How to rate numerical packages? Foundations of ratings, and their possible use in numerical mathematics, in education, and in military design

Raúl Antonio Trejo Ramírez[1] and Vladik Kreinovich[2]

[1]Sistemas de Informacion
Division de Ingeneria y Ciencias
ITESM (Instituto Technologico de Monterrey)
Campus Estado de Mexico
Apdo. Postal 2
Modulo de Servicio Postal
Atizapan, Mexico 52926
email rtrejo@campus.cem.itesm.mx

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968
email vladik@cs.utep.edu

Abstract. *One of the major human activities is solving problems. It is desirable to be able to predict whether a problem will be solved or not or, more precisely, the probability that a person can solve a problem. This probability can be estimated based on previous experience of people solving similar problems.*

Then, it is necessary to somehow describe the ability of the problem-solvers as well as the difficulty of the problems. This can be done by assigning to each problem or problem solver a rating, *i.e., a number that represents the skill of the problem solver or the difficulty of the problem.*

In this paper, we propose to use a rating system developed in sports to other real-life problems, including rating of numerical packages.

1 Introduction

Rating is important. One of the major human activities is solving problems. Students solve problems all the time; professionals (unless they are doing routine work) solve problems most of the time; even in recreation, e.g., in sports, each team is trying to solve a problem: how to win. The very word "problem" means that we are not talking about a routine task, we have a creative problem, so it is difficult to predict exactly whether it will be solved and how fast it will be solved.

However, it would be nice to be able to estimate the difficulty of the problems and the skill of the problem solvers so that they would not waste their time trying to solve a problem that is too hard, and they do not waste their professional expertise by solving problems that are too easy.

So, we need to somehow *rate* problems and problem solvers. Usually, the only information that we can use for such a rating is the past experience of solving these problems, i.e., who has been able to solve which problem.

Rating of numerical packages and algorithms is especially important. This problem is especially important for numerical packages and algorithms where so far, comparison has been very subjective, based on the ability of different methods to solve benchmark problems.

Rating in other real-life situations. This rating problem appears in other real-life situations: in education, in ergonomics, etc., but the only area where it has been successfully solved is competitive sports. It has been especially successful in chess, where a rather complicated mathematical formalism, called *Elo's rating method*, has been developed. Elo's rating has been in use for 30 years [6], [8].

What we are planning to do. It seems natural to use Elo's rating in education and in other applications.

The main drawback of this idea is that Elo's rating is based on semi-heuristic ideas, and it is not clear that these same ideas can be applied to other rating situations. To overcome this drawback, we formulate the rating problem in mathematical terms and solve this problem. As a result, we get a mathematical justification of Elo's rating, and thus, a good reason to propose this method for rating numerical packages.

Some of the material of this paper was first presented in a thesis [15].

2 Chess as a Good Working Example of Rating

Rating is important in many real-life situations. Prediction problem is important in many different areas: in education, in a workplace environment, etc. It would thus be natural to rate students, workers, etc.

Rating in workplace environment is not pure rating. For example, rating workers and marking some of them as inferior or less capable does not help, because rating encourages competition and discourages cooperation, and cooperation is vital for a workplace. So, although some rating and evaluation methods are applied in many companies, their goal is not so much to predict the behavior, but mainly to help the workers overcome their weak points and become better problem-solvers. For these applications, the main result of an evaluation is not the rating itself, but so to say, the list of deficiencies of a given worker that have to be overcome.

Rating in education is also not pure rating. Similarly, in education: the main goal of the existing evaluation systems is not to condemn or praise a student for being bad or good, but to give him or her the idea of what he or she needs to improve.

Summarizing. In all these areas, the problem is much more complicated than the one that we initially wanted to attack, and it is no wonder that in all these areas, no ultimate solutions of the complicated rating system are known.

Rating in sports. Sports are different. Here, the clear goal is exactly to rate teams, players, etc. Finding the drawbacks that can be improved in a player or team is also important, but it only has a secondary value, relevant mainly for the team itself. For the outside world, rating is all. An "almost win" is as bad as a plain loss. So, it is no wonder that mathematical rating methods have been first developed for sports.

In most sports, rating is relatively easy. In many sports, rating is based on an objective criterion: the time to run 100 meters, the height of the jump, etc. [9]. For these sports, rating is not a problem. It becomes a problem only when we consider competitive sports, in which the players play against each other.

In the majority of these sports, a stronger player usually wins. For example, the Olympic boxing or soccer champions are determined by the so-called *Olympic system*, in which once you've lost, this means you are weak, and then you're out. Sometimes, this system is not fair but mainly it is. In this case, rating is also not a problem, and prediction based on that rating is easy: a player or a team who is placed higher usually wins.

Sports where rating is a problem. Rating is a problem only for sports in which strengths are so close that no one expects a world champion to win in all his games. The most well known of such games is *chess*. So, it is natural that chess has long been a training field for various rating methodologies.

We need a method that would, given the results of several players, rate them, and predict how they will play in the future.

We have already mentioned that in chess, even if we rated a person A as the best (world champion), and B as not the best, we cannot guarantee that A will always defeat B. We can only expect that in the majority of the cases, he will. The bigger the difference between A and B, the more frequent the wins are. So the only thing that we can potentially predict is a probability of a player A winning over the player B.

Rating in chess. In 1959, a Hungarian mathematician Arpad E. Elo proposed a new rating method [5, 6] that has been used in chess for a few decades [8, 9]. This method is, crudely speaking, based on the fact that to any player A we can assign a number $r(A)$ in such a way that the probability that a player A defeats another player B is given by

$$p(A > B) \approx \frac{1}{1 + \exp(-k(r(A) - r(B))} \tag{1}$$

for some constant k [13].

We would like to use chess rating in other real-life problems. It is desirable to use this idea for rating numerical packages, problems, etc., but Elo's rating system is based on the semi-empirical formula (1) that is only justified for chess. Therefore, in order to use this formula for more general rating situations, we must find a more fundamental justification for it.

3 Rating as a Mathematical Problem

3.1 Choosing a Probability Function. Informal Motivation

3.1.1 Relativity of Players' Strength

Our goal is to describe a formula that assigns to every two players A and B, a probability $p(A > B)$ that A will win over B. To be more precise, we must be able to somehow describe a *strength* $s(A)$ of each player, and then to predict $p(A > B)$ as a function of $s(A)$ and $s(B)$.

Players are of different strength. It is believed that "strength" is relative: no matter how strong a player is, there can potentially be players that are stronger. There are several known ways to assign strength. For example, we can take one of the ratings described in Section 2 as strength. So, let's assume that some method of determining strength is fixed.

This situation is similar to, say, relative height: a person who is 6 ft. tall is taller than a one who is 5 ft. tall, but this is a relative advantage: there are people who are even taller.

For tall people, it is easy to describe this "relativity" (i.e., the fact that the world looks more or less the same from the viewpoint of people of different height). Indeed, from the viewpoint of a 5 ft. tall person, the world consists of people of different height, and he himself corresponds to the value of 5. from the viewpoint of a 6 ft. tall person, the chosen value is different (it is not 5), but this difference is relative: if we used a new a different unit of length (e.g., 1 new foot = 6/5 old feet), then, in these new units, the 6 ft. person will be assigned a numerical value 5, and the whole world will look exactly the same as it looked for a 5 ft. tall guy.

In other words, although at first glance, the numerical pictures are different, we can *rescale* the values of height, and make them identical from a new viewpoint. For height, this rescaling $h_{new} = f(h_{old})$ can be described by a simple formula:

$$h_{new} = (5/6)h_{old}.$$

Consider this situation for our chess players problem. We can choose a player with strength $s(A)$ and consider a world where there exist players stronger than A and players weaker than A. But we can also choose a player B with strength $s(B)$. As in the height example, this world with player B as our reference point is similar to the world where A is the reference point: there are stronger and weaker players than B.

Let's use the notation $a = s(A)$, $b = s(B)$. As in the previous example, for every a and b there exists a transformation function $f_{a\to b}$ from real numbers to real numbers such that

$$f_{a\to b}(a) = b$$

and that preserves the probabilities of winning.

3.1.2 How Strength is Related to Ratings

Remember that the probability function we want must predict probability as a function of the strength of the players, this is

$$p(A > B) = F(a, b),$$

where $a = s(A)$ and $b = s(B)$.

We must consider that $p(A > B)$ should remain the same no matter how the strength of the players is measured; this is, for any two players with strengths x, y

$$F(x, y) = F(f_{a\to b}(x), f_{a\to b}(y)).$$

So, we impose the following requirements on the function $F : R \times R \to R$:

(R1) *For every* a *and* b*, there exists a monotonic function* $f_{a\to b}$ *for which* $f_{a\to b}(a) = b$ *and* $F(x,y) = F(f_{a\to b}(x), f_{a\to b}(y))$ *for all* x *and* y*. The function* a, b, $x \to f_{a\to b}(x)$ *is continuous in all three variables* a, b, *and* x.

It is also reasonable to conclude that for each a and b, there is only one rescaling function that transforms a into b, this is:

(R2) *For every* a *and* b*, if* $f_{x\to y}(a) = b$*, then, the functions* $f_{x\to y}$ *and* $f_{a\to b}$ *coincide:* $f_{x\to y} = f_{a\to b}$.

Suppose that we have three numbers a, b, and c. Then, a transformation $f_{a\to b}$ transforms an a-viewpoint into a b-viewpoint, and $f_{b\to c}$ transforms a b-viewpoint into a c-viewpoint. so, to transform an a-viewpoint into a c-viewpoint, we must first apply $f_{a\to b}$ and get a b-viewpoint, and then apply $f_{b\to c}$ and get a c-viewpoint. The resulting transformation is

$$x \to f_{a\to b}(x) \to f_{b\to c}(f_{a\to b}(x)).$$

On the other hand, we could as well transform an a-viewpoint directly into a c-viewpoint by applying a function $f_{a\to c} : x \to f_{a\to c}(x)$. In view of our requirement **(R2)**, it is natural to require that these two transformations coincide:

(R3) *For any three numbers,* a, b, *and* c,

$$f_{a\to c}(x) = f_{b\to c}(f_{a\to b}(x)).$$

The last requirement is that to get an a-viewpoint from itself, we do not need to do anything, this is:

(R4) *For every number* a, $f_{a\to a}(x) = x$.

We want to prove that under these conditions, instead of $s(A)$, we can use a different rating $r(A)$, for which the probability of winning depends only on the difference between the ratings.

3.2 First Result: Probability Can Be Expressed In Terms of the Difference in Ratings

Definition 1. *By a winning probability function, we mean a function* $F(x,y)$ *of two real variables, that satisfies the following four conditions:*
(R1)*: For every* a *and* b*, there exists a monotonic function* $f_{a\to b}$ *for which* $f_{a\to b}(a) = b$ *and*

$$F(x,y) = F(f_{a\to b}(x), f_{a\to b}(y)) \tag{2}$$

for all x *and* y*. The function* a, b, $x \to f_{a\to b}(x)$ *is continuous in all three variables.*

(R2): *For every* a *and* b, *if* $f_{x\to y}(a) = b$, *then, the functions* $f_{x\to y}$ *and* $f_{a\to b}$ *coincide:* $f_{x\to y} = f_{a\to b}$.
(R3): *For any three numbers,* a, b, *and* c,

$$f_{a\to c}(x) = f_{b\to c}(f_{a\to b}(x)).$$

(R4): *For every number* a, $f_{a\to a}(x) = x$.

Suppose that a finite set P *is given. Its elements will be called players. Suppose also that a function* $s : P \to R$ *is given. This function will be called a rating. By a winning probability, we mean a function* $A, B \to p(A > B)$, *that is defined for all* $A, B \in P$ *and that is determined by a formula* $p(A > B) = F(s(A), s(B))$, *where* F *is a winning probability function.*

THEOREM 1. *For every winning probability function, there exist monotonic functions* $\psi : R \to R$ *and* $f : R \to R$, *such that* $p(A > B) = f(r(A) - r(B))$, *where* $r(A) = \psi(s(A))$.

Proof.
1^o. First, let us prove that any transformation $f_{a\to b}(x)$ can be described as $f_{0\to z}(x)$ for some z.

Indeed, the transformation $f_{a\to b}(x)$ transforms 0 into $z = f_{a\to b}(0)$. We assumed in **(R2)** that there is only one transformation that gets 0 into z, therefore

$$f_{a\to b} = f_{0\to z}.$$

2^o. Let us now prove that for every two functions $f_{0\to a}(x)$ and $f_{0\to b}(x)$, their composition $f_{0\to a}(f_{0\to b}(x))$ can also be described in this way, as $f_{0\to c}$ for some c.

Indeed, the transformation $f_{0\to a}$ transforms b into a point $f_{0\to a}(b)$. Lets denote this point by c. Then, due to **(R2)**, $f_{0\to a}(x) = f_{b\to c}(x)$. Hence,

$$f_{0\to a}(f_{0\to b}(x)) = f_{b\to c}((f_{0\to b}(x)).$$

Due to **(R3)**, the right-hand side of this equation is equal to $f_{0\to c}(x)$.

So, the set $\mathcal{S}$ of all transformations $f_{0\to a}$ with $a \in R$, is closed under composition.

3^o. Let us now prove that for every a, this set contains an inverse element, i.e., there exists a z such that $f_{0\to z}(f_{0\to a}(x)) = x$.

Indeed, according to part 1^o; $f_{a\to 0}(x) = f_{0\to z}(x)$ for some $z(= f_{a\to 0}(0))$. Then,

$$f_{0\to z}(f_{0\to a}(x)) = f_{a\to 0}((f_{0\to a}(x)).$$

Due to **(R3)**, $f_{a\to 0}((f_{0\to a}(x)) = f_{a\to a}(x)$, and due to **(R4)**, this expression is equal to x. So, $f_{0\to z}(f_{0\to a}(x)) = x$.

4^o. So, the set $\mathcal{S}$ with a composition operation forms a one-dimensional group. Composition is expressed by $a * b = c = f_{0\to b}(a)$, and, since we assumed that f is continuous, the composition operation is also continuous.

It is known that every one-dimensional group with a continuous operation that is defined on a set R of all real numbers is isomorphic to the group of real numbers with addition (see, e.g., [1]). In other words, there exists a one to one function ψ such that

$$\psi(a * b) = \psi(a) + \psi(b)$$

for all a and b.

5^o. Let us define $r(A) = \psi(s(A))$, and use condition **(R1)**:

$$F(x, y) = F(f_{a\to b}(x), f_{a\to b}(y)).$$

In particular, for $a = 0$, we get

$$F(x, y) = F(f_{0\to b}(x), f_{0\to b}(y)).$$

Now, by definition of $f_{a\to b}$, we have $x = f_{0\to x}(0)$, therefore

$$f_{0\to b}(x) = f_{0\to b}(f_{0\to x}(0)).$$

From our definition of a composition, we conclude that

$$f_{0\to b}(f_{0\to x}(0)) = f_{0\to b*x}(0),$$

hence,

$$f_{0\to b}(x) = f_{0\to b*x}(0).$$

Again, from the definition of $f_{a\to b}$ (condition **(R1)**) we conclude that

$$f_{0\to b*x}(0) = b * x,$$

hence,

$$f_{0\to b}(x) = b * x.$$

Similarly, $h_{0\to b}(y) = b * y$. Therefore, equation (2) from **(R1)** takes the form

$$F(x, y) = F(b * x, b * y). \tag{3}$$

To use this equation, let us introduce an auxiliary function

$$G(x, y) \stackrel{\text{def}}{=} F(\psi^{-1}(x), \psi^{-1}(y)).$$

In this case,

$$F(x, y) = G(\psi(x), \psi(y)). \tag{4}$$

In terms of G, the above equation takes the following form:

$$G(\psi(x), \psi(y)) = G(\psi(b * x), \psi(b * y)).$$

But $\psi(b*x) = \psi(b) + \psi(x)$, and $\psi(b*y) = \psi(b) + \psi(y)$. So,

$$G(\psi(x), \psi(y)) = G(\psi(b) + \psi(x), \psi(b) + \psi(y))$$

for all x, y, and b.

The function ψ is a one to one transformation. So, for each X, Y, and B, there exists x, y, and b for which $\psi(x) = X$, $\psi(y) = Y$, and $\psi(b) = B$. for these values, we get

$$G(X, Y) = G(X + B, Y + B).$$

This is true for all B, in particular, for $B = -Y$. So

$$G(X, Y) = G(X - Y, 0).$$

So, if we denote $G(z, 0)$ by $f(z)$, we get $G(X, Y) = f(X - Y)$, and (4) becomes:

$$F(x, y) = f(\psi(x) - \psi(y)).$$

We assumed that $p(A > B) = F(s(A), s(B))$. Therefore

$$p(A > B) = F(s(A), s(B)) = f(\psi(s(A)) - \psi(s(B))).$$

since we denoted $r(A) = \psi(s(A))$, we get the desired formula

$$p(A > B) = f(r(A) - r(B)).$$

Q.E.D.

3.3 How to Choose a Dependency? Informal Motivations

We already know that the probability $p(A > B)$ can be described as a function of the difference of ratings between players A and B:

$$p(A > B) = f(r(A) - r(B)).$$

We do not know f, so, one way to rate players is to fix a player B_0, and use the probability $p(A > B_0)$ as the desired rating. The choice of B_0 is arbitrary, so if we choose another player B_1, we get a different rating $A \to p(A > B_1)$. These two scales $A \to p(A > B_0)$ and $A \to p(A > B_1)$ must be connected by a "reasonable" rescaling $\varphi : p(A > B_1) = \varphi(p(A > B_0))$. By "reasonable" we mean that this rescaling must belong to a general family $\mathcal{F}$ of "reasonable" rescaling transformations.

Let us now describe the properties that this family $\mathcal{F}$ of functions must satisfy.

- If $p(A, B_1) \to \varphi(p(A > B_2))$ is a transformation of the rating of a player with respect to player B_1 to the rating with respect to player B_2, and

$$p(A > B_2) \to \rho(p(A > B_3))$$

is a transformation of the rating of player A with respect to B_2 to the rating of A with respect to B_3, then if we want to transform $p(A > B_3)$ directly to $p(A > B_1)$, we demand that this composite transformation

$$p(A > B_1) \to \varphi(\rho(p(A > B_3)))$$

also belongs to $\mathcal{F}$.

- If $p(A > B_1) \to \varphi(p(A > B_2))$ is a transformation of the probability of winning of player A over player B_1 to the probability of winning of A over B_2, where φ belongs to $\mathcal{F}$, then an inverse function $\varphi^{-1}(p)$ that transforms probability of winning of player A over player B_2 to the probability of winning of A over B_1 must also belong to $\mathcal{F}$.

Comments:

1. These demands mean that the family $\mathcal{F}$ must be closed under *composition* and *inverse function*, or in mathematical terms, that $\mathcal{F}$ must be a *transformation group*.

2. We want transformation functions to be actually used in a computer, so we must have a finite number of parameters for the function. Then $\mathcal{F}$ is a *finite dimensional transformation group*.

- We also demand rescaling functions from $\mathcal{F}$ to be smooth. If we have a small change in ratings of the fixed players, we want a small change in winning probability. A smooth finite dimensional transformation group is called a *Lie group* (see [3]).

- There are different circumstances in chess tournaments that influence the probability value $p(A > B)$. A player may do his best to win in normal play. However, there are situations in which the outcome of a game is the result of external causes, and not the result of the player's strength. For example, if a game is not played because of the absence of one of the players, the other player is declared the winner. let's see how these forfeited games affect the probability values.

  - Suppose that player A is able to win M games out of a total of N, then:

    $$p(A > B) = M/N.$$

  - A second possibility is that A misses N_1 games, so his winning probability considering these extra N_1 games is

    $$p'(A > B) = \frac{M}{N + N_1} = c \cdot p(A > B),$$

where $c = M/(N + N_1)$.

- A third possibility is that player the B is the one who misses N_1 games. Then, the winning probability for A is given by:

$$p''(A > B) = \frac{M + N_1}{N + N_1} = \frac{N \cdot p(A > B) + N_1}{N + N_1}$$
$$= a \cdot p(A > B) + b,$$

for $a = N/(N + N_1)$ and $b = N_1/(N + N_1)$.

If we need a function transforming the values obtained without the games in which the players did not do their best, then the corresponding transformation will be a linear function $p \to a + bp$, where a and b are constants (independent of p). Thus we get a two parametric family of linear functions.

We need to make a final observation about the relation between the difference of ratings $r(A) - r(B)$ and the value that function f returns. If player A is a much better player than B, i.e., $r(A) - r(B)$ is a big positive number, it is expected that A will win all the time, so $p(A > B) = 1$. Also, if B is much better than A (so the difference $r(A) - r(B)$ is a big negative number), it is expected that B will win all the time, so $p(A > B) = 0$. Finally, if both players are of the same strength (i.e., $r(A) - r(B) = 0$), it is expected that any player wins half of the games, this is, $p(A > B) = 0.5$.

3.4 The Problem of Choosing a Probability Function: Formulation and Solution

Definition 2. *By a set of reasonable transformations, we mean a finite-dimensional Lie group that contains all linear functions.*

Definition 3. *We say that a monotonic function $f : R \to R$ is rescalable, if for every r_1, $r_2 \in R$, there exists a reasonable transformation φ such that*

$$f(s - r_1) = \varphi(f(s - r_2))$$

for all s.

Definition 4. *We say that a function $f : R \to R$ is asymptotically correct if the following conditions are true:*

- $f(x) \to 0$ *if* $x \to -\infty$,
- $f(x) \to 1$ *if* $x \to \infty$,
- $f(0) = 0.5$

Definition 5. *Suppose that a finite set P is given. Its elements will be called players. Suppose also that a function $r : P \to R$ is given. This function will be called a rating. By a winning probability, we mean a function $A, B \to p(A > B)$, that is defined for all $A, B \in P$ and that is determined by a formula $p(A > B) = f(r(A) - r(B))$, where f is a rescalable asymptotically correct function.*

THEOREM 2. *A winning probability is given by:*

$$p(A > B) = \frac{1}{1 + \exp(-k(r(A) - r(B)))}.$$

Proof: First, we prove that the group of transformations $\mathcal{F}$ consists of fractionally-linear functions; then, we prove that any function f from $\mathcal{F}$ satisfies some functional equation, whose solution is known.

1^o. By definition of $\mathcal{F}$, it is a finite-dimensional Lie group of transformations of the set of real numbers onto itself that includes all linear transformations. The problem of classifying all possible transformation groups of an n-dimensional space $R^n, n = 1, 2, 3, \ldots$, was first formulated by Norbert Wiener (see, e.g., [16]). His hypothesis was confirmed in [7] by Guillemin and Sternberg and in [14] by Singer and Sternberg.

In our case (when $n = 1$), the only possible groups are the groups of all linear transformations and the group of all fractionally-linear transformations

$$x \to \frac{a + bx}{c + dx}.$$

Since linear transformations are a particular case of fractionally linear ones, in both cases $\mathcal{F}$ consists of fractionally linear transformations.

The simplified proof for the one dimensional case is given in [12]. This result has also been used in [11] and [10].

2^o. The fact that f is rescalable, means that

$$f(a - b_2) = \varphi(f(a - b_1)).$$

In particular, for $b_1 = 0$ and $b_2 = b$, we get

$$f(a - b) = \varphi(f(a)).$$

Since $\varphi \in \mathcal{F}$, φ is fractionally linear, hence

$$f(s - r) = \frac{a + b \cdot f(s)}{c + d \cdot f(s)}$$

for some a, b, c, d that depend on r. So we arrive at a functional equation for f. Taking into consideration that fractionally linear transformations are projective transformations of a line, for which the cross ratio is preserved (see [1]), we can write, for $\rho(s) = (a + bf(s))/(c + df(s))$:

$$\frac{\rho(s_1) - \rho(s_3)}{\rho(s_2) - \rho(s_3)} \frac{\rho(s_2) - \rho(s_4)}{\rho(s_1) - \rho(s_4)}$$

$$= \frac{f(s_1) - f(s_3)}{f(s_2) - f(s_3)} \frac{f(s_2) - f(s_4)}{f(s_1) - f(s_4)}$$

for all s_i. For our purposes, this is true for $\rho(s) = f(s - r)$, therefore the following equality is true:

$$\frac{f(s_1 - r) - f(s_3 - r)}{f(s_2 - r) - f(s_3 - r)} \frac{f(s_2 - r) - f(s_4 - r)}{f(s_1 - r) - f(s_4 - r)} =$$

$$= \frac{f(s_1) - f(s_3)}{f(s_2) - f(s_3)} \frac{f(s_2) - f(s_4)}{f(s_1) - f(s_4)}$$

The most general continuous solutions of this functional equation are given by [1]:

$$f(y) = \frac{a + by}{c + dy} \tag{5}$$

$$f(y) = \frac{a + b\tan(Ky)}{c + d\tan(Ky)} \tag{6}$$

$$f(y) = \frac{a + b\tanh(Ky)}{c + d\tanh(Ky)} \tag{7}$$

for some K.

Let's show that cases (5) and (6) are impossible. In case (5), if $d = 0$, then

$$f(y) = (a/c) + (b/c)(y) = A + By$$

and $f(y) \to \infty$ as $y \to \infty$, which contradicts to our Definition 4. If $d \neq 0$, then $f(y)$ is not monotonic, because it takes an infinite value when $y = -c/d$. In similar way, in case of (6), $f(y) = \infty$ when $y = \pi/(2K)$.

So, we can conclude that $f(y)$ is described by formula (7). If $K = 0$, $f(y)$ is a constant and we assumed that $f(-\infty) \neq f(\infty)$. So K must be different from zero. If $K < 0$, then we can use the fact that $\tanh(Ky) = -\tanh(-Ky)$ and get an expression

$$f(y) = \frac{a - b \cdot \tanh(-Ky)}{c - d \cdot \tanh(-Ky)}$$

with the coefficient > 0. Therefore, in the following we can assume that $K > 0$. Multiplying both the numerator and the denominator by $\cosh(Ky)$, we have

$$f(y) = \frac{(\cosh(Ky))(a + b\tanh(Ky))}{(\cosh(Ky))(c + d\tanh(Ky))} =$$

$$\frac{a\cosh(Ky) + b\cosh(Ky)(\sinh(Ky)/\cosh(Ky))}{c\cosh(Ky) + d\cosh(Ky)(\sinh(Ky)/\cosh(Ky))} =$$
$$\frac{a\cosh(Ky) + b\sinh(Ky)}{c\cosh(Ky) + d\sinh(Ky)}.$$

Let's recall that:

$$\sinh(x) = \frac{\exp(x) - \exp(-x)}{2}\ ,\ \cosh(x) = \frac{\exp(x) + \exp(-x)}{2}.$$

We then substitute the expressions for sinh and cosh:

$$\begin{aligned} f(y) &= \frac{(a/2)(\exp(Ky) + \exp(-Ky)) + (b/2)(\exp(Ky) - \exp(-Ky))}{(c/2)(\exp(Ky) + \exp(-Ky)) + (d/2)(\exp(Ky) - \exp(-Ky))} = \\ &\frac{a\exp(Ky) + a\exp(-Ky) + b\exp(Ky) - b\exp(-Ky)}{c\exp(Ky) + c\exp(-Ky) + d\exp(Ky) - d\exp(-Ky)} = \\ &\frac{(a+b)\exp(Ky) + (a-b)\exp(-Ky)}{(c+d)\exp(Ky) + (c-d)\exp(-Ky)}, \end{aligned}$$

and conclude that:

$$f(y) = \frac{A\exp(Ky) + B\exp(-Ky)}{C\exp(Ky) + D\exp(-Ky)}$$

for $A = (a+b),\ B = (a-b),\ C = (c+d),\ D = (c-d)$. Multiplying both numerator and denominator by $\exp(-Ky)$, we arrive at

$$f(y) = \frac{A + B\exp(-2Ky)}{C + D\exp(-2Ky)}.$$

If we denote $k = 2K$, we get

$$f(y) = \frac{A + B\exp(-ky)}{C + D\exp(-ky)}. \tag{8}$$

We can use the conditions in Definition 4 to determine the value of constants A, B, C, and D in equation(8). If $y \to \infty$, then $f(y) = A/C$. According to definition $f(y) \to 1$ when $y \to \infty$, hence $A/C = 1$ and therefore $A = C$. If $y \to -\infty$ then $f(y) \to B/D$. Our definition has $f(y) \to 0$ for $y \to -\infty$, thus $B/D = 0$ and we conclude that $B = 0$. So equation (8) becomes

$$f(y) = \frac{A}{A + D\exp(-ky)}.$$

The last condition of Definition 4 has $f(y) = 1/2$ when $y = 0$. In this equation, $f(y) = A/(A+D)$ for $y = 0$. So $A/(A+D)$ must be equal to $1/2$ and therefore, we can conclude that $D = A$. If we then divide both the numerator and the denominator of the equation by A, we get

$$\frac{1}{1 + \exp(-ky)},$$

which is the desired expression. Q.E.D.

4 A Rating Algorithm

The idea: Maximum Likelihood Method. We have justified the formula (1). Therefore, we know the family of probability distributions that describe the outcomes of the games. The ratings $r(A)$ are parameters of these distributions.

So, to determine these parameters, we can use a known statistical method of determining the parameters: choose the parameters that are more likely, i.e., for which the probability is the highest possible.

This method is called a *maximum likelihood method* [4]. Its general description is as follows: Suppose we have some samples $X_1, X_2, \ldots, X_n$ of an event X with a probability distribution f. The problem is to determine the parameters $\vec{\theta}$ of distribution f.

If $\vec{\theta}$ is known, the probability of the sequence $X_1, \ldots, X_n$ can be obtained as

$$\mathcal{L}(X_1, \ldots, X_n, \vec{\theta}) = f(X_1, \vec{\theta}) f(X_2, \vec{\theta}) \cdot \ \ldots \ \cdot f(X_n, \vec{\theta}).$$

We choose an estimate that maximizes this probability:

$$\mathcal{L}(X_1, \ldots, X_n, \vec{\theta}) \to \max_{\vec{\theta}}. \tag{9}$$

Since $x \to \ln x$ is a monotonic function, and $\ln \mathcal{L}$ attains its maximum when $\mathcal{L}$ is a maximum; then (9) is equivalent to

$$\ln \mathcal{L}(X_1, \cdots, X_n, \vec{\theta}) \to \max_{\vec{\theta}}.$$

And max is known to be attained when

$$\frac{\partial}{\partial \theta_j} \ln \mathcal{L} = \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \ln f(X_i, \vec{\theta}) = 0 \tag{10}$$

for all j.

Rating of a new player. Let's describe how this method can be used to determine a rating of a player A, who has just played with players of known ratings. If A has won w games against players $X_1, \ldots, X_w$, and lost l games against players $Y_1, \ldots, Y_l$, then the probability of this outcome is equal to

$$p(A > X_1) \cdot \ \ldots, \cdot p(A > X_w) \cdot p(Y_1 > A) \cdot \ \ldots \ \cdot p(Y_l > A).$$

So, according to maximum likelihood method, we must determine $r(A)$ from the equation

$$p(A > X_1) \cdot \ \ldots \ \cdot p(A > X_w) \cdot p(Y_1 > A) \cdot \ \ldots \ \cdot p(Y_l > A) \to \max_{r(A)} \tag{11}$$

Substituting (1) for $p(A > B)$, differentiating w.r.t. $r(A)$, and dividing the result of this differentiation by k, we get the following equation with the only unknown $r(A)$:

$$\sum_{j=1}^{l} \frac{\exp(k(r(A) - r_o(Y_j)))}{1 + \exp(k(r(A) - r_o(Y_j)))} - \sum_{i=1}^{w} \frac{\exp(k(r_o(X_i) - r(A)))}{1 + \exp(k(r_o(X_i) - r(A)))} = 0. \quad (12)$$

This equation can be solved using standard numerical methods for finding equation roots, e.g., by bisection (see, e.g., [2, 15]).

Initial rating. If none of the players has been rated yet, an iterative method can be used to assign ratings to these players:

1. Assign the same initial ratings $r_o(A_i)$ to all players.
2. For each player, calculate a new rating $r(A_i)$ using (12).
3. If for all i, the values $r_o(A_i)$ and $r(A_i)$ are close enough, stop; else, assign $r_o(A_i) := r(A_i)$ and perform a new iteration.

5 Application of a Rating Method to Real-life Situations

A rating system can be applied to a variety of real life problems. Let us present some examples.

Education. In the case of education, when grading students, we don't want to use tests that are too easy, because everybody will answer such a test correctly and we will not be able to tell real student's skill. In a similar way, a test that is too hard is also of no help in determining students' abilities. So, a test must be made up of problems of different levels of difficulty. The problem is, then, to predict the students' performance, i.e., predict the probability $p(S > P)$ of a student S solving a problem P. The motivations that we used to deduce Elo's formula for chess rating did not use any chess specifics. Therefore, we can repeat these arguments for students and problems, and conclude that it is possible to assign to every student his/her rating $r(S_1), \ldots, r(S_n)$, and to each problem its complexity $c(P_1), \ldots, c(P_n)$, such that this probability $p(S > P)$ can be estimated as:

$$p(S_i > P_j) = \frac{1}{1 + \exp(-k(r(S_i) - c(P_j))}.$$

Rating for students and problems can be determined by an iterative method, using the maximum likelihood method.

Rating of Numerical Packages. Similarly, we can rate numerical packages and benchmark problems.

In solving these problems, we may be interested not only in whether a method solves a problem, but also in what *time* it took. To take that time into consideration, we can consider pairs (numerical problem,time) as problems, and assume that a package solves this problem within the given time.

Military CAD. Computer Aided Design is another application where a rating method may be useful.

Consider, for example, the design of a military combat machine (tank, plane, etc.). The problem here is to predict the probability of a fighter machine winning over another. This can be done by running battle simulations on designs that are already been used. For each pair of designs, several simulations have to be done in order to estimate the probability. If all designs have to be tested against each other, we need $n(n-1)/2$ "tournaments", or set of simulations, for n designs.

Instead of doing that, it is possible to assign to each design a rating $r(D_i)$, such that the probability $p(D_i > D_j)$ of design D_i winning over D_j is given by:

$$p(D_i > D_j) = \frac{1}{1 + \exp(-k(r(D_i) - r(D_j))}.$$

Algorithm:

- Run simulations of a fixed design against the rest. Only $n - 1$ sets of simulations are needed.
- Assign ratings to all designs using an iterative method.
- The probability of any design winning over another can be predicted using ratings.

Acknowledgments. This work was partially supported by NSF Grant No. EEC-9322370, by NASA Grant No. NAG 9-757, and by the US Agency for International Development. We are thankful to Chitta Baral and Mohamed Amine Khamsi for valuable discussions and comments.

References

[1] J. Aczél, *Lectures on Functional Equations and their Applications*, Academic Press, N.Y.-London, 1966.

[2] S. C. Chapra and R. P. Canale, *Metodos Numericos para Ingenieros*, McGraw Hill, Mexico, 1988.

[3] C. Chevalley, *Theory of Lie Groups*, Princeton University Press, Princeton, NJ, 1946.

[4] R. Deutsch, *Estimation Theory*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1965.

[5] A. E. Elo, "New USCF rating system", *Chess Life*, 1967.

[6] A. E. Elo, *The Rating of Chessplayers, Past and Present*, Arco Publishing, Inc., N.Y., 1978; 2nd edition, 1986.

[7] V. M. Guillemin and S. Sternberg, "An algebraic model of transitive differential geometry", *Bulletin of American Mathematical Society*, 1964, Vol. 70, No. 1, pp. 16–47.

[8] K. Harkness, *Official Chess Handbook*, David McKay Co., N.Y., 1973.

[9] D. Hooper and K. Whyld, *The Oxford Companion to Chess*, Oxford University Press, N.Y., 1984.

[10] V. Kreinovich and S. Kumar. "How to help intelligent systems with different uncertainty representations communicate with each other," *Cybernetics and Systems: International Journal*, 1991, Vol. 22, No. 2, pp. 217–222.

[11] V. Kreinovich and C. Quintana. "Neural networks: what non- linearity to choose?," *Proceedings of the 4th University of New Brunswick Artificial Intelligence Workshop*, Fredericton, N.B., Canada, 1991, pp. 627–637.

[12] I. N. Krotkov, V. Kreinovich and V. D. Mazin. "General form of measurement transformations which admit the computational methods of metrological analysis of measuring- testing and measuring-computing systems," *Measurement Techniques*, 1987, Vol. 30, No. 10, pp. 936–939.

[13] L. E. Sadovskii and A. L. Sadovskii, *Mathematics and Sports*, American Mathematical Society, Providence, RI, 1993.

[14] I. M. Singer and S. Sternberg, "Infinite groups of Lie and Cartan. Part 1", *Journal d'Analyse Mathematique*, 1965, Vol. 15, pp. 1–113.

[15] R. A. Trejo Ramírez, *An improved rating system: its foundations and computational problems*, Master Thesis, University of Texas at El Paso, Department of Computer Science, 1994.

[16] N. Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, MA, 1962.