Faculty Publications

2000-01-01

An Extension of Zermelo's Model for Ranking by Paired Comparisons

Christopher P. Grant
grant@math.byu.edu

Gregory R. Conner
conner@math.byu.edu

225

An extension of Zermelo's model for ranking by paired comparisons

G. R. CONNER and C. P. GRANT

Department of Mathematics, Brigham Young University, Provo, UT 84602, USA
Email: {conner,grant}@math.byu.edu

In 1929, Zermelo proposed a probabilistic model for ranking by paired comparisons and showed that this model produces a unique ranking of the objects under consideration when the outcome matrix is irreducible. When the matrix is reducible, the model may yield only a partial ordering of the objects. In this paper, we analyse a natural extension of Zermelo's model resulting from a singular perturbation. We show that this extension produces a ranking for arbitrary (nonnegative) outcome matrices and retains several of the desirable properties of the original model. In addition, we discuss computational techniques and provide examples of their use.

1 Introduction

Suppose that n objects are compared a pair at a time, and that for each comparison one of the two objects in the pair is judged superior to the other. (A common example would be athletic teams engaged in pairwise competitions.) The results of the comparisons can be summarized in the *outcome matrix* $A = (a_{ij})$, where a_{ij} is the number of comparisons in which object i is judged to be superior to object j.

If all of the off-diagonal elements in $A + A^T$ are the same (i.e. there has been *round-robin* competition) then the natural way to rank objects would be according to their *scores* $s_i = \sum_{j=1}^{n} a_{ij}$. If, on the other hand, the outcome matrix lacks this symmetry, it is reasonable to suspect that ranking by score is not necessarily the best possible choice. (Contrary to common mathematical usage, we will often use the word *tournament* when referring to this asymmetric case; when we wish to emphasize the possible lack of symmetry, we may use the phrase *generalized tournament*.)

A wide variety of methods have been proposed for ranking generalized tournaments. (See, for example, [9, 10].) In 1929, Zermelo [33] derived the functional

$$P(r) := \prod_{i,j=1}^{n} \left(\frac{r_i}{r_i + r_j} \right)^{a_{ij}}, \tag{1.1}$$

by assuming that the ith object ($i = 1, \ldots, n$) has a fixed strength (or rating) r_i and that each comparison is an independent binomial trial with the probability of finding object i superior to object j in a comparison being $r_i/(r_i + r_j)$. He then determined $r = (r_1, \ldots, r_n)$ by maximum likelihood estimation (i.e. by picking r to maximize $P(r)$).

Because $P(\lambda r) = P(r)$ for any $\lambda > 0$, it is often convenient to impose some sort of normalization constraint to confine r to a bounded and/or lower-dimensional set. Zermelo chose to confine r to the region

$$\mathscr{R} := \left\{ r \in \mathbf{R}^n : r_i \geqslant 0 \text{ for each } i, \quad \text{and} \quad \sum_{i=1}^{n} r_i \leqslant 1 \right\}.$$

Since P is undefined at some points on the boundary of $\mathscr{R}$, the compactness of $\mathscr{R}$ and the continuity of P do not guarantee the existence of a maximizer of P in $\mathscr{R}$. Thus, Zermelo worked instead with limit points $\bar{r}$ of maximizing sequences. He argued that if $\bar{r}$ is in the interior of $\mathscr{R}$ (and is, therefore, a maximizer), the objects under comparison should be ranked according to the components of $\bar{r}$.

Recall that a square matrix A is said to be *irreducible* if there is no permutation matrix M such that

$$M^T A M = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where B and D are square submatrices. A necessary and sufficient condition for P to have a maximizer in the interior of $\mathscr{R}$ is that the outcome matrix be irreducible. (For one derivation of this fact, see [20].) An equivalent condition on $\Gamma(A)$ is that it be strongly connected [16]. ($\Gamma(A)$ is the directed graph with n nodes that has an edge from node i to node j if and only if $a_{ij} > 0$. In this paper, we will often view $\Gamma(A)$ as a network or weighted graph that has the quantity a_{ij} associated with the edge from node i to node j.)

In many situations in which a ranking is desired, the outcome matrix is, however, reducible. An example of this is when all of the comparisons are consistent with a fixed pecking order. In fact, whenever there are undefeated teams or teams with no victories in an athletic league or tournament A is reducible. In such situations, $\bar{r}$ may lie on the boundary of $\mathscr{R}$. Zermelo suggested that when this happens, the objects be ordered in a way that corresponds to values of r that are close to $\bar{r}$ and in the interior of $\mathscr{R}$. For this to provide a unique ranking, there must be a directed path between every pair of nodes in $\Gamma(A)$. (A simple example when this fails is a three-team, three-game tournament in which the first team defeats the second team twice and the third team once.)

In the 70 years since it was first proposed, Zermelo's model has experienced a fair amount of popularity [4, 5, 11, 17, 20, 32]. (Often the model has been identified more closely with those who rediscovered the model independently at a later date rather than with Zermelo himself.) In an effort to obtain a unique ranking when Zermelo's original model fails to provide one, various modifications of it have been proposed. For example, in the context of competition between (American) football teams Keener [20] suggests letting a_{ij} represent something other than the number of victories of team i over team j; for instance, a_{ij} could be set equal to $\sum_k [(S_{ijk} + C)/(S_{ijk} + S_{jik} + 2C)]$, where C is a constant and S_{ijk} is the number of points scored by team i in its kth encounter with team j. In the same context, Allen [1] developed a model that assumes a priori that r_i is a random variable that is gamma distributed with parameters $p = \sigma = 1$.

In this paper, we analyse a model that is an *extension* of Zermelo's model in the following sense: If A is irreducible then the 'extended' ranking is identical with the ranking provided by the original model, with the possible exception that a tie that exists

between object k and object ℓ under the original model might be broken by the extension if node k and node ℓ are not symmetrically located in $\Gamma(A)$, in the sense that there does not exist a relabelling of the objects that interchanges k and ℓ and produces an identical outcome matrix. This extension generates a ranking for each nonnegative outcome matrix while avoiding the introduction of arbitrary constants necessary in some models. The underlying assumption behind this approach is that objects are roughly comparable in strength; however, this assumption is implemented so as to have as little effect as possible. Artificial connectivity is introduced into the graph by adding edges of vanishingly small weight between each pair of distinct nodes in $\Gamma(A)$. More precisely, for $\varepsilon > 0$ a maximizer $r = r(\varepsilon)$ of

$$P(r, \varepsilon) := \prod_{i,j=1}^{n} \left(\frac{r_i}{r_i + r_j} \right)^{a_{ij}(\varepsilon)} \tag{1.2}$$

or, equivalently, of

$$F(r, \varepsilon) := \log P(r, \varepsilon) = \sum_{i,j=1}^{n} a_{ij}(\varepsilon)(\log r_i - \log(r_i + r_j)),$$

is sought, where $a_{ij}(\varepsilon) = a_{ij} + \varepsilon(1 - \delta_{ij})$ and δ_{ij} is the Kronecker delta, and objects are then ranked according to $r(\varepsilon)$ for ε small. (Note that the presence of δ_{ij} affects the value of P, but is irrelevant to the maximization problem, so when it is convenient to do so, we will drop it.)

This approach to extending Zermelo's model seems natural. It mimics the idea of introducing artificial viscosity to regularize solutions of partial differential equations. There is, in particular, a clear analogy between this technique and the singular perturbation methods used [2, 22, 23, 25, 26, 31] to systematically select from a continuum of low-energy equilibrium solutions to phase transition problems those that are in a sense more realistic than others. (For examples of studies in which these continuum models are replaced by spatially discrete equations on finite lattices and in which the resulting perturbed energy function is, as in the current study, a function of a finite number of real variables, see [8, 13].) In these works, the extra term added is to penalize juxtaposition of dissimilar phases (i.e. to penalize large discrepancies in the *order parameters*); in the present study, the extra term is to penalize large discrepancies in the *strengths* of the objects under comparison. Just as in the phase transition problems it is sometimes helpful to use the concept of Γ-*convergence* [24] to identify a functional which captures the limiting minimization properties of the perturbed functionals, it will be useful to us to consider Γ-$\lim_{\varepsilon \downarrow 0}(-P(\cdot, \varepsilon))$, at least in an informal and heuristic sense, in the development of computational techniques.

Some senses in which the minimizers of (1.2) converge as $\varepsilon \downarrow 0$ are established in the following 3 theorems. First, under natural normalizations the strengths converge as the penalty vanishes. In particular, we have the following:

Theorem 1.1 *Given $\varepsilon > 0$, let $P(r, \varepsilon)$ be given by (1.2) and let $R(r) = r_1^2 + r_2^2 + \ldots + r_n^2 - 1$. Then there is a unique $r = r(\varepsilon)$ in the open positive orthant of $\mathbf{R}^n$ that maximizes $P(r, \varepsilon)$ subject to the constraint $R(r) = 0$, and the limit of $r(\varepsilon)$ as $\varepsilon \downarrow 0$ exists. (We denote this limit $r(0)$.)*

We will refer to the constraint $R(r) = 0$ as the *standard normalization*.

Definition The *partial ordering induced by $r(\varepsilon)$*, denoted $\preceq_\varepsilon$, is the relation between the elements of $\{1, \ldots, n\}$ satisfying

$$i \preceq_\varepsilon j \quad \text{if and only if} \quad r_i(\varepsilon) \leqslant r_j(\varepsilon).$$

If $i \preceq_\varepsilon j$ and $j \npreceq_\varepsilon i$, we write $i \prec_\varepsilon j$ (or $j \succ_\varepsilon i$).

As would be hoped, this ordering stabilizes for small ε:

Theorem 1.2 *There exists $\varepsilon_0 > 0$ such that $\preceq_\varepsilon$ is independent of the choice of $\varepsilon \in (0, \varepsilon_0)$.*

We will denote this limiting ordering $\preceq$, and its dual ordering $\succeq$. Note that $\preceq$ might not correspond to $\preceq_0$, or even to the partial ordering induced by the limiting values of the ratios r_i/r_j, since $\lim_{\varepsilon \downarrow 0}(r_i(\varepsilon)/(r_j(\varepsilon))$ may be equal to 1 without $(r_i(\varepsilon)/(r_j(\varepsilon)))$ being identically 1. Since $\preceq$ seems to us to be more useful than $\preceq_0$, our focus in this paper will be on the former, rather than the latter.

In computing, it will turn out to be useful to consider the change of variable $p_i = p_i(\varepsilon) := \log_\varepsilon r_i(\varepsilon)$ for $\varepsilon > 0$. These variables also have nice convergence properties:

Theorem 1.3 *Under the same hypotheses as in Theorem 1.1, each $p_i(\varepsilon)$ converges to some value $p_i(0)$ as $\varepsilon \downarrow 0$.*

In §2, we will prove these convergence theorems. In Section 3, we will describe some of the basic properties of the extended model. These include basic symmetry along with a sort of monotonic dependence of r on A. Furthermore, we will show that if $\Gamma(A)$ suggests that a given object dominates another, then the ordering generated by the extended model reflects that dominance. (In this aspect, the extension is compatible with the observations made by Zermelo for those reducible cases that he analyzed.) In §4, we will consider ways to compute the asymptotic ordering. For certain tournaments, standard methods of solving parametrized nonlinear systems are successful, but other tournaments demand a subtler approach. We will describe a technique combining a heuristic equivalent of Γ-convergence with power series methods. We will then present the results of a sample calculation. Finally, in §5, we will discuss some different models that are related to the extension that is analysed in this paper.

2 Proofs of convergence theorems

To prove the convergence results, we will make use of some concepts from algebraic geometry.

Definition A *Boolean* class of subsets is a class of subsets that is closed under finite intersection, finite union and set complementation.

Definition A subset of $\mathbf{R}^m$ is *semi-algebraic* if it belongs to the Boolean class of subsets of

$\mathbf{R}^m$ generated by sets of the form

$$\{x \in \mathbf{R}^m : p(x) \geqslant 0\},$$

where p is a polynomial function on $\mathbf{R}^m$.

Hironaka [15] showed that every bounded, semi-algebraic subset of $\mathbf{R}^m$ is the homeomorphic image of a disjoint finite union of simplexes in $\mathbf{R}^m$. This union is necessarily locally connected. Since local connectedness is a topological property, every bounded, semi-algebraic subset of $\mathbf{R}^m$ is locally connected. Since every semi-algebraic subset of $\mathbf{R}^m$ is locally equivalent to a bounded, semi-algebraic set, we know that all semi-algebraic subsets of $\mathbf{R}^m$ are locally connected. This fact will be used in the proofs of Theorem 1.1 and Theorem 1.2.

Parts of the proof of Theorem 1.1 (namely, well-posed optimality for fixed ε) are not original. We include them for concreteness and because it will be necessary to generalize these arguments in §4.

Proof of Theorem 1.1 Clearly, the matrix $\tilde{A}(\varepsilon) = (a_{ij}(\varepsilon))$ is irreducible, so $F(r, \varepsilon) \to -\infty$ as r approaches any point on the boundary of the orthant except for 0. Thus, the homogeniety of $F(\cdot, \varepsilon)$ with respect to scalar multiplication implies that $F(\cdot, \varepsilon)$ achieves its maximum at some point on the unit sphere.

Any maximizing value of r must be a critical point of $F(\cdot, \varepsilon)$, i.e. a solution to

$$DF(r, \varepsilon) = 0, \tag{2.1}$$

where D represents differentiation with respect to r.

We claim that if r is a solution of (2.1) in the open positive orthant then $D^2 F(r)$ is negative semidefinite with the only 0 eigenvalue corresponding to the eigenvector r. This will imply that all the critical points of F will be strict local maxima subject to the constraint $R(r) = 0$.

To verify this claim, note that

$$\frac{\partial F}{\partial r_k} = \sum_{j=1}^n \frac{a_{kj}(\varepsilon)}{r_k} - \sum_{j=1}^n \frac{a_{kj}(\varepsilon) + a_{jk}(\varepsilon)}{r_k + r_j}, \tag{2.2}$$

and

$$\frac{\partial^2 F}{\partial r_k \partial r_\ell} = \frac{a_{\ell k}(\varepsilon) + a_{k\ell}(\varepsilon)}{(r_k + r_\ell)^2} + \delta_{\ell k} \left[-\sum_{j=1}^n \frac{a_{kj}(\varepsilon)}{r_k^2} + \sum_{j=1}^n \frac{a_{kj}(\varepsilon) + a_{jk}(\varepsilon)}{(r_k + r_j)^2} \right],$$

where $\delta_{\ell k}$ is the Kronecker delta. Setting

$$s_{ij} = s_{ji} = \frac{a_{ij}(\varepsilon) + a_{ji}(\varepsilon)}{(r_i + r_j)^2},$$

we see that for any n-dimensional column vector u

$$u^T D^2 F(r) u = \sum_{i,j=1}^n s_{ij} u_i u_j + \sum_{i=1}^n \left(\sum_{j=1}^n \left[s_{ij} - \frac{a_{ij}(\varepsilon)}{r_i^2} \right] \right) u_i^2$$

$$= \sum_{i,j=1}^{n} s_{ij}(u_i^2 + u_i u_j) - \sum_{i=1}^{n} \frac{u_i^2}{r_i^2} \sum_{j=1}^{n} a_{ij}(\varepsilon). \qquad (2.3)$$

The fact that r is a critical point implies that for any i

$$\sum_{j=1}^{n} a_{ij}(\varepsilon) = r_i \sum_{j=1}^{n} s_{ij}(r_i + r_j);$$

substituting this into (2.3) yields

$$u^T D^2 F(r)u = \sum_{i,j=1}^{n} s_{ij}(u_i^2 + u_i u_j) - \sum_{i=1}^{n} \frac{u_i^2}{r_i} \sum_{j=1}^{n} s_{ij}(r_i + r_j)$$

$$= \sum_{i,j=1}^{n} s_{ij}\left(u_i u_j - \frac{u_i^2 r_j}{r_i}\right)$$

$$= \sum_{i<j} s_{ij}\left(u_i u_j - \frac{u_i^2 r_j}{r_i}\right) + \sum_{i<j} s_{ji}\left(u_j u_i - \frac{u_j^2 r_i}{r_j}\right)$$

$$= \sum_{i<j} \frac{s_{ij}}{r_i r_j}(2 r_i r_j u_i u_j - u_i^2 r_j^2 - u_j^2 r_i^2)$$

$$= - \sum_{i<j} \frac{s_{ij}}{r_i r_j}(r_j u_i - r_i u_j)^2 \leqslant 0,$$

with equality if and only if u is parallel to r. This verifies the claim.

Next, we claim that there is only one solution of (2.1) in the open positive orthant satisfying $R(r) = 0$. Suppose there were two such solutions $r^{(1)}$ and $r^{(2)}$. From the preceding claim they are both strict local maxima (subject to the constraint). Consider all the paths $\gamma(t)$ in the open positive orthant satisfying $R(\gamma) = 1$, $\gamma(0) = r^{(1)}$ and $\gamma(1) = r^{(2)}$, and consider the optimization problem

$$M = \sup_{\gamma} \min_{t \in [0,1]} F(\gamma(t)).$$

By the Mountain Pass Theorem (e.g. see [21]), there exists a path $\tilde{\gamma}$ realizing the optimum M, i.e. there exists $\tilde{t} \in [0,1]$ such that

$$M = \min_{t \in [0,1]} F(\tilde{\gamma}(t)) = F(\tilde{\gamma}(\tilde{t})).$$

Furthermore, $\tilde{\gamma}(\tilde{t})$ is a saddlepoint of F (subject to $R(r) = 0$). However, this contradicts the fact that all critical points are constrained local maxima. This verifies the claim. Call the unique maximizer of (1.2) $r(\varepsilon)$.

Now, for $k = 1, 2, \ldots, n$, define

$$p_k(r_1, r_2, \ldots, r_n, \varepsilon) = r_k \left(\prod_{j=1}^{n}(r_j + r_k)\right) \frac{\partial F(r, \varepsilon)}{\partial r_k}$$

and define

$$p_{n+1}(r_1, r_2, \ldots, r_n, \varepsilon) = R(r).$$

Also, for $k = 1, 2, \ldots, n$, define

$$q_k(r_1, r_2, \ldots, r_n, \varepsilon) = r_k,$$

and define

$$q_{n+1}(r_1, r_2, \ldots, r_n, \varepsilon) = \varepsilon.$$

Note that by (2.2) each p_k and q_k is a polynomial function of $(r_1, \ldots, r_n, \varepsilon) \in \mathbf{R}^{n+1}$. Thus, the sets

$$\mathscr{B} = \bigcap_{k=1}^{n+1} \{(r_1, \ldots, r_n, \varepsilon) \in \mathbf{R}^{n+1} : p_k(r_1, \ldots, r_n, \varepsilon) = 0\},$$

$$\mathscr{C} = \bigcap_{k=1}^{n+1} \{(r_1, \ldots, r_n, \varepsilon) \in \mathbf{R}^{n+1} : q_k(r_1, \ldots, r_n, \varepsilon) > 0\},$$

and

$$\mathscr{S}_0 = \{(r_1, \ldots, r_n, \varepsilon) \in \mathbf{R}^{n+1} : q_{n+1}(r_1, \ldots, r_n, \varepsilon) = 0\},$$

are semi-algebraic sets. Consequently, $\mathscr{S} := (\mathscr{B} \cap \mathscr{C}) \cup \mathscr{S}_0$ is also semi-algebraic and, therefore, is locally connected.

Note that $\mathscr{S}$ is the disjoint union of the $(\varepsilon = 0)$ hyperplane $\mathscr{S}_0$ and the graph of $r : (0, \infty) \to \mathbf{R}^n$, which we will call $\mathscr{S}_+$. By the compactness of the unit sphere in $\mathbf{R}^n$, $\mathscr{S}_+$ has a limit point in $\mathscr{S}_0$, and if this limit point is unique then $\lim_{\varepsilon \downarrow 0} r(\varepsilon)$ exists and we are done. Suppose, on the other hand, that there are two such limit points $x_1, x_2 \in \mathscr{S}_0$. Then by examining the intersection of $\mathscr{S}$ with disjoint neighbourhoods of x_1 and x_2, we get a contradiction to local connectedness. □

Proof of Theorem 1.2 By the continuity of $r : [0, \infty) \to \mathbf{R}^n$ and the fact that there are only finitely many i, j pairs, the only way the theorem could fail would be if there were $i, j \in \{1, \ldots, n\}$ and a sequence $\varepsilon_k \downarrow 0$ such that $r_i(\varepsilon_k) \leqslant r_j(\varepsilon_k)$ for k even and $r_i(\varepsilon_k) > r_j(\varepsilon_k)$ for k odd. Suppose that this happens, and consider the set

$$\mathscr{S}_{i,j} := \{(r_1, \ldots, r_n, \varepsilon) \in \mathbf{R}^{n+1} : r_i \leqslant r_j\} \bigcap \mathscr{S},$$

where $\mathscr{S}$ is the semi-algebraic set defined in the proof of Theorem 1.1. Note that $\mathscr{S}_{i,j}$ is semi-algebraic and that, by continuity, $(r_1(0), \ldots, r_n(0), \varepsilon) \in \mathscr{S}_{i,j}$. But our assumption implies that $\mathscr{S}_{i,j}$ cannot be locally connected at $(r_1(0), \ldots, r_n(0), \varepsilon)$, which contradicts the local connectedness of $\mathscr{S}_{i,j}$. Thus, the theorem holds. □

The following lemma will be of use in the proof of Theorem 1.3.

Lemma 2.1 *Let* $M = \max\{a_{ij} \in A\}$, *and fix* $\varepsilon > 0$. *Suppose that for some constants* $b > a > 0$, *the sets* $\mathscr{L} := \{i : r_i(\varepsilon) < a\}$ *and* $\mathscr{H} := \{i : r_i(\varepsilon) > b\}$ *are nonempty and their union is* $\{1, 2, \ldots, n\}$. *Then* $(b/a) < 1 + (M/\varepsilon)$.

Proof By the optimality of r, multiplying all the $r_i \in \mathcal{H}$ by some factor λ should not increase $F(r, \varepsilon)$. This implies that

$$\frac{d}{d\lambda}\bigg|_{\lambda=1} \sum_{i \in \mathcal{H}, j \in \mathcal{L}} \left[a_{ij} \log\left(\frac{\lambda r_i}{\lambda r_i + r_j}\right) + a_{ji} \log\left(\frac{r_j}{\lambda r_i + r_j}\right) + \varepsilon \log\left(\frac{\lambda r_i r_j}{(\lambda r_i + r_j)^2}\right) \right] = 0.$$

Simplifying, this gives

$$\sum_{i \in \mathcal{H}, j \in \mathcal{L}} \left[\frac{a_{ij} r_j - a_{ji} r_i + \varepsilon(r_j - r_i)}{r_i + r_j} \right] = 0.$$

But, then if $(b/a) \geqslant 1 + (M/\varepsilon)$ we have, for $i \in \mathcal{H}$ and $j \in \mathcal{L}$,

$$\frac{r_i}{r_j} > \frac{b}{a} \geqslant 1 + \frac{M}{\varepsilon} \geqslant 1 + \frac{a_{ij}}{\varepsilon} \geqslant \frac{a_{ij} + \varepsilon}{a_{ji} + \varepsilon},$$

so

$$a_{ij} r_j - a_{ji} r_i + \varepsilon(r_j - r_i) = (a_{ij} + \varepsilon) r_j - (a_{ji} + \varepsilon) r_i < 0,$$

which is a contradiction. So the lemma holds. $\qquad\square$

Proof of Theorem 1.3 Note that under the standard normalization $\min_i p_i$ stays bounded as $\varepsilon \downarrow 0$. Thus, by Lemma 2.1 each p_i stays bounded as $\varepsilon \downarrow 0$.

Determining $\lim_{\varepsilon \downarrow 0} p_i(\varepsilon)$ will be done using a Dedekind cut since it is r_i, not p_i, that appears naturally in a semi-algebraic system. Fix $q \in \mathbf{Q}$ and choose integers q_1 and q_2 such that $q = q_1/q_2$. By augmenting the semi-algebraic system in the proof of Theorem 1.1 with the polynomial inequality

$$r_i^{q_2} \geqslant \varepsilon^{q_1},$$

and using an argument similar to that found in the proof of Theorem 1.2, there must be an interval $(0, \varepsilon_0)$ such that $r_i = r_i(\varepsilon)$ satisfies this inequality for all $\varepsilon \in (0, \varepsilon_0)$ or it satisfies it for no $\varepsilon \in (0, \varepsilon_0)$. If the former holds, place q in $\mathcal{U}_i$; if the latter, place it in $\mathcal{L}_i$. Repeat this for every $q \in \mathbf{Q}$. Since ε^q is a decreasing function of q for $\varepsilon \in (0, 1)$, note that all the elements of $\mathcal{U}_i$ are larger than all the elements of $\mathcal{L}_i$. Furthermore, by Lemma 2.1, $\mathcal{U}_i$ and $\mathcal{L}_i$ are both nonempty. Set $p_i(0) = \sup \mathcal{L}_i (= \inf \mathcal{U}_i)$. It is not hard to see that $\lim_{\varepsilon \downarrow 0} p_i(\varepsilon) = p_i(0)$, as was desired. $\qquad\square$

3 Properties

We will first discuss properties that the extension inherits from Zermelo's original model itself. We will then discuss some properties that only have meaning within the context of the extended model.

3.1 Properties inherited from the original model

First, we note that, unlike some models, Zermelo's model (original or extended) treats wins and losses symmetrically in that the ordering corresponding to A^T is the reverse of the ordering corresponding to A.

A second important property that the extended model inherits is one of monotonicity:

If a generalized tournament is altered by an additional comparison favouring object i over object j, then the relative ranking of object i will not drop (and the relative ranking of object j will not rise). This property is not universally satisfied by all ranking models. It is not difficult, for example, to construct examples where it fails for the Ratings Percentage Index, a version of which is used by the NCAA to help select basketball teams for its postseason Division 1 tournament. The following theorem shows that the extended model has this property.

Theorem 3.1 *Suppose object k is ranked better than or equal to object ℓ by the extended model for the outcome matrix $A = (a_{ij})$, and suppose $\tilde{A} = (a_{ij} + c_{ij})$, where $c_{ij} = 0$ if $i \neq k$, $c_{kj} \geqslant 0$ for every j, and $c_{kj} > 0$ for some j. Then the extended model for $\tilde{A}$ ranks k strictly better than ℓ.*

Proof The theorem is a consequence of the following lemma and the fact that the order produced by the extended model is the limit of orders corresponding to Zermelo's original model for outcome matrices in which all off-diagonal entries are nonzero. $\qquad\square$

Lemma 3.2 *Let $B = (b_{ij})$ be an $(\ell+1) \times (\ell+1)$ outcome matrix corresponding to a strongly connected graph $\Gamma(B)$, and let $\tilde{B}$ be the matrix obtained by deleting the last row and column of B. Let $r(B)$ be the strength vector generated by Zermelo's (original) model for the outcome matrix B with those strengths normalized so that $r_{\ell+1}(B) \equiv 1$. Then $\partial r_i / \partial a_{\ell+1,\ell} \leqslant 0$ for $i = 1,\dots,\ell$. More precisely, if i is in the same component of $\Gamma(\tilde{B})$ as ℓ, then $\partial r_i / \partial a_{\ell+1,\ell} < 0$, while if i is in a different component then $\partial r_i / \partial a_{\ell+1,\ell} = 0$.*

Proof Assume without loss of generality that the objects within each connected component of $\Gamma(\tilde{B})$ are numbered consecutively. By implicit differentiation of the critical point equations satisfied by r, we see that $x_i := \partial r_i / \partial a_{\ell+1,\ell}$ is a solution of the linear system $Mx = b$, where

$$
M = \begin{bmatrix}
\sum_{j=1}^{\ell+1} m_{j1} & -m_{12} & -m_{13} & \cdots & -m_{1\ell} \\
-m_{21} & \sum_{j=1}^{\ell+1} m_{j2} & -m_{23} & \cdots & -m_{2\ell} \\
-m_{31} & -m_{32} & \sum_{j=1}^{\ell+1} m_{j3} & & -m_{3\ell} \\
\vdots & \vdots & & \ddots & \\
-m_{\ell 1} & -m_{\ell 2} & -m_{\ell 3} & & \sum_{j=1}^{\ell+1} m_{j\ell}
\end{bmatrix},
$$

$$
x = [x_1, x_2, \dots, x_\ell],
$$

$$
b = [0, 0, \dots, 0, -r_\ell/(r_\ell + 1)]^T,
$$

and

$$
m_{ij} = \frac{(a_{ij} + a_{ji})r_i}{(r_i + r_j)^2}.
$$

Note that the components of $\Gamma(M)$ are the same as the components of $\Gamma(\tilde{B})$ and that the components of $\Gamma(M)$ and the strong components of $\Gamma(M)$ coincide. Because of our assumption about the labelling of objects, M is block diagonal, with one block for each connected component of $\Gamma(\tilde{B})$. Each block is column diagonally dominant, and

because $\Gamma(B)$ is strongly connected, at least one diagonal entry in each block is larger than the absolute values of the other entries in its column. By Corollary 6.2.9 in Horn and Johnson [16], each block of M is therefore invertible, so M^{-1} exists (and is similarly block diagonal). Furthermore, because the graph of each block of M is strongly connected, all the entries of the inverse of each block are positive [3]. Since $x = M^{-1}b$, the lemma follows. $\qquad\square$

Remark Rubinstein [30] proposes three axioms for a ranking method to satisfy and then shows that, for round-robin tournaments, ranking by score is the only method satisfying all these axioms. By construction, the extended model satisfies his first axiom, which deals with 'anonymity' of the objects, for arbitrary tournaments; that this model satisfies his second axiom, which deals with 'positive responsiveness', is a consequence of Theorem 3.1. The third axiom, which deals with the comparisons affecting the relative ranking of two specific objects, is more controversial, and simple tournaments can be found for which the extended model fails to satisfy this axiom.

3.2 Properties specific to the extended model

We begin by defining a new partial ordering corresponding to $\Gamma(A)$.

Definition We write $i \trianglerighteq j$ (or $j \trianglelefteq i$) if there is a direct path from i to j in $\Gamma(A)$. (Note that $\trianglerighteq$ is a pseudo-ordering of $\{1,\ldots,n\}$.) We write $i \triangleright j$ (or $j \triangleleft i$), if $i \trianglerighteq j$ but $i \ntrianglelefteq j$.

This partial ordering is related to the total ordering generated by the extended model in the following way:

Theorem 3.3   (1) *If $i \trianglerighteq j$, then r_j/r_i remains bounded as $\varepsilon \downarrow 0$.*
   (2) *If $i \triangleright j$ then $r_j/r_i \to 0$ as $\varepsilon \downarrow 0$; in particular, $i \succ j$.*

Proof To prove part 1, it suffices to consider the case when $a_{ij} \geqslant 1$. Note that

$$F(r(\varepsilon), \varepsilon) \leqslant \log\left(\frac{1}{1 + r_j/r_i}\right),$$

while

$$F(1, \varepsilon) = -\left(n(n-1)\varepsilon + \sum_{i,j=1}^{n} a_{ij}\right) \log 2.$$

This yields a contradiction to the optimality of r if

$$\frac{r_j}{r_i} > \exp\left(\left(n(n-1)\varepsilon + \sum_{i,j=1}^{n} a_{ij}\right) \log 2\right) - 1.$$

   To prove part 2, we may (because of part 1) assume without loss of generality that $a_{ij} \geqslant 1$. Fix ε, and let i and j be objects such that $i \triangleright j$. Set $\bar{r} = \bar{r}(\lambda) = (\bar{r}_1(\lambda), \ldots, \bar{r}_n(\lambda))$,

where

$$\bar{r}_k(\lambda) = \begin{cases} \lambda r_k(\varepsilon) & \text{if } k \rhd j \\ r_k(\varepsilon) & \text{otherwise.} \end{cases}$$

By the optimality of r,

$$\frac{d}{d\lambda} F(\bar{r}(\lambda), \varepsilon)\Big|_{\lambda=1} = 0. \tag{3.1}$$

On the other hand,

$$\frac{d}{d\lambda} F(\bar{r}(\lambda), \varepsilon)\Big|_{\lambda=1} = \frac{d}{d\lambda}\Big|_{\lambda=1} \left\{ \sum_{\substack{p \rhd j \\ q \rhd j}} + \sum_{\substack{p \not\rhd j \\ q \not\rhd j}} + \sum_{\substack{p \rhd j \\ q \not\rhd j}} + \sum_{\substack{p \not\rhd j \\ q \rhd j}} \right\} (a_{pq} + \varepsilon) \log\left(\frac{\bar{r}_p}{\bar{r}_p + \bar{r}_q} \right)$$

$$= \frac{d}{d\lambda}\Big|_{\lambda=1} \left\{ \sum_{\substack{p \rhd j \\ q \not\rhd j}} (a_{pq} + \varepsilon) \log\left(\frac{\lambda r_p}{\lambda r_p + r_q} \right) + \varepsilon \sum_{\substack{p \not\rhd j \\ q \rhd j}} \log\left(\frac{r_p}{r_p + \lambda r_q} \right) \right\}$$

$$= \sum_{\substack{p \rhd j \\ q \not\rhd j}} (a_{pq} + \varepsilon) \frac{r_q}{r_p + r_q} - \varepsilon \sum_{\substack{p \not\rhd j \\ q \rhd j}} \frac{r_q}{r_p + r_q}.$$

Combining this with (3.1) yields,

$$\frac{n^2}{2}\varepsilon \geq \varepsilon \left\{ \sum_{\substack{p \not\rhd j \\ q \rhd j}} - \sum_{\substack{p \rhd j \\ q \not\rhd j}} \right\} \frac{r_q}{r_p + r_q} = \sum_{\substack{p \rhd j \\ q \not\rhd j}} a_{pq} \frac{r_q}{r_p + r_q} \geq a_{ij} \frac{r_j}{r_i + r_j} \geq \frac{1}{r_i/r_j + 1},$$

so

$$\frac{r_i}{r_j} \geq \frac{2}{n^2 \varepsilon} - 1 \to \infty$$

as $\varepsilon \downarrow 0$. $\qquad\square$

Remark One ranking method that has been proposed is to order the objects so as to minimize the number of *upsets* in the tournament. That is, an order $\preceq$ is chosen so as to minimize $\sum_{i \leq j} a_{ij}$. (Related methods, which take into account the magnitude of each upset, are discussed elsewhere [14, 18, 19].) Although the extension of Zermelo's model analyzed in this paper does not necessarily do this, the preceding result signifies that it does avoid a certain class of upsets.

Recall that two objects, i and j, will be in the same strongly connected component of $\Gamma(A)$ if and only if $i \unrhd j$ and $i \unlhd j$. For convenience we will refer to a strongly connected component of $\Gamma(A)$ as an *SCC*.

Theorem 3.4 *If i and j are in the same SCC then r_i/r_j converges to a positive constant as $\varepsilon \downarrow 0$.*

Proof For each i and j in the same SCC, append the variable s_{ij} and the equation $\varepsilon(s_{ij} r_j - r_i) = 0$ to the system of polynomial equations and inequalities considered in the

proof of Theorem 1.1. The resulting solution set (subject to these additional constraints) is a semi-algebraic set in a higher-dimensional Euclidean space. By Part 1 of Theorem 3.3, the portion of this set satisfying $\varepsilon > 0$ is confined to a compact subset of this higher-dimensional space; therefore, it follows from an argument similar to the one in the proof of Theorem 1.1 that the s_{ij} converge as $\varepsilon \downarrow 0$ (and Part 1 of Theorem 3.3 implies that these limits are nonzero). □

Theorem 3.5 *Let $\mathscr{S}$ be an SCC, and let $\tilde{A}$ be the submatrix of A corresponding to the objects in $\mathscr{S}$. If $i, j \in \mathscr{S}$, then*

$$\lim_{\varepsilon \downarrow 0} \frac{r_i(\varepsilon)}{r_j(\varepsilon)} = \frac{\tilde{r}_i}{\tilde{r}_j},$$

where $\tilde{r}_i$ and $\tilde{r}_j$ are the strengths that Zermelo's (original) model for the reduced outcome matrix $\tilde{A}$ assigns to these two objects.

Proof Let $\mathscr{S}$, i, j, and $\tilde{r}$ be as in the statement of the theorem. Set

$$\rho(\varepsilon) = \sqrt{\sum_{k \in \mathscr{S}} r_k^2(\varepsilon)},$$

and let $\bar{r} = \bar{r}(\varepsilon) = (\bar{r}_1, \ldots, \bar{r}_n)$, where

$$\bar{r}_k = \begin{cases} \tilde{r}_k/\rho(\varepsilon) & \text{if } k \in \mathscr{S} \\ r_k & \text{otherwise.} \end{cases}$$

Note that

$$\Delta F := F(\bar{r}, \varepsilon) - F(r, \varepsilon)$$

$$= \left\{ \sum_{p,q \in \mathscr{S}} + \sum_{\substack{p \rhd q \\ q \in \mathscr{S}}} + \sum_{\substack{p \in \mathscr{S} \\ p \rhd q}} \right\} a_{pq} \left(\log\left(\frac{\bar{r}_p}{\bar{r}_p + \bar{r}_q}\right) - \log\left(\frac{r_p}{r_p + r_q}\right) \right)$$

$$+ \varepsilon \sum_{p,q=1}^{n} \left(\log\left(\frac{\bar{r}_p}{\bar{r}_p + \bar{r}_q}\right) - \log\left(\frac{r_p}{r_p + r_q}\right) \right)$$

$$=: \{\mathscr{S}_1 + \mathscr{S}_2 + \mathscr{S}_3\} + \varepsilon \mathscr{S}_4.$$

Suppose the conclusion of the theorem fails; then (by Theorem 3.4) $\mathscr{S}_1$ is (positive and) bounded away from zero as $\varepsilon \downarrow 0$. Also, Theorem 3.4 implies that there exists $\lambda > 0$ such that $1/\lambda < \bar{r}_k/r_k < \lambda$ for all ε sufficiently small.

If $p \rhd q$ and $q \in \mathscr{S}$ then

$$\log\left(\frac{\bar{r}_p}{\bar{r}_p + \bar{r}_q}\right) - \log\left(\frac{r_p}{r_p + r_q}\right) = \log\left(\frac{r_p/r_q + 1}{r_p/r_q + \bar{r}_q/r_q}\right) \to 0$$

as $\varepsilon \downarrow 0$, since, by Theorem 3.3, $r_p/r_q \to \infty$ as $\varepsilon \downarrow 0$. Hence, $\mathscr{S}_2 \to 0$ as $\varepsilon \downarrow 0$. Similarly, if $p \in \mathscr{S}$ and $p \rhd q$ then

$$\log\left(\frac{\bar{r}_p}{\bar{r}_p + \bar{r}_q}\right) - \log\left(\frac{r_p}{r_p + r_q}\right) = \log\left(\frac{r_p/r_q + 1}{r_p/r_q + r_p/\bar{r}_p}\right) \to 0,$$

so $\mathscr{S}_3 \to 0$ as $\varepsilon \downarrow 0$.

For arbitrary p and q,

$$\log\left(\frac{\overline{r}_p}{\overline{r}_p + \overline{r}_q}\right) - \log\left(\frac{r_p}{r_p + r_q}\right) = \log(\overline{r}_p/r_p) + \log\left(\frac{r_p + r_q}{\overline{r}_p + \overline{r}_q}\right) \geqslant -2\log\lambda.$$

This bounds $\mathscr{S}_4$ away from $-\infty$ and, in conjunction, with the preceding estimates, shows that $\limsup_{\varepsilon\downarrow 0}\Delta F > 0$, contradicting the optimality of r. Hence, the theorem holds. $\qquad\square$

4 Computational techniques

4.1 Continuation methods

The most direct way to rank objects according to the extended model would be to fix a small value of ε and then calculate $r(\varepsilon)$ using Ford's algorithm [11], which is guaranteed to converge to the true solution, or a combination of Ford's algorithm with Newton's method (to polish the roots). Then ε could be allowed to approach zero and Davidenko's method (or a more sophisticated curve-tracking method, e.g. see [27]–[29]) could be used to follow $r(\varepsilon)$ as $\varepsilon \downarrow 0$.

A drawback of this approach is that for even moderately large tournaments this may require substantial computational effort, especially if the curve must be tracked until ε is extremely small in order for the asymptotic ordering to be achieved. (See §5 for an example of how $\preceq_\varepsilon$ depends upon ε.)

A partial remedy to this difficulty is suggested by Theorem 3.4. Since the ratio of strengths of objects in the same SCC approaches their ratio for comparisons internal to the SCC, it would be reasonable to determine the SCCs and apply Zermelo's (original) model to each SCC separately to compute the optimal strength r_{ij} of the jth object in the ith SCC based on that subset of the tournament. Since these strengths would be then only determined up to a constant multiple, constants $\lambda_1, \lambda_2, \ldots, \lambda_N$ could then be chosen so as to maximize

$$\overline{F} = \overline{F}(\lambda_1, \ldots, \lambda_N, \varepsilon) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{N}\sum_{\ell=1}^{n_k}(a_{ijk\ell} + \varepsilon)\log\left(\frac{\lambda_i r_{ij}}{\lambda_i r_{ij} + \lambda_k r_{k\ell}}\right).$$

Here N is the number of SCCs, n_q is the number of objects in the qth SCC, and $a_{ijk\ell}$ represents the number of comparisons in which the jth object in the ith SCC is judged superior to the ℓth object in the kth SCC. Depending on the normalization constraints used in each SCC, it might be reasonable to think of λ_i as the strength of the ith SCC as a whole. The following theorem shows that this optimization problem is well-posed.

Theorem 4.1 *For fixed $\varepsilon > 0$, (4.1) has a unique maximizing vector $\lambda = (\lambda_1, \ldots, \lambda_N)$ satisfying the constraint $\sum_{i=1}^{N}\lambda_i^2 = 1$.*

Proof For simplicity and without loss of generality, we absorb the ε terms into the $a_{ijk\ell}$ terms and assume that the tournament is strongly connected. This implies that as the boundary of the positive orthant (in λ-space) is approached $\overline{F} \to -\infty$, so a maximizer exists. We claim that all critical points of $\overline{F}$ (subject to the constraint) are local maximizers, which will imply (by an argument similar to that in the proof of Theorem 1.1) that the maximizer will be unique.

Now, after simplification, we find that

$$\frac{\partial \overline{F}}{\partial \lambda_m} = \frac{\alpha_m}{\lambda_m} - \sum_{j=1}^{n_m} \sum_{k=1}^{N} \sum_{\ell=1}^{n_k} \frac{A_{mjk\ell} r_{mj}}{\lambda_m r_{mj} + \lambda_k r_{k\ell}},$$

where $A_{mjk\ell} = (a_{mjk\ell} + a_{k\ell mj})/2$ and $\alpha_m = \sum_{j=1}^{n_m} \sum_{k=1}^{N} \sum_{\ell=1}^{n_k} a_{mjk\ell}$. Also,

$$\frac{\partial^2 \overline{F}}{\partial \lambda_m \partial \lambda_q} = \sum_{j=1}^{n_m} \sum_{\ell=1}^{n_q} \frac{A_{mjq\ell} r_{mj} r_{q\ell}}{(\lambda_m r_{mj} + \lambda_q r_{q\ell})^2} + \delta_{mq} \left[-\frac{\alpha_m}{\lambda_m^2} + \sum_{j=1}^{n_m} \sum_{k=1}^{N} \sum_{\ell=1}^{n_k} \frac{A_{mjk\ell} r_{mj}^2}{(\lambda_m r_{mj} + \lambda_k r_{k\ell})^2} \right].$$

Define

$$S_{uvwx} = \frac{A_{uvwx}}{(\lambda_u r_{uv} + \lambda_w r_{wx})^2}.$$

A calculation shows that

$$u^T D^2 \overline{F}(\lambda) u = \sum_{m=1}^{N} \sum_{j=1}^{n_m} \sum_{q=1}^{N} \sum_{\ell=1}^{n_q} [u_m r_{mj} S_{mjq\ell}(u_n r_{q\ell} + u_m r_{mj})] - \sum_{m=1}^{N} \frac{u_m^2 \alpha_m}{\lambda_m^2},$$

which, if λ is a critical point, becomes

$$u^T D^2 \overline{F}(\lambda) u = \sum_{m=1}^{N} \sum_{j=1}^{n_m} \sum_{q=1}^{N} \sum_{\ell=1}^{n_q} [u_m r_{mj} S_{mjq\ell}(u_q r_{q\ell} + u_m r_{mj})]$$

$$- \sum_{m=1}^{N} \sum_{j=1}^{n_m} \sum_{q=1}^{N} \sum_{\ell=1}^{n_q} \frac{u_m^2}{\lambda_m} r_{mj}(\lambda_m r_{mj} + \lambda_q r_{q\ell}) S_{mjq\ell}.$$

This can be simplified to

$$u^T D^2 \overline{F}(\lambda) u = -\frac{1}{2} \sum_{m=1}^{N} \sum_{j=1}^{n_m} \sum_{q=1}^{N} \sum_{\ell=1}^{n_q} r_{mj} r_{q\ell} \frac{S_{mjq\ell}}{\lambda_m \lambda_q} (u_m \lambda_q - u_q \lambda_m)^2 \leqslant 0,$$

with equality if and only if u and λ are parallel. Since the constraint $\sum \lambda_i^2 = 1$ is imposed, only those perturbations u that are orthogonal to λ need be considered. This verifies that critical points are local maximizers and completes the proof. $\qquad\square$

4.2 Heuristics for the limiting functional

The approach just described may significantly reduce the number of unknown quantities being calculated in the main stage. A further improvement may result from doing the equivalent calculation using the variables $p_i = \log_\varepsilon r_i$, since (as is shown by Theorem 1.3) these converge as $\varepsilon \downarrow 0$. Still, for small ε the calculations often require performing operations on numbers of widely varying magnitudes, thus leading to a potentially serious loss of accuracy.

Another serious problem is the slowness with which the approximate minimizers converge for any small fixed ε. This seems to be related to the phenomenon known variously as *dormant instability* [12] or *dynamical metastability* [7], in which a gradient flow for a singular perturbation of a functional that originally had a continuum of minimum energy states evolves extremely slowly in spite of being far from a true equilibrium state (of the

perturbed system). Such phenomena were originally studied in the context of interface motion in phase transition models in the presence of non-convex free energy, but the energy method of Bronsard and Kohn [6] indicates that the underlying mechanism is quite general.

Some insight into the limiting ordering is obtained by considering the functional $G(p, \varepsilon) := -\log P(\varepsilon^p, \varepsilon)$, where P is given by (1.2), and $\varepsilon^p := (\varepsilon^{p_1}, \varepsilon^{p_2}, \ldots, \varepsilon^{p_n})$. Some algebraic manipulation reveals that

$$
\begin{aligned}
G(p, \varepsilon) = \; & \left\{ \log\left(\frac{1}{\varepsilon}\right) \sum_{i,j=1}^{n} a_{ij}(p_i - p_j)_+ \right\} \\
& + \left\{ \sum_{|p_i - p_j| < 1} a_{ij} \log(1 + \varepsilon^{|p_i - p_j|}) \right\} \\
& + \left\{ \varepsilon \log\left(\frac{1}{\varepsilon}\right) \sum_{i,j=1}^{n} |p_i - p_j| \right\} \\
& + \left\{ \sum_{|p_i - p_j| \geqslant 1} a_{ij} \log(1 + \varepsilon^{|p_i - p_j|}) + 2\varepsilon \sum_{i<j} \log(1 + \varepsilon^{|p_i - p_j|}) \right\} \\
=: \; & Q_1 + Q_2 + Q_3 + Q_4.
\end{aligned}
$$

Since $\log(1 + \varepsilon^{|p_i - p_j|}) \approx \varepsilon^{|p_i - p_j|}$, the Q_i should be in order of decreasing importance for $\varepsilon \ll 1$. This suggests that *asymptotically* the p_i should be determined by the following sequence of steps:

(1) Minimize Q_1 by requiring that $p_i \leqslant p_j$ whenever $i \trianglerighteq j$. Equivalently, require that

$$
p \in \mathscr{S}_1 := \{p : p_i - p_j \leqslant 0 \text{ if } a_{ij} > 0\}.
$$

Note that $Q_1 \geqslant 0$ with equality if and only if $p \in \mathscr{S}_1$, and $\mathscr{S}_1$ is nonempty (since $0 \in \mathscr{S}_1$), so $\mathscr{S}_1$ is, in fact, the set of minimizers of Q_1. Note, furthermore, that $\mathscr{S}_1$ is convex. (Here and below, we ignore the particular normalization constraint imposed and consider two values of p to be equivalent if their difference is a vector with equal coordinates.)

(2) Minimize Q_2 over $\mathscr{S}_1$ by requiring that

$$
p \in \mathscr{S}_2 := \{p \in \mathscr{S}_1 : p_j - p_i \geqslant 1 \text{ if } i \triangleright j\}.
$$

Note that each term in Q_2 is nonnegative. One of the effects of the previous step was to mandate that $|p_i - p_j| = 0$ if i and j are in the same SCC; each corresponding pair of terms in Q_2 is set equal to $(a_{ij} + a_{ji}) \log 2$. Q_2 is minimal if and only if there are no other nonzero terms in Q_2. This happens precisely when $p \in \mathscr{S}_2$.

Now, clearly, $\mathscr{S}_2$ is closed and convex. To see that it is nonempty, consider the following process. Begin with $p \equiv 0$. Successively examine each ordered pair of objects (i, j). If $a_{ij} > 0$, $i \triangleright j$, and $p_j = p_i$, set p_j equal to $p_i + 1$ and simultaneously increase by 1 each p_k for which $j \trianglerighteq k$. It is not hard to see that in processing (i, j):

(a) p remains in $\mathscr{S}_1$.

(b) There is no retrogression of p away from $\mathscr{S}_2$; more precisely, there will not be a pair of objects (ℓ, m) for which $a_{\ell m} > 0$ and $p_m - p_\ell \geqslant 1$ before the operation but $p_m - p_\ell < 1$ after the operation.

The final value of p when this process is complete will be an element of $\mathscr{S}_2$.

(3) Minimize Q_3 over $\mathscr{S}_2$. Since Q_3 is a convex functional and $\mathscr{S}_2$ is a closed, convex set, we know that
$$m_3 := \inf\{Q_3(p) : p \in \mathscr{S}_2\}$$
will be achieved on $\mathscr{S}_2$ and that
$$\mathscr{S}_3 := \{p \in \mathscr{S}_2 : Q_3(p) = m_3\}$$
is a convex set. The complexity of finding a point p in $\mathscr{S}_3$ appears to depend greatly on the particular structure of $\Gamma(A)$. One situation in which it is straightforward to do this is when there is an SCC, call it $\mathscr{C}$, containing at least $n/2$ objects. In that case, given $p \in \mathscr{S}_2$ normalize it so that $p_i = 0$ for $i \in \mathscr{C}$. If $y = \min\{|p_j| : j \notin \mathscr{C}\}$, pick p_j with absolute value y and move it as close to 0 as possible without causing p to leave $\mathscr{S}_2$. (This process is guaranteed not to increase Q_3.) Then do the same to the p_j that is next closest to 0 and, successively, to all $j \notin \mathscr{C}$. This yields a vector $p \in \mathscr{S}_3$.

(4) Find the (asymptotic) minimum of Q_4 over $\mathscr{S}_3$ in the following way. Given $p \in \mathscr{S}_3$, let
$$f_p = \sum_{|p_i - p_j| \geqslant 1} a_{ij} \chi_{\{|p_i - p_j|\}} + 2 \sum_{i<j} \chi_{\{|p_i - p_j|+1\}},$$
where $\chi_{\mathscr{S}} : [1, \infty) \to \mathbf{R}$ is the characteristic function satisfying
$$\chi_{\mathscr{S}}(x) = \begin{cases} 1 & \text{if } x \in \mathscr{S} \\ 0 & \text{otherwise.} \end{cases}$$

Impose a lexicographical order on $\{f_p : p \in \mathscr{S}_3\}$ under which f_p precedes $f_{p'}$ if and only if there is a number $y \in [1, \infty)$ such that $f_p(y) > f_{p'}(y)$ and $f_p(x) = f_{p'}(x)$ for every $x \in [1, y)$, and seek $p \in \mathscr{S}_3$ that is maximal under the induced order on $\mathscr{S}_3$. As with the preceding step, the difficulty of performing this step seems to vary widely from tournament to tournament.

As an example of the use of these four steps, consider the tournament with 11 objects whose outcome matrix A has entries that are all 0, except
$$a_{1,2} = a_{1,3} = a_{1,4} = a_{1,5} = a_{2,6} = a_{3,7} = a_{3,8} = a_{4,9} = a_{5,9}$$
$$= a_{6,10} = a_{7,11} = a_{8,11} = a_{9,11} = a_{10,11} = 1.$$

The tournament graph is shown in Figure 1.

Steps 1 and 2 combined indicate that $p_1 + 1 \leqslant p_2, p_3, p_4, p_5$; $p_2 + 1 \leqslant p_6$; $p_3 + 1 \leqslant p_7, p_8$; $p_4, p_5 \leqslant p_9 - 1$; $p_6 \leqslant p_{10} - 1$; and $p_7, p_8, p_9, p_{10} \leqslant p_{11} - 1$. We claim that $m_3 = 136\varepsilon \log(1/\varepsilon)$. To see this, first note that $Q_3 = 136\varepsilon \log(1/\varepsilon))$ if
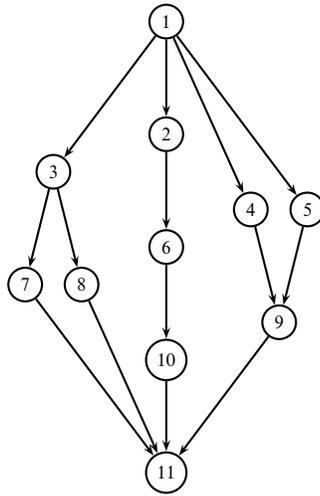$$p = p^0 := (-2, -1, -2/3, -1/3, -1/3, 0, 1/3, 1/3, 2/3, 1, 2).$$

FIGURE 1. $\Gamma(A)$.

Since Q_3 is convex, p^0 will be a minimizer of Q_3 if and only if no small perturbation (with p remaining in $\mathscr{S}_2$) lowers Q_3. Without loss of generality, we only need to consider those perturbations for which p_6 remains 0. Note that decreasing p_1 and p_2 or increasing p_{10} and p_{11} increases Q_3. By symmetry, it suffices to consider only those perturbations for which $p_7 = p_8$ and $p_4 = p_5$. A simple calculation then shows that $Q_3 = (112 + 12(p_7 - p_3) + 12(p_9 - p_4))\varepsilon \log(1/\varepsilon))$ for all perturbations under consideration, and the constraint that $p \in \mathscr{S}_2$ implies that $m_3 = 136\varepsilon \log(1/\varepsilon)$, as was claimed.

It can be seen that the set $\mathscr{S}_3$ is the convex hull of the points

$$p = (-2, -1, -1, 0, 0, 0, 0, 0, 1, 1, 2)$$
$$p = (-2, -1, -1, -1, -1, 0, 0, 0, 0, 1, 2)$$
$$p = (-2, -1, -0, 0, 0, 0, 1, 1, 1, 1, 2).$$

Proceeding to Step 4, we see that for $p \in \mathscr{S}_3$

$$f_p = g + 8(\chi_{\{2+p_4\}} + \chi_{\{2-p_7\}} + \chi_{\{1+p_7-p_4\}} + \chi_{\{2+p_4-p_7\}}) + 7(\chi_{\{1-p_4\}} + \chi_{\{1+p_7\}}) + h_p,$$

where g is independent of p and $h_p(x) = 0$ for all $x < 2$. A simple calculation then shows that p is optimal when $p_7 = -p_4 = 1/3$ (i.e. when $p = p^0$). This, in turn, implies that $i \geq j$ if and only if

- $i \leqslant j$; or
- $i = 5$ and $j = 4$; or
- $i = 8$ and $j = 7$.

4.3 Power series expansions

Let p be the vector of *asymptotic* exponents generated according to the rules of the previous section. Those heuristic rules do not necessarily determine the asymptotic ordering of objects completely. First, they need to be combined with the strengths r_{ij} internal to each SCC, as discussed in §4.1. Furthermore, if there are two objects i and j in different SCCs with $p_i = p_j$, some means needs to be provided for comparing these two objects. One idea is to use power series methods to expand r_i in terms of ε.

Two problems mitigate against using power series methods alone. First, without knowledge of the order of magnitude of the leading-order terms, the computation becomes unwieldy. Secondly, it is necessary to know what the variable of expansion should be. There are examples with as few as three objects in which the r_i cannot be expanded as Maclaurin series in ε but instead must be expanded as Maclaurin series in $\delta := \varepsilon^{1/2}$. The method of Newton polygons can be used to find the appropriate power of ε to set δ equal to; alternatively, if p has been determined by the methods of §4.2, then δ can be set equal to $\varepsilon^{1/k}$, where $1/k$ is the greatest common factor of $\{|p_i - p_j|\}_{i,j=1}^{n}$.

We set $\pi_i := p_i k$ (where p_i and k are as above), and we then set $\rho_i := r_i/\delta^{\pi_i}$, so $\rho_i(0)$ is the coefficient of the leading term in the δ-expansion of r_i. If $\pi_i = \pi_j$, we say that i and j are in the same *level*. Note that each SCC is contained in a level. Replacing k by i in (2.2), setting $\partial F/\partial r_i$ to 0, and multiplying through by r_i yield

$$n\delta^k + \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n}(a_{ij} + a_{ji} + 2\delta^k)\frac{\rho_i\delta^{\pi_i}}{\rho_i\delta^{\pi_i} + \rho_j\delta^{\pi_j}}. \tag{4.1}$$

Letting $\delta \downarrow 0$ in (4.1) yields

$$\sum_{j=1}^{n} a_{ij} = \sum_{\pi_j=\pi_i}\left((a_{ij} + a_{ji})\frac{\rho_i(0)}{\rho_i(0) + \rho_j(0)}\right) + \sum_{\pi_j>\pi_i} a_{ij}. \tag{4.2}$$

(This equation is, in fact, implied by Theorems 3.3 and 3.5.)

If we now subtract (4.2) from (4.1) and sum over all i in some SCC $\mathscr{S}$, we have

$$\#(\mathscr{S})n\delta^k = \sum_{i\in\mathscr{S}}\sum_{\pi_j=\pi_i} 2\delta^k\frac{\rho_i}{\rho_i + \rho_j} + \sum_{i\in\mathscr{S}}\sum_{\pi_j<\pi_i}(a_{ji} + 2\delta^k)\frac{\delta^{\pi_i-\pi_j}\rho_i}{\delta^{\pi_i-\pi_j}\rho_i + \rho_j}$$

$$- \sum_{i\in\mathscr{S}}\sum_{\pi_j>\pi_i} a_{ij}\frac{\delta^{\pi_j-\pi_i}\rho_j}{\rho_i + \delta^{\pi_j-\pi_i}\rho_j} + \sum_{i\in\mathscr{S}}\sum_{\pi_j>\pi_i} 2\delta^k\frac{\rho_i}{\rho_i + \delta^{\pi_j-\pi_i}\rho_j}. \tag{4.3}$$

Dividing by δ^k and letting $\delta \downarrow 0$ (and using the proof of of Theorem 3.3), we have

$$\#(\mathscr{S})n = 2\sum_{i\in\mathscr{S}}\left(\sum_{\pi_j=\pi_i}\frac{\rho_i(0)}{\rho_i(0) + \rho_j(0)}\right) + \sum_{i\in\mathscr{S}}\left(\sum_{\pi_j=\pi_i-k} a_{ji}\frac{\rho_i(0)}{\rho_j(0)}\right)$$

$$- \sum_{i\in\mathscr{S}}\left(\sum_{\pi_j=\pi_i+k} a_{ij}\frac{\rho_j(0)}{\rho_i(0)}\right) + 2\sum_{i\in\mathscr{S}}\left(\sum_{\pi_j>\pi_i} 1\right). \tag{4.4}$$

This equation relates the strengths of objects in the same level as i to the strengths of objects in levels with π-values differing from π_i by an integer multiple of k, and there are similar equations which hold at each of those levels. Together, these equations determine $\rho_j(0)$ in these levels (up to a multiplicative constant). They do not, however, provide any information about the relationship between the strengths of these objects and the strengths of any other objects outside these specific levels.

Let $\mathscr{L}$ be the union of all SCCs having π-values that are congruent to the π-value of $\mathscr{S}$ modulo k. Then multiplying (4.4) by δ^k, subtracting the result from (4.3), and summing over all $\mathscr{S}$ making up $\mathscr{L}$ yield

$$0 = \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j < \pi_i - k} a_{ji} \frac{\delta^{\pi_i - \pi_j} \rho_i}{\delta^{\pi_i - \pi_j} \rho_i + \rho_j} \right) + 2\delta^k \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j < \pi_i} \frac{\delta^{\pi_i - \pi_j} \rho_i}{\delta^{\pi_i - \pi_j} \rho_i + \rho_j} \right)$$

$$- \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j > \pi_i + k} a_{ij} \frac{\delta^{\pi_j - \pi_i} \rho_j}{\rho_i + \delta^{\pi_j - \pi_i} \rho_j} \right) - 2\delta^k \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j > \pi_i} \frac{\delta^{\pi_j - \pi_i} \rho_j}{\rho_i + \delta^{\pi_j - \pi_i} \rho_j} \right).$$

Dividing this by δ^{k+1} and letting $\delta \downarrow 0$, we have

$$0 = \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j = \pi_i - (k+1)} a_{ji} \frac{\rho_i(0)}{\rho_j(0)} \right) + 2 \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j = \pi_i - 1} \frac{\rho_i(0)}{\rho_j(0)} \right)$$

$$- \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j = \pi_i + (k+1)} a_{ij} \frac{\rho_j(0)}{\rho_i(0)} \right) - 2 \sum_{i \in \mathscr{L}} \left(\sum_{\pi_j = \pi_i + 1} \frac{\rho_j(0)}{\rho_i(0)} \right).$$

This equation (along with similar equations for other sets $\mathscr{L}$) determines the leading coefficients of all objects (again, up to a multiplicative constant). If higher-order coefficients are needed to determine which of two objects with the same leading coefficient is ranked higher, this process can be continued.

4.4 An example

As a test, the procedures outlined in §4.2 and §4.3 were applied to the outcome matrix for the 1996 NCAA Division 1-A football season. Of the 111 teams, all but five of them (Boise State, Duke, Florida, Florida State and New Mexico State) were contained in a single SCC. (There were five SCCs in all.) Two teams (Florida and Florida State) had the same largest leading coefficient, so their second-order coefficients had to be compared in order to completely determine the ordering for the extended model. A modified version of (4.4) which was not summed over an SCC provided that information. The calculations were reasonably easy and produced the top 25 given in Table 1.

Table 1. *College football ranking for 1996 from the extended model*

Rank	Team	Rank	Team
1	Florida	14	Washington
2	Florida State	15	Miami
3	Ohio State	16	Michigan
4	Arizona State	17	Syracuse
5	Nebraska	18	Northwestern
6	Tennessee	19	Iowa
7	Penn State	20	Southern Mississippi
8	LSU	21	Kansas State
9	Alabama	22	Notre Dame
10	BYU	23	East Carolina
11	Virginia Tech	24	Auburn
12	Colorado	25	West Virginia
13	North Carolina		

5 Related models

Consider the outcome matrix

$$
\begin{bmatrix}
0 & 99 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 99 \\
0 & 0 & 1 & 0
\end{bmatrix}.
$$

The extended model ranks object 2 with a 'win–loss record' of 1–99 above object 3 with a record of 99–1. If this seems undesirable, the ranking of these two objects can be reversed by selecting a suitably large fixed value for ε. We call such a model a *finite ε model*. One problem with this approach is that it is unclear how to provide an objective justification for choosing one finite ε model over another. Figure 2 illustrates the dependence of the perturbed ranking on the size of the perturbation to the outcome matrix when curve-tracking techniques are applied to the data from §4.4. Each piecewise-linear curve in this diagram corresponds to the ranking of a fixed team as ε is varied. As can be seen, for $\varepsilon \in [1/2, 1/2^{16}]$ the ordering $\preceq_\varepsilon$ depends sensitively upon ε. In particular, note that the team that is ranked 32nd for $\varepsilon = 1/2$ is ranked 88th for $\varepsilon = 1/2^{16}$.

Although the perturbation we have studied in this paper has the advantages of being simple and symmetric, other perturbations may be worth studying. In particular, what happens if the perturbation matrix is chosen to have nonzero entries precisely where $A + A^T$ has nonzero entries? In other words, what if artificial connectivity were only imposed between objects having an actual comparison between them? Note that unless $\Gamma(A)$ is connected, these perturbations will not produce a strongly connected graph and Zermelo's method fails. What if, instead of adding artificial connectivity between pairs of objects, a phantom object were introduced with which each real object was artificially connected? Finally, what if instead of considering scalar multiples of a particular perturbation, a cone $\mathscr{C}$ of perturbations were considered? Identifying each conceivable ordering with a permutation π of $\{1, \ldots, n\}$, we could let $\mathscr{S}_\pi \subset \mathscr{C}$ be the set of all
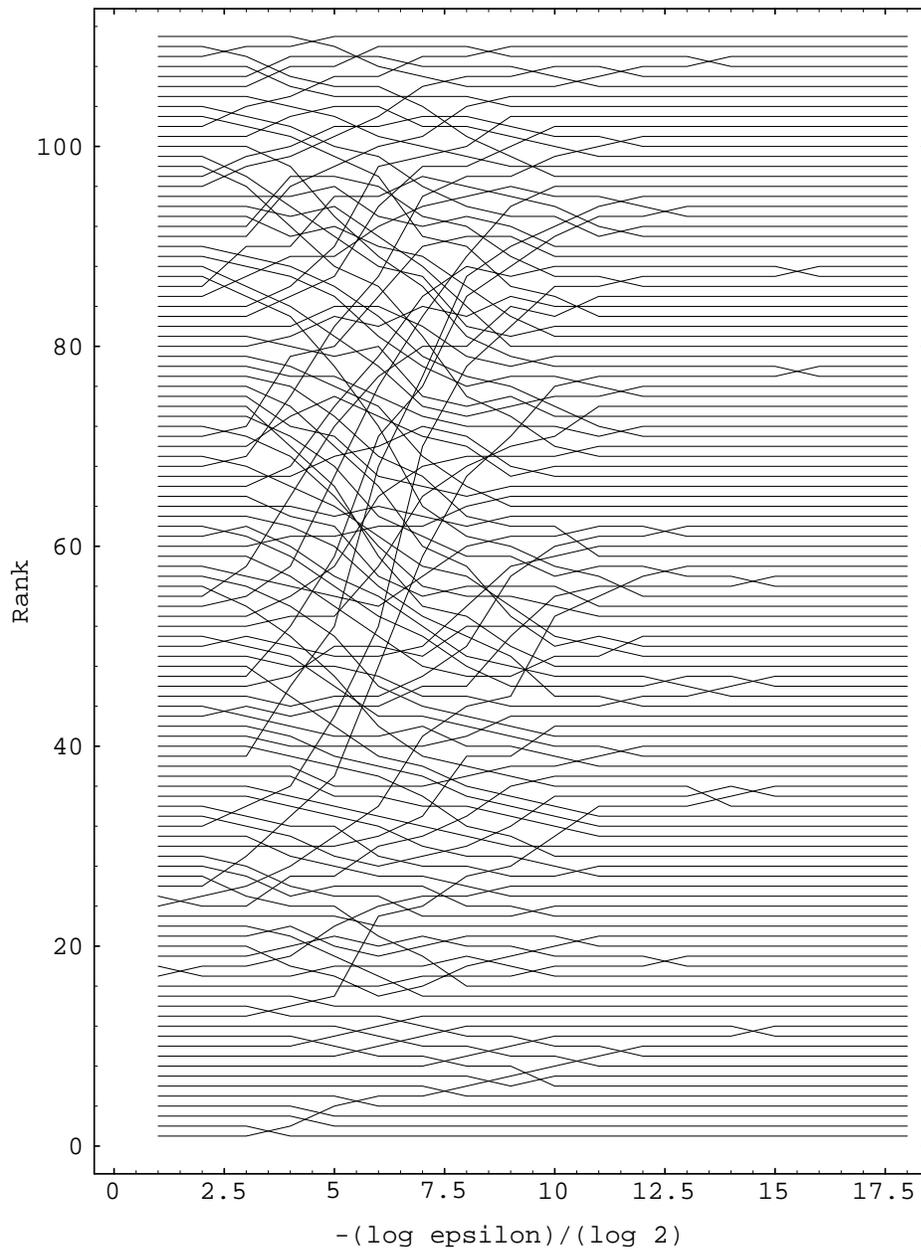
FIGURE 2. Plot of the perturbed rankings for 1996 Division 1-A NCAA football as a function of the size of the perturbation.

perturbations producing the ordering π. An ordering could then be chosen to maximize the metric density of $\mathscr{S}_\pi$ at the origin. Although such an approach is attractive in theory (for its generality), it seems unlikely to be computationally tractable.

Acknowledgements

References

[1] ALLEN, A. S. (1979) *The Ranking of Teams Competing in Pairwise Competitions.* Master's thesis, Brigham Young University.

[2] BALDO, S. (1990) Minimal interface criterion for phase transitions in mixtures of Cahn–Hilliard fluids. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **7** 67–90.

[3] BERMAN, A. & PLEMMONS, R. J. (1979) *Nonnegative Matrices in the Mathematical Sciences.* Academic Press.

[4] BRADLEY, R. A. (1954) Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics*, **10** 375–390.

[5] BRADLEY, R. A. & TERRY, M. E. (1952) The rank analysis of incomplete block designs. i. the method of paired comparisons. *Biometrika*, **39** 324–345.

[6] BRONSARD, L. & KOHN, R. V. (1990) On the slowness of phase boundary motion in one space dimension. *Comm. P. A. Math.* **43** 983–998.

[7] BRONSARD, L. & KOHN, R. V. (1991) Motion by mean curvature as the singular limit of Ginzburg–Landau dynamics. *J. Diff. Eq.* **90** 211–237.

[8] CAHN, J. W., CHOW, S.-N. & VAN VLECK, E. S. (1995) Spatially discrete nonlinear diffusion equations. *Rocky Mountain J. Math.* **25** 87–118.

[9] DAVIDSON, R. R. & FARQUHAR, P. H. (1976) A bibliography on the method of paired comparisons. *Biometrics*, **32** 241–252.

[10] FLIGNER, M. A. & VERDUCCI, J. S. (Eds.) (1993) *Probability Models and Statistical Analyses for Ranking Data: Lecture Notes in Statistics* **80**. Springer-Verlag.

[11] FORD, L. R., JR. (1957) Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly*, **64** 28–33.

[12] FUSCO, G. & HALE, J. K. (1989) Slow-motion manifolds, dormant instability, and singular perturbations. *J. Dynamics Diff. Eq.* **1** 75–94.

[13] GRANT, C. P. & VAN VLECK, E. S. (1995) Slowly-migrating transition layers for the discrete Allen–Cahn and Cahn–Hilliard equations. *Nonlinearity*, **8** 861–876.

[14] GUTIN, G. & YEO, A. (1996) Ranking the vertices of a complete multipartite paired comparison digraph. *Discrete Appl. Math.* **69** 75–82.

[15] HIRONAKA, H. (1975) Triangulations of algebraic sets. *Algebraic Geometry, Arcata 1974. Proceedings of Symposia in Pure Mathematics* **29**, pp. 165–185. AMS, Providence, RI.

[16] HORN, R. A. & JOHNSON, C. A. (1985) *Matrix Analysis.* Cambridge University Press.

[17] JECH, T. (1983) The ranking of incomplete tournaments: A mathematician's guide to popular sports. *Amer. Math. Monthly*, **90** 246–266.

[18] KANO, M. & SAKAMOTO, A. (1983) Ranking the vertices of a weighted digraph using the length of forward arcs. *Networks*, **13** 143–151.

[19] KANO, M. & SAKAMOTO, A. (1985) Ranking the vertices of a paired comparison digraph. *SIAM J. Algebraic Discrete Methods*, **6** 79–92.

[20] KEENER, J. P. (1993) The Perron–Frobenius theorem and the ranking of football teams. *SIAM Rev.* **35** 80–93.

[21] KESAVAN, S. (1989) *Topics in Functional Analysis and Applications.* John Wiley & Sons.

[22] KOHN, R. V. & STERNBERG, P. (1989) Local minimisers and singular perturbations. *Proc. Roy. Soc. Edinburgh Sect. A*, **111A** 69–84.

[23] LUCKHAUS, S. & MODICA, L. (1989) The Gibbs–Thompson relation within the gradient theory of phase transitions. *Arch. Rational Mech. Anal.* **107** 71–83.

[24] MASO, G. D. (1993) *Introduction to Γ-convergence. Progress in Nonlinear Differential Equations and Their Applications* **8**. Birkhäuser.

[25] MODICA, L. (1987) The gradient theory of phase transitions and the minimal interface criterion. *Arch. Rational Mech. Anal.* **98** 123–142.

[26] MODICA, L. (1987) Gradient theory of phase transitions with boundary contact energy. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **4** 487–512.

[27] ORTEGA, J. M. & RHEINBOLDT, W. C. (1970) *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press.

[28] RHEINBOLDT, W. C. (1974) *Methods for Solving Systems of Nonlinear Equations. Regional Conference Series in Applied Mathematics* **14**. SIAM, Philadelphia.

[29] RHEINBOLDT, W. C. (1986) *Numerical Analysis of Parametrized Nonlinear Equations. University of Arkansas Lecture Notes in the Mathematical Sciences* **7**. John Wiley & Sons.

[30] RUBINSTEIN, A. (1980) Ranking the participants in a tournament. *SIAM J. Appl. Math.* **38** 108–111.

[31] STERNBERG, P. (1988) The effect of a singular perturbation on nonconvex variational problems. *Arch. Rational Mech. Anal.* **101** 209–260.

[32] STOB, M. (1984) A supplement to "A mathematician's guide to popular sports". *Amer. Math. Monthly*, **91** 277–281.

[33] ZERMELO, E. (1929) Die Berechnung der Turnier-Ergebnisse als ein Maximum-problem der Wahrscheinlichkeitsrechnung. *Math. Z.* **29** 436–460.