

Linear Algebra on the Field: A Study of Massey's and Keener's Method on Sports Team Rankings

Belle Schmidt and Rachael Olson

Introduction

Team rankings are an integral part of sports analytics. Applications of these rankings can range from evaluation of performance and team strategy to playoff predictions and sports betting. Through these real-world applications, we have seen a profound impact on not just the teams but their fans as well. Rankings spark conversations, new ideas, and further exploration across the board on how to find the best method. The search for the best method not only improves accuracy and interpretability, but highlights the importance of linear algebra methods in a real world context.

In this paper, we focus on utilizing and comparing advanced linear algebra based methods for ranking teams, specifically teams that competed in the 2021 through 2024 NFL seasons. Our primary motivation for this investigation was to get a better understanding of how the mathematical tools we learned in class can be applied in the real world. Methods such as Massey's and Keener's use linear algebra to translate game results and points into a ranking system.

There are two goals we would like to address throughout this paper. First, we aim to present linear algebra based methods, such as Least Squares Massey's Method and Keener's Method, in a coherent way by examining how their mathematical framework shapes rankings. Secondly, we wish to compare how each method handles real NFL data and evaluate their effectiveness in producing reasonable ranking results. Through that comparison, we highlight the strengths and weaknesses of each method in regards to team rankings. Overall, our main research question was: How can linear algebra based methods be used to construct data-driven, unbiased rankings and how do the methods compare to one another and to actual playoff outcomes?

For this analysis, we used a dataset that was scraped from ESPN's Team Stats. The dataset provides game by game data for the NFL seasons between 2002 and 2025. Originally it contained 61 variables, however, we only used information about the home and away teams and the home and away teams' scores. We decided to focus on the 2021 to 2024 seasons in our analysis. This made sense to us because they were the most recent years and we wanted to draw conclusions regarding the consistency of each method over time.

Massey's Least Squares Method

Massey's Least method was developed in 1997 by professor Kenneth Massey, who was an undergraduate student at Bluefield College. An example of this method being applied in the real world is the NCAA using it to rank college football teams and then assign them to bowl games at the end of the season. The main concept of this method is that it assumes that the point differential of a single game is linearly related to the difference in rankings between the two teams. Point differential being the difference in the number of points scored by a team compared to the number of points the opposing team scored. This concept is then used to build a system of linear equations which can be approximated to make team rankings.

Justin Wyss-Gallifent, a professor at the University of Maryland, wrote the following steps to explain the process of using Massey's Least Squares Method to rank teams within a league.

1. Construct the system of equations based on games and point differentials.
 - a. For each game we assume that the point differential is equal to the difference in the rating of the two teams who were involved in the game. If team i plays team j with the point differential being p , then $r_j - r_i = p$.
2. Convert equations into a matrix of $Ar = p$.
 - a. The matrix A has a row for each game and a column for each team. In each row, the home team receives a $+1$, the away team receives a -1 , while all other entries are 0. The vector p contains the point differentials for all games.
3. Form the least squares system $A^t Ar = A^t p$.
 - a. Due to the fact that teams typically play fewer games than the number of rankings, the system $Ar = p$ is inconsistent. Massey's least squares method uses least squares such as $A^t Ar = A^t p$.
4. Replace the last row of $A^t A$ with all 1s and the last entry of $A^t p$ with 0.
 - a. Replacing the last row with all 1s and setting the corresponding entry in the vector to 0 forces the system to have a unique solution and clarifies that the sum of the ratings are equivalent to 0, such that $r_1 + r_2 + \dots r_n = 0$.
5. Solve.
 - a. The resulting vector r contains the Massey evaluated ratings for each team providing a consistent ranking based on point differentials across the season.

Application of Massey's Least Method

To illustrate the steps from Massey's Least Method, we will use our NFL data to demonstrate how this method works. Specifically we will be using the 2024 NFL season in our example.

We started this process by making equations for all games that were played in the regular NFL season. All teams were assigned a variable, for example the Buffalo Bills were represented by r_1 , the Miami Dolphins were represented by r_2 , and so on. For each game, which corresponded to one row in our data, we made the coefficient of the home team $+1$, the coefficient of the away team -1 , and all other teams had a coefficient of zero. The solution of the equation was the point differential in the final score. If the home team won, the point differential was positive. If they lost, it was negative. Below are five sample equations from the 2024 NFL Season. In reality, we had 272 equations for each of the NFL seasons other than 2022, where a game was cancelled due to a medical emergency on the field.

$$\begin{aligned}
 r_1 - r_{29} &= 6 \\
 -r_4 + r_{31} &= 13 \\
 -r_1 + r_2 &= -21 \\
 r_3 - r_{32} &= -3 \\
 r_{29} - r_{31} &= 23
 \end{aligned}$$

Next we converted these equations into a matrix A and a vector p . Our matrix had 272 rows, one for each equation (a.k.a game), and 32 columns, one for each team. The vector p contained the solutions of each of the equations (a.k.a the point differentials for each game). Below you can see the matrix equation that arose from this.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ \vdots \\ r_{29} \\ r_{30} \\ r_{31} \\ r_{32} \end{bmatrix} = \begin{bmatrix} 6 \\ 13 \\ -21 \\ -3 \\ \vdots \\ 23 \end{bmatrix}$$

After trying to solve this set of linear equations, we found that there was no solution. This inconsistency is caused by a number of factors. One of the biggest being when a first team beats a second team and that second team beats a third team, but then the first team loses to the third team. For example, the Kansas City Chiefs beat the Baltimore Ravens and the Baltimore Ravens beat the Buffalo Bills but then the Buffalo Bills lost to the Baltimore Ravens. This in fact did happen in the 2024 season. Another cause for the inconsistency is when teams play the same team multiple times in a season. This happens every year in the NFL because all teams within a division play each other twice a season. Therefore, we proceed by finding the least squares solution instead.

Finding the least-squares solution means finding the team rank values such that the rankings and game match ups remain exactly the same while the point differentials are approximated to be as close to the actual point differentials as possible. The formula for least-squares is $A^t A r = A^t p$.

$$\begin{bmatrix} 1 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ 0 & -1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & -1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \\ \hat{r}_4 \\ \hat{r}_5 \\ \vdots \\ \hat{r}_{29} \\ \hat{r}_{30} \\ \hat{r}_{31} \\ \hat{r}_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \\ -21 \\ -3 \\ \vdots \\ 23 \end{bmatrix}$$

This gives us the equation below. Let's call it $M\hat{r} = \bar{q}$.

$$\begin{bmatrix}
2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
-1 & 0 & 0 & 0 & 2 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & -1 & 0 & 2 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\hat{r}_1 \\
\hat{r}_2 \\
\hat{r}_3 \\
\hat{r}_4 \\
\hat{r}_5 \\
\vdots \\
\hat{r}_{29} \\
\hat{r}_{30} \\
\hat{r}_{31} \\
\hat{r}_{32}
\end{bmatrix}
=
\begin{bmatrix}
27 \\
-21 \\
-3 \\
-13 \\
17 \\
\vdots \\
0 \\
-10 \\
3
\end{bmatrix}$$

There are a couple of things to note about matrix M . First, the entries m_{tt} represent the number of games that team t has played. This comes from the equation

$$m_{tt} = (A^T)_{t1}a_{1t} + (A^T)_{t2}a_{2t} + (A^T)_{t3}a_{3t} + \cdots + (A^T)_{tG}a_{Gt} = a_{1t}^2 + a_{2t}^2 + a_{3t}^2 + \cdots + a_{Gt}^2$$

and for all $1 \leq g \leq G$ we have that $a_{gt} \pm 1$ only if team t participated in game g . Note that G represents the number of games played.

For m_{st} , where $s \neq t$, the entry represents negative the number of games where team t played team s . The equation that tells us this is

$$m_{st} = (A^T)_{s1}a_{1t} + (A^T)_{s2}a_{2t} + (A^T)_{s3}a_{3t} + \cdots + (A^T)_{sG}a_{Gt} = a_{1s}a_{1t} + a_{2s}a_{2t} + a_{3s}a_{3t} + \cdots + a_{Gs}a_{Gt}$$

and $a_{gs}a_{gt} = -1$ if and only if both teams s and t are involved in game g .

Most importantly to note is that these equations are linearly dependent and that taking out two equations from the system of equations would make it unsolvable. We know that the equations are because you can determine one team's season results by

References

Dataset: Cviaxmiwnptr. (2025, February 12). NFL Team Stats 2002 - Feb. 2025 (ESPN). Kaggle. <https://www.kaggle.com/datasets/cviaxmiwnptr/nfl-team-stats-20022019-espn>

Massey's method introduction. (n.d.). <https://maherou.github.io/Teaching/files/CS317/masseyMethod.pdf>

Wyss-gallifent, J. (2021, July 21). Team ranking. https://www.math.umd.edu/~immortal/MATH401/book/ch_team_ranking.pdf

Jones, M., & Gan, E. (n.d.). Applications of matrices and linear algebra in predicting sports rankings. <https://maxwelljon.es/assets/pdf/Linear.pdf>