Bounding the final rank during a round robin tournament with integer programming

Uwe Gotzes¹ · Kai Hoppmann²

Received: 25 March 2019 / Revised: 1 January 2020 / Accepted: 7 January 2020 /

Published online: 10 February 2020

© The Author(s) 2020

Abstract

This article is mainly motivated by the urge to answer two kinds of questions regarding the Bundesliga, which is Germany's primary football (soccer) division having the highest average stadium attendance worldwide: "At any point in the season, what is the lowest final rank a certain team can achieve?" and "At any point in the season, what is the highest final rank a certain team can achieve?". Although we focus on the Bundesliga in particular, the integer programming formulations we introduce to answer these questions can easily be adapted to a variety of other league systems and tournaments.

Keywords Round robin tournaments · Standings estimations · Integer linear programming · Football

Mathematics Subject Classification 90-04 · 90C27

1 Introduction

The Bundesliga consists of 18 teams and applies a system of promotion and relegation with the 2. Bundesliga. It is a double round-robin tournament where each club plays two matches against every other club. One at home and one away. Nowadays a victory is worth three points, a loss none, and a draw yields one point each. The team at the top of the league table at the end of the season is German Champion. Currently, the top six to seven clubs in the table qualify for profitable international competitions organized by the Union of European Football

✉ Uwe Gotzes
uwe.gotzes@uni-due.de

Kai Hoppmann
kai.hoppmann@zib.de

¹ Faculty of Mathematics, University of Duisburg-Essen, 47048 Duisburg, Germany

² Zuse Institute Berlin (ZIB), Takustraße 7, 14195 Berlin, Germany

Associations (UEFA). The two teams at the bottom of the table are relegated to the 2. Bundesliga, while the top two teams in the 2. Bundesliga are promoted. The 16th-placed team (third-last), and the third-placed team in the 2. Bundesliga play a two-leg play-off match. The winner of this match participates in the Bundesliga in the next season, while the loser has to play in the 2. Bundesliga. If teams are level on points, tie-breakers are applied (Liga 2019).

The Elimination Problem in modern European football, e.g., to decide whether a given team can still win the Bundesliga at some point in the season, was shown to be NP-complete independently by Bernholt et al. (1999) and Kern and Paulusma (2001). These results were further generalized to various other score allocation rules in Kern and Paulusma (2004). Schwartz (1966) was one of the first to develop necessary and sufficient conditions for a team to be eliminated in baseball using network flow techniques and algorithms. A first application of linear programming to the same problem was published by Robinson (1991). The computational complexity of determining the best and worst possible final rank of a team, among other related questions, is discussed in Schlotter and Cechlárová (2018) and Gusfield and Martel (2002). In Raack et al. (2014) a general ranking integer programming model that calculates the number of points that is needed to finish i th in any sport, including football, motor sports and ice hockey is developed. The authors also give a good overview of publications dealing with sport problems. The paper of Kendall et al. (2010) deals with the Brazilian national football championship and the calculation of the minimum number of points any team has to win to be sure it will be qualified for the play-offs. Also a method to compute the number of points each team has to win to have any chance to be qualified is presented. Again, a good overview of previous works is given.

2 Motivation

Der Postillon is a German website featuring satirical articles reporting on international, national, and local news in newspaper and TV format (Der Postillon 2019). In October 2015, it had more than 14 million visitors. It also comes as an English version titled *The Postillon* (The Postillon 2019). On 29 January 2018, *Der Postillon* announced, that the Munich based German sports club *FC Bayern München*, one of the most successful European clubs, celebrated non-relegation at an early stage of the Bundesliga season. It was reported that euphoric FC Bayern fans were singing in the town center of Munich until the early morning (Nießen 2018). A very similar version of this article was published on 3 February 2014 (Nießen 2014). At the time of publication, matchday 20 (in 2018) and matchday 19 (in 2014) were over. In February 2018 we became aware of these two articles. On the one hand we were astonished that avoidance of relegation was managed at such an early stage of the league. On the other hand we were not able to verify the statement with pencil and paper. Hence, we decided to set up a general integer programming formulation (Schrijver 1986) to obtain a mathematical proof for this kind of statements.

3 Mathematical model for round-robin tournaments

Consider the directed multigraph $G = (T, S, \mu)$ where T represents a set of n teams playing a round-robin tournament, $S \subseteq T \times T$ denotes the matches that have not been played yet, and $\mu : S \rightarrow \mathbb{N}$ states their multiplicity, i.e., $\mu(i, j) \in \mathbb{N}$ is the number of remaining home games of team i against team j . Further, let $S_h^m := \sum_{(m,j) \in S} \mu(m, j)$ and $S_a^m := \sum_{(i,m) \in S} \mu(i, m)$ denote the number of remaining home and away games of team $m \in T$, respectively. Each game $(i, j) \in S$ ends with exactly one outcome $x^\ell = (p_h^\ell, p_a^\ell) \in X := \{x^1 = (p_h^1, p_a^1), \dots, x^o = (p_h^o, p_a^o)\}$, where p_h^ℓ and p_a^ℓ are the points that are awarded to the home and the away team in case of outcome x^ℓ , respectively. In theory, there may exist an indefinite number of outcomes, but we only consider the case that $|X| = o \in \mathbb{Z}_{\geq 0}$ is finite in this article. Additionally, w.l.o.g. we assume that no points are deducted, i.e., that $p_h^\ell, p_a^\ell \in \mathbb{Z}_{\geq 0}$, and that there exists an outcome for each team such that no points are awarded. Finally, M_h and M_a denote the maximum number of points that the home and the away team can win and $p_m \in \mathbb{Z}_{\geq 0}$ denotes the number of points that team $m \in T$ has already collected during the tournament.

The described tournament model is directly applicable to the top level of European football league systems including the Bundesliga, *La Primera División* (Spain), the *Premier League* (England), and the *Serie A* (Italy). They all are double round-robin tournaments, i.e., we have a simple directed graph with the multiplicity of each arc being at most one and the set of outcomes is given by $X = \{(3, 0), (1, 1), (0, 3)\}$.

Next, we model all possible courses of the ongoing tournament together with an implicated ranking for a fixed team $t \in T$. For each $(i, j) \in S$ and $k \in \{1, \dots, \mu(i, j)\}$ we introduce o binary variables $x_{ijk}^1, \dots, x_{ijk}^o \in \{0, 1\}$ indicating whether the corresponding outcome is realized or not. Equations (1) guarantee that each game ends with exactly one outcome:

$$\sum_{\ell=1}^o x_{ijk}^\ell = 1, \quad \forall (i, j) \in S, \quad \forall k \in \{1, \dots, \mu(i, j)\}. \quad (1)$$

Additionally, for each team $m \in T$ participating in the tournament we introduce an integer variable

$$p_m^* \in \{p_m\} \oplus \bigoplus_{j=1}^{S_h^m} \{p_h^\ell : \ell = 1, \dots, k\} \oplus \bigoplus_{j=1}^{S_a^m} \{p_a^\ell : \ell = 1, \dots, k\}, \quad (2)$$

denoting its final number of points at the end of the tournament. The Minkowski sum, $A \oplus B$, of two sets A and B is defined as a set of pairwise sums of points from A and B . In other words, $A \oplus B := \{a + b : a \in A, b \in B\}$. In case of the Bundesliga, the computation of the Minkowski sum (2) yields

$$p_m^* \in \{p_m, \dots, p_m + 3 \cdot (S_h^m + S_a^m)\} \setminus \{p_m + 3 \cdot (S_h^m + S_a^m) - 1\}. \quad (3)$$

Equations (4) ensure the correct final score of team $m \in T$ depending on the results of its remaining games:

$$p_m^* = p_m + \sum_{(m,j) \in S} \sum_{k=1}^{\mu(m,j)} \sum_{\ell=1}^o p_h^\ell x_{mjk}^\ell + \sum_{(i,m) \in S} \sum_{k=1}^{\mu(i,m)} \sum_{\ell=1}^o p_a^\ell x_{imk}^\ell, \quad \forall m \in T. \quad (4)$$

Finally, we aim on modelling the ranking of team $t \in T$. Team t is ranked higher in the final table than team $m \in T \setminus \{t\}$ if it has collected more points. In case that both teams are even on points, we assume that we can arbitrarily choose their order. Therefore, for each team $m \in T \setminus \{t\}$ we introduce a binary variable c_m indicating whether t is ranked higher than team m in the final table, i.e., $c_m = 1$, or if it has a lower rank, i.e., $c_m = 0$. The following two classes of constraints ensure a correct ranking of team t :

$$p_t^* - p_m^* \leq c_m \cdot ((p_t + M_h S_h^t + M_a S_a^t) - p_m), \quad \forall m \in T \setminus \{t\}, \quad (5a)$$

$$p_t^* - p_m^* \geq (1 - c_m) \cdot (p_t - (p_m + M_h S_h^m + M_a S_a^m)), \quad \forall m \in T \setminus \{t\}. \quad (5b)$$

In case that $p_t^* - p_m^* > 0$ we have that $c_m = 1$ due to (5a) and if $p_t^* - p_m^* < 0$ we have $c_m = 0$ due to (5b). The bounds used on the right hand sides are tight and describe the biggest possible point differences: t always wins the maximum number of points and m receives no more points in their remaining matches (5a), and vice versa (5b). If $p_t^* = p_m^*$ both values for c_m are feasible, i.e., we can arbitrarily choose the order of t and m .

Using the variables and constraints introduced above, we are able to formulate integer linear programs answering the following two questions:

1. What is the lowest rank team $t \in T$ can reach in the final table?
2. What is the highest rank team $t \in T$ can reach in the final table?

The model answering the first question, which we call the Lower Bound for Final Rank Model (LBFR), can be stated as

$$\begin{aligned} \max \quad & n - \sum_{m \in T \setminus \{t\}} c_m \\ \text{s.t.} \quad & (1), (4), (5a), \end{aligned} \quad (6a)$$

while the model answering the second question, which we call the Upper Bound for Final Rank Model (UBFR), is given by

$$\begin{aligned} \min \quad & n - \sum_{m \in T \setminus \{t\}} c_m \\ \text{s.t.} \quad & (1), (4), (5b). \end{aligned} \quad (6b)$$

Due to the absence of constraints (5b), $c_m = 1$ for any $m \in T \setminus \{t\}$ is always feasible for model LBFR even though the final number of points of team t may be smaller than or equal to the number of points of team m . But since we maximize the expression $-\sum_{m \in T \setminus \{t\}} c_m$, i.e., we minimize the number of teams that end up with strictly less points than t , $c_m = 1$ iff $p_t^* > p_m^*$ holds in each optimal solution. Hence, using

(6a) as objective, LBFR yields the worst case final position of team t . An analogous argument shows that UBFR with objective (6b) yields the best case final position of team t even though Eq. (5a) is not included.

Using the idea from Robinson (1991), the model can also easily be adapted to other league systems such as the “Big Four” MLB, NBA, NFL and NHL in the United States and Canada. Here, we additionally have to account for cross-divisional matches. If we want to determine the lowest possible rank of a team $t \in T$ in its division, we simply assume that t does not gain any points in its remaining games against teams outside of its division and that all other teams $m \in T \setminus \{t\}$ score the maximum possible number of points in their remaining cross-divisional games. To determine the best possible rank we assume the opposite. In case of the four mentioned tournaments, it is important to note that our model does only determine the best possible rank of a team within its division and not whether it is eliminated from the tournament or not. For example, in case of the MLB, besides the six winners of the divisions, additionally the two teams having the best statistics within the two leagues, which consist of three divisions each, qualify for the postseason play-offs. Hence, even a team finishing third in its division may not be eliminated.

4 Computational results

We solved 14,688 instances of the integer programming models defined above originating from the Bundesliga standings during seasons 2003/2004 – 2017/2018. That is, we solved the models LBFR and UBFR for each participating team and for each matchday of these seasons. We omitted matchdays with obvious solutions, such as matchdays one to eight. To translate the data into .lp-format files we used the software ZIMPL (Koch 2004). As MIP solver we used SCIP 5.0.1 (Achterberg 2009; Gleixner et al. 2019) with standard settings. All instances were solved on a standard Linux PC with an Intel® Core™ i5-3570 CPU @ 3.40 GHz and 16 GB RAM.

The average solution time of all instances was 0.024 s, while 11.84 s was the longest solution time that we observed and it took 909 Branch-and-Bound nodes, i.e. subproblems with an increasing number of fixed binary variables, to prove optimality for the respective instance. Surprisingly, this “hard” instance is just problem LBFR for *FC Bayern* on matchday 21 in 2014. Only 0.89% of the problems were not solved in the root node, i.e., 99.11% of the models could be solved just by using preprocessing techniques, heuristics, and the examination of the linear relaxation of the original MIP and without entering the branch and cut phase of the MIP solver.

4.1 Problems emerging from the motivational chapter

Now we are ready to get back to our motivating problem from Sect. 2. Is it true that *FC Bayern* managed to avoid relegation already on matchday 20 in 2018? To answer this questions, we set up model LBFR such that $t = \text{“Bayern_Muenchen”}$, $(p_m)_{m=1}^{18}$ as listed in the Pts-column of Table 1 and $S = \{\{\text{“Eintracht_Frankfurt”}$, $\text{“Bor_Moenchengladbach”}\}, \{\text{“Bayern_Muenchen”}$, $\text{“TSG_Hoffenheim”}\}, \dots$,

Table 1 Bundesliga standings after matchday 20 of season 2017/2018. P = position, GP = games played, W = won, D = drawn, L = lost, F = goals for, A = goals against, GD = goal difference, Pts = points

P	Team	GP	W	D	L	F	A	GD	Pts
1	FC Bayern München	20	16	2	2	49	16	33	50
2	Bayer 04 Leverkusen	20	9	7	4	41	27	14	34
3	FC Schalke 04	20	9	7	4	32	25	7	34
4	Eintracht Frankfurt	20	9	6	5	26	20	6	33
5	RB Leipzig	20	9	5	6	32	29	3	32
6	Borussia Dortmund	20	8	7	5	42	27	15	31
7	Borussia Mönchengladbach	20	9	4	7	30	32	− 2	31
8	FC Augsburg	20	7	7	6	29	26	3	28
9	TSG Hoffenheim	20	7	6	7	31	32	− 1	27
10	Hannover 96	20	7	6	7	28	30	− 2	27
11	Hertha BSC	20	6	8	6	27	27	0	26
12	SC Freiburg	20	5	9	6	22	35	− 13	24
13	VfL Wolfsburg	20	4	11	5	23	24	− 1	23
14	VfB Stuttgart	20	6	2	12	16	26	− 10	20
15	1. FSV Mainz 05	20	5	5	10	24	35	− 11	20
16	Werder Bremen	20	3	8	9	16	25	− 9	17
17	Hamburger SV	20	4	4	12	16	29	− 13	16
18	1. FC Köln	20	3	4	13	15	34	− 19	13

{“VfL_Wolfsburg”, “FC_Koeln”}, {“FC_Schalke_04”, “Hannover_96”}} according to the remaining schedule, see Olympia-Verlag GmbH (2019b). It turned out, that after matchday 20 was over, it was still possible to construct a course of the league such that *FC Bayern* would be second to last with 50 points, while *1. FC Köln* becomes German Champion with 53 Points. To double-check our model, we put its results into an online standings calculator (Olympia-Verlag GmbH 2019a), see Fig. 1. Analogously, we found out that in 2014 *FC Bayern* managed to avoid relegation not until matchday 20.

4.2 Further illustrative problems

As already mentioned, we solved the models LBFR and UBFR for each team t and for every matchday of the Bundesliga season 2017/2018. Figure 2 contains four charts that visualize the results obtained by the models LBFR and UBFR during the season for the exemplarily chosen teams *FC Bayern München*, *FC Schalke 04*, *Bayer 04 Leverkusen* and *Hamburger SV*. In the first round of the season, we configured the optimization models in a way such that the final position is understood as the final position after the first round of the double round-robin tournament. This is the reason why the lines worst- and best-case final position in Fig. 2 are getting closer and closer and finally meet at the vertical dashed line at matchday 17 (end

TABELLE

Platz	Team	Spiele	g.	u.	v.	Tore	Diff.	Pkt.	Platz	Team	Spiele	g.	u.	v.	Tore	Diff.	Pkt.
1	1. FC Köln	34	16	5	13	28:34	-6	53	10	RB Leipzig	34	15	6	13	38:36	2	51
2	Borussia Dortmund	34	15	7	12	49:34	15	52	11	TSG Hoffenheim	34	15	6	13	39:38	1	51
3	Bayer 04 Leverkusen	34	15	7	12	47:35	12	52	12	Hannover 96	34	15	6	13	36:36	0	51
4	FC Schalke 04	34	15	7	12	38:33	5	52	13	VfB Stuttgart	34	16	3	15	26:29	-3	51
5	Bor. Mönchengladbach	34	16	4	14	37:39	-2	52	14	1. FSV Mainz 05	34	15	6	13	34:38	-4	51
6	FC Augsburg	34	14	9	11	36:31	5	51	15	SC Freiburg	34	14	9	11	31:40	-9	51
7	Eintracht Frankfurt	34	15	6	13	32:28	4	51	16	Bayern München	34	16	2	16	49:30	19	50
8	VfL Wolfsburg	34	13	12	9	32:28	4	51	17	Werder Bremen	34	4	8	22	17:38	-21	20
9	Hertha BSC	34	14	9	11	35:32	3	51	18	Hamburger SV	34	5	4	25	17:42	-25	19

Fig. 1 Screenshot: verification of computed final standing with an online standings calculator (Olympia-Verlag GmbH 2019a). Platz = position, Spiele = games played, g. = won, u. = drawn, v. = lost, Tore = goals for/against, Diff. = goal difference, Pkt. = points

Fig. 2 Results obtained by the optimization models on every matchday of the Bundesliga season 2017/2018 for four exemplarily chosen teams. The left y-axes are used to depict the current rank, the best-case and the worst-case final rank according to the respective matchday (x-axes). The right y-axes shows the number of points

of the first round). On matchday 18, the corridor between the two lines abruptly widens, because with respect to the end of the tournament now “everything” is possible again. In the chart for *FC Bayern München* the triangle symbolizing the fact that non-relegation was certain only on matchday 21 is encircled. By the way, 14 points from seven games marks the worst start to the season in seven years. For the first time in the club’s history, *FC Bayern München* lost a two-goal lead twice in a row (this cannot be seen from the graphic). However, *FC Bayern München* was top of the table for the rest of the season as of matchday 10 and became German Champion with 72 points already on matchday 29, as the worst- and best-case lines meet at position one on that matchday. As the only team in the Bundesliga *Bayer 04 Leverkusen* scored at least one goal in all 17 first-round matches. The team has worked its way up in the first-round matches and showed continuity in the second round. *Hamburger SV*, being the only constant member of the Bundesliga since its establishment, came second to last and was finally relegated to 2. Bundesliga. Due to a win of *VfL Wolfsburg* against 1. *FC Köln*, relegation became inevitable on the very last matchday, as the chart shows.

Remark 1 The models and data of all instances examined in the article can be downloaded from https://cloud.zib.de/s/bundesliga_data/download.

Acknowledgements Open Access funding provided by Projekt DEAL. Thanks for the inspiration to René Henrion and Thorsten Koch. And of course to our moms Karin and Anke for their endless support.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Achterberg T (2009) Scip: solving constraint integer programs. *Math Program Comput* 1(1):1–41
- Bernholt T, Gülich A, Hofmeister T, Schmitt N (1999) Football elimination is hard to decide under the 3-point-rule. In: *Mathematical foundations of computer science*, pp 410–418
- Der Postillon (2019) <https://www.der-postillon.com/>. Accessed 27 Dec 2019
- Deutsche Fußball Liga (2019) Spielordnung. <https://www.dfl.de/de/spielordnung-spol-2019-08-22-stand/>. Accessed 27 Dec 2019
- e-quadrat Communications Malta Limited and HEIM:SPIEL Medien GmbH & Co KG (2019) Worldwide stadium attendances. <http://www.weltfussball.de/zuschauer/bundesliga-2016-2017/1/>. Accessed 27 Dec 2019
- Gleixner A, Eifler L, Gally T, Gamrath G, Gemander P, Gottwald RL, Hendel G, Hojny C, Koch T, Miltenberger M, Müller B, Pfetsch ME, Puchert C, Rehfeldt D, Schlösser F, Serrano F, Shinano Y, Viernickel JM, Vigerske S, Weninger D, Witt JT, Witzig J (2017) The SCIP optimization suite 5.0. Technical report, Optimization online. http://www.optimization-online.org/DB_HTML/2017/12/6385.html. Accessed 27 Dec 2019

- Gusfield D, Martel C (2002) The structure and complexity of sports elimination numbers. *Algorithm N Y* 32(1):73–86
- Kendall G, Knust S, Ribeiro CC, Urrutia S (2010) Scheduling in sports: an annotated bibliography. *Comput Oper Res* 37:1–19
- Kern W, Paulusma D (2001) The new fifa rules are hard: complexity aspects of sports competitions. *Discret Appl Math* 108:317–323
- Kern W, Paulusma D (2004) The computational complexity of the elimination problem in generalized sports competitions. *Discret Optim* 1:205–214
- Koch T (2004) Rapid mathematical prototyping. PhD thesis, Technische Universität Berlin. <http://zimpl.zib.de/>. Accessed 27 Dec 2019
- Nießén D (2014) Nichtabstieg perfekt: FC Bayern München feiert vorzeitigen Klassenerhalt. *Der Postillon*. <https://www.der-postillon.com/2014/02/nichtabstieg-perfekt-fc-bayern-munchen.html>. Accessed 27 Dec 2019
- Nießén D (2018) Nichtabstieg perfekt: FC Bayern München feiert vorzeitigen Klassenerhalt. <https://www.der-postillon.com/2018/01/klassenerhalt-bayern.html>. Accessed 27 Dec 2019
- Olympia-Verlag GmbH (2019a) Online standings calculator. <https://www.kicker.de/1-bundesliga/tabel-lenrechner/2017-18/1>. Accessed 27 Dec 2019
- Olympia-Verlag GmbH (2019b) Tables and pairings. <https://www.kicker.de/1-bundesliga/spieltag/2017-18/1>. Accessed 27 Dec 2019
- Raack C, Raymond A, Schlechte T, Werner A (2014) Standings in sports competitions using integer programming. *J Quant Anal Sports* 10:131–137
- Robinson LW (1991) Baseball playoff eliminations: an application of linear programming. *Oper Res Lett* 10(2):67–74
- Schlotter I, Cechlárová K (2018) A connection between sports and matroids: how many teams can we beat? *Algorithmica* 80(1):258–278
- Schrijver A (1986) *Theory of linear and integer programming*. Wiley, Hoboken
- Schwartz BL (1966) Possible winners in partially completed tournaments. *SIAM Rev* 8(3):302–308
- The Postillon (2019) <https://www.the-postillon.com/>. Accessed 27 Dec 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.