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

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 tea
```

```
There are two goals we would like to address throughout this paper. First, we aim to present linear alg
```

```
For this analysis, we used a dataset that was scraped from ESPN's Team Stats. The dataset provides game
```

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 $A^tAr = A^tp$.
   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 $A^tAr = A^tp$.
4. Replace the last row of A^tA with all 1s and the last entry of A^Tp 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 $r_1 + r_2 + ...r_n = 0$.

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.

```
$-  r_5 +  r_{14} = 7$
$r_{19} - r_{23} = 5$
$- r_{8}  + r_{25} = -8$
$r_1 - r_{29} = 6$
$- r_{12} + r_{21} = 7$
```

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 is a sample of what part of our matrix looks like.

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