

# Ranking Participants in Tournaments by means of Rating Functions\*

**Miguel Brozos-Vázquez**

Dept. of Geometry and Topology  
Universidade de Santiago de Compostela

**Marco Antonio Campo-Cabana**

Dept. of Applied Mathematics  
Universidade de Santiago de Compostela

**José Carlos Díaz-Ramos**

Dept. of Mathematics  
University College Cork

**Julio González-Díaz<sup>†</sup>**

Kellogg School of Management  
Northwestern University

Accepted in Journal of Mathematical Economics (2008) 44, 1246-1256

Published version available at <http://www.sciencedirect.com>

DOI 10.1016/j.jmateco.2008.03.002

## Abstract

In this paper we bring a novel approach to the theory of tournament rankings. We combine two different theories that are widely used to establish rankings of populations after a given tournament. First, we use the statistical approach of paired comparison analysis to define the performance of a player in a natural way. Then, we determine a ranking (and rating) of the players in the given tournament. Finally, we show, among other properties, that the new ranking method is the unique one satisfying a natural consistency requirement.

## 1 Introduction

When there is a competition among the members of a population, the fundamental problem is to rank these members according to their performance.<sup>1</sup> In certain cases this confrontation takes the form of a tournament in which contestants play against themselves obtaining a certain score in each match. The aim is to determine a final ranking after all the matches have been played. Because of the wide range of applications of ranking theory, this problem has already been widely studied in management science, economic theory and statistics. The books by Moon (1968) and Laslier (1997) and the paper Iqbal et al. (1986) discuss several applications of the theory of tournament rankings.

---

\*We would like to thank the referee for his useful comments and suggestions to improve the paper. Support from the Spanish Ministry for Science and Education through projects SEJ2005-07637-C02-02 and MTM2006-01432 and from the Xunta de Galicia under projects PGIDT03PXIC20701PN and PGIDT06PXIB207054PR is gratefully acknowledged. The research of Julio González-Díaz and José Carlos Díaz-Ramos has also been supported by two Marie Curie International Fellowships within the 6th European Community Framework Programme.

<sup>†</sup>Corresponding author: Kellogg School of Management (CMS-EMS), Northwestern University. 5100 Leverone Hall, 2001 Sheridan Road. Evanston, IL 60208-2014. Phone number: +1 847-467-1745. Fax: +1 847-491-2530. E-mail address: julio@northwestern.edu

<sup>1</sup>We usually refer to the members of our population as contestants or players, but they may also be other objects such as scientific journals, political options or products to be tested.

Formally, a ranking of a population  $N$  is a complete, reflexive and transitive relation on  $N$ . To fix notation, we use the word rating when we have a cardinal ranking, that is, not only do we have an ordering of the contestants, but also a measure of the intensities of the differences among them.

In this paper, we define a measure of the strength exhibited by the players of a given tournament that we call *performance*. More specifically, we assume that there is a distribution function that governs the random process associated with the competitive environment in which the given tournament takes place. Within this setting, the performance of a player is naturally defined as a function of his results and the strength of his opponents. Then, we define a new ranking method, the *recursive performance*, and show that it is the unique one that is consistent with the notion of performance.

The most natural attempt to associate a ranking to a given tournament is to use the *scores ranking*. That is, rank the players according to their total scores. Rubinstein (1980) provides an axiomatic characterization of the scores ranking for a certain class of tournaments. Nonetheless, the scores ranking does not generally deliver sensible rankings when the tournaments have unbalanced schedules, that is, tournaments that do not correspond with the commonly known as round-robin tournaments (all the players have played exactly once against each other). Hence, most ranking methods determine the ranking of a player according to the results obtained in the tournament and the strength of the opponents the player has played against, being this last feature missing when using the scores ranking.<sup>2</sup> One widely used ranking method that takes the previous considerations into account is the minimum violations ranking, where one violation consists of two players whose relative ranking differs from the one induced by their result against each other. This ranking method is discussed, for instance, in Goddard (1983) and Iqbal et al. (1986). However, Stob (1985) is quite critical with it and sticks up for the statistical approach of paired comparison analysis initiated in Zermelo (1929) for chess tournaments and rediscovered by Bradley and Terry (1952).

In Bradley and Terry (1952), each player  $i$  is assumed to have a strength parameter  $r_i$ . It is also assumed that there exists a distribution  $F$  such that  $F(r_i, r_j)$  is the probability that  $i$  beats  $j$ . The objective is to fix a distribution  $F$  that properly fits the available data and then, using statistical tools, calculate the most likely values of the strength parameters  $r_i$ . In particular, Bradley and Terry (1952), whose approach has become the most widely used within this literature, use maximum likelihood estimation to derive the  $r_i$  parameters. Once these values are calculated, they may be used to define a rating for each player. As compared with the minimum violations ranking, Stob (1985) emphasizes that this approach endogenizes the relevance that must be given to the different victories (losses) of the players, a feature that we also consider very desirable. Two classic references within this framework are Kendall and Smith (1940) and David (1988).

Finally, a third approach comes from economic theory, where it is often the case that an axiomatic analysis is taken to determine a ranking of the population  $N$ . First, it is assumed that there is a matrix containing the relevant information about the paired results of the different players; this matrix is usually referred to as the tournament matrix. Then, a ranking method is defined as a function that ascribes a ranking to each tournament matrix. Next, the properties of the different methods are studied. Finally, a ranking method, whose properties are suitable for a given competitive environment, is chosen. To deepen into the economic literature on rankings refer to Rubinstein (1980), Liebowitz and Palmer (1984), Amir (2002), Palacios-Huerta and Volij (2004) and Slutzki and Volij (2005).

---

<sup>2</sup>Indeed, this is the reason why the scores ranking is mainly used in round-robin tournaments, where each player faces the same opponents.

In this paper we consider a competitive environment with an associated function  $F$  that describes the underlying random process. This function can be chosen among the ones whose properties have already been studied in the literature (for instance, the one used by Bradley and Terry (1952)) or determined from the data of the historical confrontations of the players in a population. Throughout this paper we take the rating function as given. Then, we assume that we have one more tournament and we want to rank the players of that tournament according to their results. Next, we bring together two widely used ideas. On the one hand, our ranking method, the *recursive performance*, is defined using a recursive formula that resembles the Liebowitz-Palmer method (Liebowitz and Palmer, 1984) and other similar methods studied in economic theory (Palacios-Huerta and Volij, 2004). On the other hand, our recursive formula uses the rating function  $F$ , so basic in the statistical approach to paired comparison analysis. Moreover, we show one property of our ranking method that is crucial for its applicability. Namely, our ranking method is robust with respect to the function  $F$ , that is, small changes in the function  $F$  lead to small changes in the proposed rating. Finally, concerning the computation of our ranking method, we show that it reduces to solving a linear system. The last two properties represent a notable advantage with respect to the maximum likelihood approach, which, quite generally, is not robust in the above sense and requires solving a nonlinear system of equations.

Within the literature in which this paper is enclosed, there is a family of tournaments that has received special attention: the round-robin tournaments. In these tournaments each player plays exactly once against any other player. The minimum violations ranking is essentially thought for this class of tournaments. Stob (1985) showed that, under quite general assumptions, the ranking methods defined using the approach of Zermelo (1929) lead to the same ordering as the scores ranking. We briefly discuss the behavior of our approach in round-robin tournaments at the end of Section 4.

As an immediate application, the results obtained in this paper can be applied to define new tie-breaking rules for disciplines such as chess and Othello.<sup>3</sup> In our opinion, these new tie-breaking rules improve the existent ones (see Sections 4 and 5 for details).

We briefly describe the contents of this paper. In Section 2 we present the concept of tournament and comment on the assumptions used throughout this paper. In Section 3 we formally define and discuss the notion of performance. Section 4 is the core of our study; we introduce the so-called recursive performance and give examples of tournaments in which it might be immediately applied. In Section 5 we discuss two properties of the recursive performance ranking method: robustness and a consistency property. We devote Section 6 to compare the recursive performance with the rating delivered by the maximum likelihood estimates, specially when the rating function is the one used by Bradley and Terry (1952). Finally, in Section 7 we prove the results presented in Sections 4 and 5.

## 2 Basic Definitions: Tournaments

We have a competitive environment in which confrontations between the different players of a population take place along time. For such an environment, there is a *rating function*  $F$  that accurately describes the probabilities associated with the different results of each match between any two given players. This rating function is such that, given two players  $i$  and  $j$  with strength parameters  $r_i$  and  $r_j$ , the probability that  $i$  beats  $j$  is  $F(r_i, r_j)$ . Thus,

---

<sup>3</sup>Remarkably, ties are always present in tournaments in which pairings are drawn following the Swiss pairing system, which is, along with the round-robin system, the most widely used in these disciplines.

$F(r_i, r_j) = 1 - F(r_j, r_i)$ .<sup>4</sup> We refer to the strength parameters  $r_i$  as *ratings*.

We work within the *linear paired comparison model* (David, 1988). More specifically, we assume that there is a strictly increasing continuous distribution function  $F_\ell : \mathbb{R} \rightarrow (0, 1)$  such that  $F(r_i, r_j) = F_\ell(r_i - r_j)$ , that is, the result of a game between any two players depends only on their rating difference. The probability that  $i$  beats  $j$  goes to 1 as  $r_i - r_j$  grows and the probability that  $i$  beats  $j$  is positive regardless of the rating difference. Moreover, since  $F(r_i, r_j) = 1 - F(r_j, r_i)$ ,  $F_\ell$  is symmetrically distributed about zero. Also, note that the function  $F_\ell^{-1}$  is well defined.

The result of a confrontation between two players  $i$  and  $j$  may be not only a win or a loss but any pair  $(a_i, a_j)$  with  $a_i, a_j \geq 0$ ,  $a_i + a_j = 1$ . Within this scenario we might interpret  $F(r_i, r_j)$  as the expected score of player  $i$  when facing player  $j$ . An important example of linear rating function is the one used in Bradley and Terry (1952), where  $F_\ell$  is taken to be the (standard) logistic distribution given by  $F_\ell(x) = \frac{1}{1+e^{-x}}$  and hence,  $F(r_i, r_j) = F_\ell(r_i - r_j) = \frac{e^{r_i}}{e^{r_i} + e^{r_j}}$ .

We denote by  $\mathcal{M}_{k \times l}$  the vector space of real  $k \times l$  matrices. A *tournament* is a pair  $(N, A)$ , where  $N = \{1, \dots, n\}$  is the set of players and  $A \in \mathcal{M}_{n \times n}$  is the *tournament matrix*. The matrix  $A$  is non-negative and its main diagonal consists of zeros. The entry  $A_{ij}$  contains the aggregate score of  $i$  against  $j$ . Note that we do not restrict the non-zero entries in our matrix  $A$  to be natural numbers as in Slutzki and Volij (2005). To each tournament  $(N, A)$  we can naturally associate a (symmetric) *matches matrix*  $M(A) := A + A^t$ , i.e.,  $M(A)_{ij}$  is the number of matches between  $i$  and  $j$ . For notational simplicity, when no confusion arises we denote  $M(A)$  by  $M$ . Thus, if  $M_{ij} \neq 0$ ,  $A_{ij}/M_{ij}$  denotes the *average score of  $i$  against  $j$* , i.e., the aggregate score of  $i$  against  $j$  divided by the total number of matches between them. Since the  $n$  players participate in the tournament, each row of  $M$  must have a nonzero entry. For each player  $i$ , let  $m_i := \sum_{j=1}^n M_{ij}$  denote the total number of matches played by  $i$ . Let  $r \in \mathbb{R}^n$  be a vector of exogenously given ratings of the players in the tournament  $(N, A)$ . We refer to  $r$  as the *vector of initial ratings*.

Given a tournament  $(N, A)$  and a vector of ratings  $r$ , we define the *total strength of the tournament* by  $\sigma^r := \sum_{i=1}^n m_i r_i$ , that is, the contribution of each player to the total strength of the tournament is weighted by the number of matches he has played. Similarly, we define the *average strength of the tournament* by  $\bar{\sigma}^r := \sigma^r / \sum_{i=1}^n m_i$ . Note that, if all the players have played the same number of games, the average strength of the tournament is just the average of the vector of initial ratings.

Summarizing, the primitives of our model are a tournament  $(N, A)$ , a rating function  $F$ , and a vector of initial ratings  $r$ . In this paper we present a ranking and rating method for the tournament in question. Our ranking method reallocates among the players the total strength of the tournament,  $\sigma^r$ , with two important features: first, the proposed ranking does not depend on  $r$  and, second, the difference between the proposed ratings for any two players is also independent of  $r$ . That is, our method is endogenous to  $(N, A)$  and the vector  $r$  is used just for the sake of exposition.

Since the ranking method we define in this paper is anonymous, the labels chosen for indexing the players are irrelevant. Based on this fact, two tournaments that are equal up to labeling are said equivalent. We make this definition precise. Denote by  $L_{\alpha\beta} \in \mathcal{M}_{n \times n}$  the *transposition matrix* that swaps the  $\alpha$ th and  $\beta$ th entries of a vector. A transposition matrix satisfies  $L_{\alpha\beta}^{-1} = L_{\alpha\beta}$  and, given  $B \in \mathcal{M}_{n \times n}$ , the product  $L_{\alpha\beta} B$  is the same matrix  $B$  but with rows  $\alpha$  and  $\beta$  interchanged. Similarly,  $B L_{\alpha\beta}$  interchanges columns  $\alpha$  and  $\beta$  of  $B$ . The group  $\Pi_n$  generated by the composition of transposition matrices  $L_{\alpha\beta}$  is isomorphic to

<sup>4</sup>For instance, chess and Othello use rating systems based on functions that have already been widely tested.

the group of permutations of  $n$  elements. Given two populations  $N$  and  $N'$  with  $n$  players, we say that two tournaments  $(N, A)$  and  $(N', A')$  are *equivalent* if there exists  $L \in \Pi_n$  such that  $A = LA'L^t$ . Note that this also implies that  $M = LM'L^t$ . Since for each  $L \in \Pi_n$  we have  $L^{-1} = L^t$ , two equivalent tournaments have similar tournament matrices and similar matches matrices.

A matrix  $B \in \mathcal{M}_{n \times n}$  is *block diagonal*, respectively *block anti-diagonal*, if

$$B = \left( \begin{array}{c|c} * & 0 \\ \hline 0 & * \end{array} \right), \text{ respectively } B = \left( \begin{array}{c|c} 0 & * \\ \hline * & 0 \end{array} \right).$$

Note that, given a tournament  $(N, A)$ , if  $M(A)$  is block diagonal (anti-diagonal), then  $A$  is block diagonal (anti-diagonal). We assume that our tournaments satisfy the following two natural assumptions:

**A1.** *The tournament  $(N, A)$  is not equivalent to a tournament  $(N', A')$  such that  $M(A')$  is block diagonal.*<sup>5</sup>

If the matches matrix  $M(A')$  is block diagonal, the tournament has an internal division: there are two disjoint subsets of players such that none of the players of one set has played against anyone of the other set. This is a standard assumption in the ranking's literature since each block may be considered as the matrix of an independent tournament.

**A2.** *The tournament  $(N, A)$  is not equivalent to a tournament  $(N', A')$  such that  $M(A')$  is block anti-diagonal.*

If  $M(A')$  is block anti-diagonal, then the tournament may be considered as a team-tournament. There are two disjoint subsets (teams) such that the players of each team have played only against the players of the other, but not among themselves. Although similar to A1, this property is more subtle and has different implications. In this case, in order to calculate the strength of the players of one team, we need the strength of the players of the other team that can only be calculated using the strength of the players in the first team. This cyclic feature of team-tournaments is the reason why, if A2 is not met, the iterative method we describe in Section 4 does not necessarily converge.

### 3 The Notion of Performance

Let  $(N, A)$  be a tournament. The *vector of average scores*,  $s$ , is defined by  $s_i := \sum_j A_{ij}/m_i$ , *i.e.*, the sum of the aggregate scores of player  $i$  divided by his total number of matches. Hereafter, we assume that  $s \in (0, 1)^n$ .<sup>6</sup> We define  $D^M := \text{diag}(m_1, \dots, m_n) \in \mathcal{M}_{n \times n}$  and  $\bar{M} := (D^M)^{-1}M \in \mathcal{M}_{n \times n}$ , that is,  $\bar{M}_{ij} = M_{ij}/m_i$  is the number of confrontations between  $i$  and  $j$  divided by the total number of matches played by  $i$ .

Let  $r \in \mathbb{R}^n$  be the vector of initial ratings and let  $F$  be the distribution function of the linear paired comparison model. The *vector of  $r$ -performances*,  $p^r \in \mathbb{R}^n$ , is defined as

$$p^r := \bar{M}r + c, \quad \text{where } c_i := F_\ell^{-1}(s_i).$$

Note that  $(\bar{M}r)_i$  coincides with the average rating of the opponents of player  $i$ . Hence, the  $r$ -performance of player  $i$  is the unique rating  $p_i^r$  such that  $F(p_i^r, (\bar{M}r)_i) = s_i$ . Then, we

<sup>5</sup>This is equivalent to saying that  $M(A')$  is *irreducible*.

<sup>6</sup>This assumption is standard when working with rating functions, since it allows to apply function  $F_\ell^{-1}$  to the observed scores; for instance, it is also needed for those ranking methods based on the maximum likelihood approach.

might say that  $s_i$  is  $i$ 's expected score against a player of rating  $(\bar{M}r)_i$  if and only if  $i$  has a rating  $p_i^r$ . This justifies the name performance.

The vector  $p^r$  depends on the tournament and the rating function associated with the competitive environment which the tournament belongs to. Nevertheless, it also depends on the vector of initial ratings, which is exogenous to the tournament. On the other hand, the  $r$ -performance of player  $i$  takes into account the strengths of his opponents; however we consider that also the opponents of the opponents of  $i$  should be used to calculate the strengths of the opponents of  $i$ , and the  $r$ -performances do not do so. We devote Section 4 to address these two problems.

*Example 1.* The World Chess Federation (FIDE) has an official rating of players called Elo. Let  $r$  denote the vector of Elo ratings of the players participating in a given tournament. Elo's formula considers the distribution  $F_\ell$  given by  $F_\ell(\lambda) = 1/(1 + 10^{-\frac{\lambda}{400}})$ , which is, as the rating function used in Bradley and Terry (1952), a logistic distribution. Hence, the  $r$ -performance of player  $i$  in a tournament is defined as  $p_i^r = (\bar{M}r)_i - 400 \log_{10}(1/s_i - 1)$ , that is, the average of the Elos of his opponents plus a correcting factor depending on the percentage of points achieved by the player. Remarkably, the  $r$ -performance is one of the tie-breaking rules recommended by the FIDE for chess tournaments.

The assumption  $s \in (0, 1)^n$  is needed in order to define the vector of performances correctly, but this situation holds in most tournaments. Indeed, the  $r$ -performance ranking method is already used as a tie breaking rule for chess tournaments, as we have just seen. The idea of this method is to use the strengths of the opponents of the players to define the rankings. Note that the vector  $r$  is the "historical" strength of the players whereas the vector  $c$  is, essentially, the average score of each player in the tournament. Hence,  $p^r$  measures the results of a player in relation to the strength of his opponents.

## 4 Recursive Performance

In the definition of  $r$ -performance, the vector  $c$  depends crucially on the rating function  $F$  and, although our notation does not make this dependence explicit, the rating function keeps being an essential element of this paper.

The ranking associated with the vector of  $r$ -performances is not a bad ranking for the tournament, but it heavily depends on the initial ratings  $r_1, \dots, r_n$ . The latter measure the historical strength of the players, which might be different from the strength exhibited by the players in the tournament. Moreover, in the paired comparison literature, these ratings are often calculated using the method of maximum likelihood, and thus, they are subject to statistical errors.

The objective of this section and the next one is to formally introduce a new ranking method, the *recursive performance*, and discuss some of its properties. Refer to Section 7 for the technical results and their corresponding proofs.

As we have already discussed in Section 3, even though  $(p^r)_i$  is a better indicator of the strength of  $i$  in the given tournament than his initial rating, using the vector  $p^r$  as a ranking method has relevant shortcomings. Nonetheless, it is natural to calculate a new performance by replacing the initial ratings with the vector of performances. This would have two main effects. Namely, it would shade the dependence on  $r$  and, given a player  $i$ , the new rating would take into account, not only the opponents of  $i$ , but also their results (*i.e.*, using the opponents of the opponents of  $i$  as well). This suggests the iterative method  $p^{(0)} := p^r (= \bar{M}r + c)$ ,  $p^{(l+1)} := p^{p^{(l)}} (= \bar{M}p^{(l)} + c)$ . Unfortunately, this method does not necessarily converge. On the other hand, the total strength of the tournament,  $\sigma^r$ , is not

preserved by the  $r$ -performance, that is,  $\sum_i m_i r_i \neq \sum_i m_i p_i^r$  in general. The inflation or deflation factor turns out to be  $\sum_i m_i c_i$ . By distributing this factor among the players, we define a new iterative method that does preserve the total strength of the tournament.

Let  $e \in \mathbb{R}^n$  be the vector  $e = (1, \dots, 1)$ . Consider the following rescaling of  $c$ ,

$$\hat{c} := c - \left( \frac{\sum_i m_i c_i}{\sum_i m_i} \right) e,$$

which we discuss below. We define the *iterated performance* as the iterative method

$$\begin{aligned} \hat{p}^{(0)} &:= \bar{M}r + \hat{c} \\ \hat{p}^{(l)} &:= \bar{M}\hat{p}^{(l-1)} + \hat{c}, \quad l \in \mathbb{N}. \end{aligned}$$

At each step  $l$ , this iterative method gives the same rating as the previous one up to a constant proportional to  $e$  and, hence, *the two proposed rankings are always the same*. Indeed, the adjustment of  $c$  via  $\hat{c}$  is the only one that preserves the total strength of the tournament while leading to a rating that is *essentially identical* under  $F_\ell$  (refer to the consistency property in Section 5). This is proved in the following lemma.

**Lemma 1.** *For each  $l \in \mathbb{N}$ ,  $p^{(l)} - \hat{p}^{(l)} = (l + 1) \left( \frac{\sum_i m_i c_i}{\sum_i m_i} \right) e$ .*

*Proof.* Since by definition of  $\bar{M}$ ,  $\bar{M}e = e$ , the result follows by an induction argument.  $\square$

In Section 7 we prove that, under our assumptions, the iterated performance converges. More specifically, we have the following result.

**Theorem 2.** *Let  $(N, A)$  be a tournament satisfying assumptions A1 and A2 and such that  $s \in (0, 1)^n$ . Let  $r \in \mathbb{R}^n$  be a vector of initial ratings and  $F_\ell$  the distribution of the linear paired comparison model. Then the iterated performance converges.*

Motivated by Theorem 2, we define the main concept of this paper, the *recursive performance*, as the limit of the iterated performance:  $\hat{p} := \lim_{l \rightarrow \infty} \hat{p}^{(l)}$ . Taking limits in the equality  $\hat{p}^{(l)} = \bar{M}\hat{p}^{(l-1)} + \hat{c}$ , we get that  $\hat{p}$  is a solution to the linear system

$$(I - \bar{M})x = \hat{c}, \tag{1}$$

where  $I \in \mathcal{M}_{n \times n}$  is the identity matrix. If A1 holds, by Theorem 5 (iii) below, the matrix  $I - \bar{M}$  has rank  $n - 1$  (even if A2 does not hold). Then, since  $\bar{M}e = e$ , the whole set of solutions to (1) is given by  $\hat{p} + \text{span}\{e\}$ . When using the iterated performance method, the different solutions to (1) arise from different initial vectors of ratings  $r$ . It is important to note that *all the solutions propose the same ranking*. When working with linear rating functions, it is often the case that a normalization is needed to pin a unique rating of the players; for instance, this is also so in Bradley and Terry (1952).

Following the previous discussion, even if A2 does not hold, since all the solutions to the linear system lead to the same ranking, we can still unambiguously associate a ranking to each linear system (1) as far as A1 is met.

*Example 2.* *An ideal chess-like tournament* is a tournament in which all the players play the same number of rounds, say  $k$ .<sup>7</sup> Thus, for an ideal chess-like tournament  $(N, A)$  we have  $D^M = kI$  and, hence,  $\bar{M} = M/k$ . By Corollary 7 below,  $\sum_i m_i \hat{p}_i^{(l)} = \sum_i m_i r_i = \sigma^r$  for all  $l \in \mathbb{N}$ . Since  $m_i = k$  for all  $i \in N$ , in an ideal chess-like tournament we have  $\sum_i \hat{p}_i^{(l)} = \sum_i r_i$ .

<sup>7</sup>Most tournaments in disciplines such as chess and Othello have this property.

This shows that, using  $\hat{c}$  instead of  $c$ , we adjust the vectors after each iteration to ensure that the sum of the ratings after each step remains constant. The average strength of an ideal tournament,  $\bar{\sigma}^r$ , coincides with the average of the components of  $r$ . In each iteration the method proposes a way to divide the total strength of the tournament,  $\sigma^r$ , among the players. That is, by working with  $\hat{c}$  instead of  $c$ , we avoid inflation or deflation in the vectors of iterated performances. Since the recursive performance is the limit of such vectors, it also provides a way of dividing the total strength of the tournament among the players. In a general tournament, the same property holds, but in this case the average strength of the tournament is calculated as a weighted average (by the  $m_i$ 's) of the components of  $r$ .

*Example 3.* A *round-robin tournament*  $(N, A)$  is a tournament in which all the players have played exactly once against each other. That is,  $M_{ij} = 1$  if  $i \neq j$  and  $M_{ii} = 0$ .<sup>8</sup> Within this family of tournaments, the ranking proposed by the recursive performance has an appealing feature: it coincides with the ranking proposed by the vector of average scores (the standard scores ranking method). To see this, let  $x$  be a solution to the system (1). Then, the claim follows from the equality  $\hat{c}_i - \hat{c}_j = x_i - \sum_{k \neq i} \frac{x_k}{(n-1)} - x_j + \sum_{k \neq j} \frac{x_k}{(n-1)} = \frac{n}{n-1}(x_i - x_j)$ . This is not surprising since the ranking proposed by the recursive performance uses both the scores of the players and the scores of the opponents, but all the players have the same opponents.

## 5 Main Properties: Robustness and Consistency

Since all the solutions to (1) propose the same ranking, the recursive performance ranking method does not depend on the vector of initial ratings. Moreover, the recursive performance ranking may unambiguously be calculated for tournaments in which there are unrated players (players with no historical results). Thus, if there is an unrated player, we can assign him an arbitrarily chosen rating and this election does not affect the final ranking. The recursive performance proposes a way to divide the total strength of the tournament among the players. Therefore, when used as a rating method, there is neither inflation nor deflation with respect to the initial ratings.

### Robustness in $F$

The rating function  $F$  is a crucial element of the definition of the recursive performance. As such, this ranking method and any other that depends on the specification of a rating function, are sensitive to how close the chosen rating function is to the one that really governs the distribution of results in the given discipline. Therefore, it would be convenient for a ranking method that similar rating functions proposed similar ratings. In this sense, we say that a rating method is *robust* in  $F$  if small changes in  $F$  lead to small changes in the proposed ratings. Remarkably, the robustness of the recursive performance follows immediately from the fact that it is the solution to the linear system (1).

### Consistency with $F$

Given two vectors of ratings  $r^1, r^2 \in \mathbb{R}^n$ , we say that they are *essentially identical for function  $F$*  if  $F(r_i^1, r_j^1) = F(r_i^2, r_j^2)$  for all  $i \neq j$ . That is, two vectors are essentially identical if they lead to the same predictions for any given match between players in  $N$ .

---

<sup>8</sup>Round-robin tournaments have a special structure and different approaches have been taken to define ranking methods within this family of tournaments. See for instance Daniels (1969), Stob (1985) and Chapter 6.1 in David (1988).



In linear paired comparisons analysis this is equivalent to the existence of  $\lambda \in \mathbb{R}$  such that  $r^1 = r^2 + \lambda e$ .

**Lemma 3.** *Let  $(N, A)$  be a tournament and  $F_\ell$  a linear rating function. Then, all the solutions to (1) are essentially identical for function  $F_\ell$ .*

*Proof.* It follows from the fact that all the solutions to (1) are of the form  $\hat{p} + \text{span}\{e\}$ .  $\square$

Because of this property, we make a slight abuse of language and use the words recursive performance to encompass all the solutions of (1). Then, the total strength of the tournament suffices to pin a unique rating.

Now, we are ready to introduce a consistency property. We say that a rating  $r \in \mathbb{R}^n$  is *consistent with  $F$  for tournament  $(N, A)$*  if the  $r$ -performance rating vector is essentially identical to  $r$ . This consistency property can also be extended to any ranking (not necessarily cardinal). A ranking  $\succeq$  is *consistent with  $F$  for the tournament  $(N, A)$*  if there is a rating  $r \in \mathbb{R}^n$  that is consistent with  $F$  and whose induced ranking coincides with  $\succeq$ .

At the end of Section 7 we give a proof of the following result.

**Proposition 4.** *Let  $(N, A)$  be a tournament and  $F_\ell$  a linear rating function. Then, the ranking induced by the recursive performance is the unique one that is consistent with  $F_\ell$ .*

## 6 Recursive Performance Vs. Maximum Likelihood

Ideally, given a tournament  $(N, A)$ , one would like to associate with it a rating of the players that explains all the observed results. More formally, we say that a rating  $r \in \mathbb{R}^n$  is *pairwise-consistent* with  $F$  if, for each pair of players  $i$  and  $j$ ,  $F(r_i, r_j) = \frac{A_{ij}}{M_{ij}}$ , *i.e.*, the observed proportion of victories of  $i$  against  $j$  is exactly what one would predict using  $F$  and the ratings  $r_i$  and  $r_j$ . Unfortunately, trying to find a pairwise consistent rating reduces to solving a system with far more equations than variables, which, for most tournaments, will have no solution. We informally discuss now how the recursive performance (RP) and the rating obtained by maximum likelihood (ML) differ in how they circumvent the above problem and comment some of the features of the two approaches.

One can use the consistency property defined in Section 5 to reinterpret RP in the following way. Given a tournament  $(N, A)$ , a rating  $r \in \mathbb{R}^n$ , and a player  $i$ , the *average opponent* of  $i$  in the tournament is  $(\bar{M}r)_i$ , *i.e.*, the average rating of the opponents of  $i$  (weighted by the number of matches played against each of them). Suppose now that  $(N, A)$  is such that  $\sum_i m_i c_i = 0$ , *i.e.*, there is no inflation ( $c = \hat{c}$ ). Since finding a pairwise-consistent rating is impossible in general, RP looks for a rating that is consistent with the average opponents, namely, for each player  $i \in N$ ,  $F(r_i, (\bar{M}r)_i) = \sum_j A_{ij}/m_i = s_i$ . By Proposition 4, any rating that is consistent in this sense has to be essentially identical to the RP rating. Hence, by imposing the consistency requirement in the average opponent instead of doing it in all the comparisons, the problem reduces to a system of linear equations as the one in (1), for which we can always find an exact solution.

On the other hand, what ML does is to find the ratings under which the probability of the observed results is maximized. One advantage of ML is that the estimated ratings are consistent in the sense of statistical inference. That is, as the number of matches between every pair of players goes to infinity, the ratings obtained with ML converge to the true ratings. Therefore, we can say that these ratings are pairwise-consistent asymptotically.

From the above re-interpretations of the RP and the ML ratings, one may lean toward the latter, since asymptotic pairwise-consistency seems more desirable, and maybe more

natural, than consistency with the average opponent. However, under ML, even mild misspecifications of the rating function  $F$  can lead to severe asymptotic bias.<sup>9</sup> One reason for this is that ML estimates usually come from solving non-linear systems of equations and the robustness property we presented above for RP is therefore lost under ML. At the same time, non-linearity poses some problems to the computation of the ML estimates, since the corresponding iterative algorithms can be very slow; Dykstra (1956) showed that, under the logistic distribution assumed by Bradley and Terry (1952), even in the case of round-robin tournaments, it is important to get good starting values to get a reasonably fast convergence. In contrast, the linearity of RP ensures that the error in an iterative method like the one we described in Section 4 will converge to 0 at a speed given by  $|\beta|^l$ , where  $l$  is the number of iteration and  $\beta$  is the second largest eigenvalue of  $\bar{M}$ , which, by Theorem 5, is smaller than 1.

We present now the special cases of the two methods when the rating function is the (standard) logistic distribution, which, in the case of ML, is the most common one and was introduced by Zermelo (1929) and rediscovered by Bradley and Terry (1952). Then,  $F_\ell(x) = \frac{1}{1+e^{-x}}$  and  $F_\ell^{-1}(y) = \log(\frac{y}{1-y})$ .

In this setting, the rating  $\hat{p}$  given by the RP is such that, for each player  $i$ ,

$$\hat{p}_i = \sum_j \frac{M_{ij}}{m_i} \hat{p}_j + \log\left(\frac{s_i}{1-s_i}\right) - \frac{\sum_j m_j \log(\frac{s_j}{1-s_j})}{\sum_j m_j}.$$

On the other hand, ML reduces to calculate a positive vector  $\pi$  such that  $\sum_i \pi_i = 1$  and, for each player  $i$ ,<sup>10</sup>

$$\pi_i = \frac{s_i}{\sum_j \frac{M_{ij}}{m_i(\pi_i + \pi_j)}}.$$

Then, for each player  $i$ , his rating according to maximum likelihood is given by  $\hat{q}_i = \log(\pi_i)$ . Despite the apparent extra simplicity of the expression for ML, note that it defines a nonlinear system of equations. Analogously to the RP ratings, any translation of the ML ratings by a vector proportional to  $e$  leads to the same predictions for any given match between players in  $N$  and hence it is essentially identical to  $\hat{q}_i$ .

From the two equations above, one can easily check that both ranking methods use the information of the tournament in a similar way. In both of them the rating of a player is increasing in his average score and in the ratings of his opponents, which are weighted by the number of matches played against each of them. Therefore, both ranking methods will, quite generally, lead to similar rankings. We illustrate now some differences in the behavior of both RP and ML in a series of simple examples. In all of them we present both the RP and ML ratings normalized so that they add up to 0. Let  $N = \{1, 2, 3, 4\}$  and consider the tournament  $(N, A_1)$ , where

$A_1$				<small>(av. scores)</small>	RP	ML	$c$	$\hat{c}$
				$s$				
0	1	0.9	0.9	0.7	1.010	1.099	0.847	1.004
1	0	0.9	0	0.633	1.037	1.099	0.547	0.703
0.1	0.1	0	0	0.1	-1.017	-1.099	-2.197	-2.041
0.1	0	0	0	0.1	-1.031	-1.099	-2.197	-2.041

In this tournament, players 1 and 2 have achieved an average score of 0.5 when playing against each other and of 0.9 when facing players 3 and 4. For ML, the symmetries between players 1 and 2 are such that they get the same rating, and the same happens with players 3

<sup>9</sup>This is a well known problem of ML in general, not only in this context.

<sup>10</sup>Refer, for instance, to Ford (1957) or David (1988).

and 4. However, for RP, players 1 and 2 are not symmetric because they have different average scores and different average opponents. More specifically, we have that  $s_1 > s_2$  and, at the same time, player 1 has played more against the weak players than player 2, so his average opponent is weaker than that of player 2. According to RP, player 2 gets a slightly higher rating. Hence, for this tournament, having a lower average score is compensated by having faced harder opponents. On the other hand, the rating provided by ML for this tournament is pairwise consistent. Note that, when a pairwise consistent ranking exists, the only thing that matters is the average scores achieved in the pairwise confrontations of the different players; the total number of matches in each pairwise confrontation plays no role. As we argue below, it is not clear that this last property is desirable.

Consider now the tournament  $(N, A_2)$  below. The only difference with respect to  $(N, A_1)$  is that the number of matches between players 1 and 3 is now 100 instead of 1, but the results remain such that player 1 achieved an average score of 0.9 against player 3. This change leads to no change in the ML ratings, which are still pairwise consistent, whereas, according to RP, player 1 is now significantly better than player 2. The reason for this change is that, in this tournament, beating player 3 with an average score of 0.9 is a good result (relative to the other results of player 1) and doing it over 100 games, instead of doing it over only 1 game, reinforces this good result; being this fact captured by the recursive performance through the weights of the different pairwise results of player 1.<sup>11</sup> In a sense, it would be much more likely that the result 0.9-0.1 between players 2 and 3 was just good luck on the side of player 2 than the result 90-10 between 1 and 3 being just good luck on the side of player 1. Hence, in this tournament RP is rewarding player 1 for having his good results sustained over more matches. Moreover, tournament  $(N, A_3)$  shows that, for the RP ratings, even if player 1 had achieved an average score slightly below 0.5 in his personal matches against player 2, the fact that his good results against player 3 are much more repeated than those of player 2 would still overcome the small loss in the personal confrontation. However, according to ML, player 2 is better than player 1 in this tournament.

$$\begin{array}{ccc} & A_2 & & RP & ML & & A_3 & & RP & ML \\ \left( \begin{array}{cccc} 0 & 1 & 90 & 0.9 \\ 1 & 0 & 0.9 & 0 \\ 10 & 0.1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 \end{array} \right) & & & \begin{array}{cc} 1.127 & 1.099 \\ 0.971 & 1.099 \\ -1.049 & -1.099 \\ -1.048 & -1.099 \end{array} & & & \left( \begin{array}{cccc} 0 & 0.9 & 90 & 0.9 \\ 1.1 & 0 & 0.9 & 0 \\ 10 & 0.1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 \end{array} \right) & & & \begin{array}{cc} 1.087 & 1.055 \\ 1.082 & 1.227 \\ -1.085 & -1.140 \\ -1.084 & -1.142 \end{array} \end{array}$$

Hence, we have seen that the RP ratings tend to reward a player who has achieved his pairwise good results over a large number of matches (with an opposite behavior with respect to bad results). On the other hand, the ML ratings do not exhibit this behavior in this specific example and just rank higher the player with the better pairwise results, regardless of the number of games over which these results were obtained. We consider that both behaviors are sensible and which one is more desirable depends on the context.

<sup>11</sup>It is important to note that, to determine if the results of a player  $i$  against a player  $j$  are good for  $i$  relative to his other results, one has to look at two things: i) the average score of  $i$  in his matches against  $j$  relative to his average score in all his matches and ii) the strength of  $j$  relative to the average strength of the opponents of  $i$ . For instance, in the two matrices below, the result of player 2 against player 1, despite being inferior to his overall average score, is a good result because player 1 is much better than all the other players. Hence, when player 2 sustains his good result against player 1 over 100 matches, his rating becomes better than the one of player 3. Moreover, the behavior of RP and ML in these two tournaments is analogous. In particular, in contrast to what happened with tournaments  $(N, A_1)$  and  $(N, A_2)$ , in this case ML also rewards a player for having repeated more times a good result:

$$\begin{array}{ccc} & & & RP & ML & & & & RP & ML \\ \left( \begin{array}{cccc} 0 & 0.55 & 9 & 9 \\ 0.45 & 0 & 5 & 5 \\ 1 & 5 & 0 & 9 \\ 1 & 5 & 1 & 0 \end{array} \right) & & & \begin{array}{cc} 1.466 & 1.570 \\ -0.476 & -0.476 \\ -0.049 & -0.050 \\ -0.941 & -1.044 \end{array} & & & \left( \begin{array}{cccc} 0 & 55 & 9 & 9 \\ 45 & 0 & 5 & 5 \\ 1 & 5 & 0 & 9 \\ 1 & 5 & 1 & 0 \end{array} \right) & & & \begin{array}{cc} 0.604 & 0.615 \\ 0.273 & 0.279 \\ 0.007 & 0.010 \\ -0.885 & -0.904 \end{array} \end{array}$$

Consider now the tournament  $(N, A_4)$  below. In this case, the only change with respect to  $(N, A_1)$  is that players 1 and 2 have played against each other 100 times instead of only 1. Again, this change does not affect the ratings proposed by ML but it does affect those proposed by RP. The change in RP is such that the difference between players 1 and 2 with respect to players 3 and 4 increases. The reason for this change is similar to the one before: players 1 and 2 are the strong players and hence, an average score of 0.5 in their personal matches is a good result for both of them. Therefore, for RP, the more repeated this result is, the higher the difference between the good and the bad players become; in other words, the more times players 1 and 2 have played against each other (with 0.5 being the resulting average score), the less likely it is that the results they achieved against players 3 and 4 were just good luck.

Finally, consider the tournament  $(N, A_5)$ , which only differs from  $(N, A_1)$  in that player 1 has played twice as many matches, his pairwise average scores being the same as before. Once more, ML remains unchanged and the RP ratings vary. Indeed, the difference between players 1 and 2 is larger now. The reason for this is that, with respect to  $(N, A_1)$ , the average score of player 2 goes down, but this is largely compensated by the strengthening of his average opponent in the tournament. Yet, for player 1 everything remains essentially unchanged: his average score is the same and his average opponent is almost the same as well.

$$\begin{array}{ccc}
 & A_4 & \text{RP} \quad \text{ML} \\
 \left( \begin{array}{cccc}
 0 & 100 & 0.9 & 0.9 \\
 100 & 0 & 0.9 & 0 \\
 0.1 & 0.1 & 0 & 0 \\
 0.1 & 0 & 0 & 0
 \end{array} \right) & & \begin{array}{cc}
 1.096 & 1.099 \\
 1.097 & 1.099 \\
 -1.096 & -1.099 \\
 -1.097 & -1.099
 \end{array}
 \end{array}
 \quad
 \begin{array}{ccc}
 & A_5 & \text{RP} \quad \text{ML} \\
 \left( \begin{array}{cccc}
 0 & 2 & 1.8 & 1.8 \\
 2 & 0 & 0.9 & 0 \\
 0.2 & 0.1 & 0 & 0 \\
 0.2 & 0 & 0 & 0
 \end{array} \right) & & \begin{array}{cc}
 1.006 & 1.099 \\
 1.067 & 1.099 \\
 -1.027 & -1.099 \\
 -1.047 & -1.099
 \end{array}
 \end{array}$$

As a concluding observation, just recall that both RP and ML coincide with the scores ranking for round-robin tournaments and hence, they will often fail to give a strict ranking of the players in tournaments within this class. However, the examples above suggest that, for the rest of the tournaments, RP will tend to deliver a strict ranking more often than ML, which in many disciplines is a desirable property.

## 7 Mathematical results

In this section we prove the technical results we have used throughout Sections 4 and 5. Although these results are stated for tournaments, they may be written just in terms of linear algebra. We follow Ciarlet (1989).

A linear iterative method is (globally) convergent if and only if the eigenvalues of the corresponding matrix are, in absolute value, less than 1. For any tournament  $(N, A)$  we have  $\bar{M}e = e$  and thus 1 is an eigenvalue of  $\bar{M}$ . In this section we prove that, under the assumptions A1 and A2, the iterated performance restricts to a vector subspace where the absolute values of the eigenvalues of  $\bar{M}$  are less than 1 and, hence, the method converges.

Let  $(N, A)$  be a tournament. For each  $v, w \in \mathbb{R}^n$ , we define  $\langle v, w \rangle := v^t D^M w = \sum_{i=1}^n m_i v_i w_i$ . Since  $D^M$  is a diagonal matrix and  $m_i > 0$  for all  $i$ ,  $\langle \cdot, \cdot \rangle$  is an inner product in  $\mathbb{R}^n$ , which generalizes the Euclidean inner product. The former, which depends on the tournament, is referred to as the *inner product associated with  $(N, A)$* .

If  $(N, A)$  is a tournament, then  $M$  is a symmetric matrix but  $\bar{M}$  is not symmetric in general. However, there is a kind of symmetry in  $\bar{M}$ , namely,  $\bar{M}_{ij} = 0$  if and only if  $\bar{M}_{ji} = 0$ . Motivated by this fact, we say that two matrices  $B \in \mathcal{M}_{k \times l}$  and  $C \in \mathcal{M}_{l \times k}$  are *null-transpose* if for each  $i \in \{1, \dots, k\}$ , and each  $j \in \{1, \dots, l\}$ ,  $B_{ij} = 0$  if and only if

$C_{ji} = 0$ . With a slight abuse of notation we denote by  $B^{nt}$  a matrix that is null-transpose of  $B$ . Note that, although  $B^{nt}$  is not unique,  $B^{nt} = 0$  if and only if  $B = 0$ .

**Theorem 5.** *Let  $(N, A)$  be a tournament and  $\langle \cdot, \cdot \rangle$  its associated inner product. Then*

- (i) *The matrix  $\bar{M}$  is self-adjoint with respect to  $\langle \cdot, \cdot \rangle$ . Moreover, it is diagonalizable, its eigenvalues are real and the eigenspaces are orthogonal with respect to  $\langle \cdot, \cdot \rangle$ .*
- (ii) *If  $\lambda$  is an eigenvalue of  $\bar{M}$ , then  $|\lambda| \leq 1$ .*
- (iii)  *$(N, A)$  satisfies A1 if and only if the multiplicity of the eigenvalue 1 is one.*
- (iv)  *$(N, A)$  satisfies A2 if and only if  $-1$  is not an eigenvalue of  $\bar{M}$ .*

*Remark 1.* For those familiar with the literature on nonnegative matrices, some of the statements of the result above may resemble those in the so-called Perron-Frobenius Theorem,<sup>12</sup> hereafter PFT. Indeed, we could make use of the latter to get a shorter proof of Theorem 5. However, in order to make this paper self-contained, we present below a simple algebraic proof. The connections with PFT are as follows. The main idea of Theorem 5 is to show that 1 is an eigenvalue of  $\bar{M}$  with multiplicity one and that there is no other eigenvalue of modulus 1. The matrix  $\bar{M}$  is stochastic and hence, PFT already implies statement (ii) in Theorem 5. Now, A1 is equivalent to saying that  $\bar{M}$  is *irreducible* so, by PFT, we have (iii). Now, if  $\bar{M}$  were aperiodic, then PFT would imply that no eigenvalue other than 1 has modulus 1. We argue now that statements (i) and (iv) imply that  $\bar{M}$  is indeed aperiodic. PFT says that if  $\bar{M}$  is aperiodic, then 1 is the unique eigenvalue with modulus 1 and moreover, if  $\bar{M}$  is periodic with cycle  $d > 1$ , then there are exactly  $d$  distinct eigenvalues of modulus 1, namely the  $d$ th roots of unity. Thus, an eigenvalue of modulus 1 different from 1 has to be either  $-1$  or a (non-real) complex number. Therefore, (i) and (iv) imply that  $\bar{M}$  is aperiodic. By (i) we get that  $\bar{M}$  has no complex eigenvalues (of whatever modulus) and hence, if  $\bar{M}$  is periodic, the period has to be 2, with eigenvalues 1 and  $-1$ . But, under A2, the latter case is ruled out by (iv).

*Proof.* Let  $(N, A)$  and  $(N', A')$  be equivalent tournaments with matches matrices  $M$  and  $M'$  respectively. Then, there exists  $L \in \Pi_n$  such that  $M = LM'L^{-1}$ . Clearly,  $D^M = LD^{M'}L^{-1}$  and thus  $\bar{M} = L\bar{M}'L^{-1}$ . Hence,  $\bar{M}$  and  $\bar{M}'$  are similar matrices and their eigenvalue structure is the same. Therefore, we may make, without loss of generality, any assumption regarding the ordering of the indices. We also recall that  $\bar{M}_{ij} \geq 0$  for all  $i, j \in \{1, \dots, n\}$ . For each  $k \in \mathbb{N}$ , we define  $e^k := (1, \dots, 1) \in \mathbb{R}^k$ . Note that  $e^n = e$ .

Claim (i): Since  $M$  and  $D^M$  are symmetric,  $\langle v, \bar{M}w \rangle = v^t D^M \bar{M}w = v^t M w = w^t M v = w^t D^M \bar{M}v = (\bar{M}v)^t D^M w = \langle \bar{M}v, w \rangle$ . The second part follows from the spectral theorem.

Claim (ii): The matrix norm  $\|\cdot\|_\infty$  is defined as  $\|B\|_\infty := \max_{i \in \{1, \dots, n\}} \sum_{j=1}^n |B_{ij}|$  for any  $B \in \mathcal{M}_{n \times n}$ . By definition we have  $\|\bar{M}\|_\infty = 1$  and hence (ii) follows from Theorem 1.4-3 in Ciarlet (1989).

Claim (iii): Assume that  $(N, A)$  does not satisfy A1. Then,  $\bar{M}$  may be written as

$$\bar{M} = \left( \begin{array}{c|c} C & 0 \\ \hline 0 & E \end{array} \right), \quad \text{with } C \in \mathcal{M}_{k \times k} \text{ and } E \in \mathcal{M}_{l \times l}.$$

Hence,  $\bar{M}(e^k|0) = (e^k|0)$  and  $\bar{M}(0|e^l) = (0|e^l)$ . Thus, 1 has multiplicity at least 2.

Conversely, assume that 1 has multiplicity greater than one. Then, since  $\bar{M}$  is diagonalizable there exists  $v \in \mathbb{R}^n$ , linearly independent from  $e$ , such that  $\bar{M}v = v$ . Assume that

<sup>12</sup>See, for instance, Seneta (1981).

$v_1 = 1$  and that the components of  $v$  are decreasingly ordered, that is,  $v_i \geq v_j$  for  $i > j$ . Let  $k \in \mathbb{N}$  be such that  $v_k = 1$  and  $v_{k+1} < 1$ . Since  $v$  and  $e$  are linearly independent,  $k < n$  and  $\bar{M}$  may be decomposed as

$$\bar{M} = \left( \begin{array}{c|c} C_1 & E \\ \hline E^{nt} & C_2 \end{array} \right), \text{ where } C_1 \in \mathcal{M}_{k \times k}, C_2 \in \mathcal{M}_{(n-k) \times (n-k)} \text{ and } E \in \mathcal{M}_{k \times (n-k)}. \quad (2)$$

Now, if  $E$  has a nonzero row, namely row  $i$ , we get

$$1 = v_i = (\bar{M}v)_i = \sum_{j=1}^k \bar{M}_{ij} + \sum_{j=k+1}^n \bar{M}_{ij}v_j < \sum_{j=1}^n \bar{M}_{ij} = 1,$$

contradiction. This proves  $E = 0$  and  $E^{nt} = 0$ , which is a contradiction with A1.

Claim (iv): Assume  $(N, A)$  does not satisfy A2. Then we may write  $\bar{M}$  as

$$\bar{M} = \left( \begin{array}{c|c} 0 & C \\ \hline E & 0 \end{array} \right), \text{ with } C \in \mathcal{M}_{k \times l} \text{ and } E \in \mathcal{M}_{l \times k}.$$

Hence,  $\bar{M}(e^k | -e^l) = -(e^k | -e^l)$  and  $-1$  is an eigenvalue of  $\bar{M}$ .

Conversely, assume that  $-1$  is an eigenvalue of  $\bar{M}$ . Let  $v \in \mathbb{R}^n$  be such that  $\bar{M}v = -v$  and  $1 = v_1 \geq v_2 \geq \dots \geq v_n \geq -1$ . Again, there exists  $k \in \mathbb{N}$ ,  $k < n$  such that  $v_k = 1$  and  $v_{k+1} < 1$ . Hence,  $\bar{M}$  may be decomposed as in (2).

We show that  $C_1 = 0$ . Let  $i \in \{1, \dots, k\}$ . Since  $\sum_{j=1}^n \bar{M}_{ij} = 1$ , we have

$$-1 = -v_i = (\bar{M}v)_i = \sum_{j=1}^k \bar{M}_{ij} + \sum_{j=k+1}^n \bar{M}_{ij}v_j \geq \sum_{j=1}^k \bar{M}_{ij} + \left(1 - \sum_{j=1}^k \bar{M}_{ij}\right)(-1) = 2 \sum_{j=1}^k \bar{M}_{ij} - 1.$$

Hence,  $\sum_{j=1}^k \bar{M}_{ij} \leq 0$ . Since  $\bar{M}_{ij} \geq 0$ , the  $i$ th row of  $C_1$  is zero. Therefore,  $C_1 = 0$ .

Note that  $v_n = -1$  since, otherwise, taking  $i \in \{1, \dots, k\}$  we get  $-1 = -v_i = (\bar{M}v)_i = \sum_{j=k+1}^n \bar{M}_{ij}v_j > -1$ . Let  $l \in \mathbb{N}$  be such that  $v_{n-l} > -1$  and  $v_j = -1$  for  $j \geq n-l$ . Clearly,  $l > 0$ . Then, we may further decompose  $\bar{M}$  as

$$\bar{M} = \left( \begin{array}{c|c|c} 0 & E_1 & E_2 \\ \hline E_1^{nt} & C_{21} & C_{22} \\ \hline E_2^{nt} & C_{22}^{nt} & C_{23} \end{array} \right) \text{ with } E = (E_1 | E_2), C_2 = \left( \begin{array}{c|c} C_{21} & C_{22} \\ \hline C_{22}^{nt} & C_{23} \end{array} \right) \text{ and } E_2 \in \mathcal{M}_{k \times l}.$$

If  $k+l = n$ , then this second decomposition is trivial ( $E = E_2$  and  $C_2 = C_{23}$ ). Otherwise, we claim that  $E_1 = 0$ . If the  $i$ th row of  $E_1$  is nonzero, then there is  $j \in \{k+1, \dots, n-l\}$  such that  $v_j > -1$ . Hence,  $-1 = -v_i = (\bar{M}v)_i = \sum_{j=k+1}^n \bar{M}_{ij}v_j > -1$ , contradiction. Therefore,  $E_1 = 0$ .

Now, we show that  $(C_{22}^{nt} | C_{23})$  is zero. If the  $i$ th row of  $(C_{22}^{nt} | C_{23})$  is nonzero, then there is  $j \in \{k+1, \dots, n\}$  such that  $v_j < 1$ . Hence,  $1 = -v_i = (\bar{M}v)_i = \sum_{j=1}^n \bar{M}_{ij}v_j < 1$ , contradiction. Therefore, the above decomposition reduces to

$$\bar{M} = \left( \begin{array}{c|c|c} 0 & 0 & E_2 \\ \hline 0 & C_{21} & 0 \\ \hline E_2^{nt} & 0 & 0 \end{array} \right),$$

which contradicts A2.<sup>13</sup> □

<sup>13</sup>Note that, if the previous decomposition happens to be nontrivial (presence of  $C_{21}$ ), we also violate A1.

**Corollary 6.** *Let  $(N, A)$  be a tournament satisfying assumptions A1 and A2. Then,*

$$\lim_{l \rightarrow \infty} \bar{M}^l = \frac{1}{\sum_{i=1}^n m_i} \begin{pmatrix} m_1 & \cdots & m_n \\ \vdots & & \vdots \\ m_1 & \cdots & m_n \end{pmatrix}.$$

*Proof.* Let  $v \in \mathbb{R}^n$ . By Theorem 5 there exists a basis of eigenvectors  $\{w_1 = e, w_2, \dots, w_n\}$ , orthogonal with respect to the inner product associated with  $(N, A)$ , which we denote by  $\langle \cdot, \cdot \rangle$ . For each  $i \in \{1, \dots, n\}$ , let  $\lambda_i$  be the eigenvalue associated with  $w_i$ . Now, it suffices to show that  $(\lim_{l \rightarrow \infty} \bar{M}^l)v = (\frac{1}{\text{tr} \bar{M}} ee^t D^M)v$  for all  $v \in \mathbb{R}^n$ . For each  $i \in \{1, \dots, n\}$ , we have  $\langle (\lim_{l \rightarrow \infty} \bar{M}^l)v, w_i \rangle = \lim_{l \rightarrow \infty} \langle v, \bar{M}^l w_i \rangle = (\lim_{l \rightarrow \infty} \lambda_i^l) \langle v, w_i \rangle$ . Since  $\lambda_1 = 1$  and  $|\lambda_i| < 1$  for each  $i \in \{2, \dots, n\}$ , we get  $(\lim_{l \rightarrow \infty} \bar{M}^l)v = (\langle v, e \rangle / \langle e, e \rangle)e$ . By definition of  $\langle \cdot, \cdot \rangle$ , we have  $\langle e, e \rangle = \text{tr} D^M$  and  $\langle v, e \rangle = e^t D^M v$ , from where the result follows.  $\square$

**Corollary 7.** *Let  $(N, A)$  be a tournament satisfying A1 and A2. Let  $\langle \cdot, \cdot \rangle$  be its associated inner product. Then, for all  $l \in \mathbb{N}$ ,  $\langle \hat{p}^{(l)}, e \rangle = \langle r, e \rangle$ , or equivalently,  $\sum_i m_i \hat{p}_i^{(l)} = \sum_i m_i r_i = \sigma^r$ .*

*Proof.* Recall that  $\bar{M}$  is self-adjoint with respect to  $\langle \cdot, \cdot \rangle$  by Theorem 5 (i). By Corollary 6,  $\lim_{l \rightarrow \infty} \bar{M}^l$  exists and  $(\lim_{l \rightarrow \infty} \bar{M}^l)c = ((\sum_i m_i c_i) / (\sum_i m_i))e$ . Hence,  $\langle \hat{c}, e \rangle = \langle c, e \rangle - \langle (\lim_{l \rightarrow \infty} \bar{M}^l)c, e \rangle = \langle c, e \rangle - \lim_{l \rightarrow \infty} \langle c, \bar{M}^l e \rangle = 0$ . Then,  $\langle \hat{p}^{(0)}, e \rangle = \langle r, \bar{M}e \rangle + \langle \hat{c}, e \rangle = \langle r, e \rangle$ . By induction,  $\langle \hat{p}^{(l)}, e \rangle = \langle \hat{p}^{(l-1)}, \bar{M}e \rangle + \langle \hat{c}, e \rangle = \langle r, e \rangle$ .  $\square$

We are now ready to prove the main result of this paper.

*Proof of Theorem 2.* Defining  $q^{(l)} := \hat{p}^{(l)} - r$  for  $l \geq 0$  we have the equivalent iterative method  $q^{(0)} = \bar{M}r + \hat{c} - r$  and  $q^{(l)} = \bar{M}q^{(l-1)} + q^{(0)}$ ,  $l \in \mathbb{N}$ . Let  $\langle \cdot, \cdot \rangle$  be the inner product associated with  $(N, A)$ . By Corollary 7,  $\langle q^{(l)}, e \rangle = 0$  for all  $l \in \mathbb{N}$ . Therefore, the iterative method restricts to the vector subspace  $e^\perp$ . By Theorem 5, the absolute values of the eigenvalues of  $\bar{M}|_{e^\perp}$  are smaller than 1. Hence, the iterative method converges (Ciarlet, 1989).  $\square$

*Proof of Proposition 4.* Let  $\succeq$  be a ranking consistent with  $F_\ell$ . Then, there exists a rating  $r \in \mathbb{R}^n$  and a constant  $\lambda \in \mathbb{R}$  such that  $r = p^r + \lambda e = \bar{M}r + c + \lambda e$ . Taking inner product with  $e$  (as in Corollary 6) we get  $\langle r, e \rangle = \langle r, e \rangle + \langle c, e \rangle + \lambda \langle e, e \rangle$ . Hence,  $\lambda = -\langle c, e \rangle / \langle e, e \rangle$  and thus  $c + \lambda e = \hat{c}$ . This implies that  $r$  is a solution to (1) and the result follows.  $\square$

## References

- AMIR, R. (2002): "Impact-Adjusted Citations as a Measure of Journal Quality," Tech. rep., CORE Discussion Paper 2002/74.
- BRADLEY, R. A. AND M. E. TERRY (1952): "Rank analysis of incomplete block designs. I. The method of paired comparisons," *Biometrika*, 39, 324–345.
- CIARLET, P. G. (1989): *Introduction to numerical linear algebra and optimisation*, Cambridge: Cambridge University Press.
- DANIELS, H. (1969): "Round-Robin Tournament Scores," *Biometrika*, 56, 295–299.
- DAVID, H. A. (1988): *The method of paired comparisons*, London: Chapman and Hall.

- DYKSTRA, O. (1956): “A Note on the Rank Analysis of Incomplete Block Designs – Applications beyond the Scope of Existing Tables,” *Biometrics*, 3, 301–306.
- FORD, L. R. (1957): “Solution of a Ranking Problem from Binary Comparisons,” *American Mathematical Monthly*, 64, 28–33.
- GODDARD, S. T. (1983): “Ranking in tournaments and group decision making,” *Management Science*, 29, 1384–1392.
- IQBAL, A., W. COOK, AND M. KRESS (1986): “On the minimum violations ranking of a tournament,” *Management Science*, 32, 660–672.
- KENDALL, M. G. AND B. B. SMITH (1940): “On the method of paired comparisons,” *Biometrika*, 31, 324–345.
- LASLIER, J. (1997): *Tournament Solutions and Majority Voting*, Springer-Verlag. Berlin.
- LIEBOWITZ, S. AND J. PALMER (1984): “Assesing the Relative Impacts of Economic Journals,” *Journal of Economic Literature*, 22, 77–88.
- MOON, J. W. (1968): *Topics on Tournaments*, New York.
- PALACIOS-HUERTA, I. AND O. VOLIJ (2004): “The Measurement of Intellectual Influence,” *Econometrica*, 72, 963–977.
- RUBINSTEIN, A. (1980): “Ranking the Participants in a Tournament,” *SIAM Journal on Applied Mathematics*, 38, 108–111.
- SENETA, E. (1981): *Nonnegative Matrices and Markov Chains*, Springer. New York, 2nd ed.
- SLUTZKI, G. AND O. VOLIJ (2005): “Ranking Participants in Generalized Tournaments,” *International Journal of Game Theory*, 33, 255–270.
- STOB, M. (1985): “Rankings from Round Robin tournaments,” *Management Science*, 31, 1191–1195.
- ZERMELO, E. (1929): “Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung,” *Mathematische Zeitschrift*, 29, 436–460.