

# Using Linear Algebra to Model Sports Rankings: WNBA 2025

A Least Squares Approach to Modeling Basketball Game Outcomes

Paityn Richardson

## Introduction

Within team sports, rankings play a large role in determining processes and outcomes throughout an athletic season. Whether it be determining tournament schedules or deciding draft pick order, a team's place in the league's standings can have a large impact on how their current and future season(s) play out. This is why it is critical to ensure that these standings are correctly calculated based on a team's performance throughout the season. Schedules are rarely balanced correctly and point differentials vary widely between games and teams, making it extremely difficult to get an accurate placement of a team's overall rank. Furthermore, the most widely used model of ranking based on win-loss records fails to account for important information such as how teams play against another specific team and how close or spread out the point differential was in the game. For example, blowout and narrow wins both still count as a win in the standings, which would consider two teams comparable, but in reality, the skill difference between these teams may be much larger. These facts are the driving motivator behind this project, where I use specific tools of linear algebra, such as systems of equations and least squares calculations, to build a more expansive and accurate ranking model of the 2025 WNBA season.

The basic process of this analysis involves translating each game into an equation that represents the strength differential between the two teams. If Team A beats Team B by 10 points in a game, then we can interpret Team A's "strength" as being 10 points higher than Team B's. When modeling all games, each game will be represented in a linear equation of the form:

$$(\text{score of team A}) - (\text{score of team B}) = \text{score differential (strength)}$$

Collecting all games from the 2025 WNBA season produces a large system of equations, which I write in matrix form as

$$Ax = b$$

where  $x$  represents the unknown team strengths. Solving this system becomes the foundation for an analytically justified ranking method.

## Mathematical Background

To conduct this computation, it is important that we have knowledge of a few foundational concepts in linear algebra. The first is the formation of the system  $Ax = b$ , where  $A$  is the design matrix representing the matchups,  $x$  is a vector containing the teams' "strength" ratings, and  $b$  is the vector of observed score margins. Because, in a season, there are more games than teams, the system will be considered "overdetermined", meaning there are more equations than unknowns in the entire system. This is why we will utilize the concept of least squares, as it does not make sense to attempt to satisfy each equation exactly. With least squares, we seek the strengths,  $x$ , that minimize the total squared error  $\|Ax - b\|^2$  (difference between what the model predicts and what actually happened in the game). The resulting normal equations  $(A^\top A)x = A^\top b$  form the basis of the Massey Method, which I will also utilize throughout the rest of the calculations.

There is one issue with this model alone, and that is that the matrix  $A^\top A$  is always **rank-deficient**, because the linear system cannot determine an absolute rating for each team. Conceptually, this means there is an entire line of solutions that would give identical predictions, which is not ideal. Mathematically, matrix  $A^\top A$  is considered rank-deficient because it has a non-trivial null space. This means there exists a non-zero vector,  $v$ , such that

$$A^\top Av = 0$$

To account for this, we need to add a constraint to the system. I will impose a constraint which requires the average team strength to be zero. Doing this makes the solution to the system unique and allows the strengths to be centered on a relative scale.

Together, I will use concepts from systems of equations, overdetermined models, least squares, rank deficiency, and constraint enforcement to provide the mathematical backbone for constructing and interpreting a linear-algebraic ranking system.

## Method/Model

### Massey Method

The specific model used in this project is based on the Massey method and the equivalent **game-row design matrix** formulation. The Massey method expresses each game as a linear equation relating the two teams' strengths to the final score margin. For a game where Team i plays Team j, and the margin is  $m = \text{score}_i - \text{score}_j$ , the resulting equation would be:

$$x_i - x_j = m$$

Stacking the equations from each game results in the matrix  $Ax = b$  where:

- $A$  is the **Massey Matrix**, which contains +1 for teamA, -1 for teamB, and 0 elsewhere.

- $x$  is the vector of team strengths
- $b$  is the vector of observed score margins

This matrix is typically tall (more games than teams), so the system is **overdetermined**, and we solve it using least squares.

### Constraint Enforcement

Because adding a constant to all team strengths leaves every predicted margin unchanged, the system has infinitely many solutions unless we add a constraint. In this project I will set the mean strength to zero such that:

$$\sum_i x_i = 0$$

### Data Overview

This project is centered around data from the WNBA 2025 regular season, which can be found [here](#). For each game, the dataset includes the two participating teams, their final scores, and the date of the game. From these records, I extract a set of matches among teams, compute the score differential (margin), and compile the full-season game list into a design matrix for linear-algebraic analysis. In the next section, I apply the methods described earlier to a small hand-worked example before turning to the full 2025 WNBA dataset.

### Massey Method & Least Squares Example

To model the mathematical processes utilized in this project, we will construct a small version of the WNBA dataset using only three games among three teams. Observe the following data that represents 3 rows out of the full regular-season data.

Team A	Score A	Team B	Score B
Minnesota Lynx	99	Dallas Wings	84
Minnesota Lynx	89	Los Angeles Sparks	75
Los Angeles Sparks	93	Dallas Wings	79

We can arbitrarily index the teams as follows:

- 1 = Minnesota Lynx
- 2 = Los Angeles Sparks

- 3 = Dallas Wings

Each game yields an equation of the form “score of Team A minus score of Team B equals margin.” Writing these equations in matrix form gives:

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} 15 \\ 14 \\ 14 \end{bmatrix}$$

Thus,

$$Ar = b$$

where vector  $r = (r_1, r_2, r_3)^\top$  is the unknown vector which represents the strength (ranking) of each team. Because the system is underdetermined (three games, three teams, but the matrix is singular), we impose the constraint  $\sum_i r_i = 0$  to obtain a unique solution. To solve this system we will use the following process of least squares. As stated before, minimizing our system using least squares gives us the normal equation:

$$(A^\top A)r = A^\top b$$

We can compute each piece starting with the transpose:

$$A^\top = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & -1 \end{bmatrix}$$

Now multiply,

$$A^\top A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

This matrix is symmetric and singular (its determinant is zero), which is why we need the constraint. Now let's compute  $A^\top b$

$$A^\top b = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & -1 \end{bmatrix} * \begin{bmatrix} 15 \\ 14 \\ 14 \end{bmatrix} = \begin{bmatrix} 29 \\ 0 \\ -29 \end{bmatrix}$$

Finally, this gives us the structure of the normal equations which is as follows:

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} * \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 29 \\ 0 \\ -29 \end{bmatrix}$$

To finish the solution, we must enforce the mean-zero constraint such that:

$$r_1 + r_2 + r_3 = 0$$

We will do this by replacing one of the equations with the constraint. That gives us a non-singular system:

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} * \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 29 \\ 0 \\ 0 \end{bmatrix}$$

Now this matrix is invertible, so the system has a unique solution. We can solve the system by elimination. From the constraint we see that:

$$r_3 = -(r_1 + r_2)$$

Now we can substitute this back into the first two equations and simplify to get:

$$r \approx \begin{bmatrix} 9.67 \\ 0 \\ -9.67 \end{bmatrix}$$

We would then interpret this to mean that the Minnesota Lynx are 9.67 “strength” points **above** the average, earning them the number 1 rank in this simulation. The Los Angeles Sparks serve as our middle (mean) reference point and are ranked second, while the Dallas Wings sit 9.67 strength points **below** the average and are ranked third.

This three-game example illustrates the core mathematical structure of the Massey method: each game contributes one equation, the unknowns are the team strength ratings, and the full system is solved by least squares with an appropriate constraint. To analyze the complete 2025 WNBA season, we apply this exact same procedure but at a much larger scale.

## Full Season Analysis

We will use the statistical and computational tools provided in the R programming language to extend this project full-scale.

## Data Wrangling

```
library(tidyverse)

game_data <- read_csv("wnba_2025.csv", col_names = c("Date", "teamA", "scoreA", "teamB", "scoreB"))

# Data Cleanup
game_data$scoreA <- as.numeric(game_data$scoreA)
game_data$scoreB <- as.numeric(game_data$scoreB)
game_data <- game_data |>
  slice(-1) |>
  select(teamA, scoreA, teamB, scoreB) |>
  mutate(margin = scoreA - scoreB)

## Vector of team names for design Matrix
teams <- game_data |>
  select(teamA, teamB) |>
  pivot_longer(cols = everything(), values_to = "team") |>
  distinct(team) |>
  arrange(team) |>
  pull(team)
```

## Building Design Matrix

```
# Build initial matrix with team names as col names
A <- matrix(0, nrow = nrow(game_data), ncol = length(teams))
colnames(A) <- teams

# Fill in the matrix with teamA/teamB values for each row(game)
for (i in 1:nrow(game_data)){
  A[i, game_data$teamA[i]] <- 1
  A[i, game_data$teamB[i]] <- -1
}

# Build Margin Matrix
b <- game_data$margin

## Apply the mean-zero constraint
A_constraint <- matrix(1, nrow = 1, ncol = length(teams)) # 1s across all teams
b_constraint <- 0
```

```
# Combine with original matrix/vector
A <- rbind(A, A_constraint)
b <- c(b, b_constraint)
```

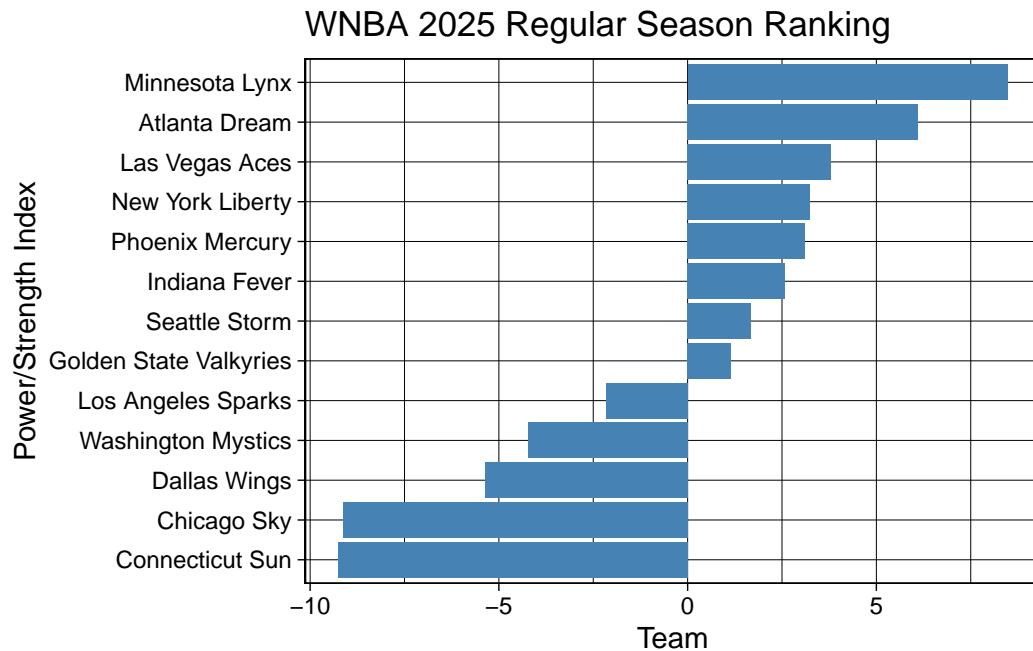
## Solution with Least Squares

```
# Solve Least Squares
team_strengths <- solve(t(A) %*% A, t(A) %*% b)
# Include team names
names(team_strengths) <- teams

# Sort in Descending Order
strength_data <- tibble(
  team = names(team_strengths),
  strength = as.numeric(team_strengths)
)
```

## Results

After this analysis, we can visualize the teams' strength rankings on a bar plot as such:



## Conclusion

In this project, I applied concepts from linear algebra to develop a data-driven method for ranking WNBA teams based on game score differentials. By constructing a design matrix and using least-squares solutions, I was able to estimate each team's relative strength in a way that accounts for all matchups across the season. This approach provides a mathematically grounded ranking system that goes beyond simple win-loss records, offering a continuous measure of team performance.

One key finding is that the linear-algebraic rankings sometimes differ from official standings where teams have identical win-loss records. For example, the Las Vegas Aces and Atlanta Dream both finished the season with 30–14 records, but the Massey-style analysis assigned slightly different strengths, ranking the Atlanta Dream higher, while the WNBA ranked the Aces higher. Similarly, the 4th and 5th place teams had tied records but were distinguishable using this method. These differences highlight how a linear-algebraic approach can provide a more nuanced and precise ordering of teams, resolving ties that standard win-loss metrics cannot.

This project demonstrates the power of linear algebra in sports analytics: systems of equations, least-squares solutions, and constraint enforcement can all be applied to real-world datasets to yield interpretable and objective rankings. Looking forward, this analysis could be further refined by incorporating additional factors, such as home-court advantage, weighted games based on recency, or capping margins to reduce the influence of outlier blowouts. Such extensions would allow for an even more accurate and flexible evaluation of team performance.

Overall, the project shows that linear algebra is not just an abstract mathematical tool, but a practical framework for understanding competitive sports and producing rankings that reflect relative team strengths more faithfully than traditional methods.