Baseball, Optimization, and the World Wide Web

Ilan Adler • Alan L. Erera • Dorit S. Hochbaum • Eli V. Olinick

Department of Industrial Engineering and Operations Research, 4135 Etcheverry Hall, University of California, Berkeley, California 94720

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205

Department of Industrial Engineering and Operations Research, 4135 Etcheverry Hall, University of California, Berkeley, California 94720

Department of Computer Science and Engineering, Southern Methodist University, P.O. Box 750122, Dallas, Texas 75275-0122

adler@ieor.berkeley.edu • alerera@isye.gatech.edu • hochbaum@ieor.berkeley.edu • olinick@engr.smu.edu

The competition for baseball play-off spots—the fabled pennant race—is one of the most closely watched American sports traditions. While play-off race statistics, such as games back and magic number, are informative, they are overly conservative and do not account for the remaining schedule of games. Using optimization techniques, one can model schedule effects explicitly and determine precisely when a team has secured a play-off spot or has been eliminated from contention. The RIOT Baseball Play-off Races Web site developed at the University of California, Berkeley, provides automatic updates of new, optimization-based play-off race statistics each day of the major league baseball season. In developing the site, we found that we could determine the first-place elimination status of all teams in a division using a single linear-programming formulation, since a minimum win threshold for teams finishing in first place applies to all teams in a division. We identified a similar (but weaker) result for the problem of play-off elimination with wildcard teams.
(*Recreation and sports*)

Fans of professional sports teams have an insatiable desire for information about the performance of their favorite teams. Fans of major league baseball (MLB) in the United States are particularly concerned about their teams' prospects for reaching the postseason play-offs: the fabled pennant race. Fans check newspapers and Web sites daily for updates on team progress (or lack thereof!).

As the end of the season nears, teams trailing the current division leader may become mathematically eliminated from first place; such teams have no chance of finishing first in their division, even if they were to win all of their remaining games. The Elias Sports Bureau, the official statistician for MLB, determines whether a particular team is eliminated using a simple criterion: if a team trails the first-place team in wins by more games than it has remaining, it is eliminated. That the San Francisco Giants had suffered this undesirable fate was announced on September 10, 1996 in the *San Francisco Chronicle* (Gay 1996); the Giants had 59 wins with 20 games left to play, while the first-place San Diego Padres had already won 80 games. The Giants, however, had actually been eliminated two days earlier: we had announced the news of their demise on September 8 on our Web site (Table 1).

The optimization community has long known that the Elias criterion is sufficient to eliminate teams from first place but not necessary (Schwartz 1966). The problem is that the criterion ignores the schedule of remaining games. Continuing the Giants example, Los

This paper was refereed.

0092-2102/02/3202/0012$05.00
1526-551X electronic ISSN

National League West

Team	Wins	Losses	Games Back	Games Left	Clinch 1st	Clinch Play	Avoid Elim 1st	Avoid Elim Play
Los Angeles	78	63	—	21	17	17	4	1
San Diego	78	65	1	19	17	17	4	0
Colorado	71	71	7.5	20	*	*	11	7
San Francisco	59	81	18.5	22	*	*	Elim	19

Table 1: The RIOT Baseball Play-off Races Web site declared the San Francisco Giants eliminated from first place on September 8, 1996, two days before an announcement was made in the newspaper. Since San Francisco has 22 games remaining and trails Los Angeles by only 18.5 games, it is not readily apparent from the traditional statistics that the team is eliminated.

Angeles and San Diego were scheduled to play each other seven more times after September 8. 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. Thus, the Giants were eliminated at this point, since they could finish with at most 81 wins. By using simple optimization techniques, one can model schedule effects explicitly to determine when teams are truly mathematically eliminated.

First-place elimination is not the fans' only interest. In baseball, teams may also reach the play-offs by securing a *wild-card* berth; the team that finishes with the best record among second-place teams in the league is assigned this berth. Thus, wild-card elimination is also important to track. In addition, teams not yet eliminated may be perilously close; a measure of closeness to first-place or wild-card elimination would be useful. Conversely, fans of teams that are performing well would like to know if their teams have clinched first place or a wild-card berth. A clinch is a guarantee; once a team has clinched first place, for example, it could lose all of its remaining games and still finish in first.

We developed a Web site to provide optimization-based MLB play-off race statistics to the general public so that fans can sort out the play-off picture with more precise information. The Berkeley Baseball Play-off Races site, a component of the Remote Interactive Optimization Testbed (RIOT), is up and running during the baseball season (April through October) at ⟨http://riot.ieor.berkeley.edu/~baseball⟩. The site provides daily updates of elimination and clinch statistics.

Although baseball fans find the site informative and entertaining, we also designed it to achieve an educational goal. Sports elimination problems are useful for teaching basic ideas in optimization; they are covered in textbooks by Schrage (1984) and Ahuja et al. (1993). Robinson (1991) argues that many students relate to the problem subject intuitively and find the results interesting. We agree and furthermore suggest that the Internet is an ideal place to present the problem and broadcast results to attract the interest of both students and those who might otherwise never be exposed to optimization concepts. The RIOT site provides links to other Web sites with an educational component, such as the Network Enabled Optimization System (NEOS) project sponsored by Argonne National Laboratory and Northwestern University. The NEOS optimization guide ⟨http://www-fp.mcs.anl.gov/otc/guide/⟩ contains interactive case studies demonstrating the application of OR models to general-interest problems, such as portfolio optimization and the diet problem. RIOT also links to Michael Trick's OR page ⟨http://mat.gsia.cmu.edu⟩, which serves as a portal for OR on the web and contains an up-to-date and comprehensive list of interactive, educational Web sites ⟨http://mat.gsia.cmu.edu/program.html⟩.

Problem Description

Because many readers may not be aficionados of America's national pastime, we will begin by describing the current major league baseball play-off structure. MLB teams are partitioned into two leagues, American and National. Each league is further subdivided into three divisions. Each team in each league plays a regular season schedule of 162 games to determine the teams that will advance to the play-off rounds. Four teams from each of the two leagues make the play-offs: the three teams that finish with the best records in their respective divisions, and a fourth team (the wild-card) that has the best record among all second place teams in the league. Ties in the final standings for a play-off spot are settled by special one-game playoffs. Each league then conducts a tournament with its four invited play-off teams to determine its pennant winner. Finally, the American and National League pennant winners play in the World Series for the MLB championship.

Now consider a particular team, say the Boston Red Sox, at some point during the regular season. Given the current win-loss records of all teams and the remaining schedule of games, are the Red Sox eliminated from finishing in a play-off position, and if not, how close are they to elimination? If the Red Sox have not been eliminated, have they clinched a play-off spot? And if they have not, how close are they to clinching?

Opponents	Games Remaining
Baltimore vs. Boston	2
Baltimore vs. New York	3
Baltimore vs. Toronto	7
Boston vs. New York	8
Boston vs. Toronto	0
New York vs. Toronto	7

Table 3: Using the remaining schedule of games given here and the standings in Table 2, it is possible to show that Detroit cannot finish with as many wins as New York under any scenario. Thus, Detroit has been eliminated from first place.

Elimination Questions

The official MLB method for determining first-place elimination for a division is somewhat naive, and often teams may be eliminated earlier than the official declaration. In the Giants example presented earlier, proving elimination was simple by inspection, but it can be much more difficult. As an example, consider the case of the Detroit Tigers on August 30, 1996. If we examine the standings in the American League East division after the completion of play that night (Table 2), it appears that Detroit has a remote chance of catching the first-place New York Yankees since they have 27 games remaining and trail New York by only 26 wins. It is possible, however, to show that Detroit is in fact mathematically eliminated from first place using some simple information regarding the remaining schedule of games between teams in the division. Using the remaining games information (Table 3), the inspired reader should try the elimination proof as an exercise; the following paragraphs detail the proof.

To prove that Detroit is eliminated, we can show that it is impossible to construct a scenario in which Detroit would win its division. If Detroit won all of its remaining games, it would finish with a record of 76 wins and 86 losses. If New York won just two more games, it would finish with 77 wins and 85 losses and therefore ahead of Detroit. Thus, we now analyze scenarios in which New York wins one or no remaining games. First, suppose that New York fails to win another game. Since Boston has eight games remaining against New York, Boston would finish with at least 77 wins in this scenario (69 + 8), and it would finish ahead of Detroit. Thus, for Detroit to have any chance of finishing first, New York would have to win exactly one of its eight games with Boston and lose all of its other games. In addition, Boston would have to lose all of the games it plays against teams other than New York. This would create a three-way tie for first place (Table 4).

Now consider Baltimore and Toronto. Baltimore has two games remaining with Boston and three with New York and therefore would finish with at least 76 wins

Team	Wins	Losses	Games Back	Games Left
New York	75	59	—	28
Baltimore	71	63	4	28
Boston	69	66	6.5	27
Toronto	63	72	12.5	27
Detroit	49	86	26.5	27

Table 2: Can the Detroit Tigers win the pennant? By examining these standings, it appears that Detroit has a (remote) chance of catching New York, since they have 26 fewer wins but 27 games remaining. In fact, Detroit is mathematically eliminated from first; can you prove it using the data in Table 3?

Team	Wins	Losses	Games Back
Detroit	76	86	—
Boston	76	86	—
New York	76	86	—
Baltimore	?	?	?
Toronto	?	?	?

Table 4: First, suppose Detroit were to win each of its 27 remaining games. Now, suppose New York were to win a single future game against Boston but were to lose all of its other remaining games. In this case, Boston would win at least seven future games (against New York). If Boston were to lose the rest of its remaining games, it would finish tied with New York and Detroit with 76 wins. But what about Baltimore and Toronto?

in our scenario. Thus, Detroit could finish in a first place tie only if Baltimore were to lose all of its remaining games to teams other than New York and Boston. Unfortunately for Detroit, Toronto has seven games remaining with Baltimore and seven remaining with New York. According to the above logic, Toronto would have to win these 14 games in any scenario in which Detroit finishes in first place. However, if Toronto were to win 14 additional games, it would finish with a record of at worst 77 wins and 85 losses and therefore ahead of Detroit. Therefore, Detroit is mathematically eliminated from first place.

Clearly, constructing such elimination proofs by hand can be a tedious endeavor. Fortunately, optimization methods can help. Researchers have previously addressed the problem of first-place elimination. Schwartz (1966) showed that a maximum-flow calculation on a small network can determine precisely when a team has been necessarily eliminated from first place. Robinson (1991) showed that such an optimization approach would have eliminated teams an average of three days earlier than the wins-based criterion during the 1987 season. Hoffman and Rivlin (1970) extended Schwartz's work, developing necessary and sufficient conditions for eliminating a team from kth place. McCormick (1987, 1999) in turn showed that determining elimination from kth place is $\mathcal{NP}$-complete. Gusfield et al. (1987) 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 (1992) showed that the minimum number of games a given team must win to avoid elimination from first place can be found by solving a parametric maximum-flow problem. By extending a result of Gallo et al. (1989) and using a binary search procedure, Gusfield and Martel proved 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. McCormick (1999) improved the time bound for solving this parametric maximum flow problem to $O(n^3)$.

Fans have an insatiable desire for information about their favorite teams.

Determining whether or not a team is eliminated from first place is only half of the story, since eliminated teams might still make the play-offs in the wild-card berth. Little research has focused on play-off elimination with wild-card teams, partially since prior to the 1994 season only the division winners advanced to the baseball play-offs. Robinson (1991) briefly discussed the complications introduced by wild-card berths in the context of applying his baseball elimination model to the National Football League (NFL) but did not provide a formulation.

For the Baseball Play-off Races Web site, we decided that the most interesting elimination information for fans would be statistics that provide a measure of how close each team is to elimination, similar to those proposed by Gusfield and Martel (1992). Therefore, we define a team's first-place elimination number to be the minimum number of remaining games that the team must win to have any chance of finishing in first place in its division. As a team's first-place elimination number approaches the number of games it has remaining, elimination becomes imminent. In addition, we define a team's play-off elimination number to be the minimum number of games the team must win to have any chance of earning a play-off spot, whether as a division winner or as the wild-card team.

Clinch Questions

Fans of the teams performing well during the regular season have a very different concern: they want to know when their team has clinched first place or a wild-card playoff spot. Currently, the media use magic numbers to determine first-place clinches. Assume that the teams in a division are ranked in order of increasing 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 given by $g_1 - (l_2 - l_1)$. 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 the top spot in the division. When the first-place team's magic number drops to zero, the team has clinched first.

Unlike the case with elimination, 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 do not specify the minimum number of future wins necessary for a team to clinch first place independent of other teams' performance, and they are typically reported only for teams in first place. To address these drawbacks, we define the first-place clinch number for each team to be the minimum number of games which, if won, guarantees that the team finishes in at least a tie for first place. Similarly, we define the play-off clinch number for each team to be the minimum number of games which, if won, guarantees that team a position in the play-offs, either as the division winner or as the wild-card team.

The Baseball Play-off Races Web Site

The baseball Web site is a component of the Berkeley RIOT Internet project, an on-line collection developed and maintained by the Industrial Engineering and Operations Research Department and the Haas School of Business at the University of California, Berkeley. The primary focus of the RIOT project has been to provide educational information about industrial engineering and operations research and to promote interest in the field via Web pages and easy-to-use, interactive Java applets. Each RIOT component application includes pages describing the details of the underlying optimization models and algorithms used in the problem solution; once visitors have played with the application and discovered its utility, they can learn about the methods used to produce the results.

To provide the most up-to-date information to fans, we designed the Baseball Play-off Races Web site to be updated each night during the baseball season. Creating the site required two primary development activities. First, we generated a set of mathematical models for calculating the new play-off statistics (Appendix). Second, we developed a software system that employs the models to produce automated nightly updates of the Web site. The system is scheduled to run in the early morning hours, creating and posting an updated HTML standings report (Table 5). The standings report is similar to those provided by newspaper sports sections, with teams grouped by league and division and sorted by win-loss record. In addition to the information traditionally reported, the report displays

National League East

						Clinch		Avoid Elim	
	Wins	Losses	Games Back	Percentage	Games Left	1st	Play	1st	Play
Atlanta	86	55	—	0.610	21	13	9	0	0
Montreal	78	63	8	0.553	21	21	17	8	0
Florida	69	74	18	0.483	19	*	*	17	9
New York	62	80	24.5	0.437	20	*	*	Elim	16
Philadelphia	58	85	29	0.406	19	*	*	Elim	Elim

Table 5: In this sample standings report from the Baseball Play-off Races Web site, the first five columns contain traditional standings information: wins, losses, games back, winning percentage, and games left to play. The two "Clinch" columns provide each team's current first-place and play-off clinch numbers, while the two "Avoid Elim" columns provide the elimination numbers. An asterisk for a clinch number indicates that a clinch is not currently possible, even if the team were to win all of its remaining games. If a team had already clinched first or a play-off spot, it would be labeled "In." While New York and Philadelphia are mathematically eliminated from finishing first, New York has a remote chance of securing a wild-card berth (by winning 16 of the remaining 20 games). Also, if Montreal were to win its remaining 21 games, it would clinch at least a tie for first place.

each team's two elimination numbers and two clinch numbers.

The software system that generates the standings reports operates as follows. Since the calculations require the current win-loss records of each team and the remaining number of games between teams, the system maintains a simple database that is updated using the results of the previous day's games. A free Internet news service called Infobeat ⟨www.infobeat.com⟩ automatically sends the system an e-mail message each night containing the final scores of all MLB games. The first component of the system initiates the update process by automatically reading and processing this e-mail message, updating the team win-loss records and games remaining in the database. Next, a program uses the database to generate text files containing the mathematical optimization models that allow calculation of the elimination and clinch numbers. The system then solves the necessary models using the CPLEX optimization package ⟨www.cplex.com⟩ and processes the results to determine each team's current numbers. Finally, a page-updating program uses the updated numbers and generates new standings reports in the HTML format required by WWW browser programs (for example, Netscape Communicator and Microsoft Internet Explorer). The process usually runs seamlessly without human intervention. Occasionally, however, the e-mail message containing game results does not arrive as expected, and we must initiate the update process manually. We implemented the software system on a Sun Microsystems SPARCstation 20, and it typically completes its various tasks in about 10 minutes.

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 particularly well suited for string manipulation. The update procedure requires parsing multiple text input files (such as the nightly e-mail message containing baseball scores and CPLEX output files), manipulating and combining text strings, and then writing out new text files (such as input files for CPLEX or the updated standings reports). These types of operations are generally much easier to code in Perl than in such languages as C or C++.

We began work on the site while the 1996 season was in progress, 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 a wealth of MLB data is available on the Web, we had difficulty obtaining the necessary information in a readily usable format. For our purposes, we would have liked 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 that lists the games to be played each day of the season. To find out how many games were left between, say, the Boston Red Sox and New York Yankees, we had to parse the list and count the number of remaining scheduled games between the teams; we easily automated this task. Since the schedules we found on the Web contained inaccuracies, and since different sources handled canceled and suspended games in different ways, producing a correct schedule of remaining games became an unexpectedly difficult chore. Eventually, however, we produced an accurate schedule. For the 1997 to 2002 seasons, generating schedules was much simpler because all of our software was ready before the start of the season.

Two new teams, the Arizona Diamondbacks and Tampa Bay Devil Rays, joined the major leagues in 1998. To accommodate these 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. While we easily adapted our system to these changes, we may have to make more significant adaptations if the play-off structure of MLB is altered. One possibility is that two additional teams will join MLB in the next two years and teams will be realigned into two leagues with two eight-team divisions. If this occurs, a second wild-card team from each league will probably be added to the play-offs, and we will need to develop new mathematical formulations for wild-card elimination and clinching.

Since the RIOT Baseball Play-off Races site went online during the 1996 season, it has been popular with Web surfers. As soon as the site was listed in several Internet directories, baseball fans started visiting the pages. During September, when the pennant race is most heated, the pages attract 100 to 200 hits each day. As further testament to its popularity, the site was featured on a 1996 broadcast of the public radio program Beyond Computers.

Mathematical Models for Elimination and Clinching

At the core of the automated system that we developed for updating the Web site are the mathematical models used for calculating the elimination and clinch numbers (Appendix).

When we first planned the site, our initial idea was to simply calculate and provide first-place elimination numbers for each team in MLB. Our initial formulation was based on the parametric maximum flow formulation given by McCormick (1999), which is an extension of the original formulation of Schwartz (1966). Using this modeling methodology, we created a separate flow formulation for each team to determine its first-place elimination number. To solve the instances, we decided to translate the flow formulation into a corresponding integer linear program. Since we had access to a fast, efficient IP solver in CPLEX and since the translation resulted in small problems, it was much easier for us to work with an integer programming formulation from an implementation standpoint.

After we had the models running using real MLB data, we noticed an interesting property in our standings reports. Adding together the first-place elimination number and the current win total for any team in a specific division k on a given date yielded a constant, v_k. Thus, we suspected that elimination numbers could be calculated using a single formulation to determine v_k for each division, instead of utilizing a separate formulation for each team. We were able to prove this suspicion and, in addition, to show that the formulations can be solved using linear programming (Appendix). As a result, we calculate first-place elimination numbers for each team in MLB by solving six small linear programs, one for each division.

Our experimentation with the integer and linear programs for first-place elimination led us naturally to consider the problem of play-off elimination with wild-card teams. Again, we began by considering a formulation for each team separately. The primary idea in first-place elimination models is to allocate wins of remaining games among teams feasibly to create an end-of-season scenario in which the team under consideration attains at least a tie for first and finishes with as few wins as possible. It was not too difficult to extend this idea to the play-off elimination setting. In this case, the idea is to create a feasible end-of-season scenario in which the team under consideration finishes either in first place or with the best record among all second place teams in its league. We again postulated that there might exist some threshold v_L that would allow us to compute the play-off elimination numbers for all teams in the league by solving a single formulation. Although this was not the case, we were able to develop a similar but weaker result that allows us to compute the numbers by solving at most $k + 1$ small integer programs for each league. For MLB, therefore, we need to solve at most eight instances (Appendix).

At this point, it seemed natural to address clinching problems with a similar mathematical-programming approach. To determine clinch numbers, we decided initially to use models that are in some sense reversed. For example, to determine a specific team's first-place clinch number, we formulated an integer program that allocated wins to create a feasible end-of-season scenario in which the team maximizes its remaining wins without finishing in first. We then find the clinch number by simply adding one to the maximized remaining wins. We developed a similar formulation for play-off clinching. After some reflection, however, we realized that solving a formal integer optimization model was not necessary to determine the first-place clinch numbers (Appendix).

Conclusions

The Baseball Play-off Races Web site broadcasts improved, optimization-based statistics to fans daily, providing information more precise than that found elsewhere. 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, they increase the value of the data. As more and more data becomes available online, the potential for more meaningful value-adding activity only increases. 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 100 stocks, the system allows users to solve a portfolio-optimization model. Both the baseball and portfolio systems should give both researchers and practitioners a glimpse at the types of opportunities that exist to use operations research to increase the value of online data.

Acknowledgments

All authors were supported by Office of Naval Research contract N00014-91-J-1241 during this research. Additional funding for Dorit

Hochbaum and Eli Olinick was provided by National Science Foundation award DMI-9713482. We extend our gratitude to these funding agencies.

Appendix: Mathematics of Elimination and Clinching

We describe the mathematical details behind the models developed for the Baseball Play-off Races site. Importantly, we show how a single linear program can be used to determine first-place elimination numbers for all teams in a single division. We then show that play-off elimination numbers for a league with k divisions can be determined by solving at most $k + 1$ small mixed-integer programs. Next, we develop a simple arithmetic calculation for determining first-place clinch numbers. Finally, we present a mixed-integer programming model that can be used to calculate a single team's play-off clinch number.

Notation and Assumptions

Let L be the set of teams in a league (for example, the American League) and suppose L is partitioned into a number of divisions. Let D_k be the set of teams in division k (for example, k could be the American League East) where $\cup_k D_k = L$. For each team i in division k, let w_i be its number of current wins, g_{ij} the number of games remaining against team j (note that $g_{ii} = 0$), and t_i the number of games remaining against nondivision opponents (that is, $t_i = \Sigma_{j \in L \setminus Dk}\, g_{ij}$). Finally, let W_i be the total number of wins attained by team i by season's end in some scenario.

In the models that follow, we assume that each team must play out its entire schedule of games (consistent with MLB rules) and that each game has a winner. In addition, we assume that finishing in a tie in the final standings with another team(s) is sufficient to secure a play-off spot; MLB typically resolves such ties with special one-game play-offs.

First-Place Elimination

Now consider a single division of teams on a given day during the season. Define the first-place-elimination problem (FEP) 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. The single-team parametric maximum-flow formulation in Gusfield and Martel (1992) could be employed to solve (FEP) by creating and solving an appropriate instance for each team. This is unnecessary, however, since we now show that there exists for each division a first-place-elimination threshold, the minimum number of wins at season-end necessary and sufficient for any team to finish in first, and that this threshold may be found by solving a single linear program. Given the threshold, the first-place-elimination number for each team in the division is simply the difference between the threshold and the team's current number of wins. Notably, Wayne (2001) concurrently proves the existence of a first-place-elimination threshold using a maximum-flow formulation.

First, consider the following minimax-type integer linear program for determining the first-place-elimination threshold for division k. Let v_k be the decision variable representing the threshold. Further, let x_{ij} represent the number of future games team $i \in D_k$ wins against team $j \in D_k$; let x denote a complete scenario of future wins, $x = \{x_{ij} \mid i, j \in D_k\}$. The following model allocates wins to teams in order to minimize the number of wins attained by the division winner by season end:

(P1) First-Place-Elimination Threshold Integer Formulation

min v_k

subject to

$$x_{ij} + x_{ji} = g_{ij} \quad \forall\, i, j \in D_k,\ i < j, \tag{1}$$

$$v_k \geq w_i + \sum_{j \in D_k} x_{ij} \quad \forall\, i \in D_k, \tag{2}$$

$$x_{ij} \geq 0 \quad \forall i, j \in D_k,\ i \neq j, \tag{3}$$

$$v_k \text{ integer}, \tag{4}$$

$$x_{ij} \text{ integer} \quad \forall\, i, j \in D_k,\ i \neq j \tag{5}$$

Constraints (1) ensure that the allocation of wins accounts for all of the remaining games between each pair of teams in the division. Constraints (2), in conjunction with the objective function, force v_k to be the minimum number of wins attained by a division-winning team at the end of the season. Games played against teams outside the division are ignored; to find the minimum number of wins necessary to win a division, it is only necessary to consider scenarios in which the teams in division k lose all remaining games against nondivision opponents.

Now suppose that the optimal objective value of (P1) is $\bar{v}_k$. We claim that $\bar{v}_k$ is the first-place-elimination threshold for division k. First, it is clear from the formulation that in the optimal solution at least one team will win exactly $\bar{v}_k$ games. Thus, no team winning fewer than $\bar{v}_k$ games can finish atop the division. To complete the proof, it can be shown that a final standings scenario can always be constructed in which any team $l \in D_k$ that can attain at least $\bar{v}_k$ wins by season end (that is, $w_l + t_l + \Sigma_{j \in D_k} g_{lj} \geq \bar{v}_k$) can win the division with exactly $\bar{v}_k$ wins. To do so, consider the optimal allocation of future wins, $\bar{x}$, in the solution to (P1) and let $\bar{v}_l = w_l + \Sigma_{j \in D_k} \bar{x}_{lk}$. If $\bar{v}_l + t_l \geq \bar{v}_k$, a division-winning scenario for l can be attained by increasing (if necessary) its number of nondivision wins such that l wins exactly $\bar{v}_k$ total games. It is also simple to construct a scenario in the alternative case when $\bar{v}_l + t_l < \bar{v}_k$, but the allocation $\bar{x}$ must change. In this case, it is assumed that l wins all of its nondivision games and an additional $\bar{v}_k - \bar{v}_l - t_l$ division games. This reallocation of division wins to l, of course, can only lower the standings of its foes, and thus team l wins the division with $\bar{v}_k$ wins.

Armed with the first-place-elimination threshold, we can determine the elimination status of each team in division k. First, team $i \in D_k$ is eliminated from first-place contention if and only if $w_i + t_i + \Sigma_{j \in D_k} g_{ij} < \bar{v}_k$. Since the left-hand side of the inequality is the maximum number of wins attainable by team i and the right-hand side is the threshold, this condition is clear. Furthermore, if team i is not eliminated, its first-place-elimination number is $\bar{v}_k - w_i$, the minimum number of future wins team i needs to reach the threshold.

Solving First-Place Elimination via Linear Programming

The integer formulation (P1) can be simplified to determine $\bar{v}_k$ using linear programming. Consider the constraint coefficient matrix of (P1), which we denote A. Ignoring the column corresponding to variable v_k, it is easy to see that the remaining submatrix, A' is totally unimodular, since A' represents the node-edge incidence matrix of an undirected bipartite graph (see, for example, Nemhauser and Wolsey 1988). Since the right-hand side of the constraint system is integer, the total unimodularity of A' guarantees that for every fixed integer v_k (including $\bar{v}_k$), a corresponding integer basic solution x exists if the system is feasible. Thus, constraints (5) in (P1) can be ignored without affecting the optimal objective, $\bar{v}_k$. The resulting formulation contains only one integer variable, v_k, and its optimal objective function value can be determined by solving the linear relaxation obtained by ignoring constraint (4), and then rounding the resulting solution $\hat{v}_k$ up to the nearest integer. Therefore $\bar{v}_k = \lceil \hat{v}_k \rceil$.

Play-off Elimination

Elimination from the play-offs occurs only 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 (thus earning the wild-card berth). A scenario can arise in which a team is eliminated from the wild card but not eliminated from first place in its division. This may happen, for example, if the teams with the two best records in a league play in the same division. In this case, a team from a weaker division may be able to reach the play-offs only by finishing first in its division. We define the play-off-elimination problem (PEP) as follows: given the current win-loss records of each team and the remaining schedule of games, determine the play-off-elimination number for each team, as previously defined.

The PEP is inherently more difficult than the FEP. For each team, the problem to be solved is similar to the problem of elimination from lth place in a division, which McCormick (1987) showed to be NP-complete. From a practical standpoint, play-off-elimination formulations must also be slightly larger than first-place-elimination models, since the remaining games between all teams in a league must be considered. It is a straightforward task to formulate an integer program to determine a single team's play-off-elimination number; solving an appropriate instance for each team would solve PEP. In the spirit of the previous section, however, we attempt to avoid solving a separate instance for each team unnecessarily.

We now show that the play-off-elimination numbers for each team in a league L with k divisions can be computed by solving at most $k + 1$ small mixed-integer programs. The first model (P2) is used to compute a wild-card-elimination threshold that applies to most teams in L. Similar to the first-place-elimination threshold, the wild-card threshold is the minimum number of wins necessary for any team finishing in a play-off spot at the end of the season. The sufficiency of the threshold, however, applies only to a certain subset of teams; at most one team from each division may require more wins than the wild-card threshold to finish in a play-off position. We explore these ideas below.

First, consider the following integer-programming formulation for determining a wild-card threshold:

(P2) Wild-Card-Elimination-Threshold Integer Formulation

min u

subject to

$$x_{ij} + x_{ji} = g_{ij} \quad \forall\, i, j \in L,\ i \neq j, \tag{6}$$

$$u \geq w_i + \sum_{j \in L} x_{ij} - M\alpha_i^k \quad \forall\, k \in \{1, 2, 3\},\ i \in D_k, \tag{7}$$

$$\sum_{i \in D_k} \alpha_i^k = 1 \quad \forall\, k \in \{1, 2, 3\}, \tag{8}$$

$$x_{ij} \geq 0 \quad \forall\, i, j \in L, \tag{9}$$

$$x_{ij} \text{ integer} \quad \forall\, i, j \in L, \tag{10}$$

$$u \text{ integer}, \tag{11}$$

$$\alpha_i^k \text{ binary} \quad \forall\, k \in \{1, 2, 3\},\ i \in D_k, \tag{12}$$

where M is a large integer (specifically, for example, greater than the number of games in a season, 162) and for illustrative purposes, we assume that league L has three divisions. This minimax model attempts to allocate future wins among all teams in L such that u is at least as large as the number of wins attained by all teams except possibly one team from each of the three divisions; thus, the optimal objective $\bar{u}$ is the minimum number of wins necessary to be a play-off team.

In the optimal solution to (P2), let F be the set of exception teams finishing with more than $\bar{u}$ wins; that is, $w_f + \sum_{j \in L} \bar{x}_{fj} > \bar{u}\ \forall\, f \in F$. The formulation guarantees that there can be at most one exception team f_k from each division D_k. We now show that teams in $L \backslash F$ that can attain $\bar{u}$ wins by season's end can always finish in a play-off position, either as a division winner or as the wild-card team. Again, we employ a scenario-construction technique for the proof given the optimal win allocation $\bar{x}$ from (P2).

Consider team $l \in L \backslash F$, where $w_l + \sum_{j \in L} g_{lj} \geq \bar{u}$. Since l can attain at least $\bar{u}$ wins, it is always possible to construct an $\hat{x}$ in which $w_l + \sum_{j \in L} \hat{x}_{lj} = \bar{u}$ by a process similar to the one described in the previous section. Again, we initialize $\hat{x} = \bar{x}$ and then increase $\hat{x}_{lj}$ (up to g_{lj}) for some opponents j and correspondingly decrease $\hat{x}_{jl}$ until team l has exactly $\bar{u}$ wins, that is, $w_l + \sum_{j \in L} \hat{x}_{lj} = \bar{u}$. The resulting end-of-season win scenario is:

$$W_i = \begin{cases} w_i + \sum_{j \in L} \hat{x}_{ij} & \forall\, i \in L,\ i \neq l, \\ w_l + \sum_{j \in L} \hat{x}_{lj} = \bar{u} & i = l. \end{cases} \tag{13}$$

Team l makes the play-offs in this scenario with $\bar{u}$ wins, since $\bar{u} \geq w_i + \sum_{j \in L} \bar{x}_{ij} \geq w_i + \sum_{j \in L} \hat{x}_{ij}$ for all $i \in L \backslash F$, where the first inequality holds from the formulation and the second from the construction of $\hat{x}$. Thus, l necessarily finishes with at least as many wins as all teams except possibly the division leaders. Team l may actually win its division in this new scenario, since the wins of the division leader given by $\bar{x}$ may be decreased to $\bar{u}$ or fewer wins during the construction of $\hat{x}$.

The exception teams in F may not be able to finish in a play-off spot by winning only $\bar{u}$ games. In the optimal solution to (P2), it is possible that $w_f + \sum_{j \in L} \bar{x}_{fj} > \bar{u}$ for $f \in F$; therefore, we would need to *decrease* f's win total to find a scenario in which f makes the play-offs with exactly $\bar{u}$ wins. Decreasing the wins of team f requires increasing other teams' wins, which may create a scenario in which f does not finish in a play-off position. Thus, there may be no scenario in which f makes the play-offs with $\bar{u}$ wins. To address this problem, we propose solving an additional integer program for each exception team f_k. This model, denoted (P2f_k) is identical to (P2), with the additional constraint:

$$\alpha_{f_k}^k = 0, \tag{14}$$

which guarantees that team f_k is not an exception team and therefore no longer finishes with more wins than the optimal objective function value $\bar{u}_k$.

We can now determine play-off-elimination numbers for each team. For $i \in D_k \setminus F$, the play-off-elimination number is $\min\{\bar{v}_k, \bar{u}\} - w_i$. For each exception team $f_k \in F$, the number is $\min\{\bar{v}_k, \bar{u}_k\} - w_{f_k}$. Since a team is not eliminated from the play-offs if it is not eliminated from first place, the elimination number is the minimum of the first-place-elimination threshold and the wild-card-elimination threshold minus the team's current win total. Again, if the play-off-elimination number for a team is greater than its number of remaining games, it is eliminated from the play-offs.

By an argument similar to that made in the previous section, constraint (10) is unnecessary in formulation (P2) (and (P2f_k)). To generate play-off-elimination numbers for each team, the RIOT system first solves two integer linear formulations of type (P2), one for both the American and National Leagues. Then, the exception teams from each division are identified, and formulations of type (P2f_k) are solved for each. Thus, play-off-elimination numbers for all MLB teams are created by solving at most eight small integer linear programs. As a final note, in the models presented here, we assume that first-place thresholds are calculated separately before calculation of the play-off-elimination numbers. It is possible alternatively to modify formulations (P2) and (P2f_k) to calculate $\min\{\bar{v}_k, \bar{u}\}$ and $\min\{\bar{v}_k, \bar{u}_k\}$ directly; we omit the details.

First-Place Clinch

Finally, we briefly address clinching problems. First, we consider the problem of determining a team's first-place clinch number on a given day during the season. For a team $i \in D_k$ to clinch first, it must win enough remaining games to guarantee that it finishes with a record at least as good as all other teams in its division. However, nothing prevents any other team from winning all of its remaining games except perhaps games against team i. Thus, first-place clinch numbers can be calculated easily without using optimization, as we now describe.

Let g_j be the total number of remaining games for each team $j \in D_k$, $g_j = t_j + \sum_{l \in D_j \setminus \{j\}} g_{jl}$. The first-place clinch number for team i, θ_i, can now be determined via the following arithmetic calculations:

$$\theta_{ij} = \min\left\{\frac{1}{2}(w_j + g_j - w_i + (g_i - g_{ij})),\ w_j + g_j - w_i\right\} \quad \forall\, j \in D_k \setminus \{i\} \tag{15}$$

$$\theta_i = \lceil \max_{j \in D_k \setminus \{i\}} \theta_{ij} \rceil. \tag{16}$$

The definition of θ_i guarantees that if team i were to win θ_i games, it would finish with a record at least as good as all of its division rivals. If $\theta_i \le 0$, team i has already clinched first place. Alternatively, if $\theta_i > g_i$, there is no way for team i to currently guarantee at least a first-place tie with all teams.

To briefly justify the above definition, let f_i be the number of future games won by team i. To guarantee a tie with some other team j, it is clearly sufficient for i to win $f_i = w_j + g_j - w_i$ future games. However, since i has g_{ij} future games with j, i may need to win fewer games. Consider the worst case for team i. If $f_i \le g_i - g_{ij}$, we assume that each future win by team i comes against teams other than team j. However, if $f_i > g_i - g_{ij}$, then in the worst case for team i, exactly $f_i - (g_i + g_{ij})$ of its future wins must come against team j, resulting in the same number of losses for j. Thus, each win for i beyond $g_i - g_{ij}$ effectively counts as two wins, justifying the expression for θ_{ij}.

Play-off Clinch

Determining the number of future wins needed to clinch a play-off spot is more complicated. We model this problem with a mixed-integer linear-programming formulation for each team similar to the play-off-elimination models presented previously. However, the formulation is reversed: instead of determining the minimum number of games a team must win to finish in a play-off position, we determine the maximum number of games a team can win *without* finishing in a play-off position.

To determine play-off clinch numbers, we solve such a problem for each team separately. Consider the following clinch formulation (P3) for some team a in division D_1. The objective is to maximize the number of additional wins v_a accrued by team a such that a finishes with fewer wins than the first-place team in its division, and at least one division contains two teams with better records. Since v_a therefore represents the maximum number of additional wins that a could accrue without finishing in a play-off position, $v_a + 1$ is the play-off clinch number for team a.

(P3) Play-off Clinch Formulation

max v_a
subject to

$$x_{ij} + x_{ji} = g_{ij} \quad \forall\, i, j \in L,\ i \ne j, \tag{17}$$

$$v_a \le w_i + \sum_{j \in L} x_{ij} + M\alpha_i^k - 1 \quad \forall\, k \in \{1,2,3\},\ i \in D_k,\ i \ne a, \tag{18}$$

$$v_a = w_a + \sum_{j \in L} x_{aj}, \tag{19}$$

$$\sum_{i \in D_k} \alpha_i^k \le |N^k| - 1 - \beta^k \quad \forall\, k \in \{1,2,3\}, \tag{20}$$

$$\sum_{k=1}^{3} \beta^k = 1, \tag{21}$$

$$x_{ij} \ge 0 \quad \forall\, i, j \in L, \tag{22}$$

$$x_{ij} \text{ integer} \quad \forall\, i, j \in L, \tag{23}$$

$$\beta^k, \alpha_i^k \text{ binary} \quad \forall\, k \in \{1,2,3\},\ i \in D_k. \tag{24}$$

Constraints (18) force team a to finish with strictly fewer wins than all teams for which $\alpha_i^k = 0$. Constraints (20) and (21) ensure that each division k contains at least one team i^k with $\alpha_i^k = 0$, and that at least one division contains 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 play-off position.

References

Ahuja, R. K., T. L. Magnanti, J. B. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications.* Prentice-Hall, Englewood Cliffs, NJ.

Gallo, G., M. D. Grigoriadis, R. E. Tarjan. 1989. A fast parametric maximum flow algorithm and applications. *SIAM J. Comput.* **18**(1) 30–55.

Gay, N. 1996. Giants officially leave the N.L. West race. *San Francisco Chronicle,* September 10, D1.

Gusfield, D., C. Martel. 1992. A fast algorithm for the generalized parametric minimum cut problem and applications. *Algorithmica* **7**(5–6) 499–519.

——, ——, D. Fernandez-Baca. 1987. Fast algorithms for bipartite network flow. *SIAM J. Comput.* **16**(2) 237–251.

Hoffman, A. J., T. J. Rivlin. 1970. When is a team 'mathematically' eliminated? H. W. Kuhn, ed. *Proc. Princeton Sympos. on Math. Programming.* Princeton University Press, Princeton, NJ.

McCormick, S. T. 1987. Two hard min cut problems. Technical report presented at the TIMS/ORSA Conference, New Orleans, LA.

——. 1999. Fast algorithms for parametric scheduling come from extensions to parametric maximum flow. *Oper. Res.* **47**(5) 744–756.

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

Robinson, L. W. 1991. Baseball playoff eliminations: An application of linear programming. *Oper. Res. Letters* **10**(2) 67–74.

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

Schwartz, B. L. 1966. Possible winners in partially completed tournaments. *SIAM Rev.* **8**(3) 302–308.

Wayne, K D. 2001. A new property and a faster algorithm for baseball elimination. *SIAM J. Discrete Math.* **14**(2) 223–229.