Mathematica Militaris

Volume 22
Issue 1 *Volume 22, Issue 1 (Spring)*

Article 3

Spring 2017

An Application of Matrix Theory to Baseball Rankings

David Joyner
United States Naval Academy, wdj@usna.edu

An Application of Matrix Theory to Baseball Rankings

David Joyner
United Stats Naval Academy
Dedicated to the memory of TS Michael.

Introduction

In the *Application* section, we give a variation on Massey's ranking, applied to the U.S. Naval Academy 2016 men's baseball team. That's not quite the same as the typical application to football, where it's used for bowl rankings by the NCAA (see, for example, [1]). In that section, we use the data from the entire Patriot league, including the Patriot league championship tournament.

In the *Pre-tournament Ranking* section, we apply the same method but only use the pre-tournament Patriot league data.

In the Final section, we follow a formula explained to me by T.S. Michael. In this "multi-graph version," we record the win-loss record (a $+1$ for a win, -1 for a loss) in a 59×6 matrix M, one for each game. While more accurate in general, we find that the final ranking is the same as in *Pre-tournament Ranking* section of this paper.

Throughout, we roughly follow the presentation of Massey's method by C. Wessell (see [3]). Further examples of matrix-theoretic ranking methods applied to baseball can be found on the author's matrix theory course page at the USNA.

Keywords: Baseball, Matrix Theory

Application: (a variation on) Massey's ranking

In this section we give an application of orthogonal projection to the ranking of team sports. Massey's method, currently in use by the NCAA, was developed by Kenneth P. Massey while an undergraduate math major in the late 1990s. We present a variation of Massey's method adapted to baseball, where teams typically play each other multiple times.

In our application, we shall consider Patriot League men's baseball:

1. Army (U.S. Military Academy),
2. Bucknell,
3. Holy Cross,
4. Lafayette,
5. Lehigh,
6. Navy (U.S. Naval Academy).

The cumulative results of the 2016 regular season[1] are collected in Figure 11. The total score (since the teams play multiple games against each other) of the team in the vertical column on the left is listed first and the team in the horizontal row second. In other words, "a - b" in row i and column j means a runs were scored by team i against team j in all their games, and b runs were scored by team j against team i over all the games. For instance if X played Y and the scores were $10-0$, $0-1$, $0-1$, $0-1$, $0-1$, $0-1$, then the table would read $10-5$ in the position of X and Y.

First, we order the 6 teams as above. There are exactly 15 pairing between these teams. These pairs are sorted lexicographically, as follows:

$$(1,2),(1,3),(1,4),\ldots,(5,6).$$

[1] We count only the games played in the Patriot league, including the Patriot league tournament.

x\y	Army	Bucknell	Holy Cross	Lafayette	Lehigh	Navy
Army	×	14-16	14-13	14-24	10-12	8-19
Bucknell	16-14	×	27-30	18-16	23-20	28-42
Holy Cross	13-14	30-27	×	19-15	27-13	43-53
Lafayette	24-14	16-18	15-19	×	12-23	17-39
Lehigh	12-10	20-23	43-53	23-12	×	12-18
Navy	19-8	42-28	30-12	39-17	18-12	×

Figure 11: Sorted/ordered as Army vs Bucknell, Army vs Holy Cross, Army vs Lafayette, ..., Lehigh vs Navy.

That is to say, we sort them as

Army vs Bucknell, Army vs Holy Cross, Army vs Lafayette, ..., Lehigh vs Navy.

In this ordering, we record their (sum total) win-loss record (a 1 for a win, -1 for a loss) in a 15×6 matrix:

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

We also record their total losses in the entries of a column vector:

$$\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 10 \\ 2 \\ 11 \\ 3 \\ 2 \\ 3 \\ 14 \\ 4 \\ 14 \\ 10 \\ 11 \\ 22 \\ 6 \end{pmatrix}.$$

The Massey ranking of these teams is a vector $\mathbf{r}$ which best fits the equation

$$M\mathbf{r} = \mathbf{b}.$$

While this is over-determined, we can look for a best approximate solution using the orthogonal projection formula

$$P_V = B(B^tB)^{-1}B^t. \tag{1}$$

Unfortunately, in this case $B = M$ does not have linearly independent columns, so (1) does not apply.

Massey's clever idea is to solve

$$M^tM\mathbf{r} = M^t\mathbf{b} \tag{2}$$

by row-reduction and determine the rankings from the parameterized form of the solution.

To this end, we compute

$$M^tM = \begin{pmatrix} 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{pmatrix}$$

and

$$M^t\mathbf{b} = \begin{pmatrix} -24 \\ -10 \\ 10 \\ -29 \\ -10 \\ 63 \end{pmatrix}.$$

Then we compute the rref of

$$A = (M^tM \mid M^t\mathbf{b}) = \left(\begin{array}{cccccc|c} 5 & -1 & -1 & -1 & -1 & -1 & -24 \\ -1 & 5 & -1 & -1 & -1 & -1 & -10 \\ -1 & -1 & 5 & -1 & -1 & -1 & 10 \\ -1 & -1 & -1 & 5 & -1 & -1 & -29 \\ -1 & -1 & -1 & -1 & 5 & -1 & -10 \\ -1 & -1 & -1 & -1 & -1 & 5 & 63 \end{array}\right),$$

which is

$$rref(M^tM \mid M^t\mathbf{b}) = \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & -1 & -\frac{87}{6} \\ 0 & 1 & 0 & 0 & 0 & -1 & -\frac{73}{6} \\ 0 & 0 & 1 & 0 & 0 & -1 & -\frac{53}{6} \\ 0 & 0 & 0 & 1 & 0 & -1 & -\frac{92}{3} \\ 0 & 0 & 0 & 0 & 1 & -1 & -\frac{73}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

If $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6)$ denotes the ratings of Army, Bucknell, Holy Cross, Lafayette, Lehigh, Navy, in that order, then

$$r_1 = r_6 - \frac{87}{6},\ \ r_2 = r_6 - \frac{73}{6},\ \ r_3 = r_6 - \frac{53}{6},\ \ r_4 = r_6 - \frac{92}{6},\ \ r_5 = r_6 - \frac{73}{6}.$$

Massey adds the condition

$$r_1 + r_2 + r_3 + r_4 + r_5 + r_6 = 0, \qquad (3)$$

so

$$r_1 = -4,\ \ r_2 = -5/3,\ \ r_3 = 5/3,\ \ r_4 = -29/6,\ \ r_5 = -5/3,\ \ r_6 = 21/2.$$

Therefore

Lafayette < Army = Bucknell = Lehigh < Holy Cross < Navy.

Pre-tournament (Massey-like) ranking

We shall use the above method to determine the ranking before the Patriot league tournament. The ranking used by the Patriot league is simply the win-loss record:

Army (6-13) < Lafayette (7-13) < Bucknell (9-11)
< Lehigh (9-10) < Holy Cross (13-7) < Navy (15-5).

The pre-tournament matrix is displayed in Figure 12.

x\y	Army	Bucknell	Holy Cross	Lafayette	Lehigh	Navy
Army	×	14-16	14-13	14-24	10-12	8-19
Bucknell	16-14	×	27-30	18-16	23-20	10-22
Holy Cross	13-14	30-27	×	19-15	17-13	9-16
Lafayette	24-14	16-18	15-19	×	12-23	17-39
Lehigh	12-10	20-23	13-17	23-12	×	12-18
Navy	19-8	22-10	16-9	39-17	18-12	×

Figure 12: Regular season results, sorted/ordered as Army vs Bucknell, Army vs Holy Cross, Army vs Lafayette, . . . , Lehigh vs Navy.

Note that the Patriot league tournament involved only three teams, so Figure 12 only differs in relatively few entries compared to Figure 11. In this case, their total losses is:

$$\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 10 \\ 2 \\ 11 \\ 3 \\ 2 \\ 3 \\ 12 \\ 4 \\ 4 \\ 7 \\ 11 \\ 22 \\ 6 \end{pmatrix}$$

and

$$M^t\mathbf{b} = \begin{pmatrix} -24 \\ -8 \\ 3 \\ -29 \\ 0 \\ 58 \end{pmatrix}.$$

We then compute the rref of

$$A = (M^tM \mid M^t\mathbf{b}) = \left(\begin{array}{rrrrrr|r} 5 & -1 & -1 & -1 & -1 & -1 & -24 \\ -1 & 5 & -1 & -1 & -1 & -1 & -8 \\ -1 & -1 & 5 & -1 & -1 & -1 & 3 \\ -1 & -1 & -1 & 5 & -1 & -1 & -29 \\ -1 & -1 & -1 & -1 & 5 & -1 & 0 \\ -1 & -1 & -1 & -1 & -1 & 5 & 58 \end{array}\right)$$

to get

$$rref(A) = \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & -1 & -\frac{82}{6} \\ 0 & 1 & 0 & 0 & 0 & -1 & -\frac{66}{6} \\ 0 & 0 & 1 & 0 & 0 & -1 & -\frac{55}{6} \\ 0 & 0 & 0 & 1 & 0 & -1 & -\frac{87}{6} \\ 0 & 0 & 0 & 0 & 1 & -1 & -\frac{58}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

If $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6)$ denotes the rankings of Army, Bucknell, Holy Cross, Lafayette, Lehigh, Navy, in that order, then

$$r_1 = r_6 - \frac{82}{6},\ \ r_2 = r_6 - \frac{66}{6},\ \ r_3 = r_6 - \frac{55}{6},\ \ r_4 = r_6 - \frac{87}{6},\ \ r_5 = r_6 - \frac{58}{6}.$$

Using this and (3), one can obtain the Massey ratings of these (this is left as an exercise). Therefore,

$$\text{Lafayette} < \text{Army} < \text{Bucknell} < \text{Lehigh} < \text{Holy Cross} < \text{Navy}.$$

"Multi-graph" pre-tournament Massey ranking

This section includes strategies initiated by T.S. Michael.

In this multi-graph version, we record the win-loss record (a 1 for a win, -1 for a loss) in a 59×6 matrix M, one for each game. The display of the matrix is omitted as it won't fit on the page, but it's similar to the incidence matrix used in the previous sections. However, we do display the product

$$M^t M = \begin{pmatrix} 19 & -4 & -4 & -4 & -3 & -4 \\ -4 & 20 & -4 & -4 & -4 & -4 \\ -4 & -4 & 20 & -4 & -4 & -4 \\ -4 & -4 & -4 & 20 & -4 & -4 \\ -3 & -4 & -4 & -4 & 19 & -4 \\ -4 & -4 & -4 & -4 & -4 & 20 \end{pmatrix}.$$

We also must record the loss vector (which has length 59) $\mathbf{b}$, but it too is omitted as it won't fit on the page. It records the (positive) difference (number of runs of winner)-(number of runs of loser), one for each game. However, we do display the augmented matrix

$$A = (M^t M \mid M^t \mathbf{b}) = \left(\begin{array}{cccccc|c} 19 & -4 & -4 & -4 & -3 & -4 & -24 \\ -4 & 20 & -4 & -4 & -4 & -4 & -14 \\ -4 & -4 & 20 & -4 & -4 & -4 & 11 \\ -4 & -4 & -4 & 20 & -4 & -4 & -29 \\ -3 & -4 & -4 & -4 & 19 & -4 & -8 \\ -4 & -4 & -4 & -4 & -4 & 20 & 64 \end{array}\right),$$

as well as its rref:

$$rref(A) = \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & -1 & -\frac{122}{33} \\ 0 & 1 & 0 & 0 & 0 & -1 & -\frac{13}{4} \\ 0 & 0 & 1 & 0 & 0 & -1 & -\frac{53}{24} \\ 0 & 0 & 0 & 1 & 0 & -1 & -\frac{31}{8} \\ 0 & 0 & 0 & 0 & 1 & -1 & -\frac{98}{33} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

If $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5, r_6)$ denotes the rankings of Army, Bucknell, Holy Cross, Lafayette, Lehigh, Navy, in that order, then

$$r_1 = r_6 - \frac{976}{264},\ \ r_2 = r_6 - \frac{858}{264},\ \ r_3 = r_6 - \frac{583}{264},\ \ r_4 = r_6 - \frac{1023}{264},\ \ r_5 = r_6 - \frac{784}{264}.$$

Using this and (3), one can obtain the Massey rankings of these:

$$\text{Lafayette} < \text{Army} < \text{Bucknell} < \text{Lehigh} < \text{Holy Cross} < \text{Navy}.$$

Acknowledgements:

I think T.S. Michael for both encouraging comments and help on this paper. I also thank the anonymous referees for numerous suggestions which improved the paper.

Bibliography

[1] K. Massey, *Statistical models applied to the rating of sports teams*, available at masseyratings.com.

[2] Sagemath Developers, **Sagemath - a mathematical software package**, version 7.5, http://www.sagemath.org/.

[3] C. Wessell, *Massey's method*, available at public.gettysburg.edu/~cwessell/RankingPage/massey.pdf.