Team Selection using Multi-/Many-Objective Optimization with Integer Linear Programming

Shelvin Chand[1], Hemant Kumar Singh[2], Tapabrata Ray[3]
School of Engineering & Information Technology, The University of New South Wales, Australia
Email: Shelvin.Chand@student.adfa.edu.au[1], H.Singh@adfa.edu.au[2], T.Ray@adfa.edu.au[3]

***Abstract*—Assembling a competitive team is a task encountered in many professional league sports such as cricket, soccer, rugby etc. Teams are assembled annually with players being bid for by competing franchises. While stochastic optimization approaches for team selection have been suggested in the past, the approximate nature of these techniques could be disadvantageous for team selection when the stakes are high. In this paper, we explore the use of multi-objective integer programming approach to alleviate this issue and deliver a set of optimal trade-off solutions (teams). We illustrate the performance of the approach using professional Twenty20 cricket league data from the Indian Premier League. We also demonstrate the ability to support partial team construction, i.e., selecting few members of the team with others unchanged. Lastly, we also present a way to rank the importance of the players within a team considering the key objectives.**

I. Introduction

Winning and losing in modern day sports is not only reflected on the leader-board but has significant consequences on potential future earnings. For example, in the English Premier League (EPL), the winning team in 2014/2015 season, Chelsea Football Club, had a total revenue close to £99 million ($140 million) [1]. Other similar examples can be drawn from cricket, baseball, rugby etc., in which teams compete for big prize money, while simultaneously receiving broadcast revenue, ticket revenue and merchandise revenue. All of these things are of course to a certain extent influenced by the team's performance on the field. Therefore, selection of a strong team is an important decision-making problem for the team management.

Cricket is a popular team sport in which two teams of 11 players compete. About 15 years ago, the game expanded to the new Twenty20 (T20) format, in which the matches are shorter i.e only 20 overs per team, compared to the classic formats of 50 overs (one-day) or test (five-day) matches. This format is designed to attract the audience who would not be inclined to watch long matches. In 2008, the Board of Control for Cricket in India (BCCI) launched a T20 format league called the Indian Premier League (IPL). As of 2010, the IPL was valued at $4.13billion and was the second highest paying league in the world [2]. The participating teams were auctioned off to high profile/celebrity owners at stellar prices. For example, during the 2010 auction, Pune and Kochi franchises were sold for over $330 million each [2]. To put this into perspective, Liverpool from the English Premier League, which is considered by many as the most successful club in English football, in 2007, was bought for around £219 million ($326.3 million) [2].

After acquiring the franchise, the owners/managers have the responsibility of not only building a winning team but also focus on the business aspects of the franchise. Ticket sales, merchandise sales, apparel and equipment sponsorship and winning prize money are all strongly related to the teams performance. The players of the team are usually "bought" through an auction process. As pointed out by Ahmed et al. [3], building a team from a large pool (120+) of available players is not a trivial task as the number of possible combinations are huge ($> 10^{15}$). Since the owners would be investing millions of dollars on buying the players, they would ideally want to have an efficient way to explore their options and identify potential trade-offs that may exist.

Multi-objective optimization (MO) [4] would be ideally suited to deal with such problems as it delivers a set of trade-off solutions that are useful from a decision making point of view. Even so, most of the methods proposed to date for team construction/selection are single-objective. In [5], a standard genetic algorithm was used to build an optimal cricket team from a set of 50 national level Indian cricket players. The teams constructed using their approach were very similar to the One Day International (ODI) team of India, with the exception of one or two players which improved the balance of the team. In [6], a simulated annealing based approach was used for generating optimal line-ups in T20 cricket. Their optimal line-up consisted of three main components, namely team selection, batting order and bowling order. The objective was to select a team line-up that produces the greatest expected difference between runs scored and runs allowed. The use of exact single-objective techniques, such as integer programming has also been reported in a few studies [7, 8, 9] for optimal cricket team selection. Besides cricket, optimization has also been used in other team sports, a bibliographic review of which can be found in [10].

To the authors' knowledge, the only existing works on multi-objective team selection are reported in [11, 3]. They use NSGA-II [4] to optimize the overall bowling and batting strength of a team with 11 players. They applied their algorithm for a set of 129 players which were auctioned for the 4th edition (2011) of the IPL. Each player's performance was represented using actual T20 data. After generating the trade-off front, they used innovization principles to uncover crucial insights to assist managers in assembling a high performing

978-1-5090-6017-7/18/$31.00 ©2018 IEEE

team. Finally, they also focused on team selection as a decision making task in which they aim to identify good teams based on various performance or economic factors. At their core, evolutionary algorithms rely on a stochastic approach with no guarantee that the output is indeed Pareto optimal. Like any other stochastic approach, several instantiations may be needed to establish if the output can be trusted. With substantial investments on the line, team owners are often reluctant to use a stochastic approaches which could rely significantly on the chosen set of algorithm parameters.

In this paper, we investigate a multi-objective formulation of the team selection problem using integer linear programming (ILP) approach for selecting optimal cricket teams for the IPL. This is motivated by the fact that for a problem formulated suitably, ILP can provide an optimal solution instead of an approximate one. We use actual player data from the 2011 season to select players based on a number of several common forms of conflicting objectives. We also demonstrate how the procedure can be used for decision making with specific user preferences. Subsequently, we also explore the means to identify the most valuable player in a given team. While in the IPL format, players cannot be bought/sold within the season, such indicators can be used to offer non-contractual incentives.

The rest of the paper is organized as follows. Section II gives the detailed problem description and discusses the proposed multi-objective integer linear programming formulation. Section III presents a discussion on the results obtained. In Section IV, a player ranking process is given to judge their importance to the team. Finally, Section V presents the concluding remarks.

II. Proposed Approach

A. Problem Description

The IPL team selection problem can formally be described as follows. Select 11 players from a pool of available players in the IPL auction. Each player has a batting performance, bowling performance and a given cost. Within the list of available players, certain players are capable of playing as a wicket keeper. Each team needs to have minimum of one player who can play as a wicket keeper. Any chosen team cannot have more than 4 international (non-Indian) players. The selected team must also have one player who is capable of being the captain of the team. In the following paragraphs we give details of the formulation.

The first objective (Equation 1) is the combined bowling average of the team. The bowling average for a given player is calculated as total number of runs conceded by the bowler divided by the number of wickets taken by the bowler [3]. A lower average is considered to be better.

$$Minimize \quad \sum BowlingPerformance \tag{1}$$

The second objective (Equation 2) is the combined batting average of the team. For a given player this is calculated as the total number of runs scored divided by the number of times the player has been out [3]. A higher batting average is preferred.

$$Minimize \quad \sum -BattingPerformance \tag{2}$$

Non-bowlers are given a penalty bowling average of 100, while non-batsman are given a penalty batting average of 0. This allows us to prevent players who have attained good averages over a few matches to gain advantage or to be ranked higher than those experienced players who have accumulated similar numbers over a greater number of matches.

The next objective is the total cost(Equation 3) of the team. Each player is assigned a monetary value which indicates the cost involved in acquiring that player. Ideally, one would aim to minimize the total team cost.

$$Minimize \quad \sum Cost \tag{3}$$

We also include two additional objectives which were briefly mentioned by [3] but not considered in a true many-objective scenario. These two objectives are combined star power and combined fielding average.

Within the given list of players there are certain players who can be considered as ‘star’ players (Equation 4) i.e. they carry a certain brand-value. They may assist in increasing merchandise sales or help in attracting more fans to the stadium. Each player has a 0 or 1 binary value assigned, where 1 indicates the player is considered a ‘star’player and 0 indicates otherwise. A team would ideally aim to increase ‘star’power.

$$Minimize \quad \sum -StarPower \tag{4}$$

After bowling and batting performance, the fielding performance (Equation 5) is the next most important performance statistic in cricket. Every player in the team has a fielding average which is calculated by dividing the total number of catches taken by the total number of matches played. Through the optimization we aim to improve fielding performance.

$$Minimize \quad \sum -FieldingPerformance \tag{5}$$

Note that there are certain objectives such as batting, fielding and star-power in which a higher value indicates better quality. However, within our ILP procedure we consider all objectives as minimization objectives and hence we have taken the negative value for those objectives. Other objective functions (provided they are linear) can also be likewise added to the formulation.

B. Solution Formulation

The multi-objective procedure involves solving multiple ILPs in order to generate the trade-off set of solutions. We use the classical ε-constraint approach, where one objective is minimized/maximized, while imposing a constraint on the other objective(s). A solution is represented as a binary vector of size NP (size of pool of players), where there are eleven 1's and remaining elements of the vector are 0's. Let us start off by considering the simple two-objective case with batting and bowling performance as the two objectives. We use batting performance as the objective to be minimized and bowling performance as the constraint. The first step is to identify the range for the bowling objective, i.e. the minimum (*min*) and maximum (*max*) achievable objective values. The given range is then divided evenly to obtain α values for bowling constraint with upper value as UB. The value of α can be varied based on the required density of the trade off solutions. We then attempt to solve α ILPs to minimize the batting objective with each ILP using one value from the given set of UB values as a constraint. This process can then be reversed with bowling as the objective to be minimized and batting as the constraint. Once a set of trade-off solutions have been obtained the final step is to eliminate any dominated(weak Pareto optimal) solutions through non-dominated sorting. Iterating through all objectives ensures that we eliminate as many weak Pareto optimal solutions as possible. However, a few may still remain depending on the resolution of the search. We are not particularly concerned though with weak Pareto solutions as any solution on the front is mainly a starting point to building a good team. This is because there are a number of factors such as player partnerships which this algorithm is not able to account for. Ideally this tool will be useful for analyzing trade-off information and providing good starting points for building strong teams. Human input will obviously be required to further enhance the quality of teams picked. For this reason the partial-team building approach discussed later in the paper will be particularly useful as it incorporate user preferences into the search.

In a similar way, more objectives can be incorporated where one objective is minimized/maximized, while others are used as constraints. In case of three or more objectives, all possible combinations of the UB values (based on α) for each constraint/objective have to be used for generating a set of trade-off solutions. It is equivalent to iterating through the range for all objective values based on α. The overall procedure is summarized in Algo. 1.

1) Note with regards to Algorithm 1::

- *other* refers to all the set of objectives which act as constraints.
- The set c for each objective k refers to the set of UB values for objective k.
- *exitFlag=1* indicates that the solver was able to find a feasible and optimal solution.

The mathematical formulation for the ILP is given below:

2) Definitions::

- NP: Total number of players.
- TS: Team Size.
- i: Represents player index ($i = \{1, ..., NP\}$).
- j: Represents constraint index ($j = \{1, ..., \alpha\}$).
- CT: Vector containing transfer cost of all players.
- BL: Vector containing the bowling average of all players.
- BT: Vector containing the batting average of all players.
- SP: Binary vector containing star-player information. $SP_i = 0$ means player i is not considered a star player and $SP_i = 1$ means player i is considered a star player.
- FD: Vector containing the fielding average of all players.
- IC: Binary vector containing captaincy information. $IC_i = 0$ means player i cannot be captain and $IC_i = 1$ means player i can be captain.
- IW: Binary vector containing wicket keeper information. $IW_i = 0$ means player i cannot play as wicket keeper and $IW_i = 1$ means player i can play as wicket keeper.
- IIP: Binary vector containing nationality information. $IIP_i = 0$ means player i is not an international player and $IIP_i = 1$ means player i is an international player.
- IBL: Binary vector containing bowler information. $IBL_i = 1$ means that the primary role for player i is to play as bowler.
- IBT: Binary vector containing batsman information. $IBT_i = 1$ means that the primary role for player i is to play as batsman.
- UB_{CT}: Upper-bound constraint value on total team cost.
- UB_{BL}: Upper-bound constraint value on combined bowling average for the team.
- UB_{BT}: Upper-bound constraint value on combined batting average for the team.
- UB_{SP}: Upper-bound constraint value on combined star power for the team.
- UB_{FD}: Upper-bound constraint value on combined fielding average for the team.
- min_{BL}/max_{BL}: Minimum/Maximum achievable values for bowling objective.
- min_{BT}/max_{BT}: Minimum/Maximum achievable values for batting objective.
- min_{CT}/max_{CT}: Minimum/Maximum achievable values for cost objective.
- min_{SP}/max_{SP}: Minimum/Maximum achievable values for star-power objective.
- min_{FD}/max_{FD}: Minimum/Maximum achievable values for fielding objective.

A decision vector $\mathbf{x}$, of size NP, with binary variables is used in which:

- $x_i = 1$: Player i is selected to be part of the team.
- $x_i = 0$: Player i is not selected to be part of the team.

Note that there may be players for which both IBL and IBT are equal to one indicating that the player is an all-rounder.

3) Two-Objective Formulation: The two objective formulation uses bowling performance and negative batting performance as the two objectives to be minimized. The formulation is given in Table I which lists out the constraints being used with each objective. In addition to the constraints given in Table I, the common constraints from Table IV are also applied to obtain a valid team composition. The whole process involves solving two sets of single-objective ILPs. First we solve a set of ILPs with bowling performance as the objective and batting performance as the constraint with the number of ILPs being solved equal to α. The same process is repeated with batting

Algorithm 1 Proposed multi-objective team selection

```
Variables: α{Resolution for each objective}; TS{Team Size};
  numWK{Number of Wicket Keepers}; numCptn{Number of Captains};
  numIP{Number of International Players}; numObj{Number of Objectives}
Find the min and max for each objective with constraint on TS, numWK,
  numCptn, numIP
for i = 1 : numObj do
    Create constraint set c for each of the other objectives based on α
      and min, max
    Create matrix m which contains all possible combinations of c for all
      other objectives
    for j=1:numRows(m) do
        [x, fval, exitFlag]=Min(Objective i) s.t. constraint set j
        {Check if solution is feasible and optimal based on exitFlag}
        if exitFlag==1 then
            Evaluate x for all other objectives
            Add solution into fset
        end if
    end for
end for
ndSet=Unique(nonDominatedSorting(fset))
```

performance as the objective and bowling performance as the constraint. The solutions from these two sets are combined and non-dominated sorting is then applied to obtain the final set of optimal solutions.

4) Three-Objective Formulation: The three objective formulation given in Table II uses cost along with bowling and negative batting performance as the objectives to be minimized. Evidently, because of the added cost component the number of constraints for each objective have increased. Just like the previous formulation, the common constraints from Table IV also apply. Three sets of single-objective ILPs are solved. First, with bowling performance as the objective and all combinations of batting performance and cost objective values as the constraints. This is then repeated for the other two objectives and finally the sets are combined and the non-dominated solutions are extracted to obtain the trade-off front.

5) Five-Objective Formulation: The five-objective formulation includes all the objectives discussed in Section II-A. When any one objective is being minimized, all other objectives are used as constraints. The common constraints from Table IV also apply. Just like with the previous two formulations, multiple sets of single-objective ILPs are solved and the solutions are then combined to obtain the final trade-off front.

III. Optimization Case Study

A. Experimental Set-up

The data used in this paper is the same as the one used in [11], consisting of 129 players from the 2011 season. To solve the ILPs we used the CPlex solver [12]. All examples are based on TS =11. All cost values are in US Dollars (USD). The code and data used in this paper can be obtained on request from the authors. The value of α used within the examples are just for demonstration purposes. It can be chosen based on the users' preference on the required density of the obtained solutions. α can be considered equivalent to the population size parameter in evolutionary algorithms. A larger value gets a better representation of the front. For reference, the total number of possible team combinations is $\binom{129}{11} \approx 2.66 \times 10^{15}$.

Table V gives details for the Chennai Super Kings (CSK) team of 2011. All references to CSK in the following sections are based on this team.

B. Two-Objective Formulation

To test our proposed formulation and solution methodology, we first generate the complete Pareto front for the two-objective formulation. Our objectives are to minimize the combined bowling average and negative combined batting average of the team. The α value is set to 100 for both objectives.

The resulting front had 81 non-dominated solutions. The computation time was 5.04 seconds. The convex front has a fairly visible knee. Knee solutions [13] are often the most preferred solutions on a Pareto front. Interestingly, the cost of each team on the front (Fig. 1 (a)) reveals that the most expensive teams lie around the knee region. Also, batting oriented teams cost more than bowling oriented teams.

We have also plotted within the figure the objective values for the 2011 season's winning team, the Chennai Super Kings (Table V). One can clearly see that the team is dominated with respect to the solutions on the front and hence there was still scope for further improvement. Fig. 1 (b) also shows the similarity of the solutions on the front to CSK. The similarity is measured in terms of the number of players that are in common with CSK. Teams in and around the knee region are mostly dissimilar.

C. Three-Objective Formulation

The three-objective formulation involves the minimization of combined bowling average, (negative) combined batting average and total team cost. α value of 50 is used for bowling and batting objectives and 20 for cost objective.

We obtained a convex front (Fig. 2) containing 982 non-dominated solutions within 88.74 seconds. The most expensive team costs \$14.6million. This means that any owner/management spending more than \$15million will not see any improvement in team quality. Options which allow for a more balanced (knee-like) trade-off between the performance objectives usually cost more than solutions which only favor one of the performance objectives. One can also see that CSK still remains a dominated solution even with the introduction of the third objective.

D. Preferred Players

Team building is a complex process in real life. Team management often have preferences on players they would want included in the team regardless of their cost or performance. This may be because of various reasons such as brand value, contractual obligations, good relationship with management, etc.

Let us assume that the owner wants to build a team around two key players, namely Virat Kohli (India–Player ID:122) and AB de Villiers (South Africa–Player ID:108). We also assume

TABLE I: Two-objective formulation

Objective 1 (Bowling)	Objective 2 (Batting)
$Minimize \quad \sum_{i=1}^{NP} BL_i x_i$ $s.t.$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$	$Minimize \quad \sum_{i=1}^{NP} -BT_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$

TABLE II: Three-objective formulation

Objective 1 (Bowling)	Objective 2(Batting)
$Minimize \quad \sum_{i=1}^{NP} BL_i x_i$ $s.t.$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$	$Minimize \quad \sum_{i=1}^{NP} -BT_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$
Objective 3 (Cost)	
$Minimize \quad \sum_{i=1}^{NP} CT_i x_i$ $s.t.$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]; \quad \sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$	

TABLE III: Five-objective formulation

Objective 1 (Bowling)	Objective 2 (Batting)
$Minimize \quad \sum_{i=1}^{NP} BL_i x_i$ $s.t.$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$ $\sum_{i=1}^{NP} -SP_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{SP}, max_{SP}]$ $\sum_{i=1}^{NP} -FD_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{FD}, max_{FD}]$	$Minimize \quad \sum_{i=1}^{NP} -BT_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$ $\sum_{i=1}^{NP} -SP_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{SP}, max_{SP}]$ $\sum_{i=1}^{NP} -FD_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{FD}, max_{FD}]$
Objective 3 (Cost)	**Objective 4 (Star-Power)**
$Minimize \quad \sum_{i=1}^{NP} CT_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$ $\sum_{i=1}^{NP} -SP_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{SP}, max_{SP}]$ $\sum_{i=1}^{NP} -FD_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{FD}, max_{FD}]$	$Minimize \quad \sum_{i=1}^{NP} -SP_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$ $\sum_{i=1}^{NP} -FD_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{FD}, max_{FD}]$
Objective 5 (Fielding)	
$Minimize \quad \sum_{i=1}^{NP} -FD_i x_i$ $s.t.$ $\sum_{i=1}^{NP} BL_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{BL}, max_{BL}]$ $\sum_{i=1}^{NP} -BT_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{BT}, max_{BT}]$ $\sum_{i=1}^{NP} CT_i x_i \leq UB_j \quad \forall j, \quad UB_j \in [min_{CT}, max_{CT}]$ $\sum_{i=1}^{NP} -SP_i x_i \leq -UB_j \quad \forall j, \quad UB_j \in [min_{SP}, max_{SP}]$	

TABLE IV: Common constraints

Constraint Description	Formulation
Number of Players in Team	$\sum_{i=1}^{NP} x_i = TS$
Maximum of 4 International Players	$\sum_{i=1}^{NP} IIP_i x_i \leq 4$
Minimum of 1 Wicket Keeper	$\sum_{i=1}^{NP} IW_i x_i \geq 1$
Minimum of 1 Captain	$\sum_{i=1}^{NP} IC_i x_i \geq 1$

that the owner has a maximum of \$12million to spend. From an ILP formulation perspective this is as simple as adding the following constraint:

$$x_{108} + x_{122} = 2 \tag{6}$$

The same three objectives as the previous example are used. α value of 50 is used for bowling and batting objectives and 20 for cost objective. The resulting convex front (Fig. 3) had a total of 1116 non-dominated solutions. The total computation

TABLE V: Chennai Super Kings 2011 team

ID	Name	Cost	Batting Avg	Fielding	Bowling Avg	Star Player?	Position
27	Wriddhiman Saha	100000	28.73	0.4821	100	No	Batsman
37	Dwayne Bravo	200000	23.94	0.4483	25.89	No	All-Rounder
42	Sudeep Tyagi	240000	0	0.2143	16	No	Bowler
66	Michael Hussey	425000	41.87	0.5667	100	Yes	Batsman
73	Albie Morkel	500000	27.88	0.2078	27.40	No	All-Rounder
85	Doug Bollinger	700000	0	0.2708	23.12	No	Bowler
94	R Ashwin	850000	0	0.2692	18.38	No	Bowler
95	S Badrinath	850000	32.44	0.2169	100	Yes	Batsman
100	M Vijay	900000	28.03	0.4262	100	No	Batsman
115	Suresh Raina	1300000	35.59	0.4545	27.08	Yes	All-Rounder
119	MS Dhoni	1800000	36.66	0.3895	100	Yes	Batsman
-	**Team Total**	**7865000**	**255.14**	**3.9463**	**637.87**	**4**	-

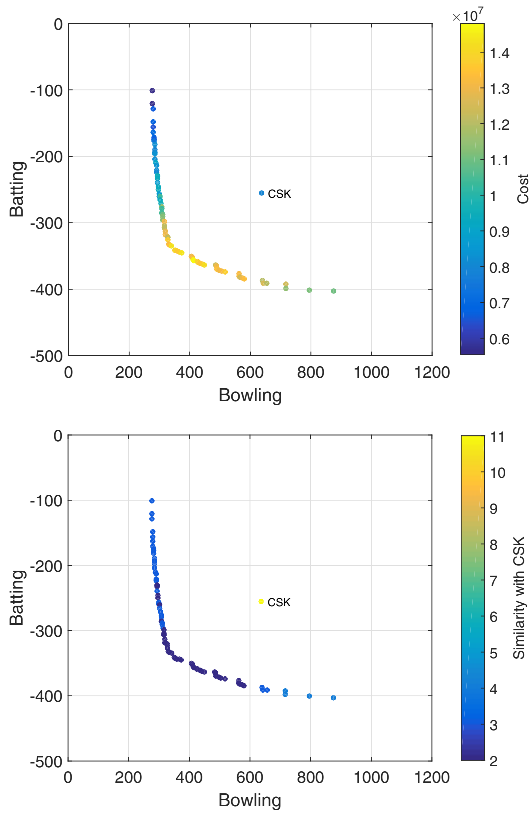

Fig. 1: Two-objective formulation: Cost analysis (top) and Similarity analysis (bottom)

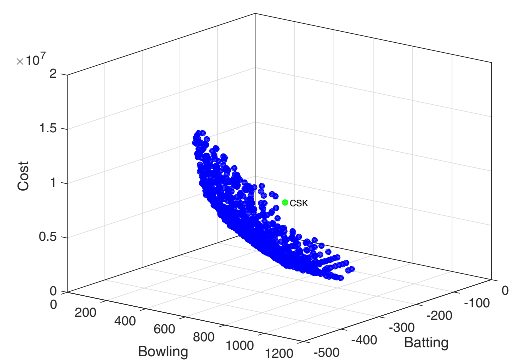

Fig. 2: Three-objective Formulations

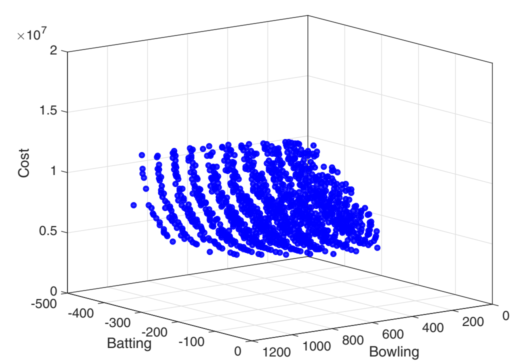

Fig. 3: The obtained front for a budget limit of 12million and two preferred players.

time was 97.78 seconds. The cost of getting de Villiers is \$1.1million and the cost of getting Kohli is \$1.8million. These two players take up one quarter of the given budget. Notice how because of this, in Fig. 3, the minimum cost team is just a little higher than 4 million. Also, since such a significant portion of the budget is being spent on two top batsmen, there will be an obvious skew in the results towards teams which have better batting performance.

This procedure for building partial teams can be especially useful in a dynamic environment such as an auction. The optimizer can be re-run after each auction round based on the players which have already been picked allowing the management to explore potential teams which can be built around these players.

E. Many-Objective Optimization

Many-objective optimization refers to optimization problems with four or more objectives. In this section we demonstrate the effectiveness of our approach on a many-objective formulation. We use bowling performance, batting performance, cost, star-power and fielding performance as the objectives. α value of 10 is used across all objectives.

As a result, we obtained a non-dominated front (Fig. 4) containing 2864 solutions within 883.78 seconds. It can be seen that the objective space is covered with good diversity among all objectives. The data in Fig. 4 is normalized using the objective range given in Table VI. Fig. 5 shows the pair-wise visualization for the objectives. A closer look at the data reveals some interesting trends. For example, the highest cost team of $16.475million had 9 star players. There were also teams which had higher number of star players but cost less. Another interesting observation is the conflict between the bowling and fielding objectives. Teams which have excellent fielding performance usually have poor bowling performance. An important thing to note is that the CSK team given in Table V becomes non-dominated when all five objectives are considered.

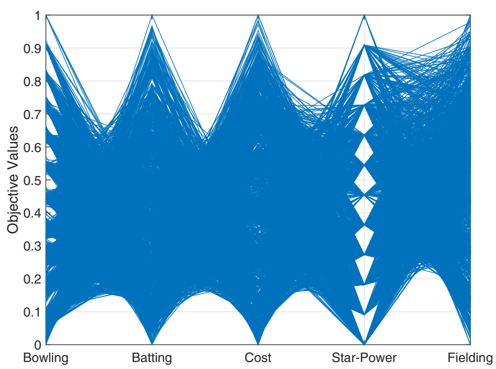

Fig. 4: The complete 5 objective front.

TABLE VI: Objective Ranges for 5-objective formulation

	Bowling	-Batting	Cost	-StarPower	-Fielding
Min.	276.93	-403.14	1960000	-11	-6.4184
Max.	1100.00	-73.23	16475000	0	-2.8064

IV. Player Ranking

Given a team of players, a manager or owner may want to identify how the players rank based on their performance statistics relative to their cost. For this we make use of the widely known hypervolume measure in this study. The hypervolume [14] of a set of solutions refers to the volume/area of objective space that is collectively dominated by those solutions.

As it can be seen from the problem formulations in Section II, all the objectives are additive in nature, i.e., they are summed up over the number of players in the team to give an overall objective value for the team. Note that for all the objectives that need to be maximized (batting average, star power, fielding average), a higher value for an individual player implies more impact. For the objectives that need to be minimized (bowling average, cost), this is reversed, since player with lower average represents a better bowler and a player with lower cost would be of better value for the

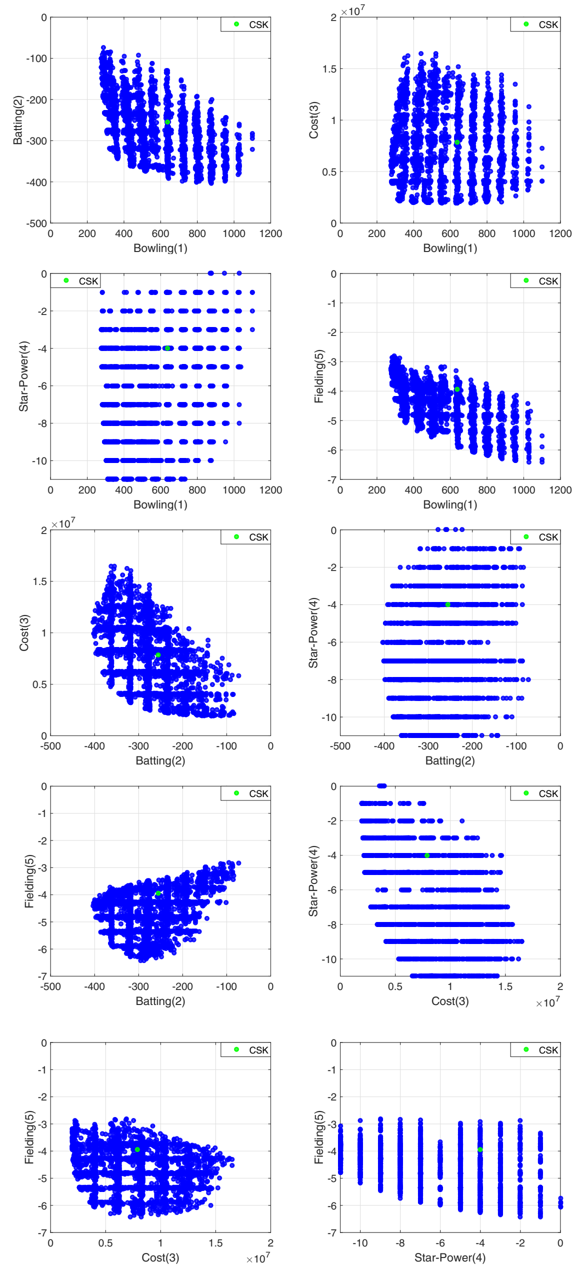

Fig. 5: Pair-wise Objective Visualization

same performance. Therefore, firstly, we negate the values of minimization objectives and keep the maximization objectives the same (note that this is opposite to that done in Section II). Next, we normalize the player data linearly between 0 and 1 (using bounds of each objective). This gives us a dataset where all objectives are in the same range [0,1], and a higher value in any objective for a player in this normalized dataset reflects a better performance with respect to that objective.

From the normalized data above, we establish the bounds for an 11 player team (maximum achievable value) and a 10 player team (minimum achievable value). This can be done in a straightforward way by sorting each of the objectives in descending order, followed by adding the top 11 values to get the upper bound and adding the bottom 10 values for the lower bound for each objective. These bounds serve as the normalization factors for the *team* objective values.

With these bounds, we can now consider a given team of 11 players, normalize its objectives, and calculate the hypervolume with respect to the extreme point $\{1,1\ldots1\}$ (in a minimization sense). This serves as the base hypervolume value H_t of the team. To determine the relative contribution of any given player i, we calculate the hypervolume of the normalized team performance without this player, H_{t-i}. Given the positive and additive nature of the objectives in the normalized form, it can be seen that a higher value of the difference $\delta_i = H_{t-i} - H_t$ would be reflective of a higher contribution (and hence a higher preference) of a given player. Thus, the overall preference order of the players can be calculated by sorting $\{\delta_i; i = 1,2,\ldots 11\}$ in a descending order. Further sub-rankings can be made within different categories such as batsmen, bowlers and all-rounders. As an example, consider the CSK team (Table V) with all 5 objectives. The hypervolume of the complete team is $H_t = 0.0212$. Table VII gives the player rankings using the above method.

The best ranked batsman is Hussey. He outperforms second placed Saha in terms of both batting performance and fielding performance. A similar comparison for the bowlers ranks Tyagi as first and Bollinger as second. Tyagi has a better bowling average while Bollinger has a better fielding average. It can be noted that Tyagi is only slightly worse off in fielding but has a significantly better bowling performance compared to Bollinger. Tyagi can also be bought for less than half the price of Bollinger. Hence, the preference order appears justified. In terms of all-rounders, it is no surprise that Raina, who is a key member of the Indian national team, is ranked higher than Bravo and Morkel. Raina significantly outperforms Bravo in terms of batting performance and is also classified as a star-player. Therefore, he is able to adequately justify his high cost. Note that these rankings could change depending on the set of objectives being considered (e.g. this particular example uses the 5-objective formulations, but rankings could be different for the 3-objective formulation).

TABLE VII: Player rankings based on hypervolume

ID	**Name**	**Team HV (player removed)** H_{t-i}	**Diff.** (δ_i)	**Rank in category**	**Overall Rank**
Batsmen					
66	**M Hussey**	0.0453	0.0241	1	1
27	**W Saha**	0.0373	0.0161	2	5
95	**S Badrinath**	0.0369	0.0158	3	6
119	**MS Dhoni**	0.0357	0.0145	4	7
100	**M Vijay**	0.0336	0.0124	5	9
Bowlers					
42	**S Tyagi**	0.0339	0.0127	1	8
85	**D Bollinger**	0.0325	0.0113	2	10
94	**R Ashwin**	0.0322	0.0110	3	11
All-Rounders					
115	**S Raina**	0.0448	0.0236	1	2
37	**D Bravo**	0.0415	0.0203	2	3
73	**A Morkel**	0.0380	0.0168	3	4

V. Conclusion

In this paper, we proposed a multi-objective approach for the selection of an IPL cricket team. We applied our approach to an actual data set of 129 players taken from the 2011 season of the IPL. Using our approach we demonstrated how a team manager or owner can go about searching for optimal teams. Not only does this approach guarantee optimality, it also produces results within a reasonable run-time.

Additionally, we have also shown examples with as many as five objectives to demonstrate the scalability of the proposed approach. Construction of partial teams has also been explored which is especially useful during auctions in assisting team management to explore potential teams which can be constructed based on existing picks. Lastly, a process of ranking the players based on their performance is also presented to further aid the decision making process. We would like to conclude with a caveat that the player performance on a given day cannot be guaranteed and may not be consistent with the player's form or track-record. Therefore, the presented approach can only serve to make informed decisions based on past performance.

References

[1] English Premier League, "Premier league announces payments to clubs in season 2014/15," 2015, http://www.premierleague.com/en-gb/news/news/2015-16/jun/020615-premier-league-payments-to-clubs-in-season-2014-15.html.

[2] S. Mitra, "The ipl: India's foray into world sports business," *Sport in Society*, vol. 13, no. 9, pp. 1314–1333, 2010.

[3] F. Ahmed, K. Deb, and A. Jindal, "Multi-objective optimization and decision making approaches to cricket team selection," *Applied Soft Computing*, vol. 13, no. 1, pp. 402 – 414, 2013.

[4] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr 2002.

[5] S. S. Sathya and M. S. Jamal, "Applying genetic algorithm to select an optimal cricket team," in *International Conference on Advances in Computing, Communication and Control*, 2009, pp. 43–47.

[6] H. Perera, J. Davis, and T. B. Swartz, "Optimal lineups in twenty20 cricket," *Journal of Statistical Computation and Simulation*, pp. 1–13, 2016.

[7] G. Sharp, W. Brettenny, J. Gonsalves, M. Lourens, and R. Stretch, "Integer optimisation for the selection of a twenty20 cricket team," *The Journal of the Operational Research Society*, vol. 62, no. 9, pp. 1688–1694, 2011.

[8] H. H. Lemmer, "Team selection after a short cricket series," *European Journal of Sport Science*, vol. 13, no. 2, pp. 200–206, 2013.

[9] G. R. Amin and S. Sharma, "Cricket team selection using data envelopment analysis," *European Journal of Sport Science*, vol. 14, no. sup1, pp. S369–S376, 2014.

[10] G. Kendall, S. Knust, C. C. Ribeiro, and S. Urrutia, "Scheduling in sports: An annotated bibliography," *Computers & Operations Research*, vol. 37, no. 1, pp. 1 – 19, 2010.

[11] F. Ahmed, A. Jindal, and K. Deb, "Cricket team selection using evolutionary multi-objective optimization," in *International Conference on Swarm, Evolutionary, and Memetic Computing*, 2011, pp. 71–78.

[12] IBM, "IBM ILOG CPLEX optimization studio," 2016, https://www.ibm.com/developerworks/downloads/ws/ilogcplex/.

[13] K. Deb and S. Gupta, "Understanding knee points in bicriteria problems and their implications as preferred solution principles," *Engineering Optimization*, vol. 43, no. 11, pp. 1175–1204, 2011.

[14] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 29–38, Feb 2006.