Fargo Ratings

A New Approach to Rating and Handicapping Players in the Game of Eight Ball

Michael Page

Fargo Billiards & Gastropub
3234 43rd Street South
Fargo, North Dakota 58104

March, 2010

I. Overview

I describe here a system for rating players in the game of 8-ball that has been piloted 2009/2010 in the Flatland Pool League at Fargo Billiards & Gastropub. Our pilot has been going for 7 months, during which time we have 324 players who between them have played over 12,500 games of 8-ball. The scheme is similar to the NPL (National Pool League) approach created by Bob Jewett, which in turn was based on the approach Arpad Elo developed for rating players in chess. Rudiments of the scheme were described in the October 2002 issue of Billiards Digest magazine (*Sizing Up With The Pros – A new rating system could rank all players at all levels, by Michael Page).* Advantages of this approach include

- Approach is mathematically sound
- Ultimately rate 8-ball players worldwide on the same scale
- No secret formulas—approach is transparent and above board
- Use information from every single game to update ratings
- Naturally resistant to sandbagging; sandbagging easy to detect and prevent
- No cumbersome inning counts required
- Lends itself to innovative easy-to-implement handicapped round-robin team league play
- Leads easily to fully or partially handicapped tournament design
- Easy to implement league-division team skill ceilings, similar to the APA 23 rule.

II. In a Nutshell

Any player in the system has at any given time a **Fargo Rating** as well as a measure of the ***robustness*** of that rating, which is an effective number of games upon which that rating is based. For the pilot group of 324 players, the Fargo Ratings range from 234 to 707. The mean rating is 423 and the median rating is 406. Fargo ratings have the following approximate correspondence

- 800 — A top world-class 8-ball player such as Corey Duell, Shane Van Boening, Darren Appleton, or Thorsten Hohman
- 700 — A top regional 8-ball player in the US –a threat to cash in the Master's Division at the BCA/VNEA Championships – a threat to run six in a row if the break is working
- 600 — Likely to cash in the BCA Open Division but probably won't make it to the top 32. –may get moved to Master's Division and then flounder –has run three-in-a-row multiple times and maybe four-in-a row a time or two

- 500 A good local league player. Runs out first time at the table in about 10% of the games

- 400 Runs out first time at the table in about 1% of the games—once or twice a league season

- 300 A beginner league player

- 200 absolute beginner- may miscue frequently

These ratings are on a logarithmic scale like the Richter scale for earthquakes. What that means is that for each gap of 100 points, the higher rated player is *twice as good* as the lower rated player in the sense that a fair match between them would be 8-4 or 10-5, i.e., the higher rated player wins twice as many games as the lower rated player. This would also be true for any other 100-point gap, such as from 550 to 650. Tables are easily constructed that show fair matches for any rating difference. For instance, when the stronger player goes to 9 games, a fair match is one for which the weaker player goes to the following number of games:

Rating difference	**Weaker player goes to**
17	**8**
36	**7**
58	**6**
85	**5**
117	**4**
158	**3**
217	**2**
317	**1**

When two players play a single league game or play a tournament match, both players' **Fargo Ratings** get adjusted based upon the game score. In addition both players **robustness** values increase by the number of games they played.

Two points are important to note:

(1) Mathematics is in this document to explain the details of the approach to any reader who is in a position to understand and wants to understand the detailed approach. A user of the system sees none of this. It is not necessary to understand this mathematics to use this system as a player, as a league operator, or as a tournament director.

(2) Though there is a lot of discussion about fairly handicapped matches and handicapped league matches, recognize this system does not require a tournament director or

league operator ever handicap anything. All of the rating updates work just fine using unhandicapped competition.

III Relationship to NPL (National Pool League) approach

The NPL approach, which is used for handicapped 9-ball tournaments in San Francisco as well as in a number of other locations around the United States, is based upon a similar logarithmic scale for which a rating difference of 30 points (rather than 100 points) represents a doubling of skill level. The NPL approach is designed for tournaments that have 100% handicapped matches; that is, matches for which each player by design has an equal chance of winning. A match win comes with a rating increase, and a match loss comes with a rating decrease. The rating adjustments can depend on the number of prior matches, such that a newer player gets adjusted more quickly. In addition the size of the rating adjustment can depend on the event entry fee, a policy that disarms the sandbagger. A skill assessment based on a progressive skill test can help enter a new player into the system.

Though the Fargo Ratings 8-ball approach retains a lot from the NPL approach, it differs in some key ways.

(1) The 100-point skill doubling allows for a finer graded scale.

(2) Rating updates are based on each individual game rather than a match outcome. Thus match scores must be reported and recorded.

(3) All league and tournament results contribute to the rating adjustments

(4) We have derived the equations necessary to determine the optimum ratings for an entire group of players based upon all of the competition results within the group. So, for example, this analysis may serve annually to couple together different league systems based on regional or national competition.

(5) Based upon the above optimization equations we have derived rating update procedures that can be used within a league division or in a geographic area. The update procedure is self correcting and uses the available information in an optimal way.

IV New Players

In general new players are assigned an initial rating League results or known ability can be used as guidance for the initial rating. For players for which there is no knowledge, a progressive skill test like that used in the NPL can be used. Or a new region can choose to start every player at the same initial rating, say 450. In any case, new players are given a **robustness** value of 0, meaning the rating is based upon 0 games played in the system. As the player logs games, the **Fargo Rating** changes and the **robustness** value grows. Ratings adjust quickly when the **robustness** is small and slowly when the **robustness** is large.

V. Relation to other Rating Schemes

I give here an approximate correspondence between Fargo Ratings and other popular or regional rating schemes. This is based solely on my understanding of these other schemes, and it's useful only as a guideline or as an aid in assigning initial ratings to players.

NPL

Though the NPL (National Pool League) scheme is 9-ball based, here is an approximate correspondence between San Francisco NPL Ratings and Fargo Ratings.

NPL Rating	Fargo Rating
135	700
105	600
75	500
45	400
15	300

APA

APA equalizer handicaps are based upon a proprietary formula that considers, amongst other things, inning counts for 8-ball games. The following is an approximate correspondence based just on anecdotal information.

APA Rating	Fargo Rating
7	>560
6	500-560
5	425-500
4	350-425
3	< 350

Minnesota Rankings

Pool player organizations in Minnesota have put significant effort toward developing a uniform list of rankings, largely for 8-ball events. In Minnesota "master" means more or less the same thing that BCAPL and VNEA mean by master. Below that is AA, then A, then B. Here are approximate Fargo Rating correspondences. Minnesota also has M8 ratings.

Minnesota Rating	M8 Master/Advanced	Fargo Rating
Master	> 125	> 630
AA	100-125	575-630
A	75-100	500-575
B	50-75	425-500
C	< 50	< 425

VI. Rating Updates

A player's rating goes up or down after every game!

When a player wins a game, his or her rating goes up.

When a player loses a game, his or her rating goes down.

It may seem intuitively wrong that an average player's rating should go down because he or she lost a game against Efren Reyes, or won a game against a blind grandmother, but that is exactly what is supposed to happen. The simplified formula for a rating update after a single game against a well established opponent is based on equation 28 below and is

$$d_+ = \frac{630}{N}[1 - p] \qquad\qquad d_- = -\frac{630}{N}[p]$$

The expression on the left is the rating increase after winning a game. The expression on the right is the rating decrease after losing a game. The N in these equations is the player's robustness factor, the effective number of games the original rating is based on. The p is the player's probability of winning the game. For an opponent of the same rating, p=1/2. For an opponent 100 points above, p=1/3. For an opponent 100 points below, p=2/3. So if you lose a game to a better player, where you were expected to likely lose, your rating goes down only a small amount. If you lose to a weaker player, your rating goes down by a greater amount.

The game win probability, p, is determined by the rating difference between the players. If Δr is your opponent's rating minus your rating, then your probability of winning a single game is

$$p = \frac{1}{1 + 2^{\Delta r / 100}}$$

These formulas are in a spreadsheet and not seen by the league operator. A league operator just inputs the results like normal. A tournament director just inputs the match score.

As an example, suppose you are playing an unhandicapped race to 7 in a tournament against an opponent who is 50 points better than you. The *expected* outcome, based on your ratings, is you win 5 games and your opponent wins 7 games and wins the match. If this is what actually happens (you lose the match 5-7), your net rating change is zero. Your rating goes up for each of the five games you won, and it goes down by a slightly smaller amount per game for each of the seven games you lost. Your net rating change, once again, is zero, as is your opponent's.

If instead you went to the hill, losing the match 6-7, your rating would increase by a small amount, and your opponent's rating would decrease by a small amount.

More generally, the games you play may sometimes be against an established opponent and other times against an unestablished opponent. The following is the update formula (derived in Appendix 2) we actually use that includes the ratings and robustness for both players.

$$d_i = 630\left[n - p_{i,j}(n+m)\right]\left(\frac{N_j - (n+m)}{N_i N_j}\right)$$

This formula is used after each league game and after each tournament match. Whenever a match is played, there is an expectation for the number of games each player is supposed to win. This formula computes the update for player i's rating, d_i,

n	# games player i won
m	# games player j won
$p_{i,j}$	probability player i wins a single game against player j
N_i	player i's robustness after the match
N_j	player j's robustness after the match

Note that if your opponent's robustness was zero before the match started, then your rating will not change (the numerator in the fraction will be zero). Note too that as your opponent's robustness gets very large, this formula reverts to the simplified formula above. Though a player's Robustness, N_i, is zero for brand new players and has no upper limit, in our implementation N_i in the rating update formula is replaces with the maximum of N_i and 50 and by the minimum of N_i and 500. That is, we don't let N_i in the rating update formula go below 50 or above 500.

VII Sandbagging

Most handicapping systems are easily manipulated. And such manipulation is a serious problem. A small number of unscrupulous players begin trying to cheat any new system. Then when other players hear about this they feel they need to join in or be played the fool. Soon it becomes an industry. Many of those who make it to national events are those most adept at working the system. The APA seems to be notorious for this based upon newsgroup discussions. While tens of thousands of people play 8-ball every week, there is another group for which 8-ball is only *one* of the games they're playing. The *secret* rating algorithm depends prominently on inning counts. And any player capable of running out against a weaker opponent is also capable bunting balls around for a couple innings like a cat plays with a mouse, padding the inning count while still winning the game. Making the secret, proprietary formula more complex to try to stem this problem is tempting, but it just fuels the game and the true gamers.

The best way to deal with this problem is to devise a system that is open, transparent, and naturally resistant to manipulation.

While most players are honest, there are three ways an unscrupulous player might attempt to manipulate the Fargo system: fraudulent entry, fraudulent reporting, and intentionally losing games. I address each in turn.

Fraudulent entry

A new unknown player enters the system through a skill test. Clearly a player could intentionally under perform on a skill test. For an extreme example, suppose a player with an actual skill of 600 successfully enters the system as a 400. Then after a single weekly tournament, the player plays 20 games performing like a 600 is expected to perform. The update procedure would put that person as a 589 the following week. The low robustness with which that new player entered the system affords rapid rating adjustments. A scenario for which fraudulent entry could be a more serious problem is when the first or an early event is a high-entry-fee, high-payout handicapped event. There are two simple ways to deal with this. Any player with a robustness under a certain threshold may not enter, or any player under a certain threshold robustness value has his or her rating adjusted after each match of the tournament. For larger unhandicapped events limited to certain skill levels, such as an *under 500* tournament, a minimum robustness value can be required to enter.

Fraudulent reporting

When an unscrupulous player who wants his rating to stay low loses a match 5-2, he has an incentive to have it reported as 5-1 or 5-0, because the latter scores would give him a slightly more favorable rating adjustment. The single best defense against fraudulent reporting is the simple fact the losing player cannot gain unfair advantage without causing direct unfair disadvantage to the winning player of the same match. The two players have opposing incentives. Further, there are simple statistical *signatures* that emerge for lost matches. A score pattern that differs significantly from expectation is straightforward to detect.

Intentionally Losing Games

Sword-fight scene from The Princess Bride (1987)

[after long arduous duel]
Inigo Montoya: *I admit it, you are better than I am.*
Man in Black: *Then why are you smiling?*
Inigo Montoya: *Because I know something you don't know.*
Man in Black: *And what is that?*
Inigo Montoya: *I... am not left-handed.*
[Moves his sword to his right hand and gains an advantage]

One strategy for the unscrupulous player is to lay down during the $10 entry fee weekly tournaments only to come alive for the $30 entry fee monthly tournaments. A simple way to deal with this situation is to have a multiplier in front of the rating adjustment equation that varies, for example, from 0.5 for casual competition to 2.0 for more serious events.

While the possibility of manipulating the system can never completely be eliminated, the fact that every single game against every opponent contributes to a player's rating makes that manipulation much more difficult. Also, there are a number of features of the system described here that mitigate the problem.

VIII Handicapped Tournaments

There are many ways to handicap tournaments using Fargo Ratings as a guide. Just to discuss one example, here is a list of match pairings along with the Fargo rating-point *difference* that gives each player an equal chance of winning the match. A pairing called "5-4" means the higher rated player must win 5 games to win the match, and the lower rated player must win 4 games to win the match. So this is the same as a one-game spot in a race to 5.

Match Pairing	"Fair" match at this Fargo Rating-point difference
5 – 5	0
5 – 4	35
6 – 4	63
5 – 3	80
6 – 3	108
7 – 3	132
6 – 2	176
7 – 2	199
8 - 2	219

An example of a highly handicapped tournament that uses these match pairings is the following:

Play this match	If the rating difference is in this range
5 – 5	0 - 34
5 – 4	35-62
6 – 4	63-79
5 – 3	80-107
6 – 3	108-131
7 – 3	132-175
6 – 2	176-198
7 – 2	199-218
8 - 2	219 - UP

This particular tournament template has a slight statistical bias in favor of higher-rated players, but really is still an *anyone-can-win* format. Take the 6-3 match for instance. This is the designated matchup if the rating difference is between 108 and 131 points. Right at a point-gap of 108, this match is a coin toss. At the far end of the range, near 131, the "fair" match is getting closer to 7-3. So playing at 6-3 gives a modest bias to the higher-rated player. The rest of the chart follows the same pattern. The match is *fair*, i.e., a coin toss, at the beginning of each range and becomes modestly biased toward the higher ranked player throughout the range.

We use the above chart for a weekly handicapped 8-ball tournament. This, following NPL, is called "Chart 10" because the number of games for each matchup adds to a number close to 10. There is a similar "Chart 8" and "Chart 12," etc.

There are many different views about handicapped tournaments. Some don't like them at all. Others like to handicap just enough that weaker players don't stay home. Still others like the idea of having a 100% handicapped format where every player has an equal chance of winning. I'm not advocating any particular approach here. I'm merely advocating that Fargo Ratings allow a tournament director to implement whatever philosophy he or she wants accurately and without the subjective skill-level judgments that lead to so much discussion in the pool room.

IX **Handicapped Round-Robin Team League Play**

Left to natural forces, teams in league divisions tend in time to become more stratified and that causes the division to become less competitive. The top several players are together on a team. The weakest several players are together on another team, etc. There are two different approaches to making a division more competitive: (1) team skill ceilings, and (2) handicapping. Fargo Ratings are useful for either one.

Flatland Pool Leagues has chosen to implement team skill ceilings. We have four-player teams. Our *Fun* division requires the sum of the players' Fargo Ratings not exceed 1400. Our *Intermediate* division requires the sum of the players' Fargo Ratings not exceed 1850. And our *Advanced* division requires the sum of the players' Fargo Ratings not exceed 2100.

Handicapping

APA format – popular, but I don't like it

The most popular league for casual players around the country is APA (American Poolplayers Association). The basic structure of the APA format is that player A on one team plays a single many-game match against player B on another team. The handicap is the lesser-rated player can win the match by winning fewer games. Besides the inferior and minipulable handicap system, immediate problems with this format are individual matches take too long, and there's no incentive for team members not playing in a match to stay engaged.

A format I like

The following is the best format I've seen. There are four-player teams. While five-player teams are somewhat more common, four is a good number for a few reasons. (1) The two-team group playing on two tables means half of the players are playing at any given time. So each player is either playing or he or she is up next game on one of the two tables. (2) Two "couples" with a third couple as a backup or three couples that rotate who sits out make up good teams.

Players are ordered on each team according to their current averages: So players A1, A2, A3, and A4 play against B1, B2, B3, and B4.

There are 6 four-game "rounds." The first is four games matched up as follows:

A1 – B1,
A2 – B2,

A3 – B3,
A4 – B4.

The next three rounds are the other three round-robin combinations,

A1 – B2	A1 – B3	A1 – B4,
A2 –B3	A2 – B4	A2 – B1,
A3 – B4	A3 – B1	A3 – B2,
A4 – B1	A4 – B2,	A4 – B3.

The fifth and sixth rounds are two more repeats of round 1. So each player plays six games, playing his counterpart three times and every other opponent once.

There is a winning team for each round. The games are scored one point for each ball and 3 for the 8-ball (so 10 points for a game winner and anywhere from 0 to 7 points for a game loser). A good feature of this format—scoring by *rounds*-- is players are quite aware during games 3 and 4 of each round what the score is from games 1 and 2. Sometimes the last game comes down to "winner wins," i.e., winner of the game wins the round for his or her team, and other times one of the players needs only one or a couple balls to cinch the round. Players stay engaged in this format, and players are never more than a few minutes away from drama. The winning team for total number of points for the six rounds is awarded an extra *round*. So team scores for the night are 4 rounds to 3 rounds or 6 rounds to 1 round, etc.

Games take longer for newer players. For divisions with newer players, this same format can be used with each player playing four games per night instead of six—or five—or seven.

Handicapping this format

Here is an interesting statistical observation for 8-ball games that score 10 points for a win and 1 point for each ball by the loser. The average losing score for an 8-ball game is remarkably consistent amongst a wide range of skill levels. For instance, the top seven players in the league I played in a few years ago played 414 games and won 311 (75% of them). Of these 311 wins, 147 of them were won in the first at-bat, i.e., by running the table. This is a group of players that would be the top ranked (rank 7) in the APA, and probably average above 600 for Fargo Ratings. The average score this group got for *lost* games was 4.38.

The same analysis for a group of 7 players near the bottom of the league standings is as follows: They played 432 games and won 146 (34% of them). This group is over 200 Fargo rating points below the first group, perhaps averaging 400. Of the 142 wins, 8 were first at-bats. This would be a mid-level APA group, perhaps APA rank 4. Interestingly, the average score this group got for *lost* games was 4.26, very close to that of the other group. The consistency of this number suggests a good handicapping scheme. In other words if it can be assumed a typical game loser earns 4.3 points, then this can be combined with knowledge of the probability each player wins each game to determine an expected score for each team against any other team.

Given all the Fargo ratings, it's easy to determine the expected score for each team each round. A player's expected score for a game is 10 * p + 4.3 * (1-p), where p is the player's chance of winning the game. The simplest way to handicap the situation is to spot the weaker team a certain number of points. But there is a more satisfying way. Everything stays the same, but the weaker team gets (10+n) points for a win instead of 10. It's like instead of giving a person welfare, you give the person a tax break. Rather than being given a handout, the person must earn income to realize the benefit. Similarly here, a person must win a game to earn extra points.

It is straightforward to derive a good formula for n based just on the difference in the sums of Fargo ratings between the teams. This is all transparent for the players. They just see, for example, that the weaker team might earn 13 points for a game win, while the stronger team only earns 10 points for a game win.

As an example, suppose that on average each player of the stronger team is 44 Fargo points better than the corresponding player on the weaker team. This corresponds to n=2. So each round the score starts out 0 to 0. The stronger team gets 10 points for a game win, while the weaker team gets 12 points for a game win. For these teams, this maximizes the likelihood each round comes down to the last game; it maximizes the drama, which is what team competition is all about.

Once again, the handicapping doesn't have to be 100%. It can by design be enough that the weaker team has a reasonable chance of winning while the stronger team still has a statistical edge.

Appendix I

Derivation of the Optimization Equations

The probability player i beats player j in a single game depends on the difference in the players' Fargo ratings,

$$p_{i,j} = \frac{1}{1+2^{(r_j - r_i)/f}} \tag{1}$$

where f is the rating interval that corresponds to a factor of two in game winning success. For the Fargo system, f=100. For the NPL (National Pool League) system, f=30. It follows from Eqn. (1) that

$$p_{i,j} = 1 - p_{j,i}. \tag{2}$$

That is, the sum of the two players' win probability is unity.

If players i and j play a match to n games, the probability player i wins the match by the specific score n to m is

$$\pi_{n,m} = \binom{n+m-1}{m} p_{i,j}^{n} p_{j,i}^{m} \tag{3}$$

The combinatorial prefactor reflects the many possible permutations of the first n+m-1 games. Any match outcome n to m is possible with any pair of ratings, r_i and r_j. But depending on the ratings some match outcomes are more likely than others. For instance, if a match score between two unknown players is 10 to 5, then it is more likely player 1 is near 100 points above player 2 than that they have the same rating. The object of the algorithm discussed here is to determine the set of ratings for a group of players that best correspond to a series of matches between members of a group. There is a set of ratings for the group for which the actual detailed match results were most likely. These are the optimum ratings. For np players, there are at most np-1 degrees of freedom, because it is only rating *differences* that matter.

The function to be maximized is

$$S'(r_1, r_2, \ldots r_{np}) = \prod_{i=1}^{nmatches} \pi_i \tag{4}$$

where the product goes over all the matches in the database and π_i is the probability match i turned out like it did. Because the probabilities are all less than unity and sometimes much less than unity, S' can become a very small number. It is convenient to maximize instead the logarithm of S'.

$$S(r_1, r_2, \ldots r_{np}) = \ln(S') = \ln\left(\prod_{i=1}^{nmatches} \pi_i\right) = \sum_{i=1}^{nmatches} \ln(\pi_i) \tag{5}$$

The two functions S and S' are maximized by the same values of the ratings.

The maximum of S is found by first expanding it in a Taylor series to second order about the current ratings.

$$S(r_1, r_2, \ldots r_{np}) = S(r_1^0, r_2^0 \ldots r_{np}^0) + \sum_{i=1}^{np} \left(\frac{\partial S}{\partial r_i} \right)_0 d_i + \frac{1}{2} \sum_{i=1}^{np} \sum_{j=1}^{np} \left(\frac{\partial^2 S}{\partial r_i \partial r_j} \right)_0 d_i d_j + \ldots \tag{6}$$

where $d_i = r_i - r_i^0$ is the change in rating i.
In matrix notation, this equation is

$$S = S^0 + \mathbf{g}_0^t \mathbf{d} + \frac{1}{2} \mathbf{d}^t \mathbf{F}_0 \mathbf{d} + \ldots \tag{7}$$

Taking the derivative of this equation with to the deviations and setting the result to zero leads to the following system of linear equations.

$$\mathbf{g_0} + \mathbf{F}_0 \mathbf{d}^{\mathbf{opt}} = \mathbf{0} \tag{8}$$

Such equations normally are solved through left multiplying by the inverse of the Hessian matrix. This cannot be done here because $\mathbf{F_0}$ is singular and cannot be inverted. This is a consequence of the fact the dimension of the square matrix $\mathbf{F_0}$ is the number of players, np, while the rank of $\mathbf{F_0}$ is at most np-1. The function S is invariant to shifting all the ratings up or down, and this manifests itself as a zero eigenvalue of $\mathbf{F_0}$. $\mathbf{F_0}$ can have other zero eigenvalues as well. If the players fall into two or more groups that have no matches between members of the different groups, then S is invariant to shifting the ratings of one group relative to another group. This too manifests itself as a zero eigenvalue. Fortunately Eqn. (8) can still be solved because the vector $\mathbf{g_0}$ lies entirely within the range of $\mathbf{F_0}$.

To solve Eqn (8), we use the technique of Singular Value Decomposition. The subscripts on **F** and **g,** which denote the derivatives are evaluated at the current ratings, are dropped for clarity. We begin by finding the orthogonal matrix **U** that diagonalizes $\mathbf{F_0}$,

$$\mathbf{U}^{\mathbf{t}} \mathbf{F} \mathbf{U} = F \tag{9}$$

where F is the diagonal matrix of eigenvalues of **F,** and

$$\mathbf{U}^{\mathbf{t}} \mathbf{U} = \mathbf{U} \mathbf{U}^{\mathbf{t}} = \mathbf{I} \tag{10}$$

Eqn. (8) becomes, after left multiplying by $\mathbf{U^t}$ and using Eqn.(10),

$$[\mathbf{U}^{\mathbf{t}} \mathbf{F} \mathbf{U}][\mathbf{U}^{\mathbf{t}} \mathbf{d}] = -\mathbf{U}^{\mathbf{t}} \mathbf{g} \tag{11}$$

which has the form,

$$\begin{pmatrix} F_1 & 0 & 0 & 0 \\ 0 & F_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_{np} \end{pmatrix} = - \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ 0 \end{pmatrix} \tag{12}$$

The elements of the vector D are the optimum deviations in the eigenvector basis. Because F is diagonal, they are determined simply by the scalar equations,

$$D_i = \frac{-G_i}{F_i} \tag{13}$$

D_{np} is the deviation that shifts all the ratings up or down and is in undefined because F_{np}=0. In practice, F_{np} and any other zero or near zero eigenvalues can be set to large numbers so that the corresponding null deviations vanish. The optimum deviations then become

$$\mathbf{d} = \mathbf{U}D \tag{14}$$

Evaluating Derivatives

The elements of the gradient vector, **g**, and the Hessian matrix, **F**, are, respectively, the first and second derivatives of S with respect to the rating deviations, **d**. These derivatives can be computed analytically. They are necessary for solving Eqns (9)-(14) and they also are useful for determining sensible rating update procedures for ongoing tournaments.

The First Derivative

It is convenient to introduce a matrix, **A**, such that $A_{i,j}$ is the number of games won by player i against player j. Then, from Eqns (3) and (5)

$$S = \sum_{matches} \left[\ln \binom{A_{i,j}+A_{j,i}-1}{A_{j,i}} + A_{i,j} \ln(p_{i,j}) + A_{j,i} \ln(p_{j,i}) \right] \tag{15}$$

and

$$\left(\frac{\partial S}{\partial d_k}\right) = \left(\frac{\partial S}{\partial r_k}\right) = \frac{1}{2}\sum_{i=1}^{np}\sum_{j=1}^{np}\left[\frac{A_{i,j}}{p_{i,j}}\left(\frac{\partial p_{i,j}}{\partial r_k}\right) + \frac{A_{j,i}}{p_{j,i}}\left(\frac{\partial p_{j,i}}{\partial r_k}\right)\right] \tag{16}$$

The factor of one half is necessary because the matches are double counted by summing i and j independently. Eqn (16) becomes

$$\left(\frac{\partial S}{\partial r_k}\right) = \frac{\ln(2)}{2f}\sum_{i=1}^{np}\sum_{j=1}^{np}\left[(\delta_{i,k} - \delta_{j,k})(A_{i,j}p_{j,i} - p_{i,j}A_{j,i})\right]$$

$$= \frac{\ln(2)}{f}\sum_{i=1}^{np}\left[A_{k,i}p_{i,k} - p_{k,i}A_{i,k}\right] \tag{17}$$

This element of the gradient vector vanishes when r_k is optimum. This derivative becomes easier to interpret by using Eqn (2),

$$g_k = \left(\frac{\partial S}{\partial r_k}\right) = b\sum_{i=1}^{np}[A_{k,i} - p_{k,i}(A_{k,i} + A_{i,k})] \tag{18}$$

where b=ln(2)/f = 0.00693 for the Fargo system.

The sum here for player k goes over all opponents i. For each opponent, the first term is the *actual* number of player k's wins. The second term is the *expected* number of player k's wins

(the expected fraction of wins, $p_{k,i}$ times the total number of games against opponent i, $A_{k,i}$ $+A_{i,k}$).
The gradient, then, is

$$\left(\frac{\partial S}{\partial r_k}\right) = b \sum_{opponents}[\text{expected wins} - \text{actual wins}]$$

The Second Derivative

Differentiating Eqn. (17) with respect to r_l gives an element of the Hessian (second derivative) matrix.

$$\left(\frac{\partial^2 S}{\partial r_l \partial r_k}\right) = F_{l,k} = b^2 \delta_{k,l} \sum_{i=1}^{np} [(A_{k,i} + A_{i,k}) p_{k,i} p_{i,k}] - b^2 (A_{k,l} + A_{l,k}) p_{k,l} p_{l,k}$$

$$= -b^2 \left(N_{k,l} p_{k,l} p_{l,k} - \delta_{k,l} \sum_{i=1}^{np} [N_{i,k} p_{k,i} p_{i,k}] \right) \qquad (19)$$

The off-diagonal elements consist of just the first term in parentheses, while the diagonal elements consist of just the summation.

Appendix II

Approximate Rating Update Procedures

Although the optimization algorithm described in the previous section can be run to find the new optimum ratings whenever new data are added to the database, it is useful to develop approximate update procedures that can be used for ongoing tournaments among a group of players. We begin by assuming the ratings prior to a new match are optimum for the existing data, and the new match constitutes new information. We are interested in how the existing ratings should be updated in light of the new information under a variety of circumstances. In particular, the approach is different depending upon the available information about the current match and the players' prior matches.

As a simple example involving only two players, suppose Bert and Ernie play a few matches with the total score after the matches of 10 to 10. At this point, the optimum ratings are equal, say $r_1 = r_2 = 600$. Now if the pair plays a new race to 10, and Bert wins 10 - 0, then the new information in conjunction with the old information (total score of 20 to 10) suggests the pair are100 points apart, say 650 and 550. Each player was adjusted 50 points. Clearly the size of the update depends on how robust is the original rating. For instance if Bert and Ernies' 600 ratings were based on prior record of 100 games to 100 games, then the update after the new10-0 match would be considerably smaller, about 7 points each to 607 and 593. Note that the rating adjustment for the players following the 10-0 match fell by about a factor of seven (50 to 7) when the total number of games that contributed to the ratings increased by a factor of seven (from 30 to 210). This is a general result. The rating adjustment following any new information falls off as 1/N. Typically the prior games for each player will be against a collection of other opponents rather than against the current opponent.

The best update procedure depends on what information is known about the current match as well as the prior matches of both players. Procedures can be developed that use detailed match scores as well as simply match win/loss information. We develop some of these here. The updates discussed here can be easily modified to reflect the available information.

When the ratings are optimum, **g** and **d** in Eqn. (8) are both **0**. Suppose players i and j play a match with a score n to m. Now $A_{i,j}= A^0_{i,j}+ n$ and $A_{j,i}= A^0_{j,i}+ m$. Eqn (8) then has the form

$$\begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1,np} \\ F_{21} & F_{22} & & \vdots \\ \vdots & & \ddots & \\ & & & \\ F_{np,1} & \cdots & & F_{np,np} \end{pmatrix} \begin{pmatrix} 0 \\ d_i \\ \vdots \\ d_j \\ 0 \end{pmatrix} = - \begin{pmatrix} 0 \\ g_i \\ \vdots \\ g_j \\ 0 \end{pmatrix} \qquad (20)$$

This leads to the equations

$$\begin{aligned} F_{i,i}d_i + F_{i,j}d_j + g_i &= 0 \\ F_{j,i}d_i + F_{j,j}d_j + g_j &= 0 \end{aligned} \qquad (21)$$

Recognizing that $F_{I,j}=F_{j,I}$ and $g_I=-g_j$, these equations lead to

$$d_i = -g_i\left(\frac{F_{j,j}+F_{i,j}}{F_{i,i}F_{j,j}-F_{i,j}^2}\right) \tag{22}$$

The ratio of the off diagonal element $F_{i,j}$ to the diagonal element $F_{i,i}$ is approximately the ratio of player i's games against player j to all of player i's games. For established players, the second term in the denominator can be ignored. And as we discuss in the New Players section, even for two new players playing a first match--a worst case scenario--this term can be ignored. This leads to

$$d_i \approx -\frac{g_i}{F_{i,i}}\left(1+\frac{F_{i,j}}{F_{j,j}}\right) \tag{23}$$

This expression is straightforward to compute with a spreadsheet, but it is instructive to examine some further simplifications. The diagonal elements of F are given by

$$F_{i,i} = b^2\sum_{k=1}^{np}[N_{i,k}p_{i,k}p_{k,i}] \tag{24}$$

This is a sum over all opponents of the number of games against that opponent times the product $p_{i,k}p_{k,i}$. This product varies from 0.25 when $r_i = r_j$ down to 0.19 when the rating difference is 100, 0.11 when the rating difference is 200 and ultimately to zero when the rating difference goes to infinity. If we replace this product with the quantity q, which is the product for the average rating deviation $<p_{i,j}p_{jj}>$, then it can be taken out of the sum. The sum then becomes the total number of games player i has played, N_i. The factor q for a player with an average rating depends on the average deviation from the mean rating. If the ratings form a normal distribution, then the average deviation from the mean is

$$\langle|r-\langle r\rangle|\rangle = \sqrt{\frac{2}{\pi}}\,\sigma \approx 0.80\,\sigma \tag{25}$$

If the standard deviation is about 100 Fargo points, then q=0.23. The factor q will be somewhat smaller for players whose rating is far from the mean rating. A simple expression to account for this can be derived, but for now we use q=0.23 for all players. With this approximation, the diagonal element of F then becomes

$$F_{i,i} \approx b^2 q N_i = 0.23\, b^2 N_i \tag{26}$$

and the optimum deviation for player i following a match with player j becomes

$$d_i = 630\,\left[n - p_{i,j}(n+m)\right]\left(\frac{N_j-(n+m)}{N_i N_j}\right) \tag{27}$$

Eqn. (27) is the working equation for determining the optimum update based upon a specific match score between two players when the players' ratings as well as the total number of games for both players are known. For updating NPL ratings, where f=30, the factor in front is 190 rather than 630.

It is interesting to note the dependence of player i's rating adjustment on his opponent. At one extreme, opponent j has no matches prior to the present match. In this case the numerator in parentheses is zero, and player i's rating adjustment is zero. This makes sense because player i's rating should not depend upon performance against a player of unknown ability. In practice player j's initial rating will be based upon a skill test or upon subjective knowledge of his ability. As discussed in the New Player section, the confidence in this initial rating is manifested by considering him to have played some modest number of games against a hypothetical player of known rating. At the opposite extreme, player j is so well established that the current match represents a negligible fraction of player j's game. In this case the rating adjustment becomes

$$d_i \approx \frac{630}{N_i}\left[n - p_{i,j}(n+m)\right] \quad \text{(well established opponent)} \tag{28}$$

The term in square brackets, once again, is the number of games player i actually won minus the number of games he was expected to win.

Fairly Handicapped Matches

The expressions thus far are valid for any matches, handicapped or unhandicapped, among the players. We consider here the special case of fairly handicapped matches (such as in NPL tournaments). Here the matches are played n' to m', where $p_{i,j}$=n'/(n'+m'). If we define w_I=n-n' and w_j=m-m', then Eqn (27) becomes

$$d_i = 630\left[p_{j,i}w_i - p_{i,j}w_j\right]\left(\frac{N_j - (n+m)}{N_i N_j}\right) \tag{29}$$

For the match winner, w=0. For the loser, w=-1 for a match lost on the hill, -2 for a match lost by two games, etc. If the number of prior matches for each player is known, but the number of games is unknown, then for matches with n'+m' about equal to 10, the number of matches for player i, M_i is approximately N_i/8. Eqn.(29) then becomes

$$d_i = 80\left[p_{j,i}w_i - p_{i,j}w_j\right]\left(\frac{M_j - 1}{M_i M_j}\right) \tag{30}$$

$$d_i \approx \frac{80}{M_i}\left[p_{j,i}w_i - p_{i,j}w_j\right] \quad \text{(well established opponent)} \tag{31}$$

For NPL ratings, 80 is replaced by 24.