theorem 4.1:

Corollary 1 (First Place Elimination Number) *Given the optimal solution of (P1) for a particular division N, the first place elimination number for each team $i \in N$ is given by*

$$v^* - w_i \tag{2}$$

Since v^* represents the minimum number of games that team i must win to have a chance of finishing first in the division, and since w_i is the current number of wins for team i, it is clear that the team needs to win a minimum of $v^* - w_i$ games over the remainder of the season to have a chance of finishing in first. ∎

4.2 Solving first place elimination via linear programming

The formulation (P1) can be recast as an equivalent mathematical program with a single integer variable. The optimal objective function value of this simplified integer program can then be determined by solving the linear relaxation of the problem and rounding.

To reduce (P1) to a formulation with a single integer variable, we take advantage of the special structure of the constraint coefficient matrix of (P1), which we denote A. Suppose that we treat variable v as a fixed integer constant. Moving the column of A corresponding to v

to the right-hand side of the constraint system leaves a smaller constraint matrix, A', which is totally unimodular. This can be verified by noting that A' now represents a node-edge incidence matrix of an undirected bipartite graph, which is known to be totally unimodular (see, e.g., Nemhauser and Wolsey [8]). Sincethe right-hand side of the constraint system is integer, the total unimodularity of A' guarantees that basic solutions x for every fixed integer v must be integer. Consider, therefore, the following reformulation of (P1) with a single integer variable:

(P2) First Place Elimination Single-Integer Formulation

min v

subject to

(1)	$x_{ij} + x_{ji} = g_{ij}$	$\forall\, i, j \in N,\ i < j$
(2)	$v \geq w_i + \sum_{j \in N} x_{ij}$	$\forall\, i \in N$
(3)	$x_{ij} \geq 0$	$\forall\, i, j \in N$
(4)	v integer	

Let $\bar{v}$ be the optimal objective function value for (P2). We now claim that $v^* = \bar{v}$. Since $\bar{v}$ is optimal for (P2), it is clearly integer. For a fixed, integer v, the constraint matrix of (P2) is totally unimodular. Therefore, there must exist *some* basic optimal solution, $\bar{x}'$, that is integer. Since (P2) is a relaxation of (P1), $v^* \geq \bar{v}$. But $\bar{v}, \bar{x}'$ is feasible for (P1). Since $\bar{v}$ is a lower bound on the optimal objective function for (P1), $\bar{v}, \bar{x}'$ is optimal for (P1) and $v^* = \bar{v}$.

Finally, we show that problem (P2) can be solved using a rounding procedure on a linear program. Consider the linear programming relaxation of (P2), removing constraint (4). Let $\hat{v}$ be the optimal objective value of this relaxation. First, it is clear that since we are minimizing,

if we round $\hat{v}$ up to the nearest integer, no optimal integer objective could be better than this value. Suppose that $\hat{x}$ is an optimal x vector for the relaxed problem. It is clear that a solution of $\lceil \hat{v} \rceil$, $\hat{x}$ remains feasible for the linear relaxation of (P2). Since no value of the objective function can be lower, $\lceil \hat{v} \rceil$, $\hat{x}$ is optimal for (P2). Therefore $v^* = \bar{v} = \lceil \hat{v} \rceil$, thus v^* can be determined by solving a linear program, then rounding.

4.3 Playoff Elimination Problems

Elimination from the playoffs only occurs when a team has no chance of either (1) finishing in first place in its division, or (2) finishing with the best record among all second place teams in its league. Following the framework presented earlier, there are again two goals: (1) to determine which teams have been eliminated from the playoffs, and (2) to calculate the minimum number of games a team must win to have any chance of finishing the season with a playoff berth. The models presented in this section do not specify what type of playoff berth a team earns. Sometimes circumstances arise where a team is eliminated from the wildcard berth, but *not* eliminated from first place in its division. This may happen, for example, if the teams with the two best records in the league are in the same division.

Determining playoff elimination status with wildcard teams is inherently more difficult than simple first place elimination. The problem to be solved is similar to the problem of elimination from k^{th} place in a division of teams, which McCormick [6] showed to be NP-complete. Each of the problems solved is also slightly larger, since we are now concerned with the performance of all of the teams in the league. It is a fairly simple task to formulate an integer program to determine a single team's playoff elimination number. In the spirit of the previous section, however, we attempt to avoid solving a separate formulation for each team.

Let S be the set of teams in the league (American or National). Suppose further that the league consists of three divisions, and let N^k be the set of teams in division k, $k = 1, 2, 3$. We will again use g_{ij} and x_{ij} to model games remaining and future wins, but now we define these parameters and variables for all pairs of teams $i, j \in S$, whereas in the formulations of Section 4.1 we only modeled within a single division. Finally, let M be a large integer (specifically, greater than the number of games in a season, 162).

Consider the following integer linear program which we hope will provide playoff elimination numbers for all teams in a single division, say division "1":

(P3) Playoff Elimination Formulation

min v_1

subject to

(1)	$x_{ij} + x_{ji} = g_{ij}$	$\forall\ i, j \in S,\ i < j$
(2)	$v_1 \geq w_i + \sum_{j \in S} x_{ij} - M\beta$	$\forall\ i \in N^1$
(3)	$v_1 \geq w_i + \sum_{j \in S} x_{ij} - M(1-\beta) - M\alpha_i^k$	$\forall\ k \in \{1,2,3\},\ i \in N^k$
(4)	$\sum_{i \in N^k} \alpha_i^k \leq 1$	$\forall\ k \in \{1,2,3\}$
(5)	$x_{ij} \geq 0$	$\forall\ i, j \in S$
(6)	x_{ij} integer	$\forall\ i, j \in S$
(7)	v_1 integer	
(8)	$\beta,\ \alpha_i^k$ binary	$\forall\ k \in \{1,2,3\},\ i \in N^k$

Denote the optimal objective function value of (P3) as $\bar{v}_1$. We will now show that $\bar{v}_1$ represents the minimum number of wins that any team in this division must attain to have a chance of reaching the post-season. Suppose $\bar{\beta} = 0$. In this case, (P3) reduces to (P1) and $\bar{v}_1$ represents the number of wins necessary and sufficient to win the division and make the playoffs. Now suppose $\bar{\beta} = 1$. Constraints (3) and (4) guarantee that at most one team (with $\bar{\alpha}_i^k = 1$) finishes with more wins than $\bar{v}_1$ in each division. Hence, $\bar{v}_1$ represents the minimum number of wins necessary to be a wildcard team from this division.

Consider now the following theorem:

Theorem 4.2 *Suppose $\bar{v}_1$ is the optimal objective function value of (P3) for some division N^1, and let $\bar{\beta}$ be the optimal value of β. If $\bar{\beta} = 1$, and some team in N^1 finishes with strictly more wins than $\bar{v}_1$, denote that team p, else $p = \emptyset$.*

Then, team $\ell \in N^1 \setminus \{p\}$ is eliminated from the playoffs if, and only if:

$$w_\ell + \sum_{j \in S} g_{\ell j} \ < \ \bar{v}_1 \tag{3}$$

The sufficiency of condition 3 is clear for all teams, including team p, from the formulation (P3). Again, the left side of the condition is the total potential wins that team ℓ may attain, while the right side is the minimum number needed to secure a playoff berth. If a team's potential is smaller than the minimum, it is eliminated. To prove the necessity of the condition, we simply show that each team $\ell \in N^1 \setminus \{p\}$ that satisfies the condition can finish with a playoff berth. This proof is provided in Appendix B. ■

Formulation (P3) therefore provides playoff elimination information for each team in a division, with a single possible exception. Since the exception team p finishes with more wins (denoted $\bar{v}_1^p$) than $\bar{v}_1$ in the optimal solution, there is no proof that p winning $\bar{v}_1$ games will result in a scenario where p makes the playoffs. We can, of course, solve another simple integer program to determine the minimum number of wins, $\hat{v}_1^p$, that team p needs. Note that since team p does reach the playoffs in the solution scenario to (P3), $\bar{v}_1 \leq \hat{v}_1^p \leq \bar{v}_1^p$. The extra formulation for team p, called (P3p) is identical to (P3), with the added stipulation that v_1 is set equal to $w_p + \sum_{j \in S} x_{pj}$. (We could also bound v_1 as just explained.) This formulation is included in Appendix C for completeness.

If we solve formulation (P3), and if necessary (P3p), for each division in baseball, twelve integer program solutions provide all the information necessary to construct playoff elimination information for each team:

Corollary 2 (Playoff Elimination Number) *Given the optimal solution of (P3), and the optimal solution to (P3p) for the exception team p given in Theorem 4.2, the playoff elimination number for each team $i \in N$ is given by*

$$\bar{v}_1 - w_i \qquad \forall\, i \neq p \tag{4}$$

$$\hat{v}_1^p - w_p \qquad \textit{for team } p \tag{5}$$

As a final note, if we fix values of v_1, β, and α, the constraint matrix of (P3) (and (P3p)) reduces to a totally-unimodular form. Therefore, by the same reasoning presented in the previous section, we do not need to constrain x to be integer, and we can remove constraint (6) from the formulation yielding a problem with an equivalent optimal objective value. We still have a binary integer programming problem, however. If we cannot solve an integer program, recognize that since the number of teams in any league is usually small (approximately 15), we can solve (P3) by fixing all possible values for β and α. For every feasible combination of values for β and α, we then solve a linear program with rounding as described in Section 4.2.

Proceeding in this manner, we would need to solve $1 + \prod_{k=1}^{3} |N^k|$ linear programs to determine the optimal objective function for each formulation of type (P3) In practice, of course, it is easier to solve one small integer program than all of these linear programs.

5 Clinching problems

Clinching problems are the complements of elimination problems. In this section, we describe how first place and playoff clinch statistics are calculated. First, consider the problem of determining a team's first place clinch number. For a team k to clinch first place, it must win enough games to guarantee that it finish with a record at least as good as all other teams. However, nothing prevents any other team from winning all of its remaining games except perhaps games against team k. Thus, first place clinch numbers can be calculated without using optimization, as we now describe.

Consider the set of all teams, N, in a specific division. Let g_i be the total number of games remaining against all opponents for team $i \in N$ (by earlier notation, $g_i = t_i + \sum_{j \in N} g_{ij}$). Also, define g_i^j as the number of games team i has remaining against teams other than team j, $g_i^j = g_i - g_{ij}$. We will now describe how to determine two statistics: θ_{ij}, the smallest number of wins by team i that guarantees at least a tie with team j, and θ_i, the smallest number of wins that guarantees a tie with every team in N. We establish the results in the following theorem:

Theorem 5.1 (First Place Clinch Number) *Consider team $i \in N$:*

1. *If $w_j + g_j^i - w_i - g_i > 0$, then it is impossible for team i to guarantee a finish equal or better to team $j \in N$. Let $\theta_{ij} = \infty$ in this case.*

2. If $w_j + g_j^i - w_i - g_i \leq 0$, then calculate θ_{ij} as follows:

$$\theta_{ij} = \max\left\{0, \min(g_i^j + \left\lceil \frac{1}{2}(w_j + g_j - w_i - g_i^j) \right\rceil, w_j + g_j - w_i)\right\} \tag{6}$$

3. Given the θ_{ij} for all teams $j \in N$, calculate θ_i, the first place clinch number, as follows:

$$\theta_i = \max_{j \in N} \theta_{ij} \tag{7}$$

noting that $\theta_i = \infty$ means i cannot guarantee a clinch with any number of wins, given the present circumstances.

Determining θ_{ij} is straightforward. Let f_i represent the number of future games won by team i. First, if team j can win more games against teams other than i such that its win total at season end is greater than the potential win total of team i, then we know that there is no way for i to guarantee at least a tie with j. If this is not the case, then we must use the formula in the second part of the proposition.

Assume the worst case scenario for wins by team i, i.e. let as many of its future wins come against teams other than j. In the case where no future wins of i come against team j, we know that team must win enough games such that $w_i + f_i = w_j + g_j$, or rewriting, $f_i = w_j + g_j - w_i$. If $f_i > g_i^j$, however, some of these future wins *must* come against team j. In this case, each win beyond g_i^j effectively counts as two wins, since it also means that j has lost a game. Therefore, if i must win more than g_i^j additional games, it only needs to win $g_i^j + \lceil \frac{1}{2}(w_j + g_j - w_i - g_i^j) \rceil$ to guarantee that it finishes with a record as good as j's. To guarantee a first place finish, team i must ensure a tie with every team in the division, thus justifying the equation for θ_i. ■

Determining how many games are needed to clinch a playoff spot is more complicated. We model this problem with a mathematical programming formulation for each team similar to the playoff elimination problems presented in Section 4. However, we "reverse" the formulations: instead of determining the minimum number of games the team must win to finish in a playoff position, we determine the maximum number of games the team can win **without** finishing in a playoff position. To do so, we constrain the assignment of wins such that the team of concern finishes with fewer wins than the first place team in its division, and such that at least one division contains two teams with better records. Details of the formulation are presented in Appendix D. Once we have determined the maximum number of games a team can win and still not finish in a playoff position, we add one to determine the playoff clinch number.

6 Conclusions

The Internet web site developed for broadcasting improved baseball playoff race statistics exploits both old and new optimization methods to provide fans with more precise information. Using the Internet as a public forum, we are able to disseminate the improved information without relying on traditional media to "accept" the ideas and modify the information they normally provide to fans. Furthermore, the site provides detailed information about how the calculations are performed, including an on-line copy of this paper for interested individuals. In this way, the pages fulfill one of the goals of the RIOT web site: to educate members of the on-line community about various optimization techniques through the use of interesting "real-world" problems.

The Internet can be thought of as a large, distributed, public-use database. Optimization models can be used to "add value" to data: by converting unwieldy amounts of data into a usable form, such as an optimal decision or an interesting statistic, the value of the data is increased. Since more and more data is becoming available on-line, the potential for more meaningful "value-adding" activity will increase. On the RIOT site, we have begun to explore this avenue further with the development of an on-line investment portfolio design system. Using a database that automatically tracks the daily closing prices of nearly one hundred stocks, the system allows users to solve the Markowitz portfolio optimization model. Both the baseball and portfolio systems should give researchers and practitioners both a glimpse at the type of opportunities that exist to use operations research to add value to on-line data.

Appendices

A Proving Detroit's elimination

It is not apparent from the standings in Figure 3 that Detroit has been eliminated. If we look at the schedule of remaining games in the AL East (see in Figure 4) and try to construct a scenario where the Tigers win the division, however, we find that it is impossible. By winning all of its remaining games, Detroit can finish the season with a record of 76 and 86. If the Yankees win just 2 more games, then they will finish the season with a 77 and 85 record, which would put them ahead of Detroit.

We now analyze scenarios where the Yankees win one or no remaining games. First, suppose that the Tigers go undefeated for the rest of the season and the Yankees fail to win another game. The problem with this scenario is that New York still has 8 games left with Boston. If the Red Sox win all of these games, they will end the season with at least 77 wins putting them

ahead of the Tigers. Thus, the only way for Detroit to even have a chance of finishing in first place, is for New York to win exactly one of the 8 games with Boston and lose all of its other games. Meanwhile, the Sox must lose all the games they play against teams other than New York. This would create a 3-way tie for first place as shown in Figure 5 below.

Now consider the Orioles and Blue Jays in our scenario. Baltimore has 2 games left with Boston and 3 with New York. If everything happens as described above, the Orioles will finish with at least 76 wins. Detroit can catch Baltimore only if the Orioles lose all of their games to teams other than New York and Boston. In particular, this means that Baltimore must lose all 7 of its remaining games with Toronto. The Blue Jays also have 7 games left with the Yankees and we have already seen that for Detroit to finish in first place, Toronto must will all of these games. But if that happens, the Blue Jays will win at least 14 more games giving them a final record of 77 and 85 or better which means they will finish ahead of the Tigers. Therefore, Detroit is mathematically eliminated.

Team	W	L	GB
Detroit	76	86	-
Boston	76	86	-
New York	76	86	-
Baltimore	To be determined	To be determined	To be determined
Toronto	To be determined	To be determined	To be determined

Figure 5: A scenario where Detroit finishes in a tie for first place.

B Elimination proofs

B.1 Proof of condition necessity for Theorem 1

To prove that a team is eliminated only if the maximum wins the team can achieve is less than v^*, we show that any team that can finish with more wins may finish on top of the division. Consider some team ℓ in the division of concern. Suppose that ℓ can finish with some number of wins v^ℓ where:

$$v_\ell \;=\; w_\ell + t_\ell + \sum_{j \in N} g_{\ell j} \;\geq v^* \tag{8}$$

We now show that there is always a feasible assignment of wins among the teams such that they do not finish ahead of team ℓ. Consider again (P1) and an optimal solution, v^*, x^*. It is simple to now construct a solution x^ℓ such that no teams finish ahead of team ℓ given that team ℓ wins the rest of its games:

$$x^\ell_{ij} = x^*_{ij} \qquad \forall\, i,j,\; i \neq \ell,\; j \neq \ell \tag{9}$$

$$x^\ell_{\ell j} = g_{\ell j} \qquad \forall\, j \tag{10}$$

$$x^\ell_{j\ell} = 0 \qquad \forall\, j \tag{11}$$

The constraints partitioning the wins of the remaining games remain satisfied. To show that ℓ finishes with the most wins:

$$\begin{array}{rlll} & v^* \geq w_i + \sum_{j \in N} x^*_{ij} & \forall\, i & \text{from (P1)} \\ \Longrightarrow & v^\ell \geq w_i + \sum_{j \in N} x^*_{ij} & \forall\, i & \text{since } v^\ell > v^* \\ \Longrightarrow & v^\ell \geq w_i + \sum_{j \in N} x^\ell_{ij} & \forall\, i & \text{by definition} \end{array}$$

where the last relationship holds since $x^\ell_{ij} \leq x^*_{ij}$ for every $i \neq \ell$.

Hence, a scenario partitioning the wins of the remaining games always exists such that team ℓ finishes in first, given that $v^\ell \geq v^*$. Thus, a team is eliminated from first only if $v^\ell < v^*$.

B.2 Proof of condition necessity for Theorem 4.2

Consider each team in the division, $\ell \in N^1 \setminus \{p\}$. Suppose team ℓ can win all of its remaining games and finish with v^ℓ wins, where $v^\ell \geq \bar{v}_1$. For the necessity of the condition to hold, there must always exist a scenario in which team ℓ makes the playoffs. Following the methodology in Appendix B.1, consider the following system that, given it has a solution $x^\ell, \beta^\ell, \alpha^\ell$, guarantees that team ℓ makes the playoffs:

$$\sum_{j \in S} x^\ell_{ij} - M\beta^\ell \leq v^\ell - w_i \qquad \forall\, i \in N^1 \tag{12}$$

$$\sum_{j \in S} x^\ell_{ij} - M(1-\beta^\ell) - M\alpha^{k\ell}_i \leq v^\ell - w_i \qquad \forall\, k \in \{1,2,3\},\; i \in N^k \tag{13}$$

$$x^\ell_{ij} + x^\ell_{ji} = g_{ij} \qquad \forall\, i,j \tag{14}$$

$$\sum_{i \in N^k} \alpha^{k\ell}_i \leq 1 \qquad \forall\, k \in \{1,2,3\} \tag{15}$$

$$\beta^\ell, \alpha^{k\ell}_i \quad \text{binary} \tag{16}$$

Suppose $\bar{x}, \bar{\beta}, \bar{\alpha}$ represents an optimal solution to (P3). Modifying this optimal solution such that team ℓ wins the rest of its games will satisfy the above system 12 through 16:

$$x^{\ell}_{ij} = \bar{x}_{ij} \qquad \forall\, i, j,\ i \neq \ell,\ j \neq \ell \tag{17}$$

$$x^{\ell}_{\ell j} = g_{\ell j} \qquad \forall\, j \tag{18}$$

$$x^{\ell}_{j\ell} = 0 \qquad \forall\, j \tag{19}$$

$$\alpha^{k\ell}_{i} = \bar{\alpha}^{k}_{i} \qquad \forall\, k \in \{1,2,3\},\ i \in N^k \tag{20}$$

$$\beta^{\ell} = \bar{\beta} \tag{21}$$

If $\bar{\beta} = 0$, 12 through 16 reduces to the already proven case in Appendix B.1. If $\bar{\beta} = 1$, a similar simple proof can be constructed. Since a scenario exists where any team $\ell \in N^1 \setminus \{p\}$ meeting the condition makes the playoffs, it is clear that $\bar{v}_1$ is the minimum wins necessary. The interested reader should note that if this proof were attempted for the exception team p, it could not be concluded that $\bar{v}_1$ wins would be enough to guarantee a playoff position, since by definition, team p attains strictly more wins; i.e., $\bar{v}^p_1 > \bar{v}_1$.

C Playoff elimination formulation for exception team p

Using identical notation to Section 4.3, consider the following formulation to determine $\hat{v}^p_1$, the minimum number of games team p must win to finish with a playoff berth:

(P3p) Playoff Elimination Formulation for Team p

min v^p_1

subject to

(1)	$x_{ij} + x_{ji} = g_{ij}$	$\forall\, i, j \in S,\ i < j$
(2)	$v^p_1 \geq w_i + \sum_{j \in S} x_{ij} - M\beta$	$\forall\, i \in N^1$
(3)	$v^p_1 \geq w_i + \sum_{j \in S} x_{ij} - M(1-\beta) - M\alpha^k_i$	$\forall\, k \in \{1,2,3\}, i \in N^k$
(4)	$v^p_1 = w_p + \sum_{j \in S} x_{pj}$	
(5)	$\sum_{i \in N^k} \alpha^k_i \leq 1$	$\forall\, k \in \{1,2,3\}$
(6)	$x_{ij} \geq 0$	$\forall\, i, j \in S$
(7)	$\bar{v}_1 \leq v^p_1 \leq \bar{v}^p_1$	
(8)	x_{ij} integer	$\forall\, i, j \in S$
(9)	v_1 integer	
(10)	$\beta,\ \alpha^k_i$ binary	$\forall\, k \in \{1,2,3\}, i \in N^k$

This straightforward formulation simply minimizes the number of wins that team p must attain in order to finish in a playoff position. The necessity and sufficiency of its optimal value, $\hat{v}^p_1$, for determining team p's playoff prospects is clear.

D Playoff clinch formulation

To determine the playoff clinch numbers, we solve an integer formulation for each team. The formulation maximizes the wins for a particular team such that they do not make the playoffs. Adding one to the optimal objective value clearly gives the minimum number of wins that will guarantee the team a playoff position.

Consider the following clinch formulation for some team a in division N^1. Using the notation from Section 4.3,

(P4) Playoff Clinch Formulation

$\max v_a$

subject to

(1)	$x_{ij} + x_{ji} = g_{ij}$	$\forall\ i,j \in S,\ i < j$
(2)	$v_a \leq w_i + \sum_{j \in S} x_{ij} + M\alpha_i^k - 1$	$\forall\ k \in \{1,2,3\}, i \in N^k, i \neq a$
(3)	$v_a = w_a + \sum_{j \in S} x_{aj}$	
(4)	$\sum_{i \in N^k} \alpha_i^k \leq \lvert N^k \rvert - 1 - \beta^k$	$\forall\ k \in \{1,2,3\}$
(5)	$\sum_{k=1}^{3} \beta^k \geq 1$	
(6)	$x_{ij} \geq 0$	$\forall\ i,j \in S$
(7)	x_{ij} integer	$\forall\ i,j \in S$
(8)	$\beta^k,\ \alpha_i^k$ binary	$\forall\ k \in \{1,2,3\},\ i \in N^k$

Constraints (2) force team a to finish with fewer wins than all teams with $\alpha_i^k = 0$. Constraints (4) and (5) ensure that each division has at least one team with $\alpha_i^k = 0$, and at least one division with at least two teams with the property. Therefore, v_a is the maximum number of wins that team a can attain without finishing in a playoff position. The playoff clinch number for a is then simply $v_a + 1$.

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network flows: theory, algorithms, and applications.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1993.

[2] G. Gallo, M.D. Grigoriadis, and R.E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, 18:30–55, 1989.

[3] D. Gusfield and C. Martel. A fast algorithm for the generalized parametric minimum cut problem and applications. *Algorithmica*, 7:499–519, 1992.

[4] D. Gusfield, C. Martel, and D. Fernandez-Baca. Fast algorithms for bipartite network flow. *SIAM Journal on Computing*, 16:237–251, 1987.

[5] A.J. Hoffman and T.J. Rivlin. When is a team 'mathematically' eliminated? In H.W. Kuhn, editor, *Proceedings of the Princeton Symposium on Mathematical Programming.* Princeton University Press, 1970.

[6] S.T. McCormick. Two hard min cut problems. Technical report, Presentation at TIMS/ORSA Conference, New Orleans, 1987.

[7] S.T. McCormick. Fast algorithms for parametric scheduling come from extensions to parametric maximum flow. *Operations Research*, to appear.

[8] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization.* John Wiley and Sons, New York, NY, 1988.

[9] L.W. Robinson. Baseball playoff eliminations: An application of linear programming. *Operations Research Letters*, 10:67–74, March 1991.

[10] L. Schrage. *Linear, Integer, and Quadratic Programming with LINDO.* Scientific Press, Palo Alto, CA, 1984.

[11] B.L. Schwartz. Possible winners in partially completed tournaments. *SIAM Review*, 8:302–308, 1966.