IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 2, 2019, accepted September 7, 2019, date of publication September 11, 2019,
date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940893

A Novel Space-Based Orbit Determination Method Based on Distribution Regression and Its Sparse Solution

FEI FENG[1], YASHENG ZHANG[1], HENGNIAN LI[2], YUQIANG FANG[1], QIUSHI HUANG[1], AND XUEFENG TAO[1]

[1]Space Engineering University, Beijing 101400, China
[2]State Key Laboratory of Astronautic Dynamics, Xi'an 710000, China

Corresponding authors: Fei Feng (fengfei90s@yeah.net) and Yuqiang Fang (fangyuqiang@nudt.edu.cn)

This work was supported by the National Outstanding Youth Fund under Grant 2017-JCJQ- ZQ-005.

ABSTRACT There is an inherent need to track and catalog space debris (objects) in geosynchronous earth orbits (GEO) based on space-based surveillance networks and a large amount of observation data. However, for objects in GEO, angle-only measurements containing noises have been regarded as difficult for short-arc orbit determination (OD) when pursuing high accuracy. In this paper, from a data-driven perspective, we propose a novel method for space-based OD based on distribution regression (DR), which is called the weighting distribution-regression OD (WDR-OD) method. The OD is treated as a regression process, which is learned from abundant observation data and the corresponding orbits of known objects. First, we propose the structure of space-based OD samples, wherein the feature variables with a weighting matrix are introduced to enhance prediction accuracy. Second, a two-stage sampled learning theory is employed to learn the mapping from measurements to objects' orbit through kernel mean embedding. The proposed method is experimentally compared with the improved Laplace method and shows greater robustness in measurements with white Gaussian noise (WGN) and colored noise. The positional RMSE reaches 0.8793 km with WGN and 1.6972 km with colored noise, which are significantly smaller than the corresponding Laplace method's 5.0804 km and 14.8132 km. Furthermore, we propose a RIP-based ROMP algorithm to provide the theoretical bound of sparsity and then to pursue a sparse solution. Although the positional RMSE increases to 1.6554 km in the sparse method, it shrinks the 90% to 93% nonzero elements of the coefficients matrix to zero, which is helpful in reducing the computing load, and it meaningfully extends the application domain of the WDR-OD method.

INDEX TERMS Space situation awareness, orbit determination, machine learning, distribution regression, kernel mean embedding, sparse representation.

I. INTRODUCTION

For the purposes of space debris avoidance and due to the requirements of space situation awareness (SSA), scientific explorations, missions, etc., there is an inherent need to track and catalog space objects, especially space debris in the increasingly crowded belt of geosynchronous (GEO) orbits. Compared with ground-based space surveillance systems, space-based space surveillance (SBSS) systems perform with more flexible observation conditions, better observation geometry and a large amount of observation data. Orbit determination (OD) is a nonlinear problem that estimates the orbital elements from observations in one or more segments of the trajectory with Newton's laws and perturbations. Star sensors mounted on satellites can be utilized to perform OD [1] through angular measurements (right ascension and declination) of the large number of captured objects. Since these observations are ''unintentional'', there will be a mass of short-arc measurements swarming into the data processing center. However, for objects in GEO, classical initial orbit determination (IOD) methods (e.g., Laplace [1] and Gooding [2]) have been regarded as problematic for evaluating hundreds of meters or pursuing higher accuracy in short-arc (no more than 18 minutes in this paper) OD with angle-only measurements that are polluted with noise (noise of 1-3 seconds of arc). The results of these methods are

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

sensitive to noise and can easily converge to trivial solutions. An idiomatic method is to hypothesize the form of noise [3], but it is hard to conjecture it precisely in practice. Most precision orbit determination (POD) methods, such as the Kalman filter, rely on a prior initial state and dense observations of an arc period over a long time span.

Before reviewing the existing techniques in the literature, let us start with an exposition of the core idea of this paper. We solve the OD problem without establishing the dynamical and geometrical model used in classical OD methods; instead, from a data-driven perspective, we treat it as a regression problem, taking measurements as the input values and taking the object's orbit as the output values. Thus, in this way, the framework of the distribution regression (DR) technique from machine learning (ML) can be adopted. In a standard regression process, we need to predict a real-valued response variable Y from a vector valued feature variable $X \in \mathbb{R}^d$. Recently, there has been interest in extending the standard regression from finite dimensional Euclidean spaces to functional spaces. DR is a typical representation of that and focuses on the regression from a probability measure P to real-valued responses. DR differs from standard regression in two crucial ways. First, $P \in \mathbb{R}^k$ is a probability measure instead of a determinate value in X corresponding to Y. Second, practically, we cannot observe the probability measure P directly; instead, we observe the samples from P with an unknown probability measure. Similarly, the orbit of a space object is estimated through the angle-only measurements derived from one or multiple sensors over a period of time. Position vectors of observers and their angular measurements together constitute the observations. There are multiple combinations of observing sensors in space over one object, but different sensors measuring the same object must output the same orbital elements at a given moment. Thus, the conversion of these observations to orbital elements constitutes a many-to-one relationship. The observations can be regarded as being sampled from an unknown distribution P, and in this way, this approach provides a novel method to estimate the orbit of an object through DR. In this paper, we propose a weighting distribution-regression OD (WDR-OD) method for the space-based OD problem.

II. RELATED WORK

In this section, according to the proposed framework, we briefly review the related work in terms of OD and distribution regression.

A. ORBIT DETERMINATION

OD is an old problem with hundreds of years of history that existed throughout the development of astronomy. G. F. Gronchi [4] proposed an OD method for modern asteroid surveys using several different short arcs collected in different nights, based on the first integrals of Kepler's motion. However, in satellite OD problem, many differences from astronomy exist. OD can be divided into IOD and POD. Much literature is devoted to improving the classical OD methods to fit different missions [5], to compute with high accuracy [2], to solve the problem with shorter angle-only observation arc [6], [7], to consume less time [8], and so on. However, in practice, with its weak observation geometry, which is different from that of low earth orbit (LEO), the results of objects in GEO can easily converge to trivial solutions. In short-arc IOD, especially for the angle-only problem, many of the state-of-the-art methods show low robustness to noise. Meanwhile, short-arc and fragmentary observations make it difficult to perform POD through a Kalman filter. In [9], F. M. Fadrique compared three different IOD algorithms: Gauss, Baker-Jacoby and Gooding algorithm to illustrate their different performances on wide variety of orbital scenarios. In recent years, D. P. Lubey *et al*. [10] proposed a new IOD method based on Polynomial Chaos algorithm without an initial guess of the object's state. Instead, users need to provide bounds on the size of the object's possible orbit and some parameters. Additionally, [11] generally introduced the development and contributions of the first Space-Based Visible (SBV) program. T. Flohrer *et al*. [12] discussed the surveillance performance of a single small space-based optical instrumentation and illustrated the relationship between OD results and astrometric accuracy. Since the pinch point regions in GEO belt proposed by the US Lincoln Laboratory are becoming more and more widely, Y.P Hu *et al*. [13] proposed a pseudo-fixed latitude observation mode. With more and more observations obtained from space assets, it is feasible to estimate orbital elements from a data-driven perspective. In these cases, the proposed OD method in this paper fills the gap between IOD and POD. It performs much more robustly and accurately than classical IOD methods. Compared with POD, this method does not need an initial value and performs well in the short-arc problem. [14] also made contributions to explore the applications of kernel mean embedding in OD, but our work is different from their contributions in three main aspects. First, we put forward the structure of angle-only OD samples and show the effectiveness of the proposed weighting matrix in improving the accuracy of OD. Second, and more importantly, to save computing cost, we propose a RIP-based ROMP algorithm to pursue a sparse solution. This is valuable since it extends the application domain of the WDR-OD method. Finally, the proposed method is experimentally compared with the improved Laplace method with different noises.

B. DISTRIBUTION REGRESSION

Embedding the distribution to a Hilbert space is an efficient approach to perform DR, which introduces kernels between the mapping, and utilizes a kernel machine to solve the regression or classification problem. These methods can be divided into two parts, parametric and nonparametric. Parametric methods usually assume a probabilistic density and fit the data with a parametric family (a family of objects depending on a set of parameters to define), such as the Weibull translation model, the finite Gaussian mixture model and the exponential model [15]–[17], and generalization will

occur by introducing some new family, such as in [18]. However, when the true probability distribution does not belong to the assumed parametric family, biases are unavoidable. In contrast, nonparametric methods do not rely on the data belonging to any specific parametric family. Extensively used nonparametric methods include the support vector machines (SVMs), the decision tree method, neural networks, etc. Recently, a popular method referring to DR is mapping the distribution to a reproducing kernel Hilbert space (RKHS) through kernel mean embedding, proposed by Smola *et al*. [19]. The kernel approach has indicated successful in many ML applications. The idea behind kernel mean embedding is to extend the feature map to the probability distribution space. In this way, the kernel mean captures all features about the distribution P, and also, a defined metric over the probability distribution space will be introduced. In this framework, [20] generalized the regularization theorem to the probability distribution space. [21] developed a distribution-free method for DR through the kernel- kernel estimator and proved the upper bound on the risk of the estimator. However, this approach is restricted to compact domains of finite dimensional Euclidean spaces. More recently, Szabó Zoltán *et al*. [22] proposed a two-stage sampled theory with the following two contributions. First, this theory generalized the regression in any probability measure defined on a separable, topological domain endowed with kernels, where the probability measure is unobservable, instead, we can only observe samples from the probability measure. In other words, it regresses a real-valued response variable $\mathbf{y}_i$ from samples $\mathbf{x}_{i,j}, j = 1, \ldots, N_i$, where the samples are not random, but governed by an unknown probability measure P_i. Second, it worked directly on distribution embedding rather than estimating the probability density in high dimensional spaces. This theory lays a theoretical foundation for the following research in this study.

III. OUTLINE

In this study, we solve the space-based OD problem with a distribution regression technique. First, we propose the structure of space-based orbital samples with a weighting matrix. Second, the regression is performed in two-step mapping: the samples from distribution are first mapped to a RKHS H through kernel mean embedding μ, and then they are mapped to the response variables through the linear kernel K. Third, to reduce the computing load, we propose a RIP-based ROMP algorithm to pursue a sparse solution. Finally, to demonstrate the efficiency of this technique in OD, experiments are performed with simulated space objects in the GEO belt, and surveillance sensors in LEO orbits. For a better comparison, we introduce the improved Laplace method to show the different features with measurements in WGN and colored noise.

IV. ASSUMPTIONS

In this section, we detail the notations and assumptions on the existing distribution regression work. We define the following: $(\mathcal{X}, \tau)$ is a topological space, $\mathcal{B}(\mathcal{X}) := \mathcal{B}(\tau)$ is the Borel σ-algebra induced by the topology τ.$\mathcal{M}_1^+(\mathcal{X})$ is the set of Borel probability measures defined on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and $\tau_w = \tau_w(\mathcal{X}, \tau)$ is a weak topology defined on $\mathcal{M}_1^+(\mathcal{X})$. Let $H = H(k)$ be a reproducing kernel Hilbert space (RKHS) with nonlinear kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and let $\mathcal{H} = \mathcal{H}(K)$ be the other RKHS with linear kernel $K : X \times X \to \mathbb{R}$. Let X denote a subset of H, which is constructed with the kernel mean embedding of the distributions,

$$X = \mu(\mathcal{M}_1^+(\mathcal{X})) = \left\{\mu_x : x \in \mathcal{M}_1^+(\mathcal{X})\right\} \subseteq H \tag{1}$$

where μ_x denotes the mean embedding of the distribution $x \in \mathcal{M}_1^+(\mathcal{X})$ to the RKHS $H = H(k)$ determined by kernel k.

The detailed assumptions are described as follows:

(1) $(\mathcal{X}, \tau)$ is a separable topological space.

(2) kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is bounded and continuous: $\exists B_k < \infty$ satisfies

$$\sup_{\mathbf{x}\in\mathcal{X}} k(\mathbf{x}, \mathbf{x}) \leq B_k$$

(3) kernel $K : X \times X \to \mathbb{R}$ is bounded (take $\exists\ B_K < \infty$ as bound) and $\Phi(\mu_a) := K(\cdot, \mu_a) : X \to \mathcal{H}$ is Hölder continuous, that is, $\exists L > 0$ and $h \in (0, 1]$, for $\forall(\mu_a, \mu_b) \in X \times X$ satisfies

$$\|\Phi(\mu_a) - \Phi(\mu_b)\|_{\mathcal{H}} \leq L \|\mu_a - \mu_b\|_H^h$$

(4) $\mathbf{y}$ is bounded: $\exists B_{\mathbf{y}} < \infty$ satisfies $|\mathbf{y}| \leq B_{\mathbf{y}}$.

(5) $X = \mu(\mathcal{M}_1^+(\mathcal{X})) \in \mathcal{B}(H)$.

V. AN ORBIT DETERMINATION METHOD MODEL BASED ON WEIGHTING DISTRIBUTION REGRESSION

With ML, the distribution regression problem can be solved by learning samples $\mathbf{Z} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$ with $\mathbf{x}_i \in \mathcal{M}_1^+(\mathcal{X})$ and $\mathbf{y}_i \in \mathbb{R}$. The goal is to predict a new $\mathbf{y}_{l+1}$ from a new batch of samples $\mathbf{x}_{l+1}$, where l is the training sample size. For the space-based OD problem, the feature variable $\mathbf{x}_i = \left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}$ indicates the set of sensors' angular measurements during the observed arc to the i th object, and the position vectors of sensors, where N_i is the total quantity of observed points with respect to this object. The response variable $\mathbf{y}_i$ indicates the set of orbital elements of the i th object, where $\mathbf{y}_i \in \mathbb{R}^{1\times e}$, and e denotes the number of orbital elements for each object. Here, the object's position vector $\mathbf{r}_0$ and velocity vector $\mathbf{v}_0$ at the initial observation time t_0 are used to denote $\mathbf{y}_i$. In fact, for certain observers, the values of all the measurements during the observation depend not only on the geometrical relationship but also on the dynamic model with perturbations, because it is the dynamic model and its perturbations that govern where the object will appear next, and further impact its measurement. Regardless of how they are described in the analytical model, we assume these measurements are *i.i.d* (independent and identically distributed) and sampled from an unknown distribution P_i.

Briefly, the two-stage sample [22] can be described as follows: first, the distributions $P_i, i = 1, \ldots, n$ are *i.i.d* sampled from a meta-distribution $\mathcal{M}$ defined on a measurable space

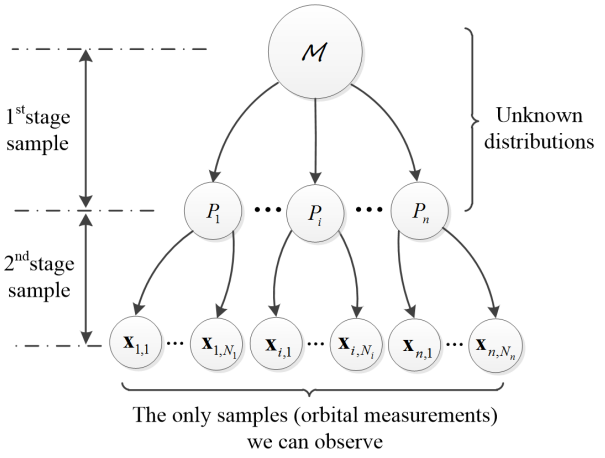


FIGURE 1. Schematic of the two-stage sample.

$\left(\mathcal{M}_1^+(\mathcal{X}) \times \mathbb{R}, \mathcal{B}(\tau_w) \otimes \mathcal{B}(\mathbb{R})\right)$; second, the observed samples $\mathbf{x}_{i,j}, j = 1, \ldots, N_i$ are *i.i.d* sampled from distribution P_i, as shown in Figure 1. We denote this process by:

$$\begin{cases} P_1, \ldots, P_n \overset{i.i.d}{\sim} \mathcal{M} \\ \mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,N_i} \overset{i.i.d}{\sim} P_i \end{cases}$$

Note that the samples $\mathbf{x}_{i,j}, j = 1, \ldots, N_i$ are the only ones that can be observed, which represent the orbital measurements. Beyond that, distributions $\mathcal{M}$ and P_i are both unknown.

Based on the preceding assumptions, the kernel mean embedding of probability measures in $\mathcal{M}_1^+(\mathcal{X})$ into $X \subseteq H(k)$ is defined by a mapping as follows [23]:

$$\mu_x : \mathcal{M}_1^+(\mathcal{X}) \to X \subseteq H(k), \quad P \to \int k(x, \cdot) \mathrm{d}P(x)$$

The integral above is a Bochner integral, which is defined for functions in Banach space.

More importantly, because we do not have access to the true distribution P, and thereby we must rely on the *i.i.d* sample $\mathbf{X} = \left\{\{\mathbf{x}_{i,n}\}_{n=1}^{N_i}\right\}_{i=1}^{l}$, the empirical estimation of kernel mean embedding μ_x can be denoted by $\hat{\mu}_x$:

$$\hat{\mu}_{x_i} := \frac{1}{N_i} \sum_{n=1}^{N_i} k(\mathbf{x}_{i,n}, \cdot) \tag{2}$$

Obviously, $\hat{\mu}_x$ is an unbiased estimation of μ_x, and $\hat{\mu}_x \to \mu_x$ when $l \to \infty$.

Generally, the OD regression process can be divided into two steps. First, the measurements $\mathbf{x}_i$ are mapped to $X \subseteq H$ through the empirical mean embedding $\hat{\mu}_x$. Second, the results in X are then mapped to $\mathbf{y}_i \in \mathbb{R}^{1\times e}$ by the linear kernel K and the function $f \in \mathcal{H} = \mathcal{H}(K)$, as shown in Figure 2.

Given samples $\mathbf{Z} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{l}$, the classical regression with regularization approach is a supervised method where

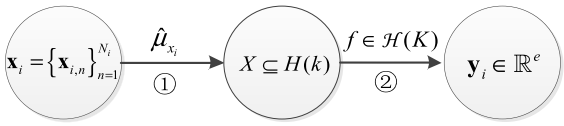


FIGURE 2. The two steps of OD regression process.

we solve the loss function as,

$$f^* = \arg\min_{f \in \mathcal{H}(K)} \frac{1}{l} \sum_{i=1}^{l} (\mathbf{y}_i - f(\mathbf{x}_i))^2 + \gamma \|f\|_{\mathcal{H}}^2 \tag{3}$$

where $\gamma \|f\|_{\mathcal{H}}^2$ is the regularization term, γ is the regular coefficient, and l is the training sample size. $\mathbf{y}_i$ represents the true value and $f(\mathbf{x}_i)$ represents the predicted value. In DR, owning an empirical estimation of the kernel mean embedding, the loss function above can be rewritten as:

$$f^* = \arg\min_{f \in \mathcal{H}(K)} \frac{1}{l} \sum_{i=1}^{l} \left(\mathbf{y}_i - f\left(\hat{\mu}_{x_i}\right)\right)^2 + \gamma \|f\|_{\mathcal{H}}^2 \tag{4}$$

According to the Representer Theorem proven by the reproducing property of RKHS, the function in RKHS $\mathcal{H}(K)$ can be given in the following expression,

$$f^* = \sum_{i=1}^{l} \alpha_{m,i}^* K\left(\cdot, \hat{\mu}_{x_i}\right) \tag{5}$$

where the symbol "$*$" denotes the optimal solution. To determine the value of $\alpha_{m,i}^* \in \mathbb{R}$, we assume a coefficient vector $\boldsymbol{\alpha}_m = [\alpha_{m,1}, \ldots, \alpha_{m,l}]^{\mathrm{T}}$, and a coefficient matrix $\mathbf{A}$ consisting of the $\boldsymbol{\alpha}_m$, where $m = 1, \ldots, e$, and $\mathbf{A}=[\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_e]$. Introducing the form of f^* above to Eq. (4), we arrive at the following convex differentiable objective function of coefficient matrix $\mathbf{A}$,

$$\mathbf{A}^* = \arg\min_{\mathbf{A}} \left\{ \frac{1}{l} (\mathbf{Y} - \mathbf{G}\mathbf{A})^T (\mathbf{Y} - \mathbf{G}\mathbf{A}) + \gamma \mathbf{A}^T \mathbf{G} \mathbf{A} \right\} \tag{6}$$

where $\mathbf{A}^*=[\boldsymbol{\alpha}_1{}^*, \ldots, \boldsymbol{\alpha}_e{}^*]$, $\mathbf{Y}=[\mathbf{y}_1, \ldots, \mathbf{y}_l]^T$ is the response matrix, and $\mathbf{G} = [K(\hat{\mu}_{x_i}, \hat{\mu}_{x_j})] \in \mathbb{R}^{l\times l}$ is the Gram matrix. For the linear kernel K,

$$K(\hat{\mu}_{x_i}, \hat{\mu}_{x_j}) = \langle \hat{\mu}_{x_i}, \hat{\mu}_{x_j} \rangle = \frac{1}{N_i N_j} \sum_{n,m=1}^{N_i, N_j} k(\mathbf{x}_{i,n}, \mathbf{x}_{j,m}) \tag{7}$$

where N_i and N_j are total quantity of observed points respect to the i th object and the j th object, respectively. It reflects another advantage of kernel mean embedding, that is, it does not restrict the size of every sample bag to be equal, which is crucial in OD, since the number of points observed varies from object to object in practice.

Taking the derivative of the objective function in Eq. (6) with respect to the coefficient matrix $\mathbf{A}$, we can get the expression for the optimal coefficient matrix,

$$\mathbf{A}^* = (\mathbf{G} + \gamma l \mathbf{I}_{l\times l})^{-1} \mathbf{Y} \tag{8}$$

Substituting Eq. (8) into Eq. (5), we can derive the analytical solution of the prediction for new test distributions, according to the given samples $\mathbf{Z}$,

$$\mathbf{y}_t = f^*(\mu_t) = \mathbf{g}(\mathbf{G} + \gamma l \mathbf{I}_{l\times l})^{-1}\mathbf{Y} \tag{9}$$

where $\mathbf{g} = [K(\hat{\mu}_{x_1}, \mu_t), \ldots, K(\hat{\mu}_{x_l}, \mu_t)] \in \mathbb{R}^l$ is calculated by the inner product of every training $\hat{\mu}_{x_l}$ and test μ_t, and the subscript t denotes the test data.

The derivation process above is solved through the loss function in Eq. (4) with a L_2 regularization term, which is the well-known ridge regression.

So far, the OD model based on distribution regression has been built. However, the space-based OD problem with angle-only measurements has some specific issues. The feature variables consist of angular measurements and positional measurements (the position vectors of observers). As is known, angular measurements play a very important role in short-arc and angle-only OD results. Even a few seconds of arc errors in angular measurements will lead to a hundreds-kilometer deviation in OD results. By contrast, the deviation caused by hundred-meter errors of the observer positions will remain on the order of tens of kilometers [1]. However, since the value of position vectors of LEO observers are several orders of magnitude larger than the angle values, their effect is lessened. To improve it, we introduce a weighting matrix into the measurements, which is described as follows,

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_a & 0 & \cdots & & 0 \\ 0 & \omega_a & & 0 & \vdots \\ \vdots & & \ddots & & \\ & 0 & & \omega_p & 0 \\ 0 & \cdots & & 0 & \omega_p \end{bmatrix} \tag{10}$$

The utilized OD sample set with the weighting matrix has the following form,

$$\mathbf{Z} = \left\{ \boldsymbol{\omega} \cdot \left\{ \mathbf{x}_{i,n} \right\}_{n=1}^{N_i}, \mathbf{y}_i \right\}_{i=1}^{l} \tag{11}$$

In the matrix, $\omega_a > 1$ is the angle weighting coefficient, and $\omega_p > 0$ is the position weighting coefficient. The physical significance of the weighting matrix is to increase the weight of angular measurements, as in the real OD dynamic model. In the following experiment section, we will calculate and show the changes in the RMSE of OD with the increase of ω_a, and choose the candidate interval for ω_a. Because ω_a has more important physical significance than ω_p, the following work focuses on the effects of ω_a, and assigns ω_p as a constant. In fact, we find that ω_a plays the role of regulating the weight between the angular measurements and the positional measurements. If more measurement types are introduced, such as range, ω_p can be introduced at that point to analyze their interactions.

VI. SPARSE SOLUTIONS OF THE WDR-OD

We note that the solution of ridge regression is a full representation of the Gram matrix $\mathbf{G} \in \mathbb{R}^{l\times l}$, the size of which depends on the dimension of the training data. Based on the ridge regression, WDR-OD can obtain higher regression accuracy, but when processing a large amount of training data, memory overflow may result, along with heavy computing resource consumption. Note that Eq. (5) can be rewritten in a vector form as follows,

$$\mathbf{Y} = f^*(\hat{\mu}_x) = \mathbf{G} \cdot \mathbf{A}^* \tag{12}$$

where $\mathbf{A}^*$ is calculated through Eq. (8), based on the ridge regression with the L_2 regularization term.

To solve the preceding problem, our idea is to find the sparse solution $\tilde{\mathbf{A}}$ of the coefficient matrix $\mathbf{A}^*$. Sparsity means most nonzero coefficients of the matrix shrink to zeros. Our aim is to shrink as many as possible nonzero coefficients of $\mathbf{A}^*$ to zeros under the premise of OD accuracy. In this way, we are able to store and call the elements of the large matrix selectively, corresponding to the nonzero coefficients in $\tilde{\mathbf{A}}$.

Consider the vector $\tilde{\boldsymbol{\alpha}}_m$ as one of the columns of the coefficient matrix $\tilde{\mathbf{A}}$. In the sparse representation theory, to obtain the sparse solution is to find a suitable "dictionary" $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_l]$, where $\mathbf{d}$ is called an atom, and for a sparse solution $\tilde{\boldsymbol{\alpha}}_m$, let $\mathbf{D} \cdot \tilde{\boldsymbol{\alpha}}_m$ close to $\boldsymbol{\alpha}_m$, that is

$$\min_{\tilde{\boldsymbol{\alpha}}} \left\{ \|\boldsymbol{\alpha}_m - \mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2 \right\} \quad \text{s.t. } \forall m \in 1, \ldots, e, \ \|\tilde{\boldsymbol{\alpha}}_m\|_0 \le S_0 \tag{13}$$

where $\|\cdot\|_0$ denotes the L_0 norm, and S_0 is the sparsity of $\tilde{\boldsymbol{\alpha}}_m$. By solving Eq. (13), we obtain the sparse solution $\tilde{\mathbf{A}} = [\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_e]$. Substituting $\mathbf{A}^* \approx \mathbf{D} \cdot \tilde{\mathbf{A}}$ into Eq. (12), we obtain,

$$\mathbf{Y} = \mathbf{G} \cdot \mathbf{A}^* \approx \mathbf{G} \cdot \mathbf{D}\tilde{\mathbf{A}} = \boldsymbol{\Lambda} \cdot \tilde{\mathbf{A}} \tag{14}$$

where $\boldsymbol{\Lambda}$ is called the sensing matrix. Comparing Eq. (14) with Eq. (12), we can replace $\mathbf{G}$ with $\boldsymbol{\Lambda}$, and replace $\mathbf{A}^*$ with $\tilde{\mathbf{A}}$. In this way, we no longer need to store the large matrix $\mathbf{G}$; instead, we need only to store and call the very few elements in $\boldsymbol{\Lambda}$ corresponding to the nonzero coefficients in $\tilde{\mathbf{A}}$.

Note that solving Eq. (13) is NP hard. Classical orthogonal matching pursuit (OMP) is an iterative greedy algorithm which is efficient and has been widely applied to the sparse representation of a high-dimensional signal. Regularized orthogonal matching pursuit (ROMP) [24] algorithm selects the first S_0 largest atoms at one time, that further improves the efficiency of OMP.

In OMP or ROMP, the sparsity S_0 is considered as prior knowledge, but it is unknown in our problem. We try to select S_0 as small as possible, but the lower bound is unknown. The sparsity adaptive matching pursuit (SAMP) algorithm proposed in [25] does not need the sparsity as the prior knowledge, instead, it uses a fixed step-size s to reconstruct the original signal. But different values of s may lead to different performance in efficiency and accuracy. It also lacks theoretical support. Hence, we utilize the restricted isometry property (RIP) [26], [27] and propose a RIP-based ROMP algorithm. First, we use RIP to theoretically restrict the sparsity S_0, and then, we perform ROMP under this constraint.

To state our method, we first recall the concept of RIP.

Definition: For any $\tilde{\mathbf{x}}$ *with sparsity* S_0*, define the isometry constant* $\delta \in (0, 1)$ *of a matrix* Φ *as the smallest number such that*

$$(1-\delta)\|\tilde{\mathbf{x}}\|_F^2 \le \|\Phi\tilde{\mathbf{x}}\|_F^2 \le (1+\delta)\|\tilde{\mathbf{x}}\|_F^2 \tag{15}$$

holds for all $\tilde{\mathbf{x}}$ *with sparsity* S_0.

For the problem in our study, Eq. (15) can be rewritten as:

$$-\delta\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \le \left|\|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2 - \|\tilde{\boldsymbol{\alpha}}_m\|_2^2\right| \le \delta\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \tag{16}$$

where $m = 1, \ldots, e$.

The term $\left|\|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2 - \|\tilde{\boldsymbol{\alpha}}_m\|_2^2\right|$ in Eq. (16) can be further derived as follows:

$$\begin{aligned}
&\left|\|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2 - \|\tilde{\boldsymbol{\alpha}}_m\|_2^2\right| \\
&= \left|\left\|\sum_{j=1}^{l}\mathbf{d}_j\tilde{\alpha}_{mj}\right\|_2^2 - \sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|^2\right| \\
&= \left|\sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|^2 + \sum_{j\neq q}\tilde{\alpha}_{mj}\tilde{\alpha}_{mq}\left\langle\mathbf{d}_j, \mathbf{d}_q\right\rangle - \sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|^2\right| \\
&= \left|\sum_{j\neq q}\tilde{\alpha}_{mj}\tilde{\alpha}_{mq}\left\langle\mathbf{d}_j, \mathbf{d}_q\right\rangle\right| \le \zeta\sum_{j\neq q}\left|\tilde{\alpha}_{mj}\tilde{\alpha}_{mq}\right| \\
&= \frac{1}{2}\zeta\left(\left(\sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|\right)^2 - \sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|^2\right) \\
&\le \frac{1}{2}\zeta(S_0-1)\sum_{j=1}^{l}\left|\tilde{\alpha}_{mj}\right|^2 \\
&= \frac{1}{2}\zeta(S_0-1)\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 < \zeta(S_0-1)\|\tilde{\boldsymbol{\alpha}}_m\|_2^2
\end{aligned} \tag{17}$$

where $\zeta = \max_{q\neq j}\left|\left\langle\mathbf{d}_j, \mathbf{d}_q\right\rangle\right|$.

Substituting Eq. (17) into Eq. (16), we obtain,

$$-\delta\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \le \zeta(S_0-1)\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \tag{18}$$

that is,

$$\frac{\zeta-\delta}{\zeta} \le S_0 \tag{19}$$

In this way, we derive the lower bound of sparsity S_0.

In ROMP algorithm, it selects the first S_0 largest atoms in $\left|\left\langle\mathbf{d}_j, \varepsilon_{n-1}\right\rangle\right|$ at one time, rather than one biggest atom as OMP did. The complete RIP-based ROMP procedure is shown in Table 1. Step (4) is the regularized procedure. In step (5), $\mathbf{D}_{\hat{\Gamma}_n}$ denotes the atoms (columns) in $\mathbf{D}$ corresponding to the index $\hat{\Gamma}_i$. In step (6), $(\mathbf{Sup}_n^T\cdot\mathbf{Sup}_n)^{-1}\cdot\mathbf{Sup}_n^T\cdot\boldsymbol{\alpha}_m$ is the least square solution of

$$\arg\min_{\tilde{\boldsymbol{\alpha}}_i}\left\|\boldsymbol{\alpha}_i - \mathbf{Sup}_n\cdot\tilde{\boldsymbol{\alpha}}_i\right\| \tag{20}$$

In this way, the RIP integrates into ROMP algorithm, and the sparsity S_0 in prior knowledge is replaced by a step size t. There is no additional constraint on the step size t. In practice, we recommend setting t to 1 in the initial step so that the optimal sparsity value will not be missed. Although we also need to provide a value of t first, by contrast, it is much less restrictive. Note that, the procedure above only involves one column vector $\tilde{\boldsymbol{\alpha}}_m$ of $\tilde{\mathbf{A}} = [\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_e]$, repeat this procedure for $m = 1, \ldots, e$ we arrive at the sparse solution $\tilde{\mathbf{A}}$.

In the problem addressed in this study, we first determine the optimal solution $\mathbf{A}^*$ according to the preceding WDR-OD method; then, we employ the RIP-based ROMP algorithm to pursue the sparse solution $\tilde{\mathbf{A}}$. However, since $\mathbf{D}\cdot\tilde{\mathbf{A}}$ is approximately equal to $\mathbf{A}^*$, the sparse representation will lead to an OD accuracy loss to some extent. Nevertheless, the sparse solution is still valuable, since it extends the application domain of the WDR-OD method to the case with

TABLE 1. The RIP-based ROMP algorithm.

Input: $\boldsymbol{\alpha}_m$, $\mathbf{D}$, step size t

(1) Initialize: residual $\varepsilon_0 = \boldsymbol{\alpha}_m$, expected residual $\hat{\varepsilon}$, support vector $\mathbf{Sup}_0 = \varnothing$, index set $\Gamma_0 = \varnothing$ and counter variable n=0.

(2) n=n+1.

(3) Calculate $\mathbf{u} = \left|\left\langle\mathbf{d}_j, \varepsilon_{n-1}\right\rangle\right|, j = 1,...,l$, find the first t largest atoms in $\mathbf{u}$, and save the corresponding index of these atoms in Γ_n.

(4) Find a subset $\hat{\Gamma}_n$ in Γ_n, which satisfies $|\mathbf{u}(k)| \le 2|\mathbf{u}(v)|$ for $\forall k, v \in \hat{\Gamma}_n$. On this basis, find the $\hat{\Gamma}_n$ with maximal energy, that is $\max\sum|\mathbf{u}(k)|^2, k \in \hat{\Gamma}_n$.

(5) Update $\mathbf{Sup}_n$ with $\mathbf{Sup}_n = \mathbf{Sup}_{n-1} \cup \mathbf{D}_{\hat{\Gamma}_n}$ according to the atoms corresponding to the index in $\hat{\Gamma}_i$.

(6) $\tilde{\boldsymbol{\alpha}}_m = (\mathbf{Sup}_n^T\cdot\mathbf{Sup}_n)^{-1}\cdot\mathbf{Sup}_n^T\cdot\boldsymbol{\alpha}_m$, $\zeta = \max_{q\neq j}\left|\left\langle\mathbf{d}_j, \mathbf{d}_q\right\rangle\right|$, where $\mathbf{d}_j, \mathbf{d}_q \in \mathbf{Sup}_n$.

(7) If $(1-\delta \le \|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2/\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \le 1+\delta)$ && $(\frac{\zeta-\delta}{\zeta} \le t)$
t=t+1;
update residual $\varepsilon_n = \boldsymbol{\alpha}_m - \mathbf{Sup}_n\cdot\tilde{\boldsymbol{\alpha}}_m$;
jump to step (2) and continue.
else if $(1-\delta \le \|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2/\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \le 1+\delta$ is false $||$ $\frac{\zeta-\delta}{\zeta} > t)$ && $(t \neq 1)$
$t = t-1$;
update residual $\varepsilon_n = \boldsymbol{\alpha}_m - \mathbf{Sup}_n\cdot\tilde{\boldsymbol{\alpha}}_m$;
jump to step (3) and continue.
else if $(1-\delta \le \|\mathbf{D}\tilde{\boldsymbol{\alpha}}_m\|_2^2/\|\tilde{\boldsymbol{\alpha}}_m\|_2^2 \le 1+\delta$ is false && $\frac{\zeta-\delta}{\zeta} > t)$ && (n>1)
if $\varepsilon_n > \varepsilon_{n-1}$ $||$ $\varepsilon_n \le \hat{\varepsilon}$
terminate;
else
jump to step (2) and continue;
end
end

Output: sparsity S_0=t and sparse solution $\tilde{\boldsymbol{\alpha}}_m$.

TABLE 2. The ranges of orbital elements of candidate objects in GEO.

Orbit Epoch	Semimajor axis a/km	Eccentricity e	Inclination i /°	RAAN Ω /°	Argument of Perigee ω /°	True Anomaly θ /°
27 Mar 2020 00:00:00 UTC	[41000,44000]	[0,0.1]	[0,10]	[0,360)	[0,360)	[0,360)

TABLE 3. The orbital elements of the five observation satellites.

Observation Satellites	Orbit Epoch	Semimajor Axis a/km	Eccentricity e	Inclination i /°	RAAN Ω /°	Argument of Perigee ω /°	True Anomaly θ /°
Sat_1	27 Mar 2020 00:00:00 UTC	7078.14	0	98.19	184.88	0	0
Sat_2	27 Mar 2020 00:00:00 UTC	7078.14	0	98.09	256.87	0	71.91
Sat_3	27 Mar 2020 00:00:00 UTC	7078.14	0	98.15	328.85	0	143.79
Sat_4	27 Mar 2020 00:00:00 UTC	7078.14	0	98.28	40.86	0	215.80
Sat_5	27 Mar 2020 00:00:00 UTC	7078.14	0	98.30	112.87	0	287.93

limited computing resources, e.g., performing the WDR-OD method onboard for an online OD mission. It means that, the time-consuming training process of WDR-OD can be carried out in the ground center in advance, and then the obtained sparse solution with its dictionary can be uploaded to space-based surveillance network to further perform the WDR-OD utilizing observed measurements in a real-time way.

VII. EXPERIMENTS AND DISCUSSION

In this section, we attempt to perform experiments in different environments through simulations. First, as a preliminary work before estimating the orbit, we simulate 14520 random objects as a data set, covering the majority of GEO belt. The ranges of orbital elements for these candidate objects are shown in Table 2. In System Tool Kit (STK) software, we set the propagator as the HPOP model, considering earth gravity model 21×21 degree, three-body gravity and solar radiation pressure. Let the coefficient Cr=1 and the area-mass ratio be 0.02 m^2/kg. We simulate five nearly uniformly distributed satellites in LEO to act as observers to obtain angle-only measurements of these objects. The orbits of the five observation satellites are shown in Table 3. We set the field of view size to 8°×8° and the sampling interval to 5 seconds. In the simulation, the time span of the observed orbit arc is random but is no more than 18 minutes in practice. For every candidate object, there are 1∼5 observable satellites during the observation arc, which is determined by their geometrical relationships. Each response variable (object's orbit) and corresponding feature variable (measurements) constitutes a sample 'bag'; that is, the i th 'bag' consists of the i th object's orbit and its measurements. It has the form $\mathbf{Z}_i = \left\{\boldsymbol{\omega} \cdot \left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}, \mathbf{y}_i\right\}$, where $\left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}$ indicates the set of angular measurements and $\mathbf{y}_i$ indicates the orbit of the i th object at the initial observation time t_0, which is denoted by a 1×6 vector consisting of the position and velocity vectors of the object in inertial system. Let Q_i denote the number of observable satellites and N_i denote the number of observation points during the orbit arc in the i th bag. Therefore, the first $2 \times Q_i$ columns of the $\left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}$ are the angular measurements, and the last $3 \times Q_i$ columns are the corresponding observation satellite position vectors. Each row contains all of the observed values of one point. The number of rows in $\left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}$ is N_i. Note that the $\left\{\mathbf{x}_{i,n}\right\}_{n=1}^{N_i}$ of different bags may have different sizes (i.e., different arc lengths and numbers of observation satellites). All of the 14520 'bags' are order disrupted; that is, each of the 'bags' is introduced to the WDR-OD method in a random order.

Note that, in the response variables, we use the position and velocity vectors of the object in the J2000.0 Earth Centered Inertial system to describe the orbit instead of classical orbital elements.

The Cauchy kernel is employed in our work because of its long-range influence and sensitivity over the high dimension space. More information on the kernel can be found in [28].

A. WEIGHTING MATRIX EXPERIMENT

In the following work, we carry out the experiment with different angle weighting coefficients ω_a in 5-fold cross validation (CV). The relationship between the root-mean-square error (RMSE) and ω_a is shown in Figure 3. Let ω_a be a random value in the [1], [25] interval. The overall trend is that the RMSE decreases significantly when ω_a is within the [1], [5] interval, which indicates the beneficial effect of introducing the angle weighting coefficient. However, with the increase of ω_a, the RMSE shows some fluctuations. In fact, it is difficult and unnecessary to find a specific value of ω_a to fit all the OD problems; instead, we search for a common candidate interval for ω_a wherein the relatively small RMSE can be obtained in every fold of CV. So, from this perspective,

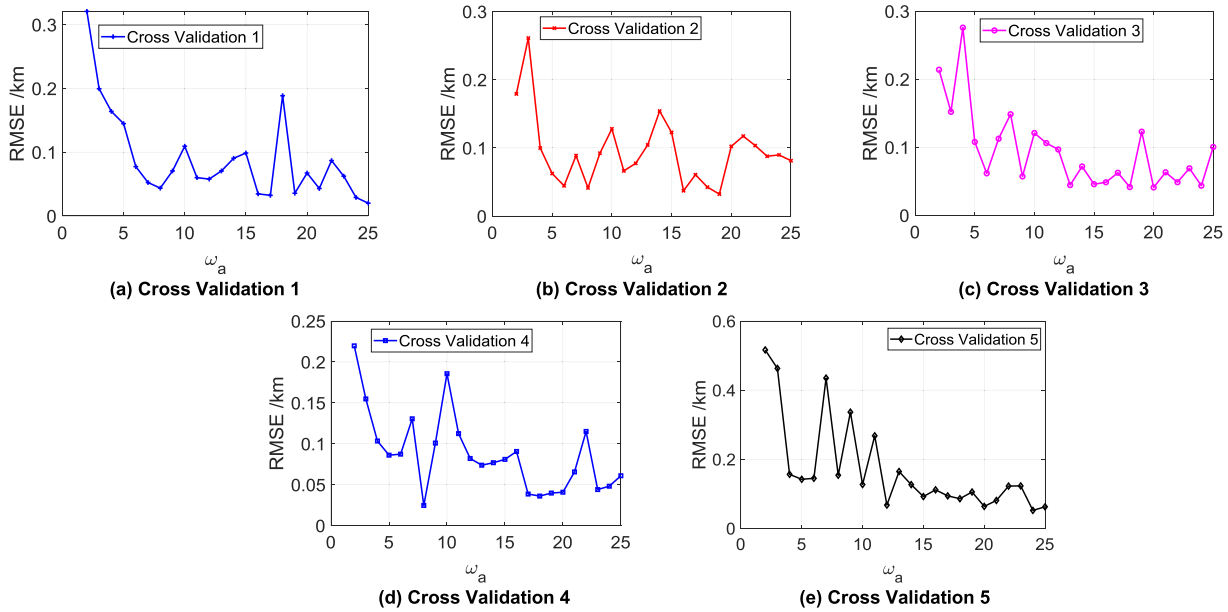


FIGURE 3. The tendency of RMSE with different ω_a in each cross validation.

according to the result in Figure 3, we use the candidate interval $\omega_a \in [15, 20]$ for the following experiments. Compared with the RMSE without weighting coefficients (equivalent to $\omega_a = 1$), the RMSE values within the $\omega_a \in [15, 20]$ significantly decrease.

B. THE WDR-OD METHOD EXPERIMENT AND DISCUSSION

Then, we can perform the WDR-OD method from Eq. (2) to Eq. (9), based on the constructed sample set $\mathbf{Z}$ with the form in Eq. (11) and the candidate interval of ω_a. First, we implement Eq. (9) derived by the L_2 regularization in the loss function. A 5-fold cross validation is employed in the following experiments. In each fold, we randomly draw 80% of the samples for training, 10% of the samples for validation and the remaining 10% of the samples for test. Figure 4 shows the OD errors in position and velocity vectors from noise-free measurements. The horizontal axis denotes the number of test objects (for the sake of visualization, it only shows the first 300 test objects). Note that the most errors in the X, Y and Z directions are concentrated within the $[-0.2$ km, 0.2 km$]$ interval. With respect to the velocity, few objects' errors are greater than 2×10^{-5} km/s; the results in V_z are even better, within the $[-0.1$ km, 0.1 km$]$ interval, except that a few errors reach a peak of 2×10^{-5} km/s.

Next, in order to test the performance in a noisy environment, we add two different kinds of noise to the measurements. White noise is a random noise with the same energy density at all frequencies, and white Gaussian noise (WGN) is a specific white noise with an amplitude represented by the Gaussian distribution. First, we add WGN ε into the measurements with a variance $(2'')^2$ (seconds of arc) in the angular measurements and $(0.1$ km$)^2$ in the positional measurements (1σ), where $\varepsilon \sim N(0, \sigma^2)$. However, the noise in practical observation is much more complex than WGN, wherein the measurements are full of colored noise, whose power spectral density is not constant. Thus, second, we model colored noise with an autoregressive-moving average (ARMA) process [29]. ARMA is a combination of the autoregressive (AR) process and the moving average (MA) process, where the AR process performs the autocorrelation of noise in a time series, and MA is a linear combination of WGN. The model of the ARMA process can be described as follows,

$$\xi_n = \varepsilon_n + \sum_{i=1}^{p} \lambda_i \xi_{n-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{n-j} \tag{21}$$

where ξ is the colored noise in the time series, and the subscript n denotes the nth moment. λ and θ are coefficients in the AR and the MA processes, respectively, and p and q denote the order of the AR and MA processes, respectively. In this paper, we employ a second-order ARMA process. Let $\lambda_1 = 0.35$, $\lambda_2 = 0.55$, $\theta_1 = 0.5$ and $\theta_2 = 0.6$.

The corresponding OD errors with WGN measurements are shown in Figure 5. Errors of most of the test objects in the X and Y directions are within the $[-1$ km, 1 km$]$ interval, except for a few errors that reach a peak of 2 km. The concentrated interval in the Z direction is $[-0.5$ km, 0.5 km$]$. Additionally, the velocity errors of most of the objects are less than 1×10^{-4} km/s. The RMSE of the OD results of different folds of cross validation are shown in Table 4, where the mean values are shown in bold font. Note that in the noise-free experiment, the RMSE of the position vectors remains on the order of tens of meters, and the RMSE of the velocity vectors performs even better. It also shows a few differences

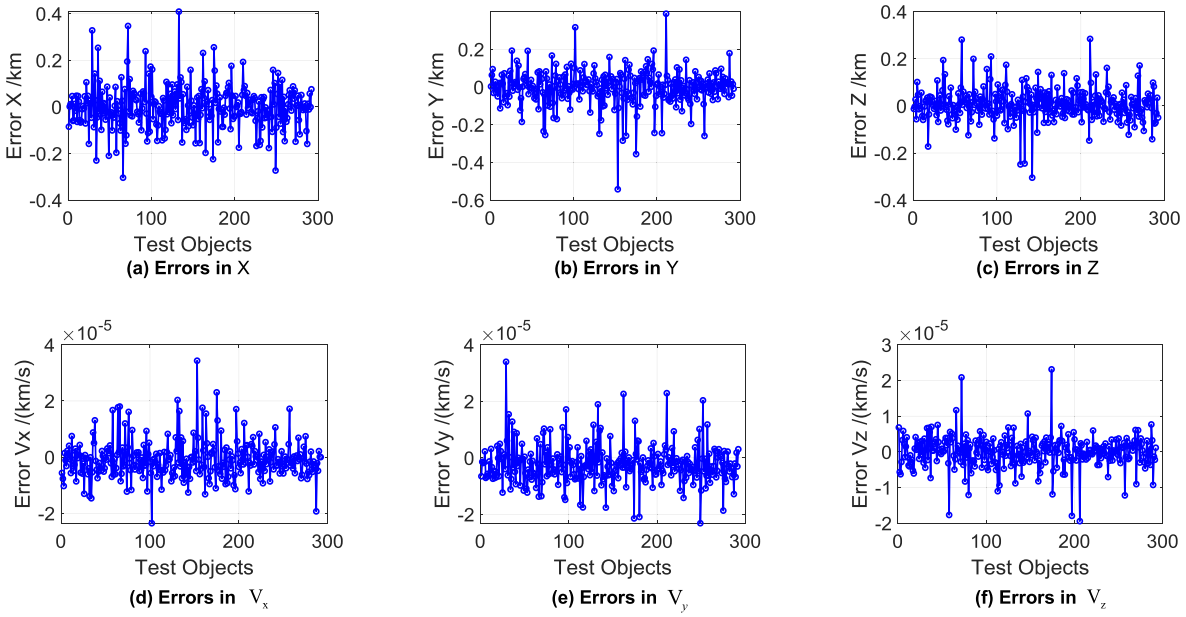


FIGURE 4. Errors of WDR-OD in the position and velocity vectors from noise-free measurements.

TABLE 4. The RMSE of WDR-OD in 5-fold cross validation.

	Cross Validation 1	Cross Validation 2	Cross Validation 3	Cross Validation 4	Cross Validation 5	Mean value
RMSE in position /km (noise-free)	0.0607	0.0566	0.0537	0.0401	0.0307	**0.0484**
RMSE in position /km (WGN)	0.4765	1.3872	1.5444	0.4834	0.5050	**0.8793**
RMSE in position /km (colored noise)	1.0610	2.3158	2.8762	1.1573	1.0756	**1.6972**
RMSE in velocity /(km/s) (noise-free)	4.4596×10^{-6}	4.6336×10^{-6}	4.3117×10^{-6}	4.0128×10^{-6}	1.6282×10^{-6}	$\mathbf{3.8092\times10^{-6}}$
RMSE in velocity /(km/s) (WGN)	3.1912×10^{-5}	1.0873×10^{-4}	1.1181×10^{-4}	3.2425×10^{-5}	3.3234×10^{-5}	$\mathbf{6.3622\times10^{-5}}$
RMSE in velocity /(km/s) (colored noise)	7.5017×10^{-5}	2.2060×10^{-4}	2.0725×10^{-4}	7.4617×10^{-5}	7.2514×10^{-5}	$\mathbf{1.2999\times10^{-4}}$

between folds of cross validation. When WGN is introduced, the mean RMSE of the 5-fold cross validation in position vectors reaches 0.8793 km. Meanwhile, the corresponding result in velocity increases to 6.3622×10^{-5} km/s. Generally, these results are all within an acceptable scope. Subsequently, we add the colored noise to the measurements, the results of which are shown in Figure 6 and Table 4. The RMSE of both position and velocity increased twice as much than the results with WGN. The largest test object error reaches -6 km in the X and Y directions, but the errors of most of the objects are within the $[-2$ km, 2 km$]$ interval. A common feature is that the results in Z and V_z are more accurate than those in other directions, whether with WGN or colored noise.

For comparison, we introduce the improved Laplace method. The improved Laplace method reduces its dependence on a relatively reliable initial value by adding a constrained term in the normal vector of orbit plane, and iterates in geometric and dynamic model with perturbations. It performs well in noise-free measurements, as discussed in [1]. Note that we utilize the consistent measurement data with the same orbit arc and the same noise level in the improved Laplace and the proposed WDR-OD method. The results are shown in Figure 7 $\sim$ 8 and Table 5. With regard to the WGN, the RMSE of position and velocity are 5.0804 km and 0.0051 km/s, respectively, while with colored noise, the values are 14.8132 km and 0.0129 km/s. Apparently, when measurements are polluted by noise, the errors of the improved Laplace method increase significantly and are remarkably greater than the RMSE in the WDR-OD method. It is not difficult to understand the cause of the differences between these two methods. Because the observation model that the improved Laplace method relies on is established

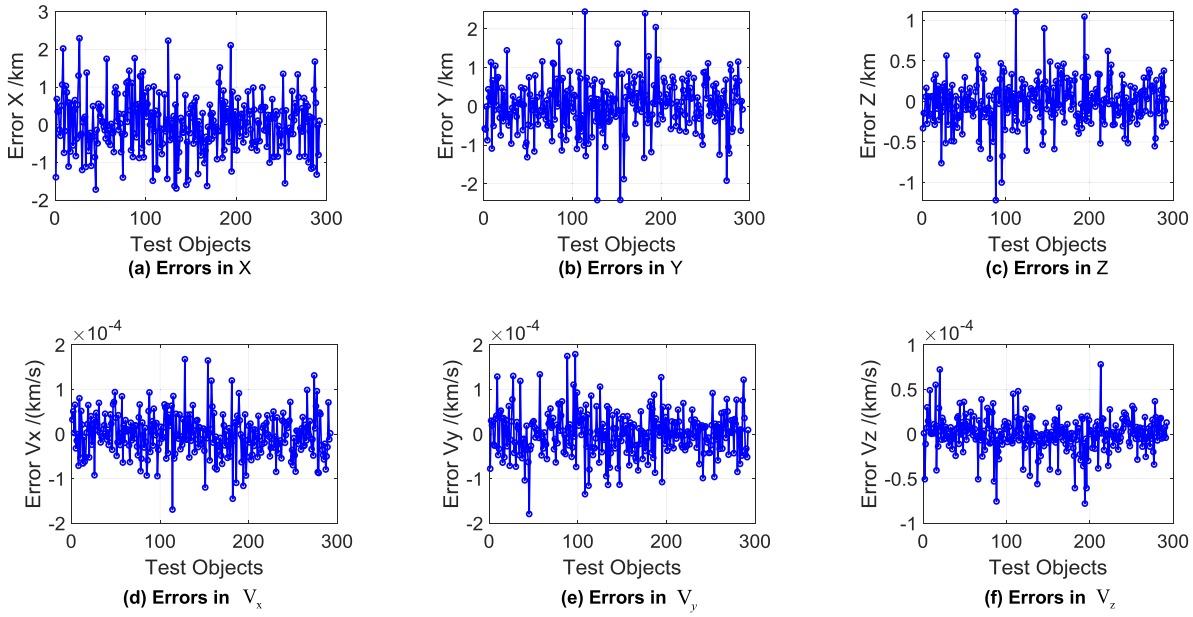


FIGURE 5. Errors of WDR-OD in the position and velocity vectors from measurements with WGN noise.

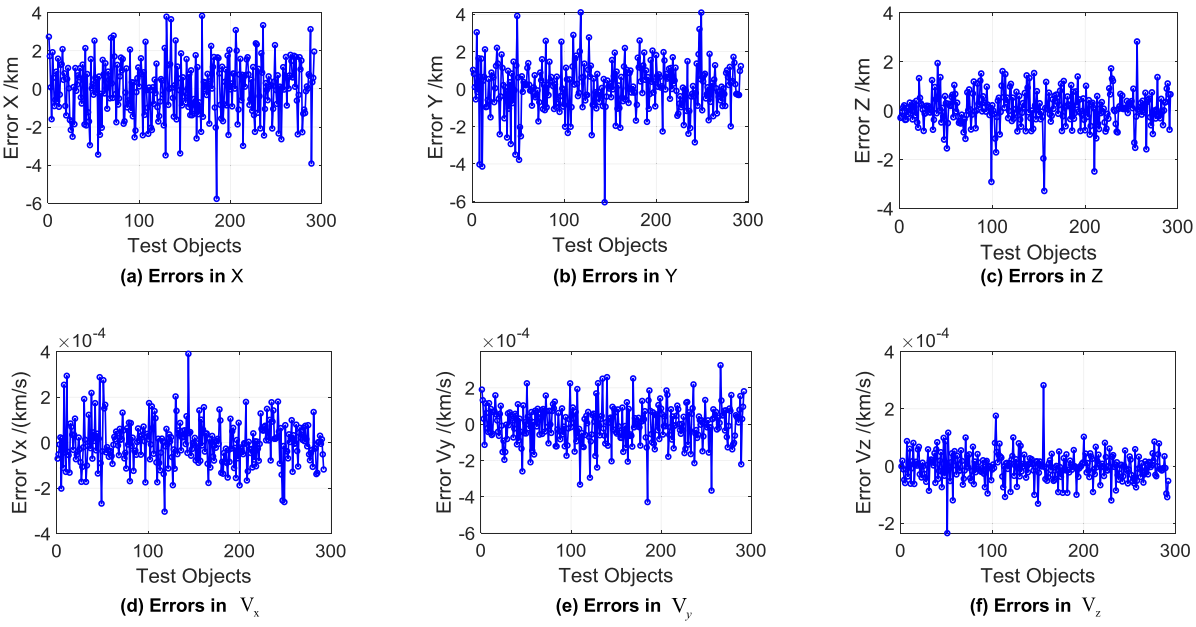


FIGURE 6. Errors of WDR-OD in the position and velocity vectors from measurements with colored noise.

on a precise geometric relationship from observer to object, any slight errors in angular measurements will disturb the true position of the object we observe. On the other hand, the proposed WDR-OD method is based on learning theory; in other words, the model it regresses from relies on the training data. Although the measurements are noisy, if the training data and test data are sampled from similar distributions, the predicted results will not deviate from the expected results dramatically. Thus, the WDR-OD method based on the weighting DR theory shows great robustness in a noisy environment. Figure 4 to Figure 8 show the fluctuations of the OD errors from the first 300 test objects. A common feature is that most test object errors in Figure 4∼6 are centered around zero, and a small number of test objects have spikes. However, the corresponding results in Figure 7∼8 show few regularities; they are more strongly dominated by random noise in each test object. In other words, the results in WDR-OD are more converged, in contrast to the more

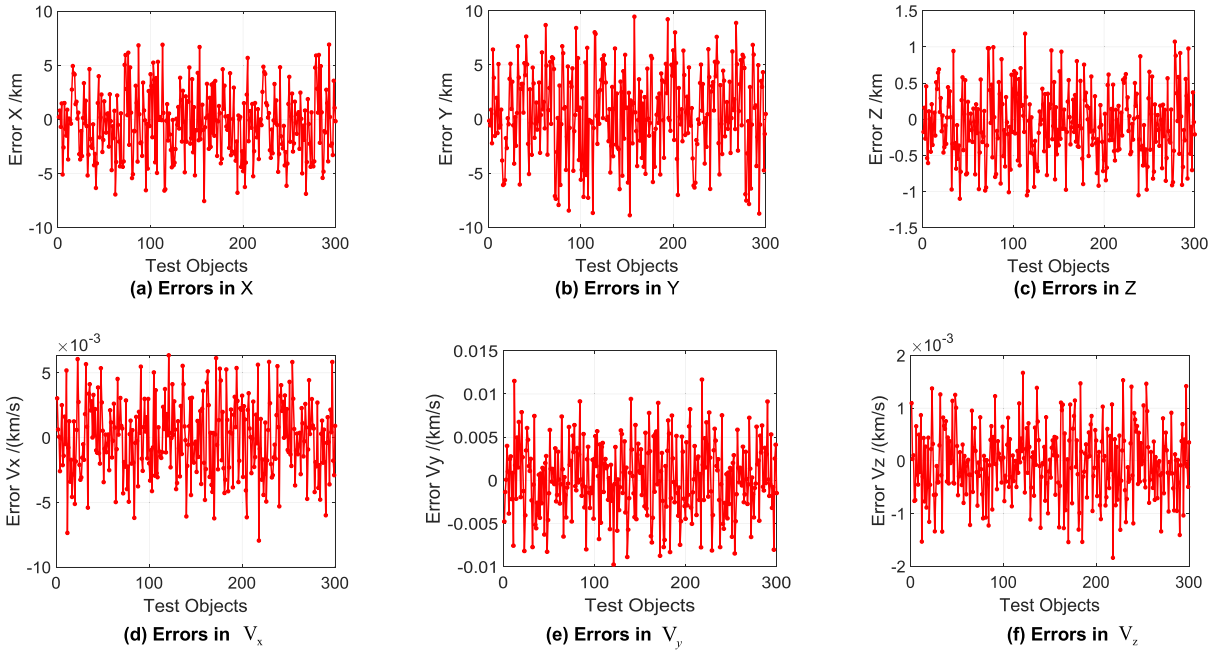


FIGURE 7. Errors of laplace OD results in the position and velocity vectors from WGN measurements.

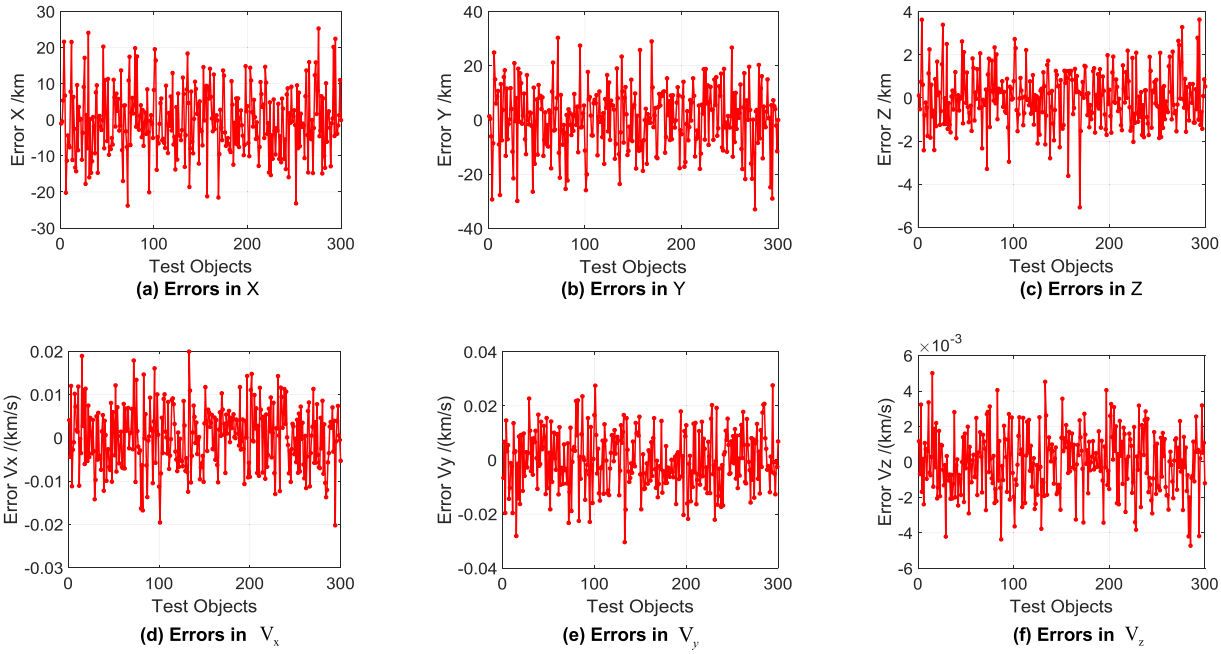


FIGURE 8. Errors of laplace OD results in the position and velocity vectors from colored noise measurements.

divergent results obtained with the improved Laplace method. This conclusion indicates that the results of WDR-OD have a higher confidence and are less affected by the fluctuations of random noise.

C. EXPERIMENT EXPLORING THE SPARSE SOLUTIONS OF THE WDR-OD

Then, in the following experiment, we employ the RIP-based ROMP algorithm to obtain sparse solution $\tilde{\mathbf{A}}$, according to the optimal solution $\mathbf{A}^*$ in the preceding WDR-OD method. In each iteration step, it selects the atom $\mathbf{d}$ of dictionary in Eq. (13) with the highest correlation to the current residual, and the signal is projected to the span of the selected atoms orthogonally. Figure 9 and Table 6 show the corresponding results. The format in Figure 9 is different from the earlier figures; the vertical coordinate represents the true values rather than the errors. The test results and the predicted results are represented by blue circles and black

TABLE 5. The RMSE of Laplace OD results in the position and velocity vectors from noisy measurements.

	X/km	Y/km	Z/km	Position/km	V_x /(km/s)	V_y /(km/s)	V_z /(km/s)	Velocity /(km/s)
WGN	3.0881	4.0043	0.4896	**5.0804**	0.0029	0.0042	6.5854×10^{-4}	**0.0051**
Colored noise	9.1600	11.5716	1.2743	**14.8132**	0.0072	0.0106	0.0017	**0.0129**

TABLE 6. The RMSE of WDR-OD results with sparse solutions from measurements with WGN.

X/km	Y/km	Z/km	Position/km	V_x /(km/s)	V_y /(km/s)	V_z /(km/s)	Velocity /(km/s)
1.1854	0.9440	0.6664	**1.6554**	4.4677×10^{-4}	5.0409×10^{-4}	1.5111×10^{-4}	$\mathbf{6.9032\times10^{-4}}$

TABLE 7. The sparsity of coefficients in each direction (shown as percentages).

	X	Y	Z	V_x	V_y	V_z
Sparsity S_0	10%	7%	8%	10%	7%	7%

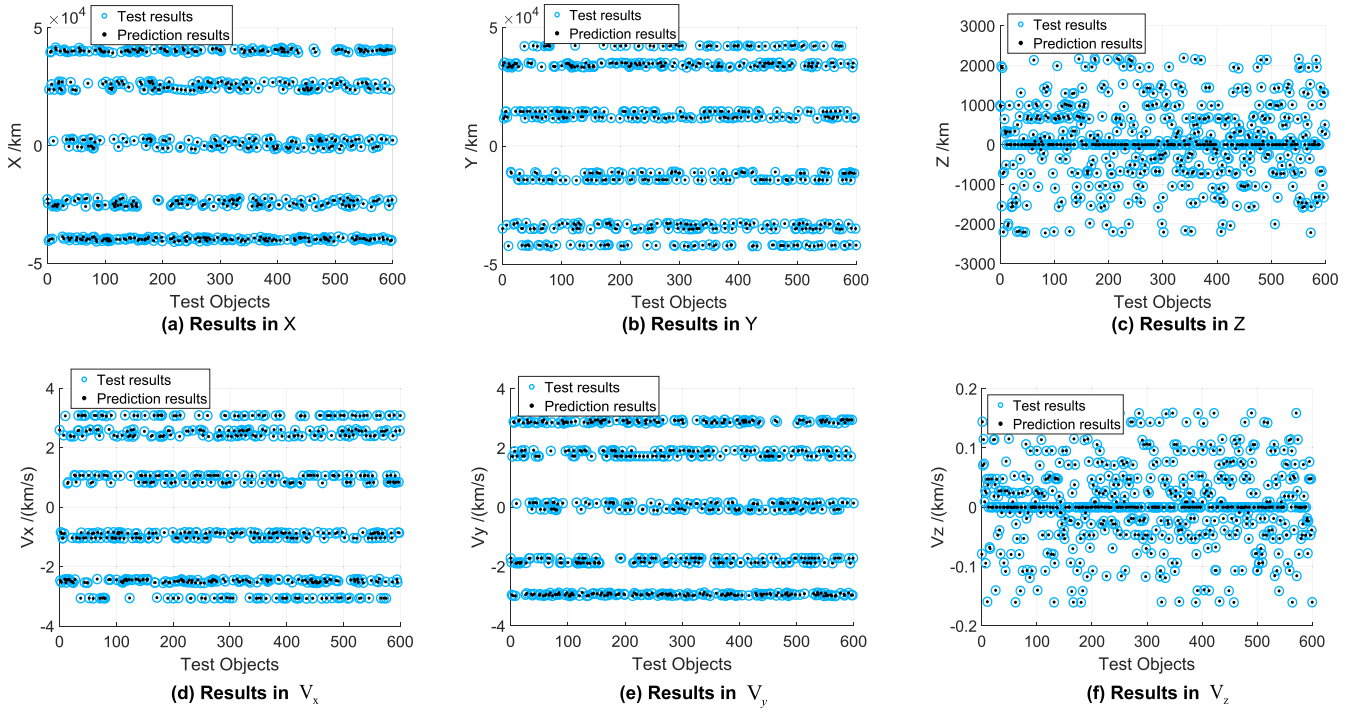
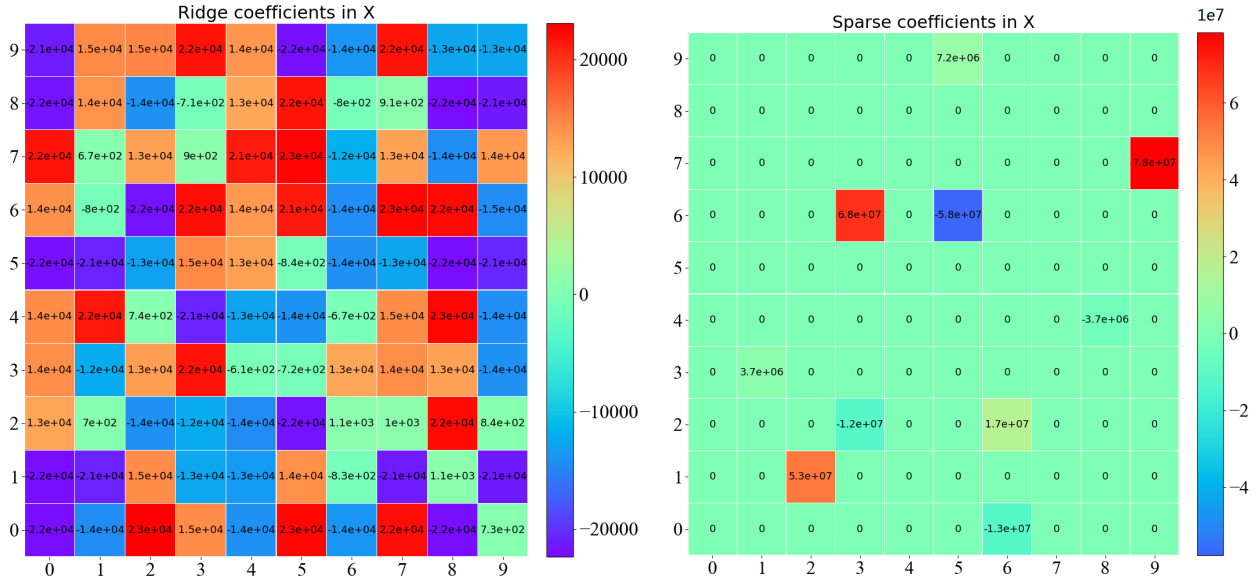


FIGURE 9. The WDR-OD results with sparse solutions from measurements with WGN.

dots, respectively. The closer a dot is to the center of a circle, the better the prediction is. We find that most predictions perform well, and also we note that, because of an inherent characteristic of objects in GEO belt, most samples in Z and V_Z are distributed around zero. The positional RMSE in Table 6 has increased from the 0.8793 km in WDR-OD to the 1.6554 km in the sparse method. However, it is still acceptable and smaller than the improved Laplace method's 5.0804 km.

For the sake of visualization, we reshape the first 100 coefficients of the 1-dimensional coefficient vectors $\boldsymbol{\alpha}_1$ and $\tilde{\boldsymbol{\alpha}}_1$ to a 10×10 dimensional matrix view and display it in a form of heat map in Figure 10 (for simplicity, Figure 10 only shows the coefficient vectors in the X-direction, as an example). On the left is the view of the coefficient vector from the ridge regression, and on the right is the view of the corresponding sparse solutions. The darker the color is, the larger the value of each element is. Apparently, most coefficients shrink to



(a) Full solutions of coefficients vector in X

(b) Sparse solutions of coefficients vector in X

FIGURE 10. The visualization of the coefficients vector from ridge regression method (left) and its sparse solution (right) (reshaped results of the first 100 coefficients of the 1-dimensional vectors α_1 and $\tilde{\alpha}_1$).

zero in the right figure. The number of nonzero coefficients of the first 100 coefficients is only ten (a 90% reduction). The sparsity in each direction is shown in Table 7; the results in other directions can also achieve 90%∼93% reductions. This sparsity is worthwhile in practice; it means that only 7%∼10% of the elements of the sensing matrix $\mathbf{\Lambda}$ need to be stored. It can be rewritten as a sparse expression corresponding to the nonzero elements of $\tilde{\mathbf{A}}$. These improvements make contribution to save computing cost.

For the same computer (ThinkPad X1, with 2.5GHz Intel Core i7 processor and 8G RAM), except for the same training process, the time consumed by the test process (with the number of 1452 objects) of the WDR-OD method is 2.1 minutes, and that of the sparse method is 32 seconds.

Especially, it is suitable for the cases owning large sample size but with limited computing resources; e.g., performing the WDR-OD method onboard for recognizing space debris with a space-based surveillance network in the future.

VIII. CONCLUSION

In this paper, from a ML perspective, we apply the two-stage sampled distribution regression to the domain of space-based orbit determination and propose the WDR-OD method for the future short-arc and angle only OD problem with big data. The following conclusions are drawn:

First, we propose the structure of angle-only OD samples consisting of the angular measurements from observers to objects and the inertial position of the observers. We show the effectiveness of a weighting matrix in improving the accuracy of orbit estimations. Second, and more importantly, we successfully apply the distribution regression technique to space-based OD. The preceding experiments show that the RMSE of position vectors reaches 0.8793 km with WGN and 1.6972 km with colored noise, which are significantly smaller than the corresponding Laplace method's 5.0804 km and 14.8132 km. The velocity RMSE performs even better. Compared with the improved Laplace method, the WDR-OD method shows great robustness and is more accurate in noisy environments. Furthermore, based on the solutions of WDR-OD method, we also introduce RIP theory in ROMP to provide the bound of sparsity theoretically, and then to pursue sparse solutions. The experimental results show that this method achieves 1.6554 km RMSE in position and 6.9032×10^{-4} km/s in velocity with a 90%-93% reduction of nonzero coefficients. This approach is helpful for reducing computing load for onboard computation.

Note that, there needs to be consistent between the training data and the test data. In other words, the training sample needs to be large enough to cover the possible range of the test sample. In the foreseeable future, with the construction of space-based surveillance networks, there will be a mass of short-arc measurements swarming into the data processing center. Many of these measurements belong to known spacecraft, and many of them belong to unknown objects. The purpose of the proposed method from a ML perspective is not only to improve the OD accuracy compared with classical method, but more importantly, it takes full advantages of the large amounts of observation data. In the proposed framework, users can predict the orbits of unknown objects by learning from the known spacecraft and their corresponding measurements.

In the future work, first, we will expand the algorithm to different objects in different catalogs and perform the WDR-OD method with real space-based measurements data.

Second, we are going to research on the global optimality to further improve the precision of sparse solution.

ACKNOWLEDGMENT

The authors would like to thank the support of the members of the State Key Laboratory of Astronautic Dynamics.

REFERENCES

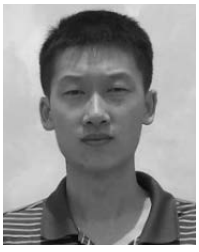
[1] F. Feng, H. Li, Y. Zhang, and Y. Huo, "An improved synchronized orbit determination method based on distributed star sensors," presented at the 69th Int. Astron. Congr., Bremen, Germany, Oct. 2018, pp. 1–5.
[2] D. A. Vallado, "Evaluating Gooding angles-only orbit determination of space based space surveillance measurements," *Paper USR*, vol. 10, pp. 1–21, Jan. 2010.
[3] X. C. Sun, P. Chen, C. Macabiau, and C. Han, "Autonomous orbit determination via Kalman filtering of gravity gradients," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2436–2451, Oct. 2016.
[4] G. F. Gronchi and G. B. A. Milani, "Keplerian integrals, elimination theory and identification of very short arcs in a large database of optical observations," *Celestial Mech. Dyn. Astron.*, vol. 127, no. 2, pp. 211–232, Feb. 2017.
[5] B. M. Hinga, "Space bourne orbit determination of unknown satellites using a stabilized-gauss-method, linear perturbation theory and angle-only measurements," presented at the Adv. Maui Opt. Space Surveill. Technol. Conf. (AMOS), Honolulu, HI, USA, Sep. 2018.
[6] M. L. Psiaki, R. M. Weisman, and M. K. Jah, "Gaussian mixture approximation of angles-only initial orbit determination likelihood function," *J. Guid., Control, Dyn.*, vol. 40, no. 11, pp. 2807–2819, 2017.
[7] B. Gong, W. Li, S. Li, W. Ma, and L. Zheng, "Angles-only initial relative orbit determination algorithm for non-cooperative spacecraft proximity operations," *Astrodynamics*, vol. 2, no. 3, pp. 217–231, Sep. 2018.
[8] A. Milani, G. F. Gronchi, D. Farnocchia, Z. Knežević, R. Jedicke, L. Denneau, and F. Pierfederici, "Topocentric orbit determination: Algorithms for the next generation surveys," *Icarus*, vol. 195, no. 1, pp. 474–492, May 2008.
[9] F. M. Fadrique, A. Á. Maté, J. J. Grau, J. F. Sánchez, and L. A. García, "Comparison of angles only initial orbit determination algorithms for space debris cataloguing," *J. Aerosp. Eng.*, vol. 4, no. 1, pp. 39–51, 2012.
[10] D. Lubey and H. Patel, "Optical initial orbit determination using polynomial chaos surrogate functions," presented at the Adv. Maui Opt. Space Surveill. Technol. Conf. (AMOS), Honolulu, HI, USA, Sep. 2017.
[11] G. Stokes, C. Vo, R. Sridharan, and J. Sharma, "The space-based visible program," in *Proc. Space Conf. Expo.*, Long Beach, CA, USA, Sep. 2000, p. 5334.
[12] T. Flohrer, H. Krag, H. Klinkrad, and T. Schildknecht, "Feasibility of performing space surveillance tasks with a proposed space-based optical architecture," *Adv. Space Res.*, vol. 47, no. 6, pp. 1029–1042, Mar. 2011.
[13] Y.-P. Hum, L. Chen, and J.-Y. Huang, "Space-based pseudo-fixed latitude observation mode based on the characteristics of geosynchronous orbit belt," *Acta Astronautica*, vol. 137, pp. 31–37, Aug. 2017.
[14] S. Sharma and J. W. Cutler, "Kernel embedding approaches to orbit determination of spacecraft clusters," Mar. 2018, *arXiv:1803.00650*. [Online]. Available: https://arxiv.org/abs/1803.00650
[15] D. N. Penenberg, "Mathematical statistics: Basic ideas and selected topics," *J. Roy. Stat. Soc.*, vol. 179, no. 4, pp. 1128–1129, 2016.
[16] D. D. Chakraborty, "Statistical decision theory. estimation, testing and selection," *Investigación Operacional*, vol. 29, no. 2, pp. 184–185, 2008.
[17] F. Nielsen and R. Nock, "A closed-form expression for the Sharma–Mittal entropy of exponential families," *J. Phys. A, Math. Theor.*, vol. 45, no. 3, 2011, Art. no. 032003.
[18] G. R. Aryal, E. M. Ortega, G. G. Hamedani, and H. M. Yousof, "The Topp-Leone generated Weibull distribution: Regression model, characterizations and applications," *Int. J. Statist. Probab.*, vol. 6, no. 1, pp. 126–141, 2017.
[19] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A hilbert space embedding for distributions," in *Proc. Int. Conf. Algorithmic Learn. Theory* Berlin, Germany: Springer, 2007.
[20] K. Muandet, K. Fukumizu, F. Dinuzzo, and B. Schölkopf, "Learning from distributions via support measure machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, 2013, pp. 10–18.
[21] B. Poczós, A. Rinaldo, A. Singh, and L. Wasserman, "Distribution-Free Distribution Regression," *J. Mach. Learn. Res.*, vol. 31, pp. 507–515, Feb. 2013.
[22] Z. Szabó, B. K. Sriperumbudur, B. Póczos, and A. Gretton, "Learning theory for distribution regression," *J. Mach. Learn. Res.*, vol. 17, no. 152, pp. 1–40, 2016.
[23] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, "Kernel mean embedding of distributions: A review and beyond," *Found. Trends Mach. Learn.*, vol. 10, nos. 1–2, pp. 1–141, 2017.
[24] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
[25] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, Oct. 2008, pp. 581–587.
[26] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comp. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008.
[27] S. Yao, T. Wang, Y. Chong, and S. Pan, "Sparsity estimation based adaptive matching pursuit algorithm," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4095–4112, Feb. 2018.
[28] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to information retrieval," *Natural Lang. Eng.*, vol. 16, no. 1, pp. 100–103, 2010.
[29] W. Wang, F. Ding, and J. Y. Dai, "Maximum likelihood least squares identification for systems with autoregressive moving average noise," *Appl. Math. Model.*, vol. 36, no. 5, pp. 1842–1853, May 2012.

FEI FENG was born in 1990. He received the B.Eng. degree from the Harbin Institute of Technology (HIT), China, in 2013. He is currently pursuing the Ph.D. degree with Space Engineering University, Beijing, China. His research interests include astronautic dynamics and machine learning.

YASHENG ZHANG was born in 1974. She received the B.Eng. and M.Eng. degrees from the National University of Defense Technology (NUDT), China, in 1996 and 1999, respectively, and the Ph.D. degree from Space Engineering University, Beijing, China, in 2006, where she is currently a Professor.

Her research interests include spacecraft orbit dynamics, design of constellation, and space mission analysis. She received the Young Science and Technology Award of China, the National Science and Technology Progress Award, and the Space Fund Award of China. She has published more than 60 articles and more than 10 monograph publications.

HENGNIAN LI received the Ph.D. degree from Xi'an Jiaotong University, China, in 2006, where he is currently a Professor and the Director of the State Key Laboratory of Astronautic Dynamics (ADL). His research interests include astronautic dynamics and aerospace TT&C.



YUQIANG FANG received the Ph.D. degree in control science and engineering from the National University of Defense Technology (NUDT), China, in 2015. He is currently a Lecturer with Space Engineering University. His research interests include machine learning, computer vision, and data mining.

QIUSHI HUANG was born in 1994. He received the B.Eng. degree from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2017. He is currently pursuing the master's degree with Space Engineering University, Beijing, China. His research interest includes multi-targets track and association.

XUEFENG TAO was born in 1992. He received the B.S. degree in aeronautical engineering and the M.S. degree in aeronautical and astronautical science and technology from the National University of Defense Technology, Changsha, China, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree with Space Engineering University, Beijing, China. His current research interests include astronautic dynamics and machine learning.