Fantasy Football Team Optimization

Aditya Joglekar
Aditya Rathi
Jaydutt Kulkarni
Jeffrey Wang

Mechanical Engineering
Engineering Optimization - 24785

October 28 2021

1 Abstract

In Fantasy Premier League, individuals manage their own squad of players, with points based on real-life player performance. Managers select a squad of players within a set budget and attempt to predict the optimal squad to maximize their team's performance. We aim to take an existing expected value model, and implement techniques from Engineering Optimization to create a model that predicts optimal squad performance. From this model, we aim to provide users with a squad which is most likely to outperform.

2 Introduction

Fantasy Premier League is a game which allows individuals to run their own, virtual team. Fantasy Premier League players (also known as managers) select 15 players from real Premier League teams and then "play" them over the course of a season to get points in-game. Points are based off real-life performance–a player who plays better in real-life will produce more points in their fantasy performance. Points are based off a number of statistics, such as goals scored, goals against (for defenders and goalkeepers) and assists. In order to have the best fantasy team, fantasy players must choose the squad which produces the most points collectively. The caveat, however, is that managers must not exceed a standard budget of $100 million each premier league player is given an associated value which fluctuates over the season, tied to overall player popularity on fantasy teams. Players who consistently produce higher point performances naturally cost more than players who do not perform as well. Thus, the goals of the fantasy manager is to maximize the return on investment of his squad by selecting players which will outperform their fellow players by providing more points at a cheaper price point.

Managers must select their squad of 15 from 4 different player positions:2 goalkeepers, 5 defenders, 5 midfielders, and 3 forwards. To clarify between fantasy teams and real Premier League teams, we will refer to a fantasy team as a squad, and Premier League team (such as Liverpool) as a club. As an additional wrinkle, managers must select their starting 11 from their overall squad of 15–only players in their starting lineup will count towards the fantasy team's performance. As players who perform better on average are inherently valued higher, the main objective of the manager is to find hidden value and maximize his performance per dollar. Some managers may elect to spend most of their budget on a few top-tier players and fill the remainder of their squad with average players, while others may try to spread the budget evenly between their squad.

While better premier league players may on average perform better in Fantasy Premier League, individual game performance is often much more nuanced. Team opponents, player health and fitness, and team strategy all may play roles in determining a player's fantasy impact. Fantasy managers need to weigh these criteria to formulate their own predictions about player performance. To help predict a player's expected value, managers can leverage tools which can take this data and more to provide an expected points value for a week. Based on these expected values along with the players' associated costs, managers can try to maximize their squad performance.

Our objective is to create a mathematical model which can provide a user with a prediction for overall squads which will maximize their performance. In the following sections, we will detail our proposed process as well as dive deeper into the various constraints and mechanisms of the problem.

3 Problem Statement

The objective of our optimization problem is to maximize the points scored by our Fantasy Premier League team while keeping in line with the numerous constraints introduced by the problem. First and foremost is the constraint of the total cost of obtaining said team which should be less than or equal to $100 Million. Secondly the total number of players in the team should be 11 in the starting lineup and 4 on the bench for a total of 15 players in the squad. Within these players there are further constraints based on the role of each player. The squad must have exactly 2 goalkeepers, 5 defenders, 5 midfielders and 3 forwards. Additionally, within the starting 11, there must be between 3 and 5 defenders, between 2 and 5 midfielders, and between 1 and 3 forwards such that the total number of players within the starting 11 is maintained. Finally, there

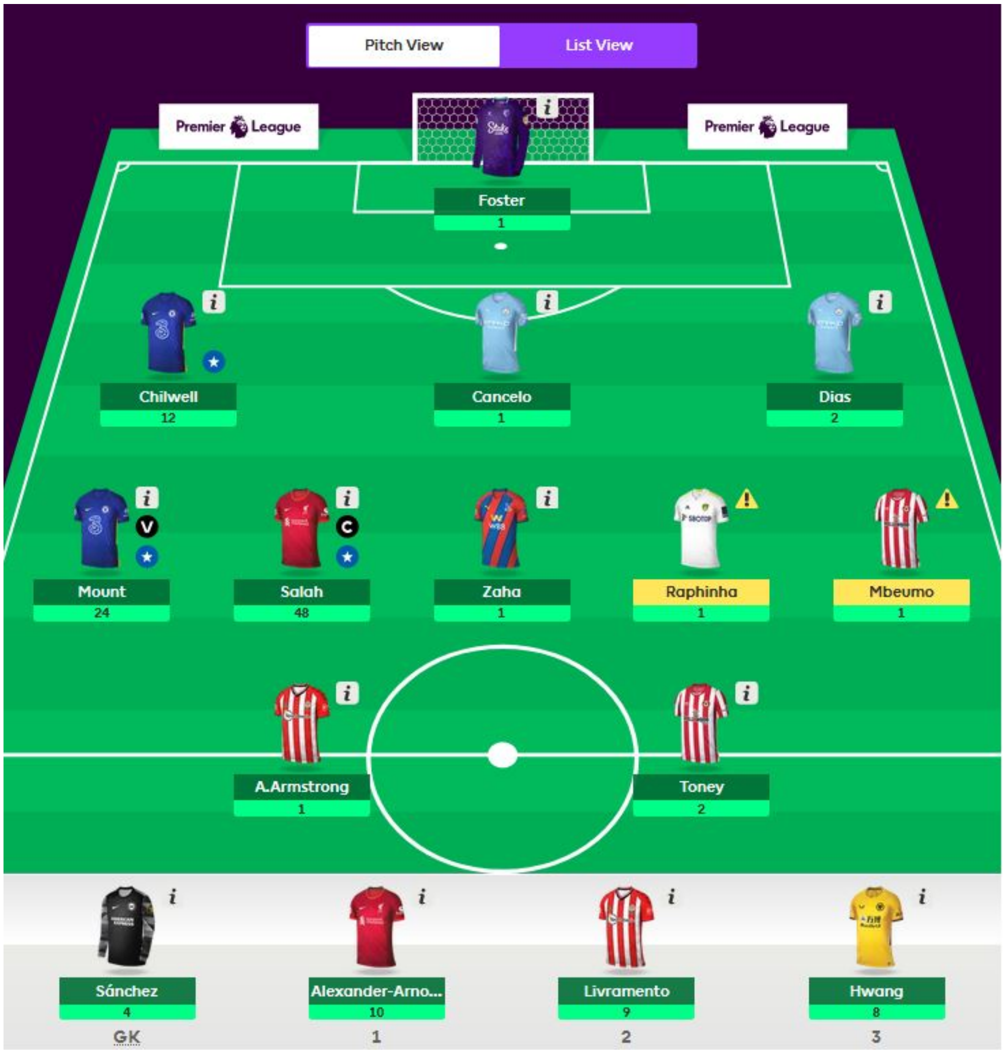

Figure 1: Sample Fantasy Premier League lineup

can be at most of 3 players from a single club (from which we pick our players) in our fantasy team. In most cases, a team's points will be the sum of the points scored by each member of his starting 11 squad. In some circumstances, there may be a scenario where a starter does not end of playing real-life and thus scores 0 points. In this case, a member of the bench will be automatically subbed on [3].

Our objective function uses weight matrices to convert the player selection into expected points. These weight matrices represent the characteristics of each player but in a team level fashion i.e. each entry is the mean value of the players from one club. As the first part of the project, we will be solving the continuous form of this problem. In our case we will assume that the number of players within a position on our squad can take continuous values within the ranges specified above. Further we will use a club based representation and not a player based representation i.e. we might have a solution wherein there are 2.3 players from club X. We will use multiple methodologies such as rounding approximations or probabilistic interpretations to create physical representations of the optimal solution. Naturally this will lead to more than one possible team and we plan to test all of them out to see which is the most optimal solution.

Mathematical Model

Minimize the negative of the expected score

$$\underset{\mathbf{X}\in\mathbb{R}^{160}}{minimize}\ f(\mathbf{X}) = -\sum_{i,j}\mathbf{W_{i,j}}\mathbf{X}_{i,j}$$

where $\mathbf{X}_{i,j} = \begin{bmatrix}\mathbf{x_{1_{i,j}}} \\ \mathbf{x_{2_{i,j}}}\end{bmatrix}$

Subject to:
Max budget = \$100 Million:

$$g_1(\mathbf{X}) = \sum(\mathbf{x_{1_{i,j}}}\mathbf{c_{i,j}} + \mathbf{x_{2_{i,j}}}\mathbf{c_{i,j}}) - b_s \le 0$$

Total Number of Players in Squad:

$$h_1(\mathbf{X}) = \sum_i\sum_j(\mathbf{x_{1_{i,j}}} + \mathbf{x_{2_{i,j}}}) - 15 = 0$$

Total Number of Players in Starting 11:

$$h_1(\mathbf{X}) = \sum_i\sum_j\mathbf{x_{1_{i,j}}} - 11 = 0$$

Number of Goalkeepers in Squad:

$$h_2(\mathbf{X}) = \sum(\mathbf{x_{1_{i,1}}} + \mathbf{x_{2_{i,1}}}) - 2 = 0$$

Number of Defenders in the squad:

$$h_3(\mathbf{X}) = \sum(\mathbf{x_{1_{i,2}}} + \mathbf{x_{2_{i,2}}}) - 5 = 0$$

Number of Midfielders in the squad:

$$h_4(\mathbf{X}) = \sum(\mathbf{x_{1_{i,3}}} + \mathbf{x_{2_{i,3}}}) - 5 = 0$$

Number of Forwards in the squad:

$$h_5(\mathbf{X}) = \sum(\mathbf{x_{1_{i,4}}} + \mathbf{x_{2_{i,4}}}) - 3 = 0$$

Maximum number Defenders in the starting 11:

$$g_2(\mathbf{X}) = \sum\mathbf{x_{1_{i,2}}} - 5 \le 0$$

Minimum number of Defenders in the starting 11:

$$g_3(\mathbf{X}) = -\sum\mathbf{x_{1_{i,2}}} + 3 \le 0$$

Maximum number of Midfielders in the starting 11:

$$g_4(\mathbf{X}) = \sum\mathbf{x_{1_{i,3}}} - 5 \le 0$$

Minimum number of Midfielders in the starting 11:

$$g_5(\mathbf{X}) = -\sum\mathbf{x_{1_{i,3}}} + 2 \le 0$$

Maximum number of Forwards in the starting 11:

$$g_6(\mathbf{X}) = \sum \mathbf{x_{1_{i,4}}} - 3 \leq 0$$

Minimum number of Forwards in starting 11:

$$g_7(\mathbf{X}) = -\sum \mathbf{x_{1_{i,4}}} + 1 \leq 0$$

Minimum number of Goalkeepers in the starting 11:

$$g_8(\mathbf{X}) = -\sum \mathbf{x_{1_{i,4}}} + 1 \leq 0$$

Maximum number of players in the squad from each club:

$$g_k(\mathbf{X}) = \sum_j \mathbf{x_{1_{k-8,j}}} - 3 \leq 0 \;\forall\; k \in [9, 28]$$

	Description	Value	Units
$\mathbf{x_{1_{i,j}}}$	Starting Lineup; i^{th} club; j^{th} position	-	-
$\mathbf{x_{2_{i,j}}}$	Bench Lineup; i^{th} club; j^{th} position	-	-
b_s	Starting Budget	100 Million	\$
$\mathbf{W_{i,j}}$	Weight matrix to convert to expected points per team per position	-	-
$\mathbf{c_{i,j}}$	Cost matrix to convert to money spent per team per position	-	\$
j=1	Goalkeeper	-	-
j=2	Defender	-	-
j=3	Midfielder	-	-
j=4	Forward	-	-
i	Club Number	-	-

Table 1: Symbols and Parameters

4 Analysis of Problem Statement

Number of Variables and Constraints

In our problem we have 160 decision variables, 5 equality constraints and 28 inequality constraints

Point in the Feasible Domain

Suppose we have club 1 in real life that has very high expected value players, but expensive players, and club 2 with lower expected value and less expensive players. Then, we cannot have all three allowed players from club1 in our team, due to the cost constraint. Also, there will be some players from club2 in our team as they are less expensive. Thus, a point in the feasible domain is:

$$
\mathbf{X_1} = \begin{array}{cc}
 & \begin{array}{cccc} Goalkeeper & Forward & Midfielder & Defender \end{array} \\
\begin{array}{c} Club1 \\ Club2 \\ Club3 \\ Club4 \\ Club5 \\ Club6 \\ Club7 \\ Club9 \\ Club10 \\ Club11 \\ Club12 \\ Club13 \\ Club14 \\ Club15 \\ Club16 \\ Club17 \\ Club18 \\ Club19 \\ Club20 \end{array} &
\begin{bmatrix}
\mathit{1.0} & \mathit{1.0} & 0.0 & 0.0 \\
0.0 & 0.0 & \mathit{1.0} & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0. \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & \mathit{1.0} & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & \mathit{1.0} \\
0.0 & 0.0 & \mathit{2.0} & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & \mathit{2.0} \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & \mathit{1.0} & \mathit{1.0} \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0
\end{bmatrix}
\end{array}
$$

$$
\mathbf{X_2} = \begin{array}{cc}
 & \begin{array}{cccc} Goalkeeper & Forward & Midfielder & Defender \end{array} \\
\begin{array}{c} Club1 \\ Club2 \\ Club3 \\ Club4 \\ Club5 \\ Club6 \\ Club7 \\ Club9 \\ Club10 \\ Club11 \\ Club12 \\ Club13 \\ Club14 \\ Club15 \\ Club16 \\ Club17 \\ Club18 \\ Club19 \\ Club20 \end{array} &
\begin{bmatrix}
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0. \\
\mathit{1.0} & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & \mathit{1.0} & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & \mathit{1.0} & 0.0 \\
0.0 & 0.0 & 0.0 & \mathit{1.0} \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0
\end{bmatrix}
\end{array}
$$

Functions not written explicitly

All functions have been explicitly defined in the problem statement.

Natural vs Practical Constraints

The rules of the real-life game of football are considered as the natural constraints. These include the total number of players that can play on the field and the minimum number of players per position. The practical constraints are the constraints implemented in Fantasy Premier League for making the game more engaging, which include the maximum amount spent on the players and the maximum number of players in the squad from each club.

Modelling Assumptions

The expected value of each player is assumed to be the one predicted by an existing machine learning model.

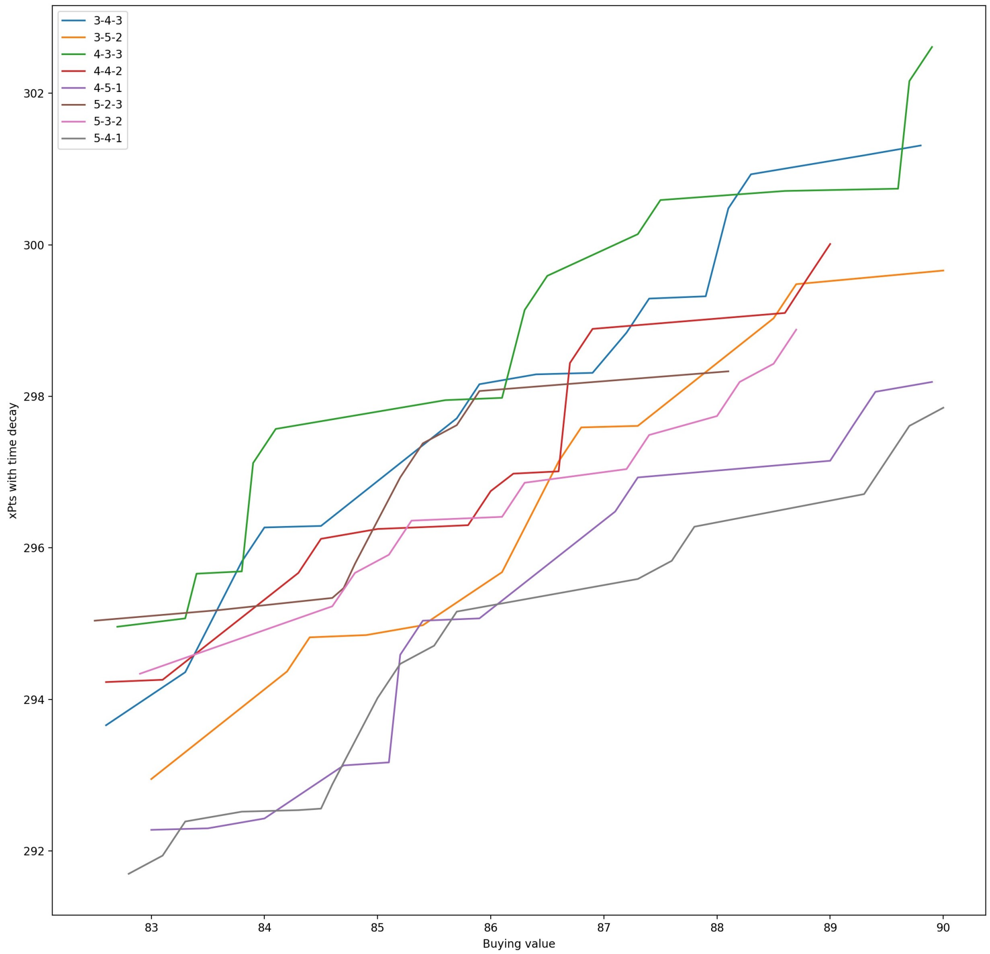

Figure 2: Parametric Study

Parametric Analysis

In the problem being addressed here, all parameters are well defined due to the rules of the game. Parameters such as number of players per position, number of players, overall budget, and number of players per team are all readily available. Cost of the squad is a parameter that could be subjected to a parametric study. If there are competing EV(Expected Value) scenarios, there is a case for selecting the squad with the lower cost.

A parametric study was conducted for all possible lineup formations, with Expected Value plotted against cost of the starting 11. This can help us make an informed decision w.r.t. squad investment.

Problem classification and nature of the functions

The first version of the problem is a Linear Programming problem. The nature of convexity is not applicable for this version of the problem. Therefore, multiple local minima will not exist. The function is smooth and continuous over the domain of the variable. It is not undefined at any point in the decision space due to continuous nature of the decision variables. Gradients are defined at all points in the decision space, which implies that the region will not lack gradient information.

References

[1] T. Matthews, S. Ramchurn, and G. Chalkiadakis, "Competing with humans at fantasy football: Team formation in large partially-observable domains," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, 2012.

[2] B. K. Kristiansen, A. Gupta, and W. Eilertsen, "Developing a forecast-based optimization model for fantasy premier league," M.S. thesis, NTNU, 2018.

[3] P. League. "Fantasy premier league rules." (), [Online]. Available: `https://fantasy.premierleague.com/help/rules`.

Appendix 1

For Phase 2 we plan to convert the problem into a nonlinear programming case by allowing the expected minutes played by each player on the squad to vary between fixed ranges for each club. Thus we will have two decision variables $\mathbf{X}$ and $\mathbf{X_m}$ multiplied in the objective function leading to the non-linearity, where $\mathbf{X_m}$ represents the expected minutes of a club position playing the given match.

To ascertain the effect of expected minutes on the selection of the squad we plan to run multiple instances of the simulation (in the form of Monte-Carlo runs). This is because the expected minutes correlates the probability of a certain player playing a match. By running Monte Carlo simulations we will be able to characterize the effects of all possible outcomes for that player ranging from him not playing at all to him playing for the entire 90 minutes.

It will be interesting to observe the effect of adding $\mathbf{X_m}$ as a variable on the bench players and the starting 11 as it will incorporate the chance of a certain player playing the game. This might result into cases where to find the optimal squad possible, it might make sense to spend a bit more on the bench players to get a stronger bench roster as they will perform better on the field if needed. Additionally, it is also possible that an excellent player has a low $\mathbf{X_m}$ value resulting in a different composition of the bench (players with higher $\mathbf{X_m}$) result into a more optimal squad.

Mathematically,

The objective function becomes:

$$\underset{\mathbf{X},\mathbf{X_m} \in \mathbb{R}^{160}}{minimize} \ f(\mathbf{X}, \mathbf{x_m}) = -\sum_{i,j} \mathbf{W_{i,j}} \mathbf{X}_{i,j} \mathbf{X}_{m_{i,j}}$$

Furthermore we will have constraints on $\mathbf{X_m}$ as follows:

$$\sum_{j} \mathbf{X_{m_{i,j}}} = 1 \quad \forall \ i$$

Appendix 2

In the third phase we hope to add multi-objective optimization to the problem at hand. In this case our primary objective will be to maximize the expected points scored by our squad and the secondary objective will be to minimize the total amount of money spent on procuring the squad.

The goal behind this is to check if there are multiple Pareto points that offer similar results with respect to the primary objective function while differing in the secondary objective, it would be beneficial from the squad management perspective to choose the cheaper squad.

Appendix 3

In the future, we may consider a multi-week approach to this problem. At the moment we are only predicting performance for the current week, but this model can be expanded on to include other mechanics from the Fantasy Premier League. Possible mechanics include variation in player costs, different game modes, and multi-week planning.