Evolutionary Multi-Objective Optimization and Decision Making Approaches to Cricket Team Selection

Faez Ahmed, Kalyanmoy Deb*, and Abhilash Jindal
Kanpur Genetic Algorithms Laboratory (KanGAL)
Indian Institute of Technology Kanpur
Kanpur, PIN 208016, India
Email: {faez,deb,ajindal}@iitk.ac.in
URL: http://www.iitk.ac.in/kangal

KanGAL Report Number 2011016

Abstract

Selection of players for a cricket team within a finite budget is a complex task which can be viewed as a constrained multi-objective optimization and a multiple criteria decision making problem. In cricket team formation, batting strength and bowling strength of a team are the major factors affecting its performance and an optimum trade-off needs to be reached in the formation of a good and successful team. We propose a novel representation scheme and a multi-objective approach using the NSGA-II algorithm to optimize the overall batting and bowling strength of a team with 11 players as variables. Fielding performance and a number of other cricketing criteria are also used in the optimization and decision-making process. Using the information from the trade-off front obtained, a multi-criteria decision making approach is then proposed for the final selection of team. Case studies using a set of players auctioned in Indian Premier League (IPL) 4-th edition are illustrated and players' current statistical data is used to define performance indicators. The proposed computational techniques are ready to be extended according to individualistic preferences of different franchises and league managers in order to form a preferred team within the budget constraints. It is also shown how such an analysis can help in dynamic auction environments, like selecting a team under player-by-player auction. The methodology is generic and can be easily extended to other sports like soccer, baseball and other league games.

1 Introduction

Formation of a good team for any sport is vital to its eventual success. Team selection in most sports is a subjective issue involving commonly accepted notions to form a good team. In this study, we have chosen the game of cricket as an example to demonstrate applicability of a multi-objective optimization methodology to a subjective issue of team formation from a set of players using available statistics. Cricket is a bat-and-ball game played between two teams of 11 players where one team bats, trying to score as many runs as possible, while the other team bowls and fields, trying to dismiss the batsmen one at a time and thus limiting the runs scored by the batting team [1, 3]. Batting and bowling strength of a team are the major criteria affecting its success along with many other objective and subjective factors like fielding performance, captaincy, home advantage etc.

Optimization studies have been done in many sports in the past [16, 13, 7], and also has been done in various issues in the game of cricket [18, 15]. Just as in most league competitions, a

*K. Deb is also an Adjunct Professor at Aalto University School of Economics, Helsinki, Finland.

pool of players is provided as an input along with their performance statistics. Each player is payed a certain amount of money by the team owners for playing for their team, which we refer to as player's cost. League organizers impose an upper limit on budget for each franchise/club to avoid giving undue advantage to rich franchises. Player cost is either fixed by organizers as salary, decided through auction or determined by some contract.

We have considered Twenty20 cricket game [6] and the Indian Premier League (IPL) [4] as a test case for our analysis. IPL is a professional league for Twenty20 cricket competition in India. As of now, not much literature is available for any team selection methodology in cricket since the concept of IPL and other such competitions is relatively new. In IPL, the franchise managers have the task of building a good team within the budget cap. Individual players are bought by the franchises during a public auction of the players. Since the total number of players in the market pool is large, the challenge of finding the optimal teams becomes increasingly complicated and common sense methods, mostly employed, may fail to provide a good team. Data used for this work (downloaded from the public domain sources) has a pool of 129 players from IPL 4-th edition. The player costs are the IPL 4-th edition auction prices for respective players. For players retained by the franchises the player price is the money deducted from their budget by the organizers. We have used other performance statistics of each player from the public domain sources on the International Twenty20 version of the game.

With so many players available for choosing a team, the need for an effective optimization technique can be justified by making a rough calculation of the apparent size of the decision space. From the given data of only 129 players from IPL 4-th edition auction, containing 10 captains and 15 wicket-keepers, the total number of possible teams under the constraints of at least one wicket-keeper and one captain are included in each team can be calculated as follows:

$$\text{Total number of teams} = \binom{10}{1}\binom{15}{1}\binom{127}{9}. \tag{1}$$

Considering the large number of different possible team combinations (to the order of 10^{15}), finding optimal teams under the added constraint of budget is not a trivial task. Currently most of the team selections are done using different heuristics, past experiences, or at most using some crude methodologies. For example, one strategy is to choose two or three high performance batsmen or bowlers and the remaining team slots are filled according to budget constraints. But, this approach may not always give an optimal or a near-optimal solution, since matches are won by a team effort and not by having one or two star players in the team. Hence, our aim in this paper is to investigate the formation of an overall team that is optimal or near-optimal from the point of multiple criteria. Since batting and bowling performances of a team are generally considered to be conflicting to each other, we use these two criteria in our optimization study, while all other objective and subjective criteria, such as fielding performance, captaincy, wicket-keeper's performance, brand value of a team and others, are used during the subsequent decision making phase. It is difficult to argue whether a batting-dominant team is better or a bowling-dominant team is better. This is an unavoidable dilemma that the team selectors face while forming a team. Due to our consideration of both bowling and batting performances in forming a team, we attempt to find a number of high-performing teams having a good trade-off between these two main objectives. Thereafter, we argue and demonstrate that a consideration of a number of other criteria applied on the obtained high-performing teams makes it convenient for the team managers to identify a preferred team of their choice.

In this paper, we have explored the problem of building a high-performing team out of a set of players given their past performance statistics and suggested, for the first time, a computational and decision-making methodology from the perspective of a multi-objective consideration. A novel representation scheme of a team of 11 players is suggested to handle different constraints associated with the team selection problem. The elitist evolutionary multi-objective optimization algorithm (NSGA-II [10]) is extended to find multiple high-performing teams. A number of realistic decision-making considerations are then used to pick a suitable team.

In the remainder of this paper, we describe the optimization problem corresponding to the cricket team selection problem in Section 2. A novel scheme to represent a 11-player team that automatically satisfies a number of constraints is described next. The computing procedures of different objective functions are then discussed. Section 3 presents the results obtained through our multi-objective optimization study. Our obtained high-performing teams are compared against the IPL 4-th edition winning team (CSK). The (theoretical) superiority of our teams is clear from the figure. The obtained solutions are analyzed for their sensitivity to the overall allowable budget and interesting conclusions are made. Section 4 then suggests a number of decision making techniques to choose a single preferred team from the set of trade-off teams. This includes the standard knee-point approach, a dynamic approach simulating the real auction procedure, and a couple of other interesting criteria. A multi-objective optimization study to find a set of teams providing a trade-off between batting and bowling performances and then a multi-criterion decision making analysis procedure to finally pick a preferred team remain as a hallmark feature of this study. The procedure is ready to be applied in practice with a minimal fine-tuning needed to suit various other rules of the IPL team selection process. Conclusions of this study are made in Section 5.

2 Proposed Methodology

In the game of cricket, player statistics have multiple parameters, like number of matches played, total runs made, batting strike rate, number of wickets taken by a bowler, number of overs bowled, etc. Importantly and interestingly, the values of these parameters for most active players are available. However, it is important to first identify the statistical parameters that reliably indicate a player's performance. The overall aim of a franchise is to build a team of 11 players with optimum bowling, batting, as well as fielding performance within budget and other rule constraints. Rule based constraints like presence of at least one player capable of wicket-keeping and one captain have also to be taken into account.

Considering the large amount of statistical data denoting various cricketing attributes that is available for each player, we first tend to reduce the dimension of data. One approach can be to use standard batting, bowling, and fielding rating of cricketers obtained after exhaustive statistical analysis. Such ratings, like the ICC world cricket rating [5], takes into account multiple factors of performance. But, such a rating system is currently available only for one-day and test match versions of this game. Since Twenty20 version of the game is very different and the performance of a player in one version of the game does not extrapolate to another version for most players, we cannot apply the available data for one-day matches and test matches to the Twenty20 format, which we are interested in this study. For simplicity, we have used batting average and bowling average of a player in the past four editions of the international Twenty20 cricket as a measure of their performance in batting and bowling. For brevity, the data is taken from the public domain sources, compiled and stored in a website [2].

Each player is assigned a tag indicating the player's unique identity. Using the available data, we then formulate the team selection problem as a multi-objective optimization problem, as follows:

$$\underset{\mathbf{t}=\{c,w,p_1,p_2,..,p_9\}}{\arg\max} \begin{cases} f_1(t) = \sum_{i=c,w,p_1,\cdots,p_9} \text{Batting Performance}(\mathbf{i}), \\ f_2(t) = \sum_{i=c,p_1,\cdots,p_9} \text{Bowling Performance}(\mathbf{i}), \\ f_3(t) = \sum_{i=c,w,p_1,\cdots,p_9} \text{Fielding Performance}(\mathbf{i}). \end{cases} \tag{2}$$

Notice that the wicket-keeper (w) does not affect the bowling performance of a team. The team is subject to the following constraints:

$$g_1(t) \equiv c \in \text{Captain list}, \tag{3}$$
$$g_2(t) \equiv w \in \text{Wicket-keeper list}, \tag{4}$$
$$g_3(t) \equiv \text{No two players are identical in a team}, \tag{5}$$
$$g_4(t) \equiv \text{Not more than four foreign players in a team}, \tag{6}$$
$$g_5(t) \equiv \sum_{i=\{c,w,p_1,\cdots,p_9\}} \text{Cost}(\mathbf{i}) \leq \text{TotalBudget}. \tag{7}$$

Here, t represents a team comprising of c (the tag of the captain of the team chosen from a captain list (CL) having of 10 names, presented in Table 6), w (the tag of the wicket-keeper of the team from a wicket-keeper list (WL) having 15 names, presented in Table 6), and $p_1, \cdots, p_9$ (tags of nine other players of the team chosen from 129 total names arranged in a ranked list (RL), presented in Tables 4 and 5, excluding the chosen captain and wicket-keeper). To have a distinct set of 11 players, no two players in a team can be identical. Players are also tagged as a foreigner (a non-resident Indian player) or not. As an IPL rule, a team should not have more than four foreign players. The final constraint indicates that overall cost of the team must be within the specified upper limit. We now describe the computational procedure for each of three objective functions.

2.1 Batting Performance

A players batting average is the total number of runs he has scored divided by the number of times he has been out [1]. Since the number of runs a player scores and how often he gets out are primarily measures of his own playing ability, and largely independent of his team mates, batting average is a good metric for an individual player's skill as a batsman. The objective function in our analysis has been taken as the summation of batting averages of all players. The problem with this approach is that some new players, even bowlers, may have a batting average comparable to few of the best established batsmen due to good performance in few matches played. Hence, to avoid this scenario, the concept of primary responsibility of a player is used. A player is identified as a designated batsman only if he has scored at least 300 runs in an international Twenty20 format. In calculation of a team's net batting performance, the batting average of players identified as designated batsmen is only added to find net batting average. This condition is used in order to exclude batsmen who have not played enough games for their skill to be reliably assessed. Thus, the overall batting average of team must be maximized, but here, we minimize the negative of the net batting average for the convenience of showing trade-off solutions.

2.2 Bowling Performance

A bowler's bowling average is defined as the total number of runs conceded by the bowler divided by the number of wickets taken by the bowler [1]. So, lower the bowling average, the better is the bowler's performance. Again to avoid including misleading high or low averages resulting from a career spanning only a few matches, we qualify a player as bowler only if he has taken at least 20 wickets in Twenty20 format. Only the bowling average of these qualified bowlers is summed to get a net bowling average of a team. Total bowling average of a team is taken as a measure of bowling performance and is minimized. Unlike in the batting performance, here, lower the net bowling average, better the bowling performance. Using such a strategy in optimization may result in a team having an exclusion of all bowlers, so that net bowling average of team is zero. Hence, instead of assigning a zero value to a non-bowler, we assign an artificial penalty of 100 as the bowling average for each non-bowler. This value (100) is chosen in such a way so as to make it worse than a designated bowler having the worst (highest) bowling average. Hence, a team's bowling performance is computed by adding the true bowling averages of designated bowlers and 100 for each non-designated bowler.

2.3 Fielding Performance

A player's fielding performance is calculated as follows:

$$\text{Player's Fielding Performance} = \frac{\text{Total catches taken}}{\text{Total number of innings played}}. \tag{8}$$

Team's net fielding performance is summation of all individual players fielding performance. The number of stumpings by a wicket-keeper is taken as his wicket-keeper's performance measure.

2.4 Multi-objective Formulation and NSGA-II

To solve the above problem, we employ a multi-objective genetic algorithm to find a set of trade-off teams. For this purpose, the elitist non-dominated sorting genetic algorithm or NSGA-II [10] is used. We provide a brief description of the NSGA-II procedure here, but readers can refer to the original study for more information.

The first task in applying a GA is to choose a representation scheme that is suited to handle constraints and also to facilitate an easier use of genetic operators. To represent a team in our proposed NSGA-II and to also take care of first three constraints (Equations 3 to 5), each player is assigned an tag, as shown in Tables 4 and 5 presented in the Appendix. To assign a tag for a player, first, we sort all 129 players according to player cost (given in appendix) and then assign a unique integer number indicating the player's rank in the range [1,129]. The sorted list is called ranked list or RL. Next, from the given 129 players, we identify the captains (marked with a 'C' in the table) and arrange them in ascending order of their price in a *captain list*, or CL. There are 10 captains in the list, thus the list contains integers in the range [1,10] (see Table 6). Similarly, the wicket-keepers are identified and put in an ascending order of their price in the *wicket-keeper list*, or WL. There are 15 wicket-keepers in the list, thereby making the list have values in the range [1,15] (see Table 6). Note that a particular value in CL may refer to the same player corresponding to a value in WL, and all values in CL and WL correspond to players represented in RL. There is a reason for this complication of maintaining three lists in our representation scheme, which will be clear in the next paragraph. We now discuss the construction procedure of a random population member in the initial generation.

Every population member is represented with two vectors: (i) a code-vector and (ii) a variable-vector. We first discuss the procedure of creating a code-vector. The first element of the code-vector is created at random from CL. Thus, this element is a random integer in the range [1,10]. The corresponding name (say, 'c') is the captain of the team. The second element of the code-vector is created at random from WL and is an integer in the range [1,15]. The selected name (say 'w') is the wicket-keeper of the team. Thereafter, the remaining nine elements in the code-vector are selected at random from RL, so that there is no repetition among these nine elements. Thus, they are nine unique integers in the range [1,129]. These are other nine team members. If 'c' or 'w' is identical to any of the other nine team members, another team member is chosen at random and the process is continued till the nine members are different from 'c' and 'w'. Thereafter, the tags of nine members are arranged in ascending order of their values.

Next, the variable-vector is constructed from the code-vector by copying the tag number of every player one by one starting with the tag of the captain, then the tag of the wicket-keeper, followed by tags of nine team members. Thus, the only difference between a code-vector and its corresponding variable-vector is in the first two elements. In the code-vector, they are ranks of captain and wicket-keeper in CL and WL, respectively, and in the variable-vector they are tags of captain and wicket-keeper. We illustrate the representation procedure by creating a random population member in the following:

Solution 1:

Position:	1	2	3	4	5	6	7	8	9	10	11
Code-vector:	3	6	10	15	36	41	59	75	88	115	125
Variable-vector:	96	48	10	15	36	41	59	75	88	115	125

In the above team represented by the code-vector, the first element (3, created at random from [1,10]) represents the third player (Adam Gilchrist) in the captain list (CL) shown in Table 6. The second element (6, created at random from [1,15]) represents the the sixth player (Parthiv Patel) from the wicket-keeper list (WL) shown in the same table. The remaining nine elements in the code-vector are chosen at random from the ranked list (RL) and presented in ascending order of the tag value. The variable-vector is constructed then from the code-vector. Interestingly, the variable-vector is identical to the code-vector except at the first two positions. The tag value of Adam Gilchrist and Parthiv Patel are taken from the RL and put in the first and second position in the variable-vector, respectively. The variable-vector is then used to compute the objective and constraint values. The code-vector will be used for the genetic operations – recombination and mutation, as discussed later.

There is a possibility of duplication that we resolve in our representation scheme. In certain population members, the chosen first two elements of the code-vector may refer to the same player which is a wicket-keeper captain. In this case, the variable-vector will indicate that its first two elements are identical and in reality there are only 10 players represented by the vectors. To resolve this issue, we simply replace the second element of the variable-vector with a random integer from RL and ensure that it is not identical to the 10 players already present in variable-vector. Let us illustrate this scenario with an example population member:

Solution 2:

Position:	1	2	3	4	5	6	7	8	9	10	11
Code-vector:	2	10	5	16	54	68	77	95	101	113	122
Variable-vector:	86	86	5	16	54	68	77	95	101	113	122
Mod. var.-vector:	86	72	5	16	54	68	77	95	101	113	122

Here, the captain and wicket-keeper is the same player (Kumar Sangakkara). Thus, although the code-vector has 11 distinct elements, there are in fact 10 players. To resolve this problem, we change the second element in the variable-vector by a random tag from RL that is not identical to any element already present in the entire variable-vector. Say, we choose the player with the tag 72 in this case. The revised variable-vector is then formed and used for computing objective function and constraint values.

In an occasion in which the chosen captain is also a wicket-keeper (such as Adam Gilchrist, Kumar Sangakkara or MS Dhoni) but the chosen wicket-keeper is a different person (say Brendon McCullum or Parthiv Patel), we still consider the situation as if the wicket-keeper is identical to captain (that is, the captain will also perform the job of wicket-keeping) and replace the chosen wicket-keeper with a random player (but non-identical to any other already chosen players) from RL.

It is interesting to note that the above procedure guarantees that the first three constraints are always satisfied in all population members. The variable-vector can now be used to compute the objective function and remaining constraints, as mentioned above. Instead of using any sophisticated methods to satisfy the fourth and fifth constraints, here, we simply count the number of foreign players in the team and calculate the total budget for hiring the team. Thereafter, we check these two values against their allowable values. If they are within limits, we declare the team as a feasible team, otherwise, we declare it as an infeasible team and compute constraint violation as follows:

$$\begin{aligned} \mathrm{CV}(\mathbf{t}) &= \min\left(0, \text{Number of foreign players}/4 - 1\right) \\ &\quad + \min\left(0, \sum_{i=\{c,w,p_1,\cdots,p_9\}} \mathrm{Cost}(\mathbf{i})/\mathrm{TotalBudget} - 1\right). \end{aligned} \tag{9}$$

The constraint violation value of a team will be used in NSGA-II's selection operator, which we shall discuss a little later. Note that for a feasible solution, the constraint violation value is zero.

From the available bowling, batting, and fielding average values of each of the 11 chosen players for a team, we can then compute the overall bowling, batting and fielding strengths for the team. NSGA-II is capable of handling three objectives and we can employ it to find a trade-off frontier that is expected to have individual optimal solutions and also a number of different compromised solutions. A little problem knowledge will reveal that a team having an extremely good fielding ability and not enough batting and bowling skills may not be what a team selection committee may be looking for in a team. Thus, instead of treating the fielding performance as an objective of its own in our optimization study, we treat it more like a secondary objective and use it to break a tie among trade-off solutions. In the tournament selection operator of comparing two solutions, if both are non-dominated to each other from the bowling and batting objectives, the team having a better fielding performance wins the tournament.

The NSGA-II procedure goes as follows. From the current set of N population members describing population P_t, a new set of N solutions (population Q_t) is created repeated use of tournament selection, recombination and mutation operators. In the tournament selection comparing two population members, a feasible solution is preferred over an infeasible solution and a lesser constraint-violated infeasible solution wins over another infeasible solution. When two feasible solutions are compared, a hierarchy of decision making is performed. First, a solution dominating its tournament competitor wins. Second, if both solutions are non-dominated to each other, the one having better fielding performance wins. Otherwise, if both solutions have the same fielding performance, the one residing on a less-crowded region in the objective space wins. This is achieved by computing the crowding distance operator [10]. Two such selected solutions are then mated under a recombination scheme (we discuss the recombination scheme in the next paragraph) using a probability p_c and two new offspring teams are created. If recombination operation is not performed, two selected solutions are saved as offspring solutions.

To achieve a meaningful recombination operation between two NSGA-II population members, we use the code-vector of two parents, instead of their variable-vectors. Thus, a code of the captain represented in one parent mates with that in the second parent so that two offspring codes in the range [1,10] are created. For example, if Solutions 1 and 2 mentioned above are recombined at the first position as parents, code values 3 and 2 will get recombined as integer numbers in the range [1,10] and two offspring integers, representing the code value of the respective captains from CL, will be created. Similarly, the code of the wicket-keeper of one parent mates with that of the second parent so that two offspring codes in the range [1,15] are created. Note here that when Solutions 1 and 2 mentioned above are recombined, code values 6 and 10 will be used as parent integers and two offspring integers will be created. Thus, in this situation, the wicket-keeper captain (Kumar Sangakkara) will be participate twice (if both positions are to be recombined) in the recombination operations – once as a captain for the first position and again as a wicket-keeper for the second position. The remaining nine code-vectors mate element-wise from the third element to 11-th element and offspring code values in the range [1,129] are created. A little thought will reveal that since the remaining nine code values are arranged in ascending order of their tag values, a variable-wise recombination becomes meaningful in the sense that a low-cost player gets a chance to mate with another low-cost player, and a high-cost player mates with another high-cost player.

We use the integer-version of the SBX operator [9] as a recombination operator. In this operator, every element in the code-vector is recombined with a probability 0.5, thus on an average 50% of the elements get recombined in a crossover operation. Note that even if one parent involves a wicket-keeper captain and another involves a different captain and a wicket-keeper, since they are recombined with their code-vectors, a meaningful recombination will always take place. Each of the offspring solutions are then mutated with a integer-version of the polynomial mutation operation [8] to slightly perturb them. This is done with a mutation probability p_m. After the mutation operation, the corresponding variable-vector of each offspring is constructed. In the event of a repetition of tag values, additional mutation operations are performed till a distinct 11-member team is formed. These operations guarantee that all created offspring solutions satisfy the first three constraints. We then rely on NSGA-II's constraint handling capabilities to satisfy fourth and fifth constraints and create feasible population members.

After the new population Q_t is created, it is combined with P_t and a combined population $R_t = P_t \cup Q_t$ is formed. The combined population is then sorted based on non-domination level (with respect to the bowling and batting objectives). Thereafter, to create the next generation population P_{t+1}, solutions from the first sorted front is selected first, then the second front members are taken, and so on. This process is continued till no more fronts can be accommodated to the new population (recall that R_t has a size twice to that of P_{t+1}). The last front that cannot be fully accommodated in P_{t+1} is undergone with more criteria. First, the usual crowding distance values are assigned to each member. Second, the individual objective-extremes are chosen straightway. Third, the front members are sorted with the third objective (fielding performance) and the remaining population slots are filled with solutions having higher fielding performance values. In the event of a tie, solutions are chosen based on their crowding distance value. This ends the operation of a generation and P_{t+1} is declared as the next generation population. NSGA-II operators are continued in this fashion till a predefined number of generations are elapsed.

3 Multi-Objective Optimization Results

Here we present the simulation results of the above-mentioned algorithm applied to the player database. The budget constraint is considered to have TotalBudget=6 million dollars. At least one wicket-keeper, one captain, and a maximum of four foreign players must be included in a team, as mentioned before. We use the following standard parameter settings in all our simulations: population size=400, maximum generations=400, crossover probability = 0.9, mutation probability = 0.05, SBX distribution index = 10, and mutation distribution index = 20.

Figure 1 shows the trade-off front obtained by our modified NSGA-II. Each point on the trade-off front represents a team of 11 players. A few teams corresponding to the trade-off points marked on the figure are shown in Table 1. Each team has a captain, a wicket-keeper and at most four foreign players. Also, they all satisfy the stipulated budget constraint. The right extreme of the front marks the team having highest overall batting average, while the left extreme shows the most bowling dominant team. The trade-off between these two objectives is clear from this figure.

To compare our results, we consider a real team of 11 cricketers who played for the Chennai Super Kings (CSK) team in the last IPL (took place in India during 8 April to 28 May 2011) and also won the final. The bowling and batting performance of this team are calculated using the same procedure as described above. The total cost of hiring the CSK team is estimated to be around 7.5 million dollars. It turns out that it is not a feasible team in our representation because of its spending more than 6 million dollars in hiring the players. Despite spending more money, CSK team has bowling and batting performances that are worse than a number of teams found by our NSGA-II. The point representing the CSK team is shown on the objective space in Figure 1. It can be seen that the team is non-optimal as well as costlier. Similar observations

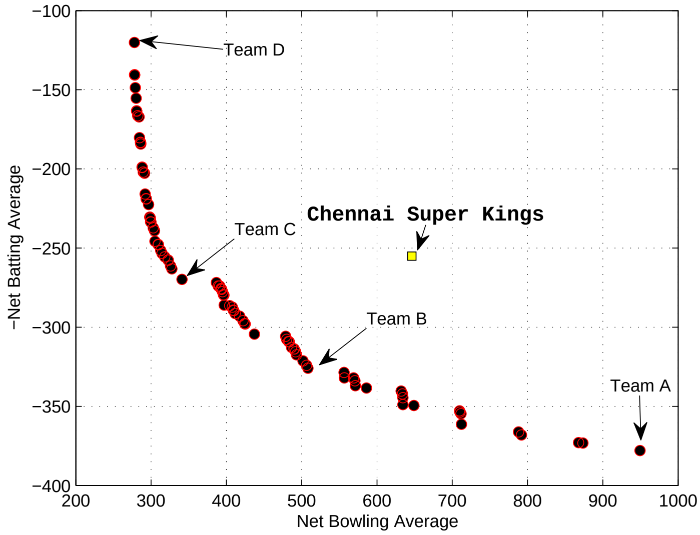

Figure 1: Multi-objective trade-off front obtained by NSGA-II is shown. CSK team is outperformed by Teams B and C (obtained in this study) on batting and bowling performances.

were found for other past IPL teams as well. For brevity, we do not discuss them here.

3.1 Budget Sensitivity Analysis

To analyze the effect of budget constraint on a team's performance we have performed a sensitivity analysis where the optimization process is run for a range of budget constraints and the corresponding trade-off front obtained by our modified NSGA-II is plotted. It can be seen from the Figure 2 that the budget constraint affects batting dominant teams more than bowling-dominant teams. This is because the price difference among batsmen with a high batting average and those having a low average is significant. The same effect does not exist among bowlers and hence the minimum bowling average region of the trade-off front is fairly unchanged due to increase or decrease in the TotalBudget value. To get an idea of the extreme trade-off front, we also perform a run in which no limit on the total budget is imposed. Such an analysis can provide the team managers an idea how much gain in batting and bowling averages is possible by spending certain amount of more money in hiring a team. Interestingly, for the chosen player's performance statistics and price, a large amount of investment need not make a better bowling team, however, a better batting team is possible to be formed by investing more money.

Table 2 shows the best batting teams found by our proposed NSGA-II procedure for different budget restrictions. It is clear how most teams include well-renowned batsmen in each of the teams. When no budget restriction is used, the team got multiple expensive batsmen, such as Sachin Tendulkar and MS Dhoni in one team. Although none of these teams do well in the bowling compartment of the game and a franchise may not choose any of these extreme teams, the purpose of presenting these teams is to demonstrate that the proposed NSGA-II is able to find such dream batting teams as best batting teams for different budgetary restrictions. It also emphasizes the motivation for choosing an intermediate trade-off solution as a preferred team which would have a good trade-off between batting and bowling aspects of the game.

Table 1: Four teams chosen from the trade-off front (Figure 1) obtained by NSGA-II. The first row marks the captain and the second row marks the wicket-keeper of a team. Foreign players in a team are shown in italics.

	Team A	**Team B**	**Team C**	**Team D**
Players	Sachin Tendulkar	Yuvraj Singh	Yuvraj Singh	Yuvraj Singh
	Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha
	Michael Hussey	*N. McCullum*	*JP Duminy*	R Ashwin
	Manoj Tiwary	Manoj Tiwary	Sudeep Tyagi	Sudeep Tyagi
	Rahul Dravid	Ravindra Jadeja	Ravindra Jadeja	*N. Rimmington*
	Suresh Raina	Suresh Raina	Suresh Raina	*Paul Collingwood*
	Shaun Marsh	*James Franklin*	*James Franklin*	*Steven Smith*
	Aaron Finch	*Brad Hodge*	*Brad Hodge*	Pragyan Ojha
	A. McDonald	*A. McDonald*	*A. McDonald*	*Shakib Al Hasan*
	Shikhar Dhawan	Shikhar Dhawan	Jaidev Unadkat	Jaidev Unadkat
	Naman Ojha	Amit Mishra	Amit Mishra	Amit Mishra
Bat. avg.	378.0	324.1	269.8	120.2
Bowl. avg.	949.5	506.6	341.0	277.9
Cost ($)	5,950,000	5,930,000	5,845,000	4,935,000

Table 2: Best batting teams chosen from the trade-off frontiers (Figure 2) obtained for different budget restrictions. First row indicates captains and second row indicates wicket-keepers. Foreign players are shown in italics.

	5 M$	**6 M$**	**7 M$**	**No limit**
Players	Sachin Tendulkar	Sachin Tendulkar	Sachin Tendulkar	Sachin Tendulkar
	Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha	MS Dhoni
	VVS Laxman	Suresh Raina	*Shaun Marsh*	*Michael Hussey*
	Rahul Dravid	Rahul Dravid	Rahul Dravid	Manoj Tiwary
	Aaron Finch	*Shaun Marsh*	Suresh Raina	*Aaron Finch*
	Shaun Marsh	*Aaron Finch*	*Aaron Finch*	Rohit Sharma
	Manoj Tiwary	*Michael Husse*	Manoj Tiwary	Saurabh Tiwary
	Mohammad Kaif	Shikhar Dhawan	*Michael Hussey*	S Badrinath
	Shikhar Dhawan	*A. McDonald*	*A. McDonald*	*A. McDonald*
	Michael Hussey	Naman Ojha	Shikhar Dhawan	*Shaun Marsh*
	Andrew McDonald	Manoj Tiwary	S Badrinath	Suresh Raina
Bat. avg.	364.7	378.0	389.3	403.1
Bowl. avg.	1022.4	949.5	949.5	874.7
Cost ($)	4,910,000	5,950,000	6,530,000	11,030,000

3.2 Player Frequency Table

Analyzing the trade-off front shown in Figure 1 not only provides us with a set of high-performing teams but can also give us a better insight in team selection strategies under a multi-criteria decision making process and in a dynamic, auction-like situation. In this section, we reveal some important features of choosing a high-performing team based on the *innovization* principle suggested elsewhere [12].

To prepare a cricket team selection strategy, identification of key players is very important. A classical approach for this purpose would be to identify the top batsmen or top bowlers as key players for the team, but it turns out that choosing such players usually come with a high cost and a team having a bunch of top domain-specific good players may not result in a high-performing and cost-effective team. To identify the key players in a team we suggest a novel idea here. All the teams appearing on the trade-off front are considered for an analysis and the frequency of each player appearing on all teams on the trade-off front is computed. For example, say, there are 50 teams on the trade-off front and the player, Sachin Tendulkar, appears in five of these teams. Hence, his frequency is 5/50 or 10%. Figure 3 shows the frequency of all players from the dataset on a bar graph. The players are represented by the tags assigned to them in the optimization

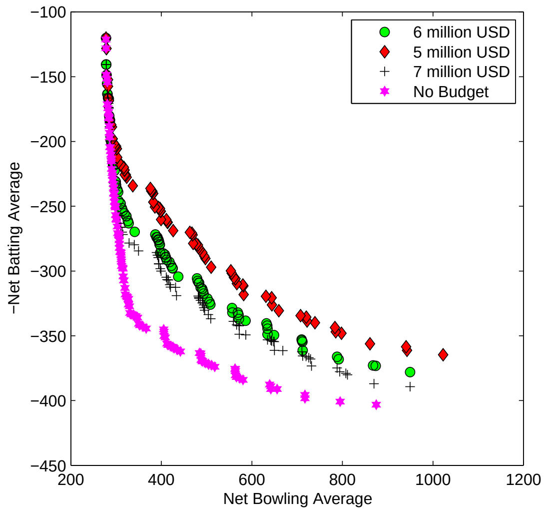

Figure 2: Budget sensitivity analysis illustrating how the trade-off front changes with the upper limit on the total budget.

process. Names of a few high frequency players are also mentioned on the bar graph. From the figure, it can be seen that many high-cost players appearing on the right half of bar graph have a zero frequency. Secondly, only 35 of the 129 players have frequency greater than zero and appears in the teams on the trade-off front, meaning that other 94 players do not cause a good trade-off between batting, bowling and fielding performances when a fixed budget is in mind. Some of these high frequency key players can be identified from the bar graph and can be given high priority in forming the team during the auction process. A close examination of the bar graph indicates that our common-sense method of forming a team with top batsmen and bowlers (mentioned above) would fail to capture these commonly-appearing high-valued players on the trade-off front and some famous and highly-priced specialist players may not have a good batting-bowling-fielding trade-off to appear as a good choice for an effective team. In a sense, the above analysis of identifying frequently appearing players on the entire trade-off front provide the names of effective players for all three compartments of the game. These players are commonly-appearing ingredients of high-performing teams and hence must be designated as most *valued players* for the overall effective performance of a team. It should be noted that a player's frequency is not a direct indicator of how good an individual player is on any one compartment of the game, but rather indicates his importance in the formation of a high-performing team considering budget constraints, all three aspects of the game (bowling, batting and fielding) and importantly the chemistry involved in other player's presence in the team. A knowledge of such players may be useful in the dynamic auction process of choosing a team as well.

Another observation that can be made by an analysis of the obtained trade-off solutions is the cost of hiring each team under the given budget restriction. An important question to ask is that although a budget of 6 M$ was the restriction, did all high-performing trade-off teams utilize all that budget? Figure 4 shows the amount of money spent to form high-performing teams on the trade-off front. The teams are arranged from high batting strength (on the left) to high bowling strength (on the right), or high to small values of bowling strength. It is clear from the figure that teams having high batting strength must spend almost all the allocated budget in hiring the teams, as most good batsman are more costly than good bowlers (see the resource containing players' statistics [2]). Only predominantly bowling teams require about 80% of cost needed to hire a good batting team. To investigate if this is a trend with other two trade-off

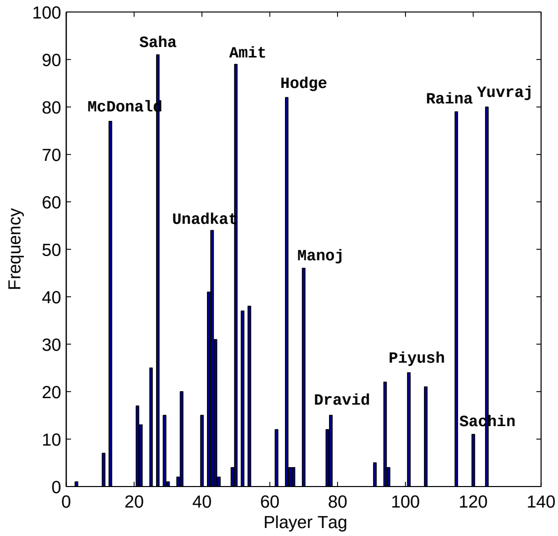

Figure 3: Frequency of players appearing on teams from the trade-off front (Figure 1).

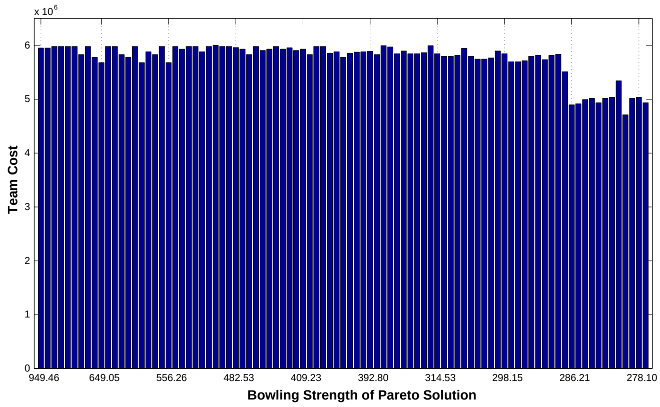

Figure 4: Total investment in hiring each team on the trade-off front with 6M$ budget restriction.

frontiers (having 7M$ and unlimited budget), we plot the total amount of money needed to hire each of the high-performing teams with these budget restrictions in Figures 5 and 6, respectively.

A similar trend can be observed from these two figures as well. In the case of 7M$ budget, most teams utilize the stipulated maximum budget, except that the optimal bowling teams use less budget. For the case of unlimited budget restriction, the result is slightly different. Optimal batting teams require higher budgets compared to optimal bowling teams, but the teams with a good trade-off in both aspects of batting and bowling (intermediate teams) requires a high budget of around 12 to 14M$. This also reveals that if a rich team owner is willing to spend more money (say to the tune of 20M$) to hire probably the best team around, the person will be on one hand disappointed that there does not exist a high-performing team that will cost so much and on the other hand the person would be happy that with a much lesser money a high-performing team can be hired. Thus, the above analysis not only helps get essential qualitative information of how to form a high-performing team, but also provides quantitative information about specifics and, above all, the names of a real team with 11 players that would do the trick.

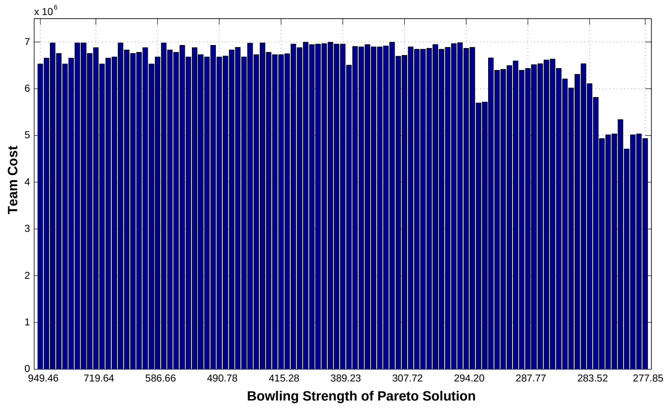

Figure 5: Total investment in hiring each team on the trade-off front with 7M$ budget restriction.

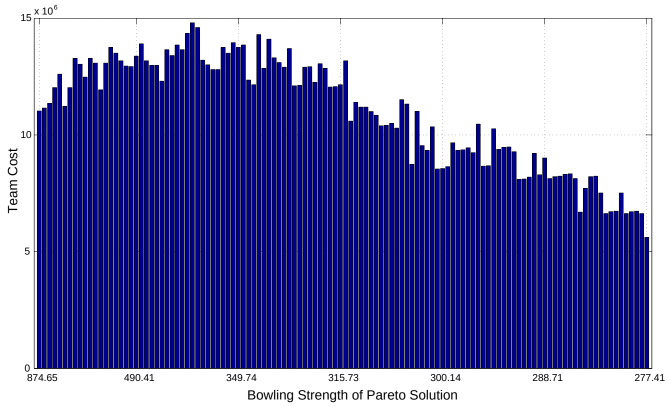

Figure 6: Total investment in hiring each team on the trade-off front with no budget restriction.

When we investigate the number of foreign players in each of these high-performing teams across all three fronts, we observe that they all have *exactly four* foreign players in each team, although there could have been zero to a maximum four foreign players present in any team. This tells us that one way to choose a high-performing team is to make sure there are four foreign players (the maximum number allowed) in a team, no matter whatever is the restriction on the overall budget.

Some of these insights are not obvious or intuitive before the optimization run is performed and resulting solutions are analyzed. Clearly, the above optimization-cum-analysis procedure reveals crucial insights that can act as thumb-rules dictating what a team-manager must do to put together a high-performing team.

4 Team Selection as a Decision Making Process

The objective of the entire process is to obtain a high-performing single team of 11 members, rather than just finding and suggesting a number of trade-off teams. Having identified a set of high-performing trade-off teams with different bowling, batting and fielding performance values, we shall now discuss a few methodologies in arriving at a choosing a particular team. Towards this goal, we use certain multi-criteria decision making (MCDM) methods [14]. After the initial optimization analysis, we have a trade-off front, similar to that shown in Figure 1. Different methods now can be adopted for this purpose, but here, we suggest the following two methods for the selection of a final team in the context of a cricket team selection of Twenty20 version of the game.

4.1 Knee Region Approach

The obtained trade-off front comprises of a set of points in the objective space representing various teams. If the trade-off front has a knee-region, it is advisable to select a solution that lies in the knee region. Such a solution is most always preferred because deviating from the knee region means that a small change in the value of one of the objectives would come from a large compromise in at least one other objective. A recent study suggested a number of viable ways of identifying a knee point in a two-objective front [11]. Although a knee region may not exist in all trade-off frontiers, a visual inspection of trade-off fronts shown in Figure 2 reveals that in general a batting-bowling trade-off front exhibits a knee in most fixed budget conditions.

Applying the knee-finding methods on our trade-off front shown in Figure 1, we observe that teams B and C lie near the knee regions, although the knee region is not visually apparent to have a knee for the obtained front. These teams are shown in Table 1. Teams C and D show a reasonable compromise between batting and bowling performances as compared to Teams A and D.

4.2 Multiple Criteria Non-Dominated Sorting Approach

The knee region approach described above does not take into account many subjective factors that may, in reality, define a good team, such as the following:

1. Direct fielding performance
2. Success rate of the captain
3. Wicket-keeper's performance
4. Team expenditure
5. Brand value of players, and others.

To take into account these factors, we consider the solution set obtained from knee-region analysis and compute the above-mentioned factors. Thereafter, a ranking of importance of the above factors are gathered from the selector's point of view. For example, the direct fielding performance may be the most important factor among the others mentioned above in our team selection strategy. So, we sort the solution set with respect to fielding performance and pick the solution having the best fielding side. In the event of a tie on direct fielding performance, the next ranked factor can be used to choose a team in a lexicographic manner. Some other MCDM techniques, such as an aggregate voting strategy [14] or analytical hierarchy process (AHP) [17] can also be used.

To another extreme, if all the above factors are equally important, a domination approach can be applied to find the non-dominated teams with respect to all the above factors. The resultant best front solutions are then considered for further analysis. To illustrate, we use here the first three criteria of fielding, captaincy, and wicket-keeping performance for a non-dominated analysis. After identifying the non-dominated teams, we sort them according to remaining factors, such as, the team expenditure or the brand value of players. Using a non-dominated sorting of the teams in the knee region using the first three factors, we obtain only the following two teams on the non-dominated front. The players in these teams are given below:

Team 1: Yuvraj Singh (C), Wriddhiman Saha (W), Suresh Raina, Manoj Tiwary, *Roelof van der Merwe*, Amit Mishra, *Brad Hodge*, Shikhar Dhawan, *Nathan McCullum*, *Andrew McDonald*, Ravindra Jadeja.

Team 2: Sachin Tendulkar (C), Wriddhiman Saha (W), Suresh Raina, Manoj Tiwary, *James Franklin*, Amit Mishra, *Brad Hodge*, Shikhar Dhawan, *Ryan ten Doeschate*, *Andrew McDonald*, Ravindra Jadeja.

To illustrate, we now use the team expenditure factor and observe that the cost of Team 1 is lesser than that of Team 2. Hence, Team 1 would be our final preference.

4.3 Customized Criteria

Using the above multiple criteria non-dominated sorting approach, finally Team 1 is proposed to be high-performing and preferred team within the six million dollar cost constraint. From Table 1, it can also be observed that the obtained teams have overall high batting and bowling average but lack famous good bowlers (such as Dimitri Mascarenhas (21), Nathan McCullum (25), and Piyush Chawla (101)). We observe that this is a problem originating due to statistical data considered where the difference in bowling average of a good bowler and a normal bowler is marginal, while the price differences are significant. In Twenty20 format of the game, it can be argued that a few well-known experienced players can play a key role towards a success in a match and just overall strength of team may not be not enough. Such 'star' players are necessary for a wide fan base of a franchise as well. Hence, a particular team management may require at least a few star players in its team which can be considered as an additional constraint. This then requires that instead of using a generic algorithmic technique, some customized criteria may be included in choosing a single preferred team.

To demonstrate the effect of customized criterion, we consider a scenario in which an emphasis of including star players is made. We define a star batsman as a player who has scored at least 1,500 runs with a batting average greater than 30 in Twenty20 format and a star bowler as one who has taken at least 70 wickets with a bowling average less than 25. In our simulation, we also assume that a franchise decides to keep at least two star batsmen and three star bowlers in its team. Other constraints imposed are six million dollar budget, a maximum four foreign players, and at least one wicket-keeper and one captain. The resultant trade-off front is shown in Figure7. Again, using the multiple criteria non-dominated sorting decision making procedure outlined above, we observe that only Team Q lies on the non-dominated front using the primary criteria of fielding, captaincy and wicket-keeping. Team Q along with a few other teams corresponding to points marked on the trade-off front are shown in Table 3. A comparison of four teams in this table with the two teams (1 and 2) listed at the end of Section 4.2, it is clear that most of these new teams have star players, thereby making these teams well-acceptable among the fans and also to not have to compromise on the overall batting-bowling performance of the teams. Interestingly, all trade-off teams have exactly four foreign players, thereby making the fourth constraint active at the trade-off solutions.

4.4 Dynamic Optimization

We now suggest a near-realistic scenario in which the player selection process goes on in a dynamic manner. In an auction based team selection system, like IPL, the entire pool of players is not available in one go for team selectors to choose. Besides, the final price of players is also not fixed beforehand. Only a minimum base price for each player is usually specified. Our proposal to handle such a scenario is as follows.

First, a pre-processing optimization analysis using the highest expected price of each player is done. The decision making procedure discussed above helps select one desired team from

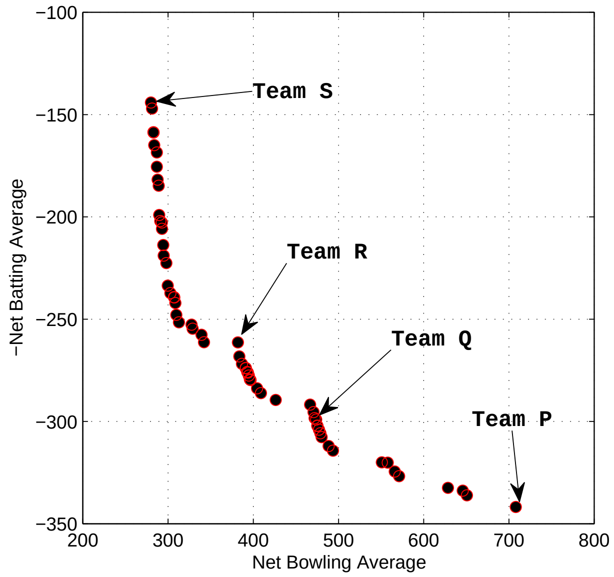

Figure 7: Trade-off front for teams with star players.

the obtained trade-off front. It may happen that in the auction process, some players from the desired team are already bought by other franchises. Thus, team owners must know trade-offs and alternative choices during the auction process. Hence in such a dynamic environment, a systematic approach needs to be followed. Using the initial bar graph analysis (shown in Figure 3), high valued players are first identified. These players must then be given priority for selection. After Phase 1 is completed, the optimization and decision-making process need to be simulated again with the remaining players. Here, the selected players must be included in each team considered during the optimization process and decision making process. The new set of high valued players can then be identified by using another bar graph analysis process and the next set of players can be identified. At each phase, the selection of players from the obtained trade-off front must be done judiciously from the high frequency (valued) players to ensure a proper balance of the team. Hence using such a procedure repeatedly in a dynamic optimization mode a team with a

Table 3: Four teams chosen from the trade-off front with customized conditions on star players. The first row marks the captain of a team and second row marks the wicket-keeper, except in Team S the captain is also a wicket-keeper. Foreign players are shown in italics.

Team P	**Team Q**	**Team R**	**Team S**
Sachin Tendulkar	Yuvraj Singh	Yuvraj Singh	MS Dhoni
Wriddhiman Saha	Wriddhiman Saha	Wriddhiman Saha	*Paul Collingwood*
S Badrinath	*Nathan McCullum*	*Kieron Pollard*	R Ashwin
Manoj Tiwary	Manoj Tiwary	Manoj Tiwary	Pragyan Ojha
Nathan McCullum	*Brad Hodge*	*Dimitri Mascarenhas*	*Brad Hodge*
Suresh Raina	Suresh Raina	Suresh Raina	Munaf Patel
Shaun Marsh	Piyush Chawla	*Brad Hodge*	*Steven Smith*
Amit Mishra	Amit Mishra	Amit Mishra	Amit Mishra
Andrew McDonald	*Andrew McDonald*	*Andrew McDonald*	*Dimitri Mascarenhas*
Shikhar Dhawan	Shikhar Dhawan	Jaidev Unadkat	Jaidev Unadkat
Dimitri Mascarenhas	*Steven Smith*	Sudeep Tyagi	Sudeep Tyagi

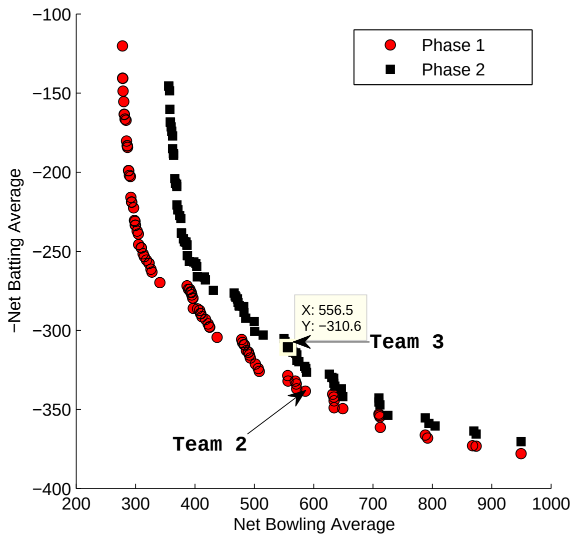

Figure 8: Trade-off front obtained in Phase 2 is compared with that in Phase 1 during the auction-based decision-making case study.

high overall performance can be obtained. We perform a case study in the following paragraph.

Suppose, in Phase 1 of auction, 11 players with tags 27, 120, 121, 122, 123, 124, 125, 126, 127, 128, and 129 are considered in the auction pool, meaning that only these players are now available to picked by any franchise. Let us also assume that after Phase 1 optimization and decision-making, the franchise X selects Team 1 mentioned in Section 4.2 to be the most preferred team. However, let us also consider a predicament where the captain of Team 1, Yuvraj Singh, has already been bought by another franchise and is no longer available for franchise X. Thus, Team 1 as a whole cannot be a preferred choice any more and franchise X now needs to look for another high-performing team from its multi-objective optimization results. Let us assume that after reviewing all the trade-off high-performing teams of Phase 1 that do not contain Yuvraj Singh, Team 2 with Sachin Tendulkar as a captain (mentioned in Section 4.2) becomes the next preferred team for franchise X. Of the above-mentioned auction pool of 11 names, Sachin Tendulkar (tag number 120), is then opted and finalized as a team member of franchise X. In the next phase, all the above 11 names in the auction pool are bought by other franchises and are no longer available for further consideration.

In this scenario, franchise X can perform another multi-objective optimization run with Sachin Tendulkar as a captain and look for a team having 10 more players. From the set of 129 players, the Phase 1 auction pool players are also eliminated and remaining 119 players become the pool from which 10 players are to be chosen. Other constraints will remain and a new optimization run can be executed. The resulting trade-off front is shown in Figure 8. It is not a surprise that with a restricted set of (119) players available to choose from, the Phase 2 front is dominated by Phase 1 front obtained with all 129 available players. Due to a slot fixed with Sachin Tendulkar (who is primarily a batsman), the team with the best bowling performance is somewhat worse that the best bowling team in Phase 1. However, the presence of a good batsman in the team makes the best batting team in Phase 2 to have a marginally worse performance than the best batting team in Phase 1. The sight deterioration in the batting performance comes from the unavailability of another batsman Wriddhiman Saha (Tag 27) in Phase 2. After obtaining the trade-off front, we perform another bar-graph analysis to investigate the most valued players in Phase 2. The bar graph is shown in Figure 9. It can be seen how the frequency of most players remains similar

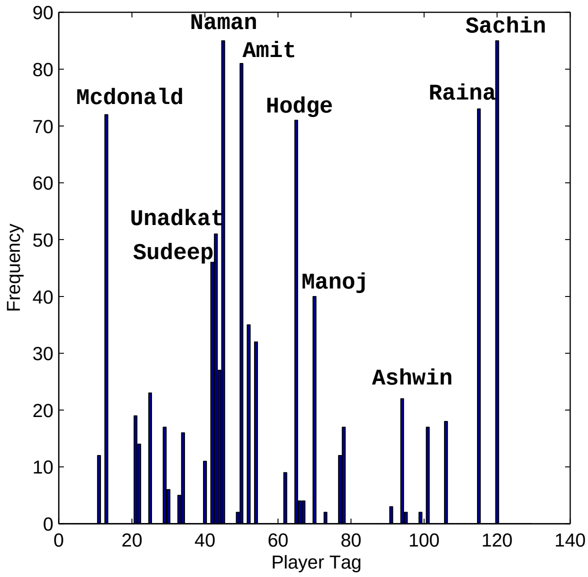

Figure 9: Frequency of players during Phase 2 auction.

to the previous bar graph (Figure 3), except for a few players for whom it instantly rises. For example, the frequency of Naman Ojha as a wicket-keeper shoots up, again due to absence of Wriddhiman Saha (Tag 27). It can be observed that in Phase 1 most trade-off teams had him as a wicket-keeper and his absence has increased the suitability of Naman Ojha as the next choice for a wicket-keeper. Again repeating the knee region and non-dominated sorting analysis with the new trade-off teams as before, we choose the following team as our next preferred team:

> **Team 3:** Sachin Tendulkar (C), Naman Ojha (W), *Roelof van der Merwe*, Shikhar Dhawan, Manoj Tiwary, Piyush Chawla, Amit Mishra, *Nathan McCullum*, *Andrew McDonald*, *Brad Hodge* and Suresh Raina.

Team 3 is shown on the objective space in Figure 8 along with Team 2. Franchise X would then choose a player from Team 3 based on the next announced auction pool of players and this process can continue till the complete team is formed. Since at every phase, the previously-chosen players are considered in both optimization and decision-making processes, the players are picked as a team and the dynamic formation of the team will produce a team that was best possible to achieve given the auction nature of the team selection process. The above procedure is flexible and if instead of one-player-at-a-time auction, if multiple-player-at-a-time auction is played, the above procedure will allow a simple way to choose a set of players to be picked at every phase. For example, if two players are allowed to be picked in every phase, in Phase 2, two top valued players based on the frequency or other decision-making criteria may be chosen.

It is worth mentioning here that the average execution time needed to perform the above dynamic optimization study took less than one minute on a standard laptop, hence this procedure is realistic to be tried in a real auction environment. We have demonstrated how the procedure can be used in a dynamic manner with certain IPL rules and decision-making criteria, but any other IPL team selection procedures and other decision-making criteria can be simulated with the procedure to make the overall task realistic for it to be used in a real scenario.

5 Conclusions

We have proposed and used for the first time a new computing methodology for an objective evaluation of cricket team selection using a multi-objective genetic algorithm and multiple criteria decision making aids. Such problems usually must consider a number of objective and subjective criteria that all must be paid attention to. In this paper, we have suggested choosing a couple of main functional criteria – batting and bowling performances – during the initial multi-objective optimization study and using other more subjective criteria during the subsequent decision making phase. For the optimization task, a novel representation scheme has allowed feasible solutions to be found in a convenient manner and enabled simple genetic operators to be employed. Starting with a list of 129 players and their available statistics, the proposed modified NSGA-II procedure has able to find a set of high-performing teams demonstrating a trade-off between their overall bowling and batting averages. In a comparison to the winning team of the IPL 4-th edition (played in April-May, 2011), the teams obtained by our study are theoretically better in both bowling and batting performances and importantly also cost less to hire the whole team. Motivated by this result, we have then proposed a number of multiple criteria decision making analysis tools to come up with a preferred team by considering a set of other subjective criteria in a systematic manner. It has been demonstrated that heuristics, such as inclusion of star players to enhance the popularity of a team, or the use of past record on captaincy of the chosen captain of the team can all be incorporated in the team selection process. To make the process more realistic, we have also demonstrated how a dynamic, auction-based player selection procedure used during the IPL 4th edition can be introduced in our approach and an effective, high-performing, and coherent team can be formed iteratively keeping in mind the different performance criteria associated with forming a cricket team.

The procedures suggested and simulated in the paper clearly demonstrate the advantage of using a multi-objective computing methodology for the cricket team selection task for major league tournaments. The consideration of multiple objectives during optimization and during the decision-making process provides team selectors a plethora of high-performing team choices before they can select a single preferred team. As an added benefit, the availability of of multiple high-performing teams can be exploited to identify key players appearing commonly to most trade-off teams. Since they appear on most trade-off teams, they bring in values from both batting and bowling aspects of the game. In a dynamic selection of players one at a time, the team selectors may put more emphasize in selecting such players. Although each idea suggested here has been demonstrated with a simulation, the proposed methodology is now ready to be used in a real scenario with more realistic team selection rules and other criteria that a particular franchise may be interested. Some fine-tuning of the proposed methodology and a GUI-based user-friendly software can be developed to customize a franchise's options, that can be used in practice without much knowledge of multi-objective optimization, genetic algorithms, or decision making aids.

References

[1] Cricket. http://en.wikipedia.org/wiki/Cricket.

[2] Cricket database. http://www.iitk.ac.in/kangal/cricket.

[3] An explanation of cricket. http://www.cs.purdue.edu/homes/hosking/cricket/explanation.htm.

[4] Indian premier league. http://en.wikipedia.org/wiki/Indian_Premier_League.

[5] International cricket council. http://icc-cricket.yahoo.net.

[6] Twenty20 cricket game. http://en.wikipedia.org/wiki/Twenty20.

[7] S. Butenko, J. Gil-Lafuente, and P. M. Pardalos. *Economics, management and optimization in sports*. Springer Verlag, 2004.

[8] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Wiley, Chichester, UK, 2001.

[9] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9(2):115–148, 1995.

[10] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[11] K. Deb and S. Gupta. Understanding knee points in bicriteria problems and their implications as preferred solution principles. *Engineering Optimization*, in press.

[12] K. Deb and A. Srinivasan. Innovization: Innovating design principles through optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 1629–1636, New York: ACM, 2006.

[13] A. Duarte, C. Ribeiro, S. Urrutia, and E. Haeusler. Referee assignment in sports leagues. In *Practice and Theory of Automated Timetabling VI*, pages 158–173, 2007.

[14] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, 1999.

[15] I. Preston and J. Thomas. Batting strategy in limited overs cricket. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(1):95–106, 2000.

[16] J. C. Régin. Minimization of the number of breaks in sports scheduling problems using constraint programming. In *Constraint programming and large scale discrete optimization: DIMACS Workshop Constraint Programming and Large Scale Discrete Optimization, 57*, pages 115–121. American Mathematical Society, 2001.

[17] T. L. Saaty. *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*. Pittsburgh, Pennsylvania: RWS Publications, 2008.

[18] T. B. Swartz, P. S. Gill, D. Beaudoin, and B. M. Desilva. Optimal batting orders in one-day cricket. *Computers & Operations Research*, 33(7):1939–1950, 2006.

A Player Data

Some statistical data of 129 cricket players considered in this study are tabulated in Tables 4 and 5. More details can be found in [2]. The captain list (CL) and the wicket-keeper list (WL) are shown in Table 6.

Table 4: Statistical data of 129 players used in this study. The 'Tag' column presents the rank number of a player in the ranked list (RL) used in the proposed representation scheme.

Tag	**Name**	**Status**	**Price** (USD)	**Tag**	**Name**	**Status**	**Price** (USD)
1	Aiden Blizzard	F	20000	34	Ryan ten Doeschate	F	150000
2	Jonathan Vandiar	F	20000	35	Wayne Parnell	F	160000
3	Nathan Rimmington	F	20000	36	Davy Jacobs	W, F	190000
4	Rilee Rossouw	F	20000	37	Dwayne Bravo	F	200000
5	Robert Frylinck	F	20000	38	Owais Shah	F, *Ba	200000
6	Travis Birt	F	20000	39	Scott Styris	F	200000
7	George Bailey	F	50000	40	Steven Smith	F	200000
8	Johan van der Wath	F	50000	41	Ajit Agarkar		210000
9	Luke Pomersbasch	F	50000	42	Sudeep Tyagi		240000
10	Moises Henriques	F	50000	43	Jaidev Unadkat		250000
11	Roelof van der Merwe	F	50000	44	Paul Collingwood	F	250000
12	Michael Klinger	F	75000	45	Naman Ojha	W	270000
13	Andrew McDonald	F	80000	46	Tim Paine	W, F	270000
14	Michael Lumb	F	80000	47	Manpreet Gony		290000
15	Suraj Randiv	F	80000	48	Parthiv Patel	W	290000
16	Thisara Perera	F	80000	49	Aaron Finch	F	300000
17	Rusty Theron	F	85000	50	Amit Mishra	*Bo	300000
18	Alfonso Thomas	F, *Bo	100000	51	Callum Ferguson	F	300000
19	Ben Hilfenhaus	F	100000	52	JP Duminy	F, *Ba	300000
20	Colin Ingram	F	100000	53	Shaun Tait	F, *Bo	300000
21	Dimitri Mascarenhas	F, *Bo	100000	54	Shikhar Dhawan		300000
22	James Franklin	F, *Ba	100000	55	Brad Haddin	W, F	325000
23	Jerome Taylor	F	100000	56	Ryan Harris	F	325000
24	Matthew Wade	W, F	100000	57	Eoin Morgan	F	350000
25	Nathan McCullum	F, *Bo	100000	58	James Hopes	F	350000
26	Nuwan Kulasekara	F	100000	59	Ashok Dinda		375000
27	Wriddhiman Saha	W	100000	60	Brett Lee	F	400000
28	Clint McKay	F	110000	61	Murali Kartik		400000
29	Faf du Plessis	F	120000	62	Shaun Marsh	F	400000
30	Mohammad Kaif		130000	63	Stuart Broad	F	400000
31	Charl Langeveldt	F, *Bo	140000	64	VVS Laxman		400000
32	Jesse Ryder	F	150000	65	Brad Hodge	F, *Ba	425000
33	Joginder Sharma		150000	66	Michael Hussey	F, *Ba	425000

Table 5: Contiuation of Table 4 showing statistical data of 129 players used in this study. The 'Tag' column presents the rank number of a player in the ranked list (RL) used in the proposed representation scheme.

Tag	**Name**	**Status**	**Price** (USD)	**Tag**	**Name**	**Status**	**Price** (USD)
67	Shakib Al Hasan	F	425000	99	Kieron Pollard	F, *Bo	900000
68	Ishant Sharma		450000	100	M Vijay		900000
69	Brendon McCullum	W, F, *Ba	475000	101	Piyush Chawla	*Bo	900000
70	Manoj Tiwary		475000	102	Sreesanth		900000
71	Morne Morkel	F	475000	103	Zaheer Khan		900000
72	Vinay Kumar		475000	104	Angelo Mathews	F	950000
73	Albie Morkel	F	500000	105	Johan Botha	F, *Bo	950000
74	Graeme Smith	F, *Ba	500000	106	Ravindra Jadeja		950000
75	L Balaji		500000	107	Ross Taylor	F, *Ba	1000000
76	Lasith Malinga	F, *Bo	500000	108	AB de Villiers	W, F	1100000
77	Pragyan Ojha	*Bo	500000	109	Cameron White	F, *Ba	1100000
78	Rahul Dravid		500000	110	Jacques Kallis	F, *Ba	1100000
79	RP Singh	*Bo	500000	111	Muttiah Muralitharan	F, *Bo	1100000
80	Daniel Vettori	C, F, *Bo	550000	112	Dale Steyn	F, *Bo	1200000
81	Dirk Nannes	F, *Bo	650000	113	Harbhajan Singh		1300000
82	Kevin Pietersen	F, *Ba	650000	114	Shane Watson	F, *Ba	1300000
83	Tillakaratne Dilshan	W, F	650000	115	Suresh Raina	*Ba	1300000
84	Cheteshwar Pujara		700000	116	David Hussey	F, *Ba	1400000
85	Doug Bollinger	F	700000	117	Mahela Jayawardene	C, F, *Ba	1500000
86	Kumar Sangakkara	C, W, F, *Ba	700000	118	Saurabh Tiwary		1600000
87	Munaf Patel		700000	119	MS Dhoni	C, W, *Ba	1800000
88	Venugopal Rao		700000	120	Sachin Tendulkar	C, *Ba	1800000
89	David Warner	F	750000	121	Shane Warne	C, F	1800000
90	Abhishek Nayar	F	800000	122	Virat Kohli		1800000
91	Praveen Kumar	F	800000	123	Virender Sehwag	C	1800000
92	Andrew Symonds	F, *Ba	850000	124	Yuvraj Singh	C	1800000
93	Ashish Nehra		850000	125	Irfan Pathan	*Bo	1900000
94	R Ashwin		850000	126	Rohit Sharma	*Ba	2000000
95	S Badrinath	*Ba	850000	127	Robin Uthappa	W	2100000
96	Adam Gilchrist	C, W, F	900000	128	Yusuf Pathan		2100000
97	Daniel Christian	F	900000	129	Gautam Gambhir	C	2400000
98	Dinesh Karthik	W	900000				

Table 6: Captain list (CL) and wicket-keeper list (WL) used in the proposed representation scheme are shown.

CL	Tag	Player	WL	Tag	Player
1	80	Daniel Vettori	1	24	Matthew Wade
2	86	Kumar Sangakkara	2	27	Wriddhiman Saha
3	96	Adam Gilchrist	3	36	Davy Jacobs
4	117	Mahela Jayawardene	4	45	Naman Ojha
5	119	MS Dhoni	5	46	Tim Paine
6	120	Sachin Tendulkar	6	48	Parthiv Patel
7	121	Shane Warne	7	55	Brad Haddin
8	123	Virender Sehwag	8	69	Brendon McCullum
9	124	Yuvraj Singh	9	83	Tillakaratne Dilshan
10	129	Gautam Gambhir	10	86	Kumar Sangakkara
			11	96	Adam Gilchrist
			12	98	Dinesh Karthik
			13	108	AB de Villiers
			14	119	MS Dhoni
			15	127	Robin Uthappa