Measures for Ranking Estimation Performance Based on Single or Multiple Performance Metrics

Hanlin Yin Jian Lan
Center for Information Engineering Science Research
School of Electronics and Information Engineering
Xi’an Jiaotong University, Xi’an 710049, P.R.China
Email: iverlon1987@stu.xjtu.edu.cn, lanjian@mail.xjtu.edu.cn

X. Rong Li
Department of Electrical Engineering
University of New Orleans
New Orleans, LA70148, U.S.A
Email: xli@uno.edu

Abstract—There are several error metrics for estimation performance evaluation. To rank the performance of estimators, a popular method is using the same error metric of performance. It is not without controversy. First, this ranking method depending on the “marginal” information without considering the “joint” information among the estimators is one-sided since different error metrics reflect different aspects of performance. Second, ranking according to different error metrics may lead to different results. Thus, we propose to use the “joint” information just like Pitman’s closeness measure (PCM) to rank the performance of estimators. However, one drawback of PCM, named nontransitivity, brings big trouble for estimation performance ranking. To rank estimators utilizing the “joint” information, we propose a new approach using a so-called estimator ranking vector (ERV). The elements of ERV reflect how good the corresponding estimators are. Order-preserving mappings are proposed to obtain ERV, which, however, may not be unique. Then we use three specific mappings (i.e., linear, contraction, and concave, respectively) to solve this problem. Linear mappings can be easily applied and the other two mappings broaden the application domain of ERV. The ranking vector can also be used in multiple-attribute ranking problem. It does not need data normalization.

Keywords—Performance ranking, estimation, ranking vector, mapping.

I. INTRODUCTION

Estimation performance evaluation has received significant attention in recent years. There are many metrics for tracking (see, e.g., [1]–[9]) and fusion (see, e.g., [10]–[12]). Numbers of performance metrics, including incomplete and comprehensive ones, enrich the theoretical work on measures for estimation performance evaluation (see, e.g., [13]–[20]). Different error metrics reflect different aspects of performance. Given multiple estimators, how to rank their performance based on single or multiple performance metrics is a worthwhile problem.

Given multiple estimators, most people may select root-mean-square error (RMSE), which is the most popular error metric for estimation performance evaluation, to determine their ranking. However, RMSE has two serious flaws, as pointed out in [15]. First, it pays too much attention on how bad the performance is, that is, it is highly large-error dominant. Second, RMSE has no clear physical interpretation. Some people may select another measure of performance, such as average Euclidean error (AEE), geometric average error (GAE), harmonic average error (HAE), median error, error mode [15], and iterative mid-range error (IMRE) [19]. In short, they rank estimators in terms of the same error metric. Such a method is not without controversy. First, this method depends on the “marginal” information without considering the “joint” information among the estimators. Here, “marginal” information means the information of each estimator and “joint” information is the relative information of the multiple estimators. Thus, this method is one-sided since different error metrics reflect different aspects of performance. Second, ranking according to different error metrics may not be consistent, that is, different error metrics may lead to different ranking results. For instance, there are three estimators ˆx1, ˆx2 and ˆx3. According to RMSE, ˆx1 is better than ˆx2 and ˆx2 better than ˆx3. However, according to AEE, ˆx1 may be better than ˆx3 and ˆx3 better than ˆx2. Thus, ranking results are not consistent. Is there any other approach?

As mentioned above, ranking based on the same measure of performance depends on the “marginal” information. If we let them compete with each other, then the “joint” information can be used. For comparing the performance of rival estimators of a parameter, Pitman [21] proposed a criterion based on the probability that one estimator is closer to the parameter than the other. If this probability exceeds 0.5 for each value of the parameter, then the first estimator can be considered to be superior to the second [22]. This criterion is known as Pitman’s closeness criterion or Pitman’s closeness measure (PCM). The choice of the first estimator over the second one based on PCM allows us to quantify “how probable” we have chosen the better one. Thus the simple philosophy of PCM can also be found in a discourse by Descartes: “...when it is not in our power to discern the truest opinion, we ought to follow the more probable.” [23]

Rao observed some peculiar phenomena about PCM [24]. He successfully argued for PCM as an intrinsic measure of acceptability and presented many diverse examples in which shrinking the MSE of an unbiased estimator to an MMSE estimator does not yield a better estimator in the sense of PCM [20], [24]. [25] discussed ways of determining the relevant probabilities for Pitman’s closeness comparison of estimators.
The main advantages of PCM were described in [20] as follows. First, it uses the “joint” information of two competing estimators, and more information can be extracted. Second, PCM is generally applicable and robust against the choice of the error metric.

However, PCM is not perfect. It has some drawbacks, as described in [24]. A basic problem is that in giving the preference to an estimator, it may take into account a subset of the sample space that occurs only with a probability slightly greater than 0.5; the “winner” can then be arbitrarily bad when the observation does not belong to this subset. Due to this peculiarity, it is not transitive. For example, if $\hat{x}_3$ is better than $\hat{x}_2$ and $\hat{x}_2$ better than $\hat{x}_3$, then by PCM, $\hat{x}_1$ is not necessarily better than $\hat{x}_3$. Nontransitivity of PCM is the most severe problem for estimation performance ranking.

To solve this basic problem, [20] proposed a relative loss measure and a relative gain measure. They can greatly alleviate the peculiarity of PCM that “the winner can be arbitrarily bad when the observation does not belong to the particular subset”. However, they are still not always transitive. To solve the problem of nontransitivity, [26] defined a Bayesian version of Pitman’s closeness called posterior Pitman closeness. It was proved that this closeness is transitive in one dimension under fairly modest assumptions, and that the posterior median is the posterior Pitman closest estimator. However, in general, transitivity is still a big problem for ranking.

To solve the nontransitivity problem of PCM and use the “joint” information, we define a PCM matrix which contains the entire pairwise competition results of all estimators based on PCM. The PCM matrix can be used to rank estimators. In this paper, by using the PCM matrix, we present two methods named simple score addition (SSA) and limit score addition (LSA), respectively. For each estimator, SSA is the sum of PCM results by competing with the other estimators one by one. LSA can be viewed as a limiting form of SSA by repeating the competition infinitely many times.

In this paper, for ranking estimation performance, we propose a new approach by using an estimator ranking vector (ERV). The elements of ERV named strength reflect how good the corresponding estimators are, that is, the magnitudes of these elements represent a rank. For example, given three estimators $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$ with ERV $[1, 1.5, 2.5]'$, the rank is: $\hat{x}_3, \hat{x}_2, \hat{x}_1$. According to the ERV, all estimators involved are ranked directly and need no consideration of the problem of nontransitivity. However, we do not have the ERV from the very beginning and what we have is the competition results. Then how can we obtain ERV? The basic idea is as follows. The competition process can be viewed as a mapping and the result should reflect the relative strengths of the estimators. Thus, it is reasonable to require that mapping be order-preserved. For example, if the ERV is $[1, 1.5, 2.5]'$ and the mapping result is $[2, 3, 4]'$, then this mapping is order-preserved. If the ERV is $[1, 1.5, 2.5]'$ and the mapping result is $[2, 1, 4]'$, then the mapping is not order-preserved. However, the obtained ERV by order-preserving mapping may not be unique. For example, if for one ERV $[1, 1.5, 2.5]'$ mapping result is $[2, 3, 4]'$ and for another ERV $[2, 1.5, 2.5]'$ the corresponding mapping result is $[3, 5, 3, 4]'$, then how should we choose? We use three specific mappings called linear mapping, contraction mapping, and concave mapping to solve this problem. Linear mapping can be easily applied and the other two mappings broaden the application domain of ERV. In short, our goal is to obtain a unique ERV according to “joint” information of the estimators. Thus, we still let the estimators compete with each other, that is, PCM or some other competition process is one step of our method.

The above proposed ranking methods are based on a single performance metric such as PCM by using “joint” information. In practice, sometimes we want to rank estimation performance based on multiple performance metrics such as RMSE, AEE, GAE, HAE, median error, error mode, IMRE, and so on. Actually, this is a multiple-attribute ranking problem. Multiple methods of comprehensive evaluation can be used to solve this problem (see, e.g., [27]–[31]). In fact, most of them are based on weighted average, that is, they convert multiple attributes into one attribute for ranking. Weighted average is not without controversy. First, data should be normalized prior to weighted averaging because different attributes may have different units and orders of magnitude. However, we do not have a perfect data normalization method and different methods of data normalization may lead to different ranking results. Moreover, improper data normalization may lead to bad results. Thus, data normalization is a big problem for methods of weighted average. Second, methods using a weighted average can also be viewed as a “marginal” information user without considering the “joint” information among the compared objects. Therefore, the existing ranking methods based on weighted average are not appropriate for us to apply in ranking estimation performance based on multiple performance metrics.

For ranking estimation performance based on multiple performance metrics, we also propose to use the ranking vectors. Just like the definition of PCM, we define multiple-attribute competition measure (MCM). To solve the nontransitivity problem of MCM and use the “joint” information, we define an MCM matrix, which contains the entire pairwise competition results of all compared objects based on MCM. Then, all the above proposed methods can be used for ranking estimation performance based on multiple performance metrics. There are two advantages. First, the problem of data normalization is naturally avoided since we just need to make comparison for the same attribute. Second, the problem of nontransitivity is avoided and “joint” information is extracted. Furthermore, in practice, different performance metrics need not have equal “weight”, that is, we can give them different “weights” based on their importance in practice. For example, if RMSE is more important, then we can count it more times.

The paper is organized as follows. Section II discusses the PCM matrix. Simple score addition and limit score addition are also presented for solving the nontransitivity problem of PCM. In Section III, order-preserving mappings are presented and linear, contraction, and concave mappings for obtaining a unique ERV are proposed. In Section IV, the MCM and MCM matrix are defined. We also propose to use the ranking vector to solve the multiple-attribute ranking problem. In Section V, a simple example is provided to illustrate the difference between PCM and ERV when transitivity of PCM does hold in some scenarios. The other example illustrates how to use the ranking vector to rank estimation performance based on multiple performance metrics. Section VI concludes this paper.
II. PCM Matrix, Simple Score Addition and Limit Score Addition

A. PCM Matrix

PCM is based on the probability of the relative closeness of competing estimators to the estimand [20]. The following definition of PCM was given in [20]. For convenience, we quote it here. Given a probabilistic model \( p(x, z) \) of the parameter \( x \) and the observation \( z \), the estimator \( \hat{x} \) is a function of the observation \( z \) alone. Let \( m(1, 2; x) \) denote the measure of the difference between two estimators \( \hat{x}_1 \) and \( \hat{x}_2 \) relative to the parameter \( x \), and

\[
m(1, 2; x) = \begin{cases} 
1 & \text{if } \hat{x}_1 \succ \hat{x}_2 \\
0.5 & \text{if } \hat{x}_1 = \hat{x}_2 \\
0 & \text{if } \hat{x}_1 \prec \hat{x}_2 
\end{cases}
\]

where \( \hat{x}_1 \succ \hat{x}_2 \) means that \( \hat{x}_1 \) is preferred to \( \hat{x}_2 \) relative to the parameter \( x \), for example, \( \| \hat{x}_1 - x \| < \| \hat{x}_2 - x \| \) for some vector norm \( \| \cdot \| \). PCM for measuring \( \hat{x}_1 \succ \hat{x}_2 \) relative to \( x \) is defined as

\[
M(1, 2; x) = E(m(1, 2; x)) = \Pr(\hat{x}_1 \succ \hat{x}_2) + 0.5 \Pr(\hat{x}_1 = \hat{x}_2)
\]

where \( E(\cdot) \) is mathematical expectation. When \( M(1, 2; x) > 0.5 \), we say \( \hat{x}_1 \) is Pitman-closer to \( x \) than \( \hat{x}_2 \) is. We have the following relation

\[
M(1, 2; x) + M(2, 1; x) = 1
\]

As mentioned in Introduction, one of the drawbacks of PCM is nontransitivity. To use the “joint” information, we define the PCM matrix as

\[
X_{PCM} \triangleq \begin{bmatrix}
M(1, 1; x) & \cdots & M(1, N; x) \\
\vdots & \ddots & \vdots \\
M(N, 1; x) & \cdots & M(N, N; x)
\end{bmatrix}
\]

The PCM matrix contains the entire pairwise competition results of all estimators based on PCM. Pairwise competition makes good use of the “joint” or relative information.

B. Simple Score Addition

Now, the above PCM matrix can be used to determine the ranking of the estimators. One simple method is simple score addition (SSA), defined as

\[
SSA = X_{PCM} [1, \ldots, 1]'
\]

Its \( i \)th component is the winning percentage of \( \hat{x}_i \) when competing with the other estimators one by one. Thus, a larger element of SSA means that the corresponding estimator is better than the estimator that corresponds to a smaller element of SSA.

The following example illustrates how SSA is calculated.

\[\text{Example 1. Assume we have the following PCM matrix:}\]

\[
X_{PCM} = \begin{bmatrix}
0.5 & 0.8 & 0.7 \\
0.2 & 0.5 & 0.7 \\
0.3 & 0.3 & 0.5
\end{bmatrix}
\]

According to (2),

\[
SSA = \begin{bmatrix}
0.5 & 0.8 & 0.7 \\
0.2 & 0.5 & 0.7 \\
0.3 & 0.3 & 0.5
\end{bmatrix} [1, \ldots, 1]'
\]

By SSA we conclude that \( \hat{x}_1 \) is better than \( \hat{x}_2 \) and \( \hat{x}_2 \) better than \( \hat{x}_3 \).

However, SSA sometimes does not reflect the strength of the estimators because of the following drawbacks. Here, strength reflects how good the performance is. First, SSA ignores the prior strength (i.e., all equal to 1) of the estimators and treats them the same from the very beginning. Second, it is not thorough, that is, the information of PCM matrix is not completely explored. For example, if we treat the result of SSA as the prior strength and calculate another SSA, then the ranking result may change.

The following example illustrates this problem.

\[\text{Example 2. Assume we have the following PCM matrix:}\]

\[
X_{PCM} = \begin{bmatrix}
0.5 & 0.6 & 0.7 \\
0.4 & 0.5 & 0.9 \\
0.3 & 0.1 & 0.5
\end{bmatrix}
\]

According to (2),

\[
SSA = \begin{bmatrix}
1.8 & 1.8 & 0.9
\end{bmatrix}'
\]

By SSA, we can not distinguish which of \( \hat{x}_1 \) and \( \hat{x}_2 \) is better.

To solve the above problem, we define simple score addition square (SSA\(^2\)) as

\[
SSA^2 = X_{PCM}^2 [1, \ldots, 1]'
\]

For the above example, according to (3),

\[
SSA^2 = \begin{bmatrix}
0.5 & 0.6 & 0.7 \\
0.4 & 0.5 & 0.9 \\
0.3 & 0.1 & 0.5
\end{bmatrix}^2 [1, \ldots, 1]'
\]

By SSA\(^2\), we conclude that \( \hat{x}_1 \) is better than \( \hat{x}_2 \). For this example, SSA\(^2\) can distinguish which one is better while SSA can not because SSA\(^2\) takes advantage of information about the prior strength of the estimators. This prior strength is obtained by the first step (i.e., the SSA). A more intrinsic reason is that SSA\(^2\) explores more information of the PCM matrix. In other words, compared with SSA, SSA\(^2\) makes better use of the “joint” information. Thus, SSA\(^2\) can make a better ranking than SSA.

C. Limit Score Addition

Since SSA\(^2\) makes better use of the “joint” information and can explore more information of the PCM matrix, a natural idea is to repeat the competition results more times, that is, to multiply the PCM matrix more times. The extreme case is
Letting where $X$ is a competition matrix and its elements are positive. Letting $r = [r_1, \ldots, r_N]'$ where $N$ is the number of estimators and $r_i > 0$ ($i = 1, \ldots, N$). The elements of ERV are all positive and stand for the performance of the estimators relative to each other. As mentioned in the previous section, the strength reflects how good the performance is. The larger the element is, the better the corresponding estimator is. Then how can we obtain ERV?

A competition process can be viewed as a mapping $\eta = F(\xi)$. For example, the calculation process of SSA can be viewed as a linear mapping, i.e., $\eta = X_{PCM}\xi$. Here, it is reasonable to assume that $F$ be an order-preserving mapping since the competition process should not change the relative strengths of estimators and competition results should show the strengths. Here, order-preserving mapping means the order/ranking of the elements of $\xi$ and that of $\eta$ are the same. Then $r$ can be obtained according to this requirement. However, $r$ may not be unique in general. To obtain a unique ERV, we consider several specific cases.

### B. Linear Mapping

Assume we have an ERV $r$ and competition matrix $X$ (e.g., the PCM matrix). The final score $y_i$ of the $i$th estimator is

$$y_i = X_{i1}r_1 + \cdots + X_{iN}r_N$$

Thus, the final scores of the estimators can be written as

$$y = Xr$$

where $y = [y_1, \ldots, y_N]'$ is the final score vector. Here, we assume that the final scores are proportional to the strengths of the estimators (i.e., ERV $r$),

$$y = \lambda r$$

where $\lambda$ is a positive real number.

In view of the above, consider a linear mapping $F$, that is,

$$F(r) = Xr$$

where $X$ is a competition matrix and its elements are positive. Letting

$$F(r) = \lambda r$$

we have

$$Xr = \lambda r$$

Now, obtaining ERV becomes finding eigenvectors of matrix $X$. The Perron-Frobenius theorem [33] shows that there exists an eigenvector $r$ with all positive entries and its corresponding eigenvalue is positive and the largest eigenvalue (in absolute value) of $X$. If $r$ is a positive eigenvector of $X$, then any scalar multiple $\alpha r$ of $r$ with a positive $\alpha$ is also an eigenvector with the same eigenvalue. Actually, the ranking is unique because all positive eigenvectors are proportional.

**Perron-Frobenius Theorem (positive matrix).** If $A_{n \times n} > 0$ (i.e., each element of $A$ is a positive number) with $\lambda = \rho(A)$, where $\rho(A)$ is the spectral radius of $A$, then the following statements are true [33].

- $\lambda > 0$
- $\lambda$ is called the Perron root.
- $\lambda$ is the only eigenvalue on the spectral circle of $A$.
- There exists an eigenvector $r > 0$ such that $Ar = \lambda r$.
- There are no other nonnegative eigenvectors except one that has all positive elements, that is, all other eigenvectors must have at least one negative or non-real element.

According to the Perron-Frobenius theorem, if we have a positive matrix $X$ (e.g., PCM matrix), then we can obtain ERV. If there are zero elements in $X$, they can be replaced by very small positive numbers.

In fact, to calculate the positive eigenvector or ERV, we can use the following limitation method [32]

$$r = \lim_{n \to \infty} \frac{X^nr_0}{\|X^nr_0\|}$$

where $r_0$ is any positive vector. Compared with LSA (4), we can see that this limit leads to LSA when $X = X_{PCM}$ and $r_0 = [1, \ldots, 1]'$. Actually, when $X = X_{PCM}$, $r$ is LSA. Thus, this method can be viewed as a steady-state result by an infinite number of identical competitions presented by matrix $X$.

**Example 3.** Assume we have the following PCM matrix

$$X_{PCM} = \begin{bmatrix} 0.5 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.84 \\ 0.4 & 0.16 & 0.5 \end{bmatrix}$$

According to (2),

$$SSA = \begin{bmatrix} 0.5 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.84 \\ 0.4 & 0.16 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.7 & 1.74 & 1.06 \end{bmatrix}$$

According to (3),

$$SSA^2 = X_{PCM}SSA = \begin{bmatrix} 0.5 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.84 \\ 0.4 & 0.16 & 0.5 \end{bmatrix} \begin{bmatrix} 1.7 \\ 1.74 \\ 1.06 \end{bmatrix} = \begin{bmatrix} 2.43 & 2.44 & 1.49 \end{bmatrix}$$
The ERV (i.e., a positive eigenvector of $X_{PCM}$) is
\[ r = \begin{bmatrix} 0.6625 & 0.6376 & 0.3931 \end{bmatrix}'.\]

SSA shows that $\hat{x}_2$ is better than $\hat{x}_1$ and $\hat{x}_1$ better than $\hat{x}_3$. Although SSA$^2$ has the same ranking result as SSA, the specific results of $\hat{x}_2$ and $\hat{x}_1$ are very close. By ERV, we conclude that $\hat{x}_1$ is better than $\hat{x}_2$ and $\hat{x}_2$ better than $\hat{x}_3$. ERV, as a steady-state result, explores much more information of the PCM matrix and makes much better use of the "joint" information. Thus, we trust ERV more.

C. Contraction Mapping

The above linear mapping is one approach to obtaining ERV. For a more general (nonlinear) mapping, can we still obtain a unique ERV? Fortunately, this question is answered by the following theorem [34].

**Banach Fixed Point Theorem.** Let $(Z, d)$ be a non-empty complete metric space with a contraction mapping $F: Z \rightarrow Z$. Then $F$ admits a unique fixed point $r$ in $Z$ (i.e., $F(r) = r$). Furthermore, $r$ can be found as follows: start with an arbitrary element $r_0$ in $Z$ and define a sequence $\{Z_n\}$ by
\[ Z_n = F(Z_{n-1}) \] (7)
and then $Z_n \rightarrow r$.

Here, $F$ is a contraction mapping, defined as follows [35]. Let $(Z, d)$ be a metric space, a map $F: Z \rightarrow Z$ is called a contraction mapping on $Z$, if there exists $q \in [0, 1)$ such that
\[ d(F(x), F(y)) \leq q d(x, y) \]
for all $x, y$ in $Z$. A contraction mapping can be a linear or nonlinear mapping. The Banach fixed point theorem [34] guarantees the existence and uniqueness of a fixed point of certain self maps of metric spaces and provides a constructive method to find that fixed point.

According to the Banach fixed point theorem, we can obtain a unique fixed point $r$, that is, the unique ERV.

**Real world illustrations** [34]. The Banach fixed point theorem has the following interesting real world illustrations that can help us understand it well. When you go into a shopping mall, somewhere, usually you can see a map of the mall. On the map, there is a cross or arrow indicating your location with the text "You are here". This point on the map coincides with the corresponding point in the shopping mall. It can be explained as follows. Drawing shopping mall onto the map can be view as a contraction mapping and that point is a fixed point. According to the Banach fixed point theorem, that unique fixed point exists.

Consider another real world illustration [34]. Take two sheets of graph paper of equal size with coordinate systems on them, and then lay one flat on the table and crumple up (without ripping or tearing) the other one and place it in any fashion on top of the first so that the crumpled paper does not reach out of the flat one. Then there will be at least one point of the crumpled sheet that lies exactly on top of its corresponding point (i.e., the point with the same coordinates) of the flat sheet. This point is a fixed point; however, it has nothing to do with the Banach fixed point theorem since that process can not be viewed as a contraction mapping. Moreover, the fix point is not unique. In fact, this real world illustration is related to another fixed point theorem named the Brouwer fixed point theorem [34].

**Brouwer Fixed Point Theorem.** Assume that $K$ is a compact convex subset of $\mathbb{R}^n$ and that $F: K \rightarrow K$ is a continuous mapping. Then $F$ has a fixed point in $K$.

The Brouwer fixed point theorem does not follow from the Banach fixed point theorem and the fixed point is not necessarily unique. The second real world illustration above is a consequence of the $n = 2$ case of the Brouwer fixed point theorem.

Furthermore, the Perron-Frobenius theorem (for a positive matrix) can be treated as an application of the Brouwer fixed point theorem [36] if let $K$ denote the set
\[ \left\{ (x_1, \ldots, x_n) : x_i \geq 0 \text{ for all } i, \sum_{i=1}^{n} x_i = 1 \right\} \]
and $F(x) = Ax/ \|Ax\|$ for $x \in K$.

Although the Brouwer fixed point theorem is well known because of its use across numerous fields of mathematics, it can not be used to obtain ERV directly since the fixed point is not unique sometimes.

D. Concave Mapping

Contraction mappings contain some nonlinear mappings. This subsection focuses on concave mapping, which also contains some nonlinear mappings. A concave mapping is defined as follows [37]. Let $\mathbb{R}^n$ be the $n$-dimensional Euclidean space with a standard cone $K = \mathbb{R}^n_+ = \{ x \in \mathbb{R}^n | x = (x_1, \ldots, x_n), x_i \geq 0 \text{ for } 1 \leq i \leq n \}$. A mapping $T: K \rightarrow K$ is concave if
\[ T(\alpha x + (1 - \alpha) y) \geq \alpha Tx + (1 - \alpha) Ty \]
for all $x, y \in K$ and all $\alpha \in [0, 1]$. A unique ERV can be obtained according to the following concave Perron-Frobenius theorem.

**Concave Perron-Frobenius Theorem** [37]. For a concave mapping $T: K \rightarrow K$ with $Tx > 0$ for $x > 0$ the following statements hold.

- The conditional eigenvalue problem $Tx = \lambda x$ with $\lambda \in \mathbb{R}, x \in K, \|x\| = 1$ has a unique solution $x = x^\star$, $\lambda = \lambda^\star$, and $x^\star > 0, \lambda^\star > 0$.
- For the normalized mapping $\hat{T}x = \frac{Tx}{\|Tx\|}, x > 0, \lim_{k \rightarrow \infty} \hat{T}^kx = x^\star$ for all $x > 0$.

In this section, we proposed to rank the performance of estimators by ERV. The "joint" information of estimators are used and the problem of nontransitivity is avoided. Linear mappings can be applied easily and directly and the unique positive ERV (eigenvector) based on the Perron-Frobenius theorem can be calculated conveniently. Contraction mappings and concave mappings are more complicated than linear mappings, but they broaden the application domain of ERV.
IV. RANKING BASED ON RANKING VECTOR

Multiple-attribute ranking is very popular and plays an important role in economy, society, etc. Multiple methods of comprehensive evaluation can be used to solve this problem, such as fuzzy comprehensive evaluation, principal component analysis, linear weighted composite method, and so on (see, e.g., [27]–[31]). Most of them are based on a weighted average. The weighted average methods differ in how to determine the weights. They can be grouped into three classes: subjective weighted, objective weighted, and a combination of the two. Just like RMSE, however, they solve the multiple-attribute ranking problem based on the “marginal” information. Moreover, data normalization is a big problem and sometimes it is the most difficult part. Thus, we propose to use our ranking vector to rank the performance. As mentioned, the ranking vector uses “joint” information and does not need to normalize data. Therefore additional information can be extracted.

A. Multiple-Attribute Competition Measure Matrix

Just like PCM, for multiple-attribute ranking problems, we can define a multiple-attribute competition measure (MCM) as follows. Let \( m_{\text{MCM}}(1, 2; a) \) denote the measure of the difference between two compared objects \( S_1 \) and \( S_2 \) relative to the \( i \)th attribute \( a_i \) (the total number of the competitions w.r.t. \( a_i \) is \( n \)), and

\[
m_{\text{MCM}}(1, 2; a_i) \triangleq \begin{cases} 
1 & \text{if } S_1 \text{ is better than } S_2 \\
0.5 & \text{if } S_1 \text{ is equal to } S_2 \\
0 & \text{if } S_2 \text{ is better than } S_1 
\end{cases}
\]

Then, the MCM is defined as

\[
M_{\text{MCM}}(1, 2; a) = \frac{1}{n} \sum_{i=1}^{n} m_{\text{MCM}}(1, 2; a_i)
\]

where \( a \) is the vector of the attributes. When \( M_{\text{MCM}}(1, 2; a) > 0.5 \), we say \( S_1 \) is MCM-better than \( S_2 \) for \( a \). We also have the following relation

\[
M_{\text{MCM}}(1, 2; a) + M_{\text{MCM}}(2, 1; a) = 1
\]

Just like PCM, MCM is nontransitive. To use the “joint” information, we define the MCM matrix as

\[
X_{\text{MCM}} \triangleq \begin{bmatrix}
M_{\text{MCM}}(1, 1; a) & \cdots & M_{\text{MCM}}(1, N; a) \\
\vdots & \ddots & \vdots \\
M_{\text{MCM}}(N, 1; a) & \cdots & M_{\text{MCM}}(N, N; a)
\end{bmatrix}
\]

where \( N \) is the number of the compared objects.

The MCM matrix contains the entire pairwise competition results of all compared objects based on MCM. Pairwise competition makes good use of the “joint” or relative information.

The following example demonstrates how the ranking vector can be used for multiple-attribute ranking.

Example 4. There are four students and ten attributes in Table I. Our goal is to rank the comprehensive quality of the four students. Let the students compete with each other just like Pitman’s closeness measure does. For example, \( S_1 \) is better than \( S_2 \) for math and arts, but \( S_2 \) is better than \( S_1 \) for the other eight attributes. Thus, \( X_{12} = M_{\text{MCM}}(1, 2; a) = 0.2 \) and

\[
X_{21} = M_{\text{MCM}}(2, 1; a) = 0.8. By pairwise competition, the following MCM matrix is obtained:
\]

\[
X_{\text{MCM}} = \begin{bmatrix}
0.5 & 0.2 & 0.2 & 0.5 \\
0.8 & 0.5 & 0.4 & 0.6 \\
0.8 & 0.6 & 0.5 & 0.55 \\
0.5 & 0.4 & 0.45 & 0.5
\end{bmatrix}
\]

Then according to the Perron-Frobenius theorem, we can obtain a positive ranking vector as

\[
\hat{r} = [0.3294 \quad 0.5558 \quad 0.6042 \quad 0.4664]'
\]

Thus, the rank is: \( S_3, S_2, S_4, S_1 \).

Moreover, our method is more proper for the case of an adequate number of attributes because the “joint” information is not adequate if there are not enough attributes, then. In this case, “marginal” information should play an important role. Therefore, for multiple-attribute ranking, we need to consider the practical case more specifically. Anyway, using the ranking vector does have the following advantages. First, “joint” information is used, that is, additional information is extracted. Second, no need to consider the severe problem of data normalization. Third, the unique positive ranking vector can be calculated easily. Fourth, the thorny problem of nontransitivity is avoided.

B. Ranking Estimation Performance Based on Multiple Performance Metrics

For estimation performance evaluation, sometimes we want to rank the performance of estimators according to RMSE, AEE, GAE, HAE, IMRE, median error, error mode and some other errors. Here, just like what we did in Example 4, the ranking vector provides an option. The following example illustrates how to use ranking vector to rank estimation performance based on multiple performance metrics.

Example 5. Assume we have four estimators \( \hat{x}_1, \hat{x}_2, \hat{x}_3 \) and \( \hat{x}_4 \). The estimation errors \( \hat{x}_1, \hat{x}_2, \hat{x}_3 \) and \( \hat{x}_4 \) are generated from the following distributions. \( \hat{x}_1 \sim \mathcal{N}(0.8, 1), \hat{x}_2 \sim 0.9\mathcal{N}(0.5, 1) + 0.1\mathcal{N}(4.1), \hat{x}_3 \sim 0.9\mathcal{N}(0.6, 1) + 0.1\mathcal{N}(2.5, 1), \) and \( \hat{x}_4 \sim 0.98\mathcal{N}(0.1, 1)+0.02\mathcal{N}(14, 1) \), where \( \mathcal{N}(a, b) \) is a Gaussian distribution with mean \( a \) and variance \( b \). Then for each estimator, 10000 errors are generated to calculate RMSE, AEE, GAE, HAE, median error, error mode and IMRE. The results are in Table II. Let the estimators compete with each other just like example 4 does. For example, \( \hat{x}_2 \) is better than \( \hat{x}_1 \) for HAE and median error, but \( \hat{x}_1 \) is better than \( \hat{x}_2 \) for the other five performance metrics. Thus, \( X_{12} = \)

<table>
<thead>
<tr>
<th>Table I. Ten Attributes of Four Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁</td>
</tr>
<tr>
<td>math</td>
</tr>
<tr>
<td>arts</td>
</tr>
<tr>
<td>long jump</td>
</tr>
<tr>
<td>English</td>
</tr>
<tr>
<td>geography</td>
</tr>
<tr>
<td>physics</td>
</tr>
<tr>
<td>music</td>
</tr>
<tr>
<td>high jump</td>
</tr>
<tr>
<td>history</td>
</tr>
<tr>
<td>shot put</td>
</tr>
</tbody>
</table>
$M_{MCM}(1, 2; a) = 5/7$ and $X_{21} = M_{MCM}(2, 1; a) = 2/7$. By pairwise competition, the following MCM matrix is obtained:

$$
X_{MCM} = \begin{bmatrix}
0.5 & 0.7143 & 0.5714 & 0.2857 \\
0.2857 & 0.5 & 0.1429 & 0.1429 \\
0.4286 & 0.8571 & 0.5 & 0.2857 \\
0.7143 & 0.8571 & 0.7143 & 0.5 \\
\end{bmatrix}
$$

Then according to the Perron-Frobenius theorem, we can obtain a positive ranking vector as

$$
r = [0.4927, 0.2374, 0.4731, 0.6907]^T
$$

Thus, we conclude that the rank is: $\hat{x}_4$, $\hat{x}_1$, $\hat{x}_3$, $\hat{x}_2$. This example shows that the ranking vector can be applied easily for ranking estimation performance based on multiple performance metrics. It uses “joint” information and does not need data normalization.

**TABLE II. RESULTS OF PERFORMANCE METRICS**

<table>
<thead>
<tr>
<th></th>
<th>$\hat{x}_1$</th>
<th>$\hat{x}_2$</th>
<th>$\hat{x}_3$</th>
<th>$\hat{x}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.2863</td>
<td>1.6928</td>
<td>1.3759</td>
<td>2.2860</td>
</tr>
<tr>
<td>AEE</td>
<td>1.0454</td>
<td>1.2126</td>
<td>1.0788</td>
<td>1.1614</td>
</tr>
<tr>
<td>GAE</td>
<td>0.7152</td>
<td>0.7246</td>
<td>0.7063</td>
<td>0.6280</td>
</tr>
<tr>
<td>HAE</td>
<td>0.1949</td>
<td>0.1656</td>
<td>0.1206</td>
<td>0.1083</td>
</tr>
<tr>
<td>median error</td>
<td>0.9102</td>
<td>0.8633</td>
<td>0.8855</td>
<td>0.7831</td>
</tr>
<tr>
<td>error mode</td>
<td>0.0214</td>
<td>0.0311</td>
<td>0.0268</td>
<td>0.0093</td>
</tr>
<tr>
<td>IMRE</td>
<td>1.0095</td>
<td>1.0838</td>
<td>1.0220</td>
<td>0.9381</td>
</tr>
</tbody>
</table>

**V. ILLUSTRATIVE EXAMPLES**

We treat the performance metrics impartially in the above example, that is, the performance metrics have equal “weight.” In fact, different metrics need not have an equal “weight.” We can give them different “weights” based on their importance in practice. For example, if RMSE is deemed more important, then we can count RMSE more times. We illustrate it by the following example.

**Example 6.** Assume we have four estimators $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, and $\hat{x}_4$ that are exactly the same as in Example 5. Here, we assume that RMSE is more important than the other performance metrics such that we count it 2.9 times. Let the estimators compete with each other. For example, $\hat{x}_2$ is better than $\hat{x}_1$ for HAE and median error, but $\hat{x}_1$ is better than $\hat{x}_2$ for the other 6.9 performance metrics (RMSE is counted 2.9 times). Thus, $X_{12} = M_{MCM}(1, 2; a) = 6.9/8.9$ and $X_{21} = M_{MCM}(2, 1; a) = 2/8.9$. By pairwise competition, the following MCM matrix is obtained:

$$
X = \begin{bmatrix}
0.5 & 0.7753 & 0.6029 & 0.4382 \\
0.2247 & 0.5 & 0.1124 & 0.3258 \\
0.3371 & 0.8876 & 0.5 & 0.4382 \\
0.5618 & 0.6742 & 0.5618 & 0.5 \\
\end{bmatrix}
$$

Then according to the Perron-Frobenius theorem, we can obtain a positive ranking vector as

$$
r = [0.5814, 0.2733, 0.5035, 0.5777]^T
$$

Thus, we can conclude that the rank is: $\hat{x}_1$, $\hat{x}_4$, $\hat{x}_3$, $\hat{x}_2$. Comparing Example 5 with Example 6, we can see that the ranking results are different. In practice, different error metrics need not have the same “weight.” We can give them different “weights” based on their importance.

If the transitivity of PCM does hold in some scenarios, then can we declare that ranking based on PCM is the same as ranking based on ERV? We answer this question by the following example.

**Example 7.** Assume we have three estimators $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$, and the PCM matrix is

$$
X_{PCM} = \begin{bmatrix}
0.5 & 0.7 & 0.8 \\
0.3 & 0.5 & 0.9 \\
0.2 & 0.1 & 0.5 \\
\end{bmatrix}
$$

Then, we have ERV (i.e., a positive eigenvector)

$$
r = [0.7691, 0.5821, 0.2637]^T
$$

From $X_{PCM}$ we can see that transitivity does hold. The ranking result based on PCM can be concluded as follows. $\hat{x}_1$ is the best and $\hat{x}_2$ is better than $\hat{x}_3$. This is the same as the ranking result of ERV.

If the PCM matrix is

$$
X_{PCM} = \begin{bmatrix}
0.5 & 0.51 & 0.51 \\
0.49 & 0.5 & 0.9 \\
0.49 & 0.1 & 0.5 \\
\end{bmatrix}
$$

then we have ERV

$$
r = [0.5996, 0.6977, 0.3922]^T
$$

From the PCM matrix we can conclude that $\hat{x}_1$ is the best and $\hat{x}_2$ is better than $\hat{x}_3$. However, ERV shows that $\hat{x}_2$ is the best and $\hat{x}_1$ is better than $\hat{x}_3$. Thus, ranking based on PCM and that based on ERV are inconsistent. In our opinion, the reason is the following. PCM cares only better or worse, and not how much better or worse. That is, 0.51 and 0.99 is the same when PCM is directly used for ranking. On the contrary, ERV does evaluate how much better or worse to some extent. Thus, they may be inconsistent sometimes. In other words, ERV uses more information than PCM although both of them use “joint” information. Thus, we should trust ERV more.

**VI. CONCLUSIONS**

This paper deals with estimation performance ranking using “joint” information. PCM, as a “joint” information user, has some drawbacks. One of its drawbacks named nontransitivity is a big problem for estimation performance ranking. We have presented simple score addition and limit score addition by using the PCM matrix. To rank estimators with “joint” information, we have proposed a new approach by using the so-called estimator ranking vector. Order-preserving mappings have been proposed to obtain ERV. However, the obtained ERV may not be unique. Then three specific mappings, named linear mapping, contraction mapping and concave mapping, have been proposed to obtain a unique ERV. Among these three mappings, linear mapping can be applied easily; contraction mapping and concave mapping broaden the application domain of ERV. These two mappings deserve more studies. For the problem of multiple-attribute ranking, just like PCM and the PCM matrix, we have defined MCM and MCM matrix. The ranking vector can also be applied to this problem. It uses the “joint” information and the troublesome data normalization is avoided. Seven examples have been presented to illustrate the following conclusions. SSA² can make a better ranking than SSA. Compared with SSA and SSA², LSA explores much
more information of the PCM matrix and makes better use of the “joint” information. Different error metrics need not have the same “weight” in practice when multiple performance metrics are used for ranking estimation performance. ERV uses more information than PCM, although both of them use “joint” information.

REFERENCES


