NOTES

A Natural Generalization of the Win-Loss Rating System

CHARLES REDMOND
Mercyhurst College
Erie, PA 16546
chad@mercyhurst.edu

How are teams in a tournament usually ranked? For most of the popular sports in the U.S., the percentage of wins is all-important, and the team with the highest winning percentage earns first place. Although this system is here to stay (and for many reasons, such as its simplicity, this is desirable), all sports fans recognize a nagging flaw: in a tournament that is not round-robin, so that each team plays only a subset of the others, a team with a weak schedule may have a considerable advantage over one facing strong opponents.

The sports world offers many possible remedies. Any reader of the sports section of *USA Today* is familiar with Jeff Sagarin. Sagarin's ratings, applied to numerous professional and collegiate sports, have enlightened fans since the mid-1980s and have officially guided both the NCAA basketball tournament selection committee and the college football Bowl Championship Series commission. It is clear that strength of schedule undoubtedly plays a role in this ranking scheme. For example, in his final ranking of NFL teams for the 2001 season, Sagarin **[8]** places Kansas City at 6-10 above Washington at 8-8. According to Sagarin, Kansas City faced the 16th most difficult schedule, whereas Washington opposed the 28th. Other ranking systems abound, including those by Richard Billingsley and Kenneth Massey, both employed by the Bowl Championship Series commission, and the time-honored Dunkel Index, which has been around since the 1920s.

The mathematical community has also tackled the problem, leaving a trail of research going back as far as Zermelo **[14]**. Driven by mathematical interest, as opposed to applicability, mathematicians tend to focus only on team performance and avoid building into their models factors such as home-field advantage, game location, recent team performance, and so on. The simplicity of this approach is preferred by many mathematicians, rather than the more complicated models used by sports professionals. Well-known authors in this genre are Keener **[4]** and Minton **[5]**. Keener uses his results to cast doubt on Brigham Young University's 1984 national football title, and Minton argues that Colorado should have stood alone in 1990, the year the Buffaloes shared the championship with Georgia Tech. For an introduction to this area of research, the reader should consult these articles as well as Stob **[10]**, which is an excellent survey of some previous advances in the area.

In the spirit of Barbeau **[1]**, Keener **[4]**, and Saaty **[7]**, at the heart of whose ranking schemes is a limiting process, and Minton **[5]**, who stresses point spread over point ratio, I wish to share with you my own system. It may not be applicable to the world of sports, but it does seem simple and natural. In my opinion, the scheme and its underlying theory are among the easiest to grasp in the literature. I hope therefore to provide a window into this mathematically intriguing subject for undergraduate students, especially those of linear algebra, as well as for anyone curious about by these questions. I am deeply indebted to the works of the authors mentioned above, especially Minton, with whom, although my approach is different, my rankings agree.

An illustrative example For the sake of clarity and to avoid tedious notation, I will limit my discussion to a simple example and leave the general formulation to the interested reader. Consider the following tournament of four teams in which each team has played two games.

Opponents	**Result**
A vs. B	5-10
A vs. D	57-45
B vs. C	10-7
C vs. D	3-10

B is 2-0, A is 1-1, D is 1-1, and C is 0-2. In other words, B's winning percentage is 1.00, A's and D's are each 0.50, and C's is 0.00. The traditional win/loss method of ranking places B in first place followed by A and D in second place and C in last place. Note that in the calculation of the winning percentages it is as if each team has been given 1 point for each win and 0 points for each loss. We have, however, ignored the possibility of a tie. The first revision, then, that we will suggest is that teams be given a score of 1 for each win, a score of 0 for each tie, and a score of -1 for each loss. Each team's rating would then be determined by the sum of its scores divided by the number of games played (2 in the case of our example). B's rating is still 1.00, but A's and D's are each now 0.00, and C's is -1.00.

With this amendment in place, we can now define the *dominance* of one team over another. Because B defeated A, we will say that B's dominance over A is 1. Conversely, A's dominance over B will be said to be -1. The average dominance of A over its opponents is its rating, 0.00, and so forth. There is still a flaw in this approach, however, since there is now an artificial limitation on one team's dominance over another. In our example, B defeats A by 5 points, but defeats C by only 3 points. Thus B's dominance of 1 over each team reflects imperfectly what has really happened. We will therefore make another change to our proposed ranking scheme by redefining one team's dominance over another to be its score in the game played minus its opponents score. B's dominance over A is then 5, whereas A's dominance over B is -5. We list each team's average dominance over the field of its competitors.

Team	**Average Dominance**
A	3.5
B	4
C	-5
D	-2.5

These new ratings are perhaps more descriptive, but they still don't factor in strength of schedule. Before tackling this problem, we need to make yet one more minor, but important, modification of our proposed ranking scheme. We will consider each team as having played one game against itself. Each team will receive a score of zero, of course, for this game. Though this requirement seems strange at first, the reason for it will be made apparent. Under this latest revision, our modified initial standings are listed in the table below.

Team	Rating
A	$7/3 = 2.33$
B	$8/3 = 2.67$
C	$-10/3 = -3.33$
D	$-5/3 = -1.67$

The fundamental idea Consider team A. It does not play C in the tournament, so we do not have any direct way of comparing the two teams. However, A plays D, and D plays C, so there is a path from A to C. A defeats D by 12 points, and D defeats C by 7 points. The fundamental idea is to consider an imaginary game to have been played between A and C where A defeats C by $12 + 7 = 19$ points. C could be called a *second-generation opponent* of A. Let us enumerate all second-generation games "played" by A by constructing a tree where each first-generation opponent of A emanates from A, and each second-generation opponent of A emanates from the first-generation opponent it plays. The appropriate dominance is assigned to each edge.

We are now considering A to have played nine games instead of three, and the scores of the games are obtained by adding down each branch of the tree as in FIGURE 1. The second-generation dominance of A is then determined by the average of these scores. We list the average second-generation dominances of all our teams.

Team	Average Second Generation Dominance
A	3.44
B	3.22
C	-4.11
D	-2.56

Note that A has moved into first place. Strength of schedule now plays a role in ranking teams since the ranking also depends on the performance of their first-generation opponents. Though a team might not have a stellar first-generational record, teams with difficult schedules will reap the benefits of their opponents' success. One interesting point is that some of these nine games are identical to an original real game. For example, the tree in FIGURE 1 lists the second-generation game $A \xrightarrow{0} A \xrightarrow{-5} B$, which is identical to the first-generation game $A \xrightarrow{-5} B$. This is one of the reasons that we require a team to be an opponent of itself. The original real comparisons are not lost.

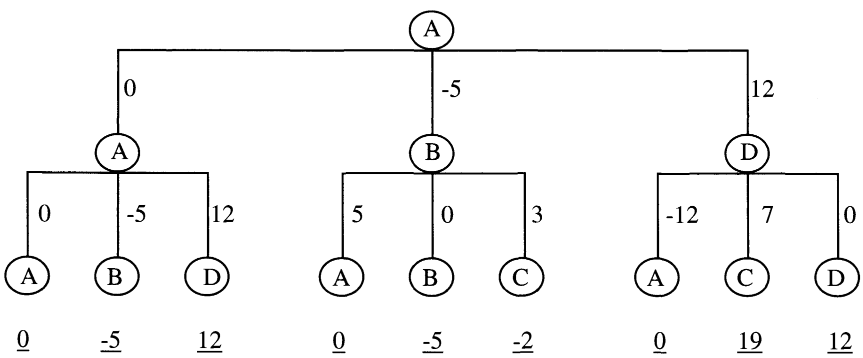

Figure 1 The average dominance of A over its second-generation opponents = 31/9 = 3.44

This process can be continued through any number of generations, and our intuition suggests that as the number of generations increases, our ranking becomes more accurate. Let $r_n(A)$ denote the average n-generation dominance of A or, equivalently, A's

n-generation rating. A natural and important question to ask at this stage is whether or not $\lim_{n\to\infty} r_n(A)$ exists. This limit, if it exists, is the ultimate rating we intend to assign to A. Here is where the mathematics becomes interesting and, I feel, surprising.

The mathematical formulation We illustrate the structure of our tournament with the graph in FIGURE 2 (note the edges from each team to itself).

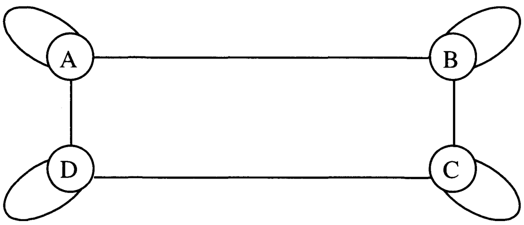

Figure 2

The incidence matrix for this graph is a 4 x 4 matrix with a row and column for each node. Each entry is a 0 or a 1, depending on whether or not the node for the entry's row is connected to the node for the entry's column. Our incidence matrix (where A corresponds to the first row and column, B to the second, etc.) is

$$M = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Now consider the square of the incidence matrix, M^2. It turns out that the (i, j) entry of M^2 is the number of distinct paths of length 2 between the node corresponding to row i and the node corresponding to column j. (The reader should think this through.)

$$M^2 = \begin{bmatrix} 3 & 2 & 2 & 2 \\ 2 & 3 & 2 & 2 \\ 2 & 2 & 3 & 2 \\ 2 & 2 & 2 & 3 \end{bmatrix}$$

Compare the first row of M^2 to FIGURE 1. The entries in the first row of M^2 indicate that A appears as a second-generation opponent of itself three times, whereas the other teams appear twice. This is exactly what FIGURE 1 shows. Other powers of M behave the same way: the (i, j) entry of M^n is the number of times the team corresponding to column j appears as an n-generation opponent of the team corresponding to row i.

Now define the vector

$$S = \begin{bmatrix} 7 \\ 8 \\ -10 \\ -5 \end{bmatrix}.$$

The coordinates of S are the net points scored by the teams in the tournament (for instance, the net number of points A scored is 7). Then the coordinates of the vector

$$\frac{1}{3}M^0 \cdot S = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 8 \\ -10 \\ -5 \end{bmatrix} = \begin{bmatrix} 2.33 \\ 2.67 \\ -3.33 \\ -1.67 \end{bmatrix}$$

are the first-generation ratings of our teams.

Using FIGURE 1, we can compute the second-generation ratings also. Observe that to calculate the second-generation rating of A, the second-generation dominances have to be added in once, whereas the first-generation dominances each have to be added in three times. The first coordinate of $3M^0 \cdot S$ will give us the latter sum, and the first coordinate of $M^1 \cdot S$ will give us the former. (The reader should pause to think this through.) Thus, the coordinates of the vector

$$\frac{1}{3^2}(3M^0 \cdot S + M^1 \cdot S) = \frac{1}{3}M^0 \cdot S + \frac{1}{3^2}M^1 \cdot S$$

yield the second-generation ratings for our teams. Likewise, the vector yielding the n-generation ratings has the formula

$$\sum_{j=1}^{n} \frac{1}{3}\left(\frac{M}{3}\right)^{j-1} \cdot S.$$

Does the limiting vector

$$\lim_{n\to\infty}\left(\sum_{j=1}^{n} \frac{1}{3}\left(\frac{M}{3}\right)^{j-1} \cdot S\right)$$

necessarily exist?

The limit To evaluate this limit (and show that it exists), we use an eigenvector decomposition of $M/3$. The eigenvalues of $M/3$ are 1, $-1/3$, and $1/3$ occurring with multiplicity 2. Because M is symmetric, we may choose a set of four linearly independent orthonormal eigenvectors corresponding to these eigenvalues. We may choose eigenvectors

$$\vec{v}_0 = \begin{bmatrix} 1/2, \\ 1/2, \\ 1/2, \\ 1/2 \end{bmatrix}, \vec{v}_1 = \begin{bmatrix} -1/\sqrt{2}, \\ 0, \\ 1/\sqrt{2}, \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0, \\ -1/\sqrt{2}, \\ 0, \\ 1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \vec{v}_3 = \begin{bmatrix} -1/2, \\ 1/2, \\ -1/2, \\ 1/2 \end{bmatrix}$$

corresponding to eigenvalues 1, $1/3$, $1/3$, and $-1/3$, respectively.

Since the eigenvectors form a basis, we may now express S as a linear combination of these eigenvectors. The coefficient of each eigenvector may be obtained by computing its scalar product with S:

$$\begin{aligned} S = \begin{bmatrix} 7 \\ 8 \\ -10 \\ -5 \end{bmatrix} &= (S \cdot \vec{v}_0)\vec{v}_0 + (S \cdot \vec{v}_1)\vec{v}_1 + (S \cdot \vec{v}_2)\vec{v}_2 + (S \cdot \vec{v}_3)\vec{v}_3 \\ &= 0\begin{bmatrix} 1/2, \\ 1/2, \\ 1/2, \\ 1/2 \end{bmatrix} - \frac{17}{\sqrt{2}}\begin{bmatrix} -1/\sqrt{2}, \\ 0, \\ 1/\sqrt{2}, \\ 0 \end{bmatrix} - \frac{13}{\sqrt{2}}\begin{bmatrix} 0, \\ -1/\sqrt{2}, \\ 0, \\ 1/\sqrt{2} \end{bmatrix} + 3\begin{bmatrix} -1/2, \\ 1/2, \\ -1/2, \\ 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 17/2 \\ 0 \\ -17/2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 13/2 \\ 0 \\ -13/2 \end{bmatrix} + \begin{bmatrix} -3/2 \\ 3/2 \\ -3/2 \\ 3/2 \end{bmatrix}. \end{aligned}$$

Note that the three vectors above are still eigenvectors of $M/3$ with eigenvalues $1/3$, $1/3$, and $-1/3$, respectively. Let us call these new eigenvectors $\vec{s_1}$, $\vec{s_2}$, and $\vec{s_3}$. In fact, for any positive integer j they are also eigenvectors of $(M/3)^j$ with associated eigenvalues $(1/3)^j$, $(1/3)^j$ and $(-1/3)^j$, respectively. Thus we can write

$$\begin{aligned}
\lim_{n\to\infty}\sum_{j=1}^{n}\frac{1}{3}\left(\frac{M}{3}\right)^{j-1}\cdot S &= \frac{1}{3}\lim_{n\to\infty}\sum_{j=0}^{n-1}\left(\frac{M}{3}\right)^{j}\cdot(\vec{s_1}+\vec{s_2}+\vec{s_3})\\
&= \frac{1}{3}\lim_{n\to\infty}\sum_{j=0}^{n-1}\left[\left(\frac{1}{3}\right)^{j}\vec{s_1}+\left(\frac{1}{3}\right)^{j}\vec{s_2}+\left(-\frac{1}{3}\right)^{j}\vec{s_3}\right]\\
&= \frac{1}{3}\sum_{j=0}^{\infty}\left(\frac{1}{3}\right)^{j}\vec{s_1}+\frac{1}{3}\sum_{j=0}^{\infty}\left(\frac{1}{3}\right)^{j}\vec{s_3}+\frac{1}{3}\sum_{j=0}^{\infty}\left(-\frac{1}{3}\right)^{j}\vec{s_3}\\
&= \frac{1}{3}\left(\frac{1}{1-1/3}\right)\vec{s_1}+\frac{1}{3}\left(\frac{1}{1-1/3}\right)\vec{s_2}+\frac{1}{3}\left(\frac{1}{1+1/3}\right)\vec{s_3}\\
&= \begin{bmatrix} 3.875\\ 3.625\\ -4.625\\ -2.875\end{bmatrix}.
\end{aligned}$$

The keys The reader can now easily see that the keys to establishing the limit are:

1. $M/3$ has an eigenvalue of 1 with a corresponding eigenvector having identical coordinates; this is because the rows of $M/3$ add to 1. In other words, the sum of each row in the incidence matrix is the number of games played by each team.
2. The coordinates of S add to zero, which is necessarily true of any vector obtained by our rankings, since the coordinates are the net number of points scored by the teams in a tournament. This fact and the key mentioned above guarantee that the coefficient of the eigenvector with eigenvalue 1 in the decomposition of S will be zero.
3. The absolute value of each of the other eigenvalues is strictly less than 1. This guarantees that the infinite series above converge. Why is this true of the other eigenvalues? This follows because $M/3$ is a *Markov matrix*: a matrix with non-negative entries and each column adding to 1. The eigenvalues of Markov matrices are very special. One eigenvalue is equal to 1, and the absolute value of each of the other eigenvalues is ≤ 1 . This inequality becomes strict if any power of the Markov matrix has all positive entries (see Strang **[11]** for more on Markov matrices). Because our tournament graph is connected (it wouldn't really make sense to consider a tournament for which this is not the case), given any team, every other team must appear as its opponent in some generation. Once a team appears as an opponent in some generation, it will appear in every generation after that because it is an opponent of itself. Consequently, some power of $M/3$ will have all positive entries. This is the primary reason we make this requirement at the outset.

Concluding remarks When I first worked on this problem, the notion of the teams opposing themselves had not occurred to me. When I used *Mathematica* to run the ranking algorithm on several tournaments, I did not always notice limiting behavior. I understand the mathematical necessity now for having teams oppose themselves, but a further explanation still eludes me.

The rankings produced by the scheme described in this note are identical to those produced by Minton's method [5], which requires the solution of a system of linear equations; in ours we calculate eigenvalues and eigenvectors. Both tasks could become difficult if the tournament is large. Ours, however, has the advantage of offering approximate rankings. One has just to run the algorithm out to a specified generation. This is not as accurate as calculating the limits, but the results would be perhaps fairer than those given by the win/loss ranking system.

What if the tournament is unbalanced in the sense that the teams do not all play the same number of games? By allowing the convention that teams may play any number of games against themselves, the number of games played can be made equal. For instance, suppose the tournament is just two games, A against B, and A against C. In this case, A has three opponents (B, C, and itself), whereas each of the others has two. To even matters, B and C can be required to play two games against themselves instead of one.

Some readers may object to the diminished significance of a win and the heightened importance of the score in our scheme. I admit that winning is part of the excitement of a game. Allotting a certain number of points to the victor just for prevailing, however, may preserve the value of a win. If many points are awarded for a victory, winning will drive the ranking of the teams.

Every point scored, though, is potentially important in the end. This is why teams, even with victory assured, must press on in each game to score as many points as possible. One may object to this feature of our system since it appears to advocate the humiliation of the loser. When only the win counts, which is the case in most tournaments today, a team losing by a large margin is embarrassed, because those extra points do not count. In our system, every point gained or lost could potentially make or break a team in the end. Even if the winner is decided early on, both teams must play as hard as possible for the entire game. A team might just qualify for the playoffs because of a heroic goal-line stand preventing a touchdown in a game it lost by 50 points. I see this as a positive aspect in that a game is never over or meaningless before its conclusion.

Finally, I would like to comment on the additive nature of the dominances we use. Since, in our example, A defeats D by a score of 57-45, we say that A is 12 points better than D. This is the additive approach, used also by Minton. A is then considered to be $(57 - 45) + (10 - 3) = 19$ points better than C, and so on. Other authors such as Barbeau [1], Keener [4], and Saaty [7] use a multiplicative approach; in such a ranking system, A is considered to be $57/45 = 1.27$ times better than D. A would then be $(57/45) \cdot (10/3) = 4.22$ times better than C. For contests in which each team plays both offense and defense, I much prefer the additive approach. In our tournament, although the scores of the respective games appear to be very different, the additive approach evaluates the strength of A against D (12 points) similarly to the way it evaluates the strength of D against C (7 points). To me this is reasonable. The game between A and D may be high scoring because they both have powerful offenses and weak defenses. On the other hand, the game between D and C may be low scoring because C is strong defensively and weak offensively. It seems unfair to allow the actual number of points scored to play a major role in comparing two teams. In the multiplicative approach, A is rated as being just 1.27 times better than D, but D is rated as being significantly better than C (3.33 times).

For teachers of introductory linear algebra and their students, I think what we've described is an interesting problem that highlights some of the topics encountered towards the end of the course. I hope that you find the time spent on it as rewarding as I have and that it enriches your enthusiasm for the subject.