Optik 127 (2016) 5084–5091

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Enhanced error spectrum for estimation performance evaluation

Weishi Peng [a,c], Yangwang Fang [a], Zhansheng Duan [b,*], Binghe Wang [c]

[a] School of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an, Shaanxi 710038, China
[b] Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[c] School of People Armed Police Engineering University, Xi'an, Shaanxi 710086, China

ARTICLE INFO

Article history:
Received 19 January 2016
Accepted 28 February 2016

Keywords:
Estimation
Estimation performance evaluation
Error spectrum
Dynamic error spectrum
Multi-objective optimization

ABSTRACT

The error spectrum is a comprehensive metric for estimation performance evaluation in that it is an aggregation of many incomprehensive measures. However, the error spectrum is a two-dimensional curve for any estimand (i.e., the quantity to be estimated) of interest. Therefore, unless one error spectrum dominates the others, it is in general not straightforward to say which one is better. Although the dynamic error spectrum (i.e., the average height of the error spectrum) was proposed to tackle this problem, it suffers from the problem of information loss due to the mapping of the whole error spectrum at a time instant into a single point. Particularly, if the average heights of two error spectrums are the same, they are still indistinguishable. To alleviate this, two new metrics called range error spectrum induced area and dynamic error spectrum induced area are proposed in this paper. Then how to combine these two new metrics, called as enhanced error spectrum, are further studied. An additive and a multiplicative form of the enhanced error spectrum are presented respectively for different scenarios. A numerical example is provided to illustrate the effectiveness of the metrics. It is shown that due to the consideration of more information, the new metrics have greater applicability than the dynamic error spectrum.

1. Introduction

In recent years, estimation performance evaluation (EPE) has received considerable attention owing to its increasing application in estimation/filtering (see, e.g. [1–3,5,6,9–12]), track fusion/tracking [13], performance analysis [14], etc. To the best of our knowledge, EPE includes mainly two components: the estimator ranking and the estimator evaluation. For the estimator ranking, Pitman proposed a criterion known as the Pitman closeness measure (PCM) [17]. Since then, most existing research has focused on the improvement of the non-transitivity problem of the PCM (see, e.g. [18–22]), which is a major obstacle for EPE. Inspired by the PCM, the authors of [21] proposed an estimator ranking vector which includes several performance metrics. Thus, a key aspect in EPE is the selection and proper interpretation of the metrics used for the estimator ranking and the estimator evaluation. The root mean square error (RMSE) is widely used in EPE, since it is the most natural finite-sample approximation of its theoretical counterpart. As pointed out in [1,2], the RMSE is easily dominated by large error terms and has no clear physical interpretation. Therefore, it was replaced with the average Euclidean error (AEE) in several applications [1]. Although the AEE has several advantages, it is still affected by extreme values. Therefore, several incomprehensive performance measures were proposed in [2] such as the harmonic average error (HAE), the geometric average error (GAE), median error, and error mode. Furthermore, the iterative mid-range error (IMRE) was presented in [6], since the above-listed metrics are not robust.

Unfortunately, all of the above-listed metrics can reflect only one aspect of the estimator performance. Thus, three comprehensive performance measures – the error spectrum (ES), desirability level, and relative concentration and deviation measures were proposed in [3–5]. Among these metrics, the ES can reveal more information about the estimation because it is an aggregation of several incomprehensive metrics.

However, the ES has some limitations and drawbacks. On one hand, its calculation without the error distribution is not easy, though in [7] (a further development of [4]), the authors provided analytical formulae for the computation of the ES when the error distribution is given. To overcome this problem, we proposed two approximation algorithms in [15,16] based on the Gaussian mixture and power means error. On the other hand, it is difficult to say which estimator performs better if their ES curves intersect with each other. To tackle this problem, a dynamic error spectrum (DES) reflect the estimation accuracy of an estimator was presented in [8,9], which is in fact the average height of the ES.

* Corresponding author. Tel.: +86 18629562612; fax: +86 2982663420.
E-mail address: zhanshengduan@hotmail.com (Z. Duan).

http://dx.doi.org/10.1016/j.ijleo.2016.02.063
0030-4026/

Although the DES does provide a solution to this problem, it still has some limitations. First, it is difficult to decide which estimator performs better, when the average height of their ES. Second, the DES provides a ruler only to measure how large the estimation error is. Recall that the least-squares (LS) estimation and minimum mean square error (MMSE) estimation differ from the maximum likelihood (ML) estimation and maximum a posteriori (MAP) estimation in their underlying ideas. The former seeks an estimator that has the smallest error, while the latter uses the "most frequently occurred" value of the estimate as the estimator [2]. Although the ML and MAP estimators may have a larger average error, they may have a higher probability of being close to the estimate, i.e., the estimation error of the ML and MAP estimators are concentrated to the estimate (the concentration of an estimator in this paper), which reflects the ES curve of the ML and MAP estimators more flatness. This has important implications while choosing an estimation method for a particular application. Thus, a worthwhile problem is how to also take into account the flatness of an ES curve in the evaluation of the performance of an estimator.

The main contribution of this work is twofold. First, two new estimation evaluation metrics, i.e., range error spectrum (RES) induced area (RESA) and the DES induced area (DESA), have been proposed, where the RESA is designed to quantify the flatness of an error spectrum curve, and the DESA is designed to measure the estimation accuracy of an estimator. Second, how to combine these two new metrics, which is called enhanced error spectrum (EES), is considered in this paper. Two forms of combination are presented for different scenarios. The first form is additive, which is dependent on prior preference between the concentration and estimation accuracy. The second form is multiplicative, which does not depend on the prior preference and is also suggested when the dynamic error spectrum induced area is dominating.

This paper is organized as follows. The ES and DES are summarized in Section 2. In Section 3, two areas of ES, called as RESA and DESA, are designed to evaluate an estimator. Furthermore, how to combine these two new metrics, i.e., the EES, is considered in Section 4. A numerical example is provided in Section 5 to illustrate the superiority of the proposed metrics. The paper is concluded in Section 6.

2. Summary of ES and DES

2.1. Error spectrum

According to [4,23], let the (possibly vector-valued) estimation error $\tilde{\theta}$ of a (point) estimator $\hat{\theta}$ be $\tilde{\theta} = \theta - \hat{\theta}$, where θ is the estimand (i.e., the quantity to be estimated). We denote $e = \|\tilde{\theta}\|$ or $e = \|\tilde{\theta}\|/\|\hat{\theta}\|$ as the absolute or relative estimation error norm, where $\|\cdot\|$ can be 1-norm or 2-norm. Then, for $r_i \in (-\infty, +\infty)$, the ES is defined as

$$S(r) = (E[e^r])^{1/r} = \left\{ \int e^r dF(e) \right\}^{1/r} = \begin{cases} \left\{ \int e^r f(e) de \right\}^{1/r} & \text{if } e \text{ is continuous} \\ \left(\sum p_i e_i^r \right)^{1/r} & \text{if } e \text{ is discrete} \end{cases} \quad (1)$$

where $F(e)$, $f(e)$, and p_i are the cumulative distribution function (CDF), probability density function (PDF), and probability mass function (PMF), respectively.

For a discrete $\{e_i\}_{i=1}^n$, ES can be approximately calculated by [15,16]

$$S(r) \approx \begin{cases} \left[\frac{1}{n} \sum_{i=1}^n (e_i)^r \right]^{1/r} & r \neq 0 \\ \left[\prod_{i=1}^n e_i \right]^{1/n} & r = 0 \end{cases}$$

From (1), it is clear that the ES includes several incomprehensive metrics as special cases when r is set to some specific values:

(1) $S(2) = (E[e^2])^{1/2}$. Thus, for a discrete e_i, $S(2) = ((1/n)\sum_{i=1}^n e_i^2)^{1/2} = \text{RMSE}$.
(2) $S(1) = E[e]$. Thus, for a discrete e_i, $S(1) = \frac{1}{n}\sum_{i=1}^n e_i = \text{AEE}$.
(3) $S(0) \triangleq \lim_{r\to 0} S(r) = \exp(E[\ln(e)])$. Thus, for a discrete e_i, $S(0) \triangleq \frac{1}{n}\prod_{i=1}^n e_i = \text{GAE}$.
(4) $S(-1) = 1/E[1/e]$. Thus, for a discrete e_i, $S(-1) = ((1/n)\sum_{i=1}^n e_i^{-1})^{-1} = \text{HAE}$.

In view of this, the notation r used in this paper is a real number that satisfies $r_i \in [-1, 2]$.

Certainly, the ES is a curve for a state estimator of a dynamic system at any time instant. Therefore, it will be a three-dimensional plot over the entire time span, which causes difficulty in the EPE of dynamic systems. Fortunately, the DES has been proposed to tackle this problem.

2.2. Dynamic error spectrum

According to [8,9], if some prior knowledge is available about the weights $\{\omega_i\}_{i=1}^n$ corresponding to each different r_i, where $\sum_{i=1}^n \omega_i = 1$, $r_i \in \{r_j\}_{j=1}^n$, $-1 \le r_i \le 2$, and n is the number of indices over $\{r_j\}_{j=1}^n$, the weighted form of the DES is given simply as

$$\text{DES}(\omega) = \sum_{i=1}^n S(r_i)\omega_i \quad (2)$$

Since it is difficult to obtain the weights, another form of the DES is given by the average height under the ES curve, as follows:

$$\text{DES} = \frac{1}{r_n - r_1} \int_{r_1}^{r_n} S(r) dr \approx \frac{1}{n} \sum_{i=1}^n S(r_i) \quad (3)$$

It can be clearly seen that the DES combines several incomprehensive metrics into a single metric. Thus, the DES reflect the estimation accuracy of an estimator the same as the incomprehensive metric. In other words, the DES suffers from the problem of information loss during this many-to-one mapping, as shown in Example 1.

Example 1. As pointed out in [2], a more complete description of estimation performance is the PDF of the estimation error. In the following example, we directly show why the DES cannot distinguish the difference among the following error PDFs. Fig. 1(a)