Optimization Models for Interpolating Overlapping Interval Data

Jing Yang* and Xuli Han
Department of Mathematics and Statistics, Central South University, Changsha, China

*Corresponding author

Abstract. We consider a general robust spline interpolating computation problem where the intervals representing the ranges in positions of knots after clustering can be overlapped. We discuss and analyze the similarities and differences of different overlapping situations, propose two definitions of overlapping types for the simple overlapping situations: PIO (partial interval overlapping) and WIO (whole interval overlapping). An optimization model called PHOM-S (position-height optimization model with a single overlapping interval) is presented and formulated as a convex semi-infinite optimization problem which is capable of solving the two overlapping types mentioned. Then a brief discussion is given on how to build PHOM-M (position-height optimization model with multiple overlapping intervals) based on PHOM-S for solving general simple overlapping problems with multiple overlapping intervals. Also, simplification of a finite optimization model called PHOM-S-3 using the cubic polynomial spline is given for the applicability of computation. A numerical example is presented to show the feasibility and effectiveness of the proposed robust model PHOM-S-3.

Keywords: Interval data; Overlapping; Interpolating computation; Optimization model.

1. Introduction
Tons of data is being generated every day in this mature internet era. For the purpose of fast acquisition to valuable information, using clustering to compress data is necessary [1,2,3,4]. The data after clustering is usually measured as an interval [5,6] which can be regarded as a general data form compared to point data, because the interval will degenerate into a point when the lower and upper bounds of the interval are the same. Interval data as one typical type of symbolic data has attracted interests from many application areas such as data mining, machine learning and decision making. No doubt the interpolation problem on interval data is an important topic that can help to visualize big data and obtain the visual character of big data more easily.

Traditional spline interpolation problems have been solved by two basic approaches. One is to construct curves or surfaces [7,8] directly by given points that are data with known values, for example, \((x_i, \phi_i)\) in a two-dimensional coordinate system. The other aims to figure out curves or surfaces [9,10,11] by approximating the optimized graphs, which is indirect but still based on the exact locations of points. However, these models based on given fixed points and knots are unable to solve the interval data interpolation problems very well. Due to the generality of interval data we mentioned, the spline interpolation of interval data would be more applicable to future interpolation problems in a broader range of actual situations.

Robust spline has been proposed to handle such situations that involve data uncertainties. Models using interval data only for height uncertainties [12] were studied. Lately, a more general robust spline model was proposed [13], in which the interpolation information was given by interval estimates for both position and height. But the range of the interval was so special that it was unclear whether this approach...
could be extended for more general settings and the application was very limited. Subsequently, two models [14] with improved interval specifications were built. So far, all the existing models were studied on the premise of non-overlapping intervals. However, the fact is that it is hard to ensure that there are no overlaps between these intervals. To enlarge the research scope and build a model that has the existing studies included, the general optimization models for interpolating overlapped interval data are discussed in this paper.

Specifically, we discuss a more general interpolation optimization problem where the intervals $I_i = [l_i, l'_i]$ ($i = 0, 1, ..., n + 1$) representing the compressed data after clustering can overlap. This is a setting that has been avoided by existing researches. Thereinto, $l_i$ and $l'_i$ represent the lower and upper bounds of interval data $l_i$, and $I_i$ will degenerate into a point when $l_i = l'_i$. The value defined on $I_i$ is an interval $H_i = [H_i, H'_i]$ as well. That is, the traditional point data $(x_i, \phi_i)$ for interpolating computation changes to $(I_i, H_i)$ in this paper. The main content could be divided into two parts. First, to make the interpolation model a general one, different overlapping types are discussed, and two definitions of simple overlapping types are proposed: PIO (partial interval overlapping) and WIO (whole interval overlapping). The proposed optimization model PHOM-S (position height optimization model with a single overlapping interval) fits for both the overlapping types and can degenerate into PHOM (with no overlapping intervals). PHOM-M (PHOM with multiple overlapping intervals) can be built based on PHOM-S and could be applied to more general overlapping situations. Second, the simplification of the proposed model by using cubic polynomial spline will be given to make the optimization problem computable.

The remainder of this paper is organized as follows. In section 2, we review PHOM with non-overlapping intervals. In section 3, definitions of two overlapping types-PIO and WIO-for the simple overlapping situations are given. In section 4, we formulate PHOM-S as a semi-infinite optimization problem, and give a brief discussion on how to build PHOM-M based on PHOM-S. Then PHOM-S is simplified to a finite optimization model PHOM-S-3 using cubic polynomial splines in section 5. A numerical example of PHOM-S-3 with PIO is presented in section 6 to show the feasibility and effectiveness of the proposed model. In the last section, we give some conclusions of this paper.

2. PHOM with Non-overlapping Intervals

Given a set of data points $\{(x_i, \phi_i), i = 0, 1, ..., n + 1\}$, where $x_i$ and $\phi_i$ represent the precise position and height data respectively, the traditional optimization problem for spline model is to find a smooth piecewise polynomial function $\Phi(x) = \Phi_i(x), x \in [x_i, x_{i+1}]$ ($i = 0, 1, ..., n$), that interpolates the given data points $\phi_i$ while minimizing some commonly used objective functions. Specifically, if one were to minimize the energy function, the $L_2$ norm of the second derivative of $\Phi(x)$ over $[a, b]$, the optimization problem can be written as follow,

$$
\min \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} [\Phi''(x)]^2 dx,
$$

subject to

$$
\Phi(x_i) = \phi_i, i = 0, 1, ..., n + 1,
$$

$$
\Phi_i(x_{i+1}) = \Phi_{i+1}(x_{i+1}), i = 0, 1, ..., n - 1,
$$

$$
\Phi'_i(x_{i+1}) = \Phi'_{i+1}(x_{i+1}), i = 0, 1, ..., n - 1,
$$

where Eq. 3 and Eq. 4 represent the continuous conditions of consecutive pieces $\Phi_i(x)$ and $\Phi_{i+1}(x)$. In the optimization problem, the decision variables are the coefficients of the polynomials.

However, as we mentioned in introduction, $x_i$ and $\phi_i$ are not known precisely after clustering and usually vary in the ranges of intervals, which motivates researchers to study PHOM [14] that considers the position data $x_i$ and height data $\phi_i$ as interval data $l_i$ and $H_i$, respectively. Specifically, every $x_i$ varies in $l_i = [l_i^L, l_i^R]$ and the function value $\phi_i$ defined on $l_i$ also turns to an interval $H_i = [H_i^L, H_i^R]$. In what follows, without losing generality, we suppose that the sequence of knots satisfies: $a < l_{ik} < l_{ik}' < l_{ik+1}$.
\[ \begin{align*}
&\cdots < I_1 < I_n < b, I_0 = a, I_{n+1} = b, \text{ and all the discussion in this paper is based on the fact}\nonumber
\end{align*} \]

\[ I_j < I_i (i = 1, 2, \ldots, n). \quad \text{Note that} \quad [a, b] = \bigcup_{i=1}^{n} I_i \bigcup \bigcup_{i=0}^{n} E_i, \quad \text{where} \quad E_i = [I_i, I_{i+1}]. \quad \text{Let} \quad f_i(x) (i = 1, 2, \ldots, n) \quad \text{be the internal polynomial piece defined on} \quad I_i \quad \text{and} \quad g_i(x) (i = 0, 1, \ldots, n) \quad \text{be the external polynomial piece defined on} \quad E_i. \quad \text{Then the spline function defined on} \quad [a, b] \quad \text{is comprised of two parts:} \nonumber
\]

\[ \text{one is the internal part} \quad \sum_{i=1}^{n} f_i(x) I_i, \quad \text{and the other is the external part} \quad \sum_{i=0}^{n} g_i(x) E_i. \quad \text{To minimize the energy of the spline function similar to Eq. 1, the objective function is formulated as} \nonumber
\]

\[ \min \sum_{i=1}^{n} \int_{I_i} [f_i''(x)]^2 dx + \sum_{i=0}^{n} \int_{E_i} [g_i''(x)]^2 dx, \quad (5) \]

\[ \text{the interpolation constraints can be represented as} \nonumber
\]

\[ H_i \leq f_i(x) \leq H_i, \forall x \in I_i, i = 1, 2, \ldots, n, \quad (6) \]

\[ H_0 \leq g_0(a) \leq H_0, H_{n+1} \leq g_n(b) \leq H_{n+1}, \quad (7) \]

\[ \text{and the continuous constraints} \nonumber
\]

\[ f_i(I_j) = g_{i-1}(I_j), f_i(I_i) = g_i(I_i), i = 1, 2, \ldots, n, \quad (8) \]

\[ f_i'(I_j) = g_{i-1}'(I_j), f_i'(I_i) = g_i'(I_i), i = 1, 2, \ldots, n. \quad (9) \]

\[ \text{Notice that the objective function Eq. 5 is convex, constraints Eq. 7, Eq. 8 and Eq. 9 are linear for their} \nonumber
\]

\[ \text{coefficients and constraint Eq. 6 represents an infinite collection of linear inequalities. Thus, Eq. 5-Eq.} \nonumber
\]

\[ \text{9 is a semi-infinite convex optimization problem. By using cubic splines for both internal and external} \nonumber
\]

\[ \text{pieces of the spline function in [13,14], this semi-infinite convex optimization problem can be simplified to a} \nonumber
\]

\[ \text{solvable finite optimization problem.} \nonumber
\]

\[ \text{3. Simple Overlapping Situations} \nonumber
\]

\[ \text{In the previous section, knots} \quad I_i \quad \text{and} \quad \tilde{I}_i \quad \text{in PHOM with non-overlapping intervals are strictly sorted in an} \nonumber
\]

\[ \text{ascending order of subscripts, i.e., for any} \quad i < j (i, j = 1, 2, \ldots, n), \quad \text{the corresponding sequence is} \nonumber
\]

\[ a < I_1 < \tilde{I}_1 < \cdots < I_j < \tilde{I}_j < \cdots < I_n < \tilde{I}_n < b \quad \text{shown in Fig. 1. We can see that every} \nonumber
\]

\[ \text{interval} \quad I_i \quad (i = 1, 2, \ldots, n) \quad \text{is independent and has no intersection with any other intervals. Because of the} \nonumber
\]

\[ \text{inevitable overlaps between intervals after clustering, however, the positions of some knots in the} \nonumber
\]

\[ \text{sequence may change and result in some overlapped parts. In other words, the value of knot with a larger} \nonumber
\]

\[ \text{subscript may be smaller than the one with a smaller subscript. Note that the overlapping situations} \nonumber
\]

\[ \text{discussed in this paper are simple overlapping situations with two prerequisites. Firstly, every} \quad I_i \quad \text{includes} \nonumber
\]

\[ \text{at most one overlapped part which means that an overlapped part associates with only two adjacent} \nonumber
\]

\[ \text{intervals. Secondly, for any three consecutive intervals, there is at least one interval which has no} \nonumber
\]

\[ \text{intersection with the other two intervals.} \nonumber
\]

\[ \text{Figure 1. Sequence of interval data with non-overlapping intervals.} \nonumber
\]

\[ \text{For the convenience of further discussion, without loss of generality, assume that} \quad I_0 < I_1 < \cdots < I_{n+1}. \nonumber
\]

\[ \text{It can be easily concluded that there are at most two overlapping types in a simple overlapping situation,} \nonumber
\]

\[ \text{one is PIO with the corresponding overlapping part} \quad [I_{i+1}, \tilde{I}_i], \quad \text{the intersection of} \quad I_i \quad \text{and} \quad I_{i+1} \quad \text{is shown in} \nonumber
\]

\[ \text{Fig. 2. The other is WIO with the corresponding overlapping interval} \quad I_{i+2}, \quad \text{which is completely covered} \nonumber
\]

\[ \text{by} \quad I_1 \quad \text{shown in Fig. 2. Denote} \quad I_{i+1} = [I_i, I_{i+1}] = [I_i, \min(\tilde{I}_i, \tilde{I}_{i+1})]. \quad \text{We notice that the bounds} \nonumber
\]

\[ I_{i+1} \quad \text{and} \quad \tilde{I}_i \quad \text{of} \quad I_{i+1} \quad \text{for PIO come from different intervals} \quad I_{i+1} \quad \text{and} \quad I_i, \quad \text{the bounds} \quad I_{i+1} \quad \text{and} \quad \tilde{I}_{i+1} \quad \text{of} \quad I_{i+1} \quad \text{for WIO come from the same interval} \quad I_{i+1}. \quad \text{Based on the differences of the subscripts, we denote} \nonumber
\]

\[ \mathcal{D}(\tilde{I}_{i+1}) \quad \text{to take the subscript of the minimum between} \quad \tilde{I}_i \quad \text{and} \quad \tilde{I}_{i+1}. \quad \text{Then the definitions for PIO and WIO can be concluded as follow,} \nonumber
\]

\[ \text{Figure 1. Sequence of interval data with non-overlapping intervals.} \nonumber
\]

\[ \text{For the convenience of further discussion, without loss of generality, assume that} \quad I_0 < I_1 < \cdots < I_{n+1}. \nonumber
\]

\[ \text{It can be easily concluded that there are at most two overlapping types in a simple overlapping situation,} \nonumber
\]

\[ \text{one is PIO with the corresponding overlapping part} \quad [I_{i+1}, \tilde{I}_i], \quad \text{the intersection of} \quad I_i \quad \text{and} \quad I_{i+1} \quad \text{is shown in} \nonumber
\]

\[ \text{Fig. 2. The other is WIO with the corresponding overlapping interval} \quad I_{i+2}, \quad \text{which is completely covered} \nonumber
\]

\[ \text{by} \quad I_1 \quad \text{shown in Fig. 2. Denote} \quad I_{i+1} = [I_i, I_{i+1}] = [I_i, \min(\tilde{I}_i, \tilde{I}_{i+1})]. \quad \text{We notice that the bounds} \nonumber
\]

\[ I_{i+1} \quad \text{and} \quad \tilde{I}_i \quad \text{of} \quad I_{i+1} \quad \text{for PIO come from different intervals} \quad I_{i+1} \quad \text{and} \quad I_i, \quad \text{the bounds} \quad I_{i+1} \quad \text{and} \quad \tilde{I}_{i+1} \quad \text{of} \quad I_{i+1} \quad \text{for WIO come from the same interval} \quad I_{i+1}. \quad \text{Based on the differences of the subscripts, we denote} \nonumber
\]

\[ \mathcal{D}(\tilde{I}_{i+1}) \quad \text{to take the subscript of the minimum between} \quad \tilde{I}_i \quad \text{and} \quad \tilde{I}_{i+1}. \quad \text{Then the definitions for PIO and WIO can be concluded as follow,} \nonumber
\]

\[ \text{Figure 1. Sequence of interval data with non-overlapping intervals.} \nonumber
\]
Definition 1. The case where $I_i \cap I_{i+1} \neq \emptyset$ $(i = 1, 2, ..., n-1)$ with $i + 1 > D(I_i \cap I_{i+1})$ is called partial interval overlapping (PIO).

Definition 2. The case where $I_i \cap I_{i+1} \neq \emptyset$ $(i = 1, 2, ..., n-1)$ with $i + 1 \leq D(I_i \cap I_{i+1})$ is called whole interval overlapping (WIO).

Figure 2. Simple overlapping situation.

Note that any simple overlapping situation is composed of the two overlapping types PIO and WIO, as shown in Fig. 2.

4. The Optimization Models for Interpolating Overlapping Intervals

In this section, we begin from the simple overlapping situation that has only one overlapping part and formulate an interpolation optimization model called PHOM-S as a semi-infinite convex optimization problem, which can be applied to either PIO or WIO. Then we discuss the PHOM-M based on PHOM-S, which is applicable to general simple overlapping situations with multiple overlapping parts.

4.1. PHOM-S

Without loss of generality, consider a set of interval data $\{I_k = [I_k, I_{k+1}] | k = 0, 1, ..., n \}$ as position data with $I_i \cap I_{i+1} \neq \emptyset$ $(i \in \{1, 2, ..., n-1\})$, where $I_0 = I_{n+1} = a$, and the corresponding interval data $H_k = [H_{k}, H_{k+1}]$ $(k = 0, 1, ..., n)$ as height data defined on $I_k$. We have two possible knot sequences: $S_1$: $a < I_1 < I_2 < ... < I_{n-1} < I_n < b$, and $S_2$: $a < I_1 < I_2 < ... < I_{n-1} < I_n < b$.

Thereinto, $S_1$ is the sequence with PIO while $S_2$ is the sequence with WIO based on the definitions in the previous section.

Unlike the internal and external polynomial pieces defined on $I_i$ and $E_i$ in [14], we define $f_k(x)$ $(k = 1, 2, ..., n)$ and $g_k(x)$ $(k = 0, 1, ..., n)$ on the intervals consisting of every two adjacent knots in knot sequence. The beginning knot of the interval determines the polynomial piece is $f_k$ or $g_k$, that is, $f_k$ is defined on the interval with beginning knot $I_k$, and $g_k$ is defined on the interval with beginning knot $I_{k+1}$.

The ending knot of the interval for $f_k$ and $g_k$ is the nearest knot on the right side of the beginning knot in the knot sequence. For a clearer explanation, two knot sequences $S_1$ and $S_2$ and the corresponding piecewise functions $f_k$ and $g_k$ defined around the overlapping intervals are shown in Fig. 3(a)-3(b). The overlapping interval is marked in bold. There are two reasons why this setting for piecewise functions...
is considered. First, this setting makes the $f_k$ is totally determined by the sorted knots $I_1$ and keeps the total number of $f_k$ and $g_k$ the same as the ones in PHOM. Second, this setting makes the degeneration of this model the same as PHOM in the existing studies [13,14]. We denote $D_1 = D(I_{n+1})$ and $D_i = i + |D(I_{n+1}) - (i + 1)|$ for the convenience of extracting the subscripts of the right side endpoints of the overlapped interval $I_{n+1}$. Then a general sequence $S_g$ for the simple overlapping situation with a single overlapping interval, which contains $S_1$ and $S_2$ as all the possibilities, is concluded in Fig. 3(c). Denote
\[ P_{k,k_1} = \int f_{k_1}''(x)^2 \, dx, \quad (10) \]
and
\[ G_{k,k_2} = \int g_{k_2}''(x)^2 \, dx, \quad \bar{G}_{k,k_2} = \int g_{k_2}''(x)^2 \, dx, \quad (11) \]
for notation simplification, where $k_1 = 1,2,\ldots,n, k_2 = 0,1,\ldots,n$. When $k_1 = k$ in Eq. 10, we further denote $F_k = P_{k,k}$ for convenience. And, when $k_2 = k + 1$ in the first equation in Eq. 11, we denote $G_k = G_{k,k+1}$. Therefore, based on $S_g$ in Fig. 3(c), PHOM-S is formulated as
\[ \min \sum_{k=1}^{n} F_{k,k_1} + \sum_{k=0}^{n} G_{k,k_2} + \bar{F}_{i+1,D_1} + \bar{G}_{D_1,D_1} + G_{D_1,D_1+2}, \quad (12) \]
with the interpolation constraints
\[ H_k \leq f_k(x) \leq \bar{H}_k, \forall x \in I_k, k = 1,2,\ldots,i,i + 2,\ldots,n, \quad (13) \]
\[ \max \{H_i, H_{i+1}\} \leq f_{i+1}(x) \leq \min \{H_i, H_{i+1}\}, \forall x \in [I_i, I_{i+1}], \quad (14) \]
\[ H_{D_1} \leq g_{D_1}(x) \leq \bar{H}_{D_1}, \forall x \in [D_1,D_1], \quad (15) \]
\[ H_0 \leq g_0(a) \leq \bar{H}_0, H_{n+1} \leq g_n(b) \leq \bar{H}_{n+1}, \quad (16) \]
the position continuous constraints
\[ g_{k-1}(I_k) = f_k(I_k), k = 1,2,\ldots,i,i + 2,\ldots,n, \quad (17) \]
\[ f_i(I_{i+1}) = f_{i+1}(I_{i+1}), \quad (18) \]
\[ f_k(I_k) = g_k(I_k), k = 1,2,\ldots,i-1,i + 2,\ldots,n, \quad (19) \]
\[ f_{i+1}(I_{D_1}) = g_{D_1}(I_{D_1}), \quad (20) \]
\[ g_{D_1}(I_{D_1}) = g_{D_1}(I_{D_1}), \quad (21) \]
and the first-order continuous constraints
\[ g_{k-1}'(I_k) = f_k'(I_k), k = 1,2,\ldots,i,i + 2,\ldots,n, \quad (22) \]
\[ f_i'(I_{i+1}) = f_{i+1}'(I_{i+1}), \quad (23) \]
\[ f_k'(I_k) = g_k'(I_k), k = 1,2,\ldots,i-1,i + 2,\ldots,n, \quad (24) \]
\[ f_{i+1}'(I_{D_1}) = g_{D_1}'(I_{D_1}), \quad (25) \]
\[ g_{D_1}'(I_{D_1}) = g_{D_1}'(I_{D_1}), \quad (26) \]
where the decision variables are the coefficients of the polynomial pieces $f_k(x)$ and $g_k(x)$. 
Note that if $I_{i\cup i+1} = \emptyset$, we take $\overline{D}_i = i = -1$ and $\overline{D}_i = i + 1 = 0$, then Eq. 12 becomes $\min \sum F_k + G_k + G_{0,1}$, which is equal to Eq. 5 with the terms $\overline{T}_{i+1,\overline{D}_i}$, $\overline{T}_{\overline{D}_i,\overline{D}_i}$ becoming meaningless. The interpolation constraints Eq. 14-Eq. 15 become meaningless, and Eq. 13 becomes Eq. 6. The continuous constraints Eq. 18, Eq. 20, Eq. 21, Eq. 23, Eq. 25 and Eq. 26 become meaningless, Eq. 17 and Eq. 19 degenerate into Eq. 8, and Eq. 22 and Eq. 24 degenerate into Eq. 9. In other words, PHOM-S will degenerate into PHOM presented previously.

4.2. PHOM-M

In this part, we will give a brief discussion on how to build PHOM-M for solving the general simple overlapping situation that has multiple overlapping intervals based on PHOM-S.

Figure 4. Simple overlapping situation with $m$ overlapping intervals.

Suppose the number of the overlapped intervals is $m (m \in \mathbb{Z}^+)$. Consider a set of interval data $\{I_k = [l_k, h_k], l_k \in \{1, 2, \ldots, n+1\}\}$ as position data after the preprocessing with $I_{i\cup i+1} \neq \emptyset (i \in \{1, 2, \ldots, n-1\}, l = 1, 2, \ldots, m)$, where $l_0 = h_0 = a$, $l_{n+1} = h_{n+1} = b$, $m < n/2$, and interval data $[I_k, H_k, \overline{H}_k]$ ($k = 0, 1, \ldots, n + 1$) as height data defined on $I_k$. For the purpose of taking advantage of PHOM-S in building PHOM-M, we select $I_{\overline{D}_i}$ as breakpoints to divide the knot sequence with $m$ overlapping intervals into $m+1$ subsequences: $S_1 (l = 1, 2, \ldots, m)$ and $S_{m+1}$, which is shown in Fig. 4. Thereinto, each $S_1$ has a single overlapping interval that makes it possible to use PHOM-S. However, we can see that since $I_{\overline{D}_i}$ is the last knot for $S_1$, the piecewise function $g_{\overline{D}_i}$ defined on the interval with the beginning knot $I_{\overline{D}_i}$ in PHOM-S becomes meaningless so that some formulations of PHOM-S should be adjusted. PHOM with non-overlapping intervals can be used for $S_{m+1}$. Then PHOM-M can be built by combining each sub-model together with some additional continuous constraints. The modeling is trivial, and the corresponding detail will not be given here.

5. Simplification of PHOM-S Using Cubic Polynomial Splines

Consider cubic polynomials for piecewise spline functions,

$$f_k(x) = \alpha_k(x - I_k)^3 + \beta_k(x - I_k)^2 + \gamma_k(x - I_k) + \delta_k, \quad k = 1, 2, \ldots, n,$$

$$g_k(x) = \alpha'_k(x - I_k)^3 + \beta'_k(x - I_k)^2 + \gamma'_k(x - I_k) + \delta'_k, \quad k = 0, 1, \ldots, n.$$

So, the objective function Eq. 12 in PHOM-S becomes

$$\min \sum_{k=1}^{n} p_k(h_k) + \sum_{k=0}^{n} p_k(h'_k) + p_{i+1} \left( \overline{h}_{i+1, \overline{D}_i} \right) + p'_{\overline{D}_i} \left( \overline{h}_{\overline{D}_i, \overline{D}_i} \right) + p''_{\overline{D}_i} \left( h''_{\overline{D}_i, i+2} \right),$$

where $p_k(x) = 12\alpha_k x^3 + 12\alpha_k x^2 + 4\beta_k x$ and $p'_k(x) = 12\alpha'_k x^3 + 12\alpha'_k x^2 + 4\beta'_k x$; $h_k = I_k - I_k$, $h'_k = I_k - I_k$, $h''_k = I_k - I_k$, $h'''_k = I_k - I_k$, where $k_1$ and $k_2$ have the same meanings as in Eq. 10 and Eq. 11. Notice that Eq. 13-Eq. 16 are of the following type,

$$S_3(x) = \alpha x^3 + \beta x^2 + \gamma x + \delta \leq C, \forall x \in [l_i, l_u],$$

for some constants $C, l_i$ and $l_u$. To make these semi-infinite constraints finite, we first denote $Q_3(\alpha, \beta, \gamma, \delta, l_i, l_u) = \max_{x \in [l_i, l_u]} S_3(x)$, that is,
\[ Q_3(\alpha, \beta, \gamma, \delta, l_1, l_u) = \begin{cases} \max\{S_3(l_1), S_3(l_u), S_3(r_1)\}, & \text{if } r_1 \in [l_1, l_u], r_2 \notin [l_1, l_u], \\ \max\{S_3(l_1), S_3(l_u), S_3(r_2)\}, & \text{if } r_1 \notin [l_1, l_u], r_2 \in [l_1, l_u], \\ \max\{S_3(l_1), S_3(l_u), S_3(r_1), S_3(r_2)\}, & \text{if } r_1, r_2 \in [l_1, l_u], \\ \max\{S_3(l_1), S_3(l_u)\}, & \text{otherwise,} \end{cases} \]

which is a simplification similar to that in [14]. Moreover, the coefficients in Eq. 28 should satisfy the conditions of 
\[ \alpha \neq 0, \beta^2 - 3\alpha \gamma \geq 0 \]
to ensure the existence of roots \( r_1, r_2 = \left( -\beta \pm \sqrt{\beta^2 - 3\alpha \gamma} \right) / 3\alpha \) for \( dS_3(x) / dx = 0 \). Then Eq. 28 is simplified to a finite constraint \( Q_3(\alpha, \beta, \gamma, l_1, l_u) \leq C \). Thus, Eq. 13-Eq. 16 can be represented as finite constraints as follows:

\[ Q_3(\alpha_k, \beta_k, \gamma_k, \delta_k, 0, h_k) \leq H_k, \]

\[ Q_3(-\alpha_k, -\beta_k, -\gamma_k, -\delta_k, 0, h_k) \leq -H_k, k = 1, 2, ... , i, i + 2, ... , n, \]  

\[ Q_3(\alpha_{i+1}, \beta_{i+1}, \gamma_{i+1}, \delta_{i+1}, 0, \bar{h}_{i+1,B_i}) \leq \min\{H_i, H_{i+1}\}, \]

\[ Q_3(-\alpha_{i+1}, -\beta_{i+1}, -\gamma_{i+1}, -\delta_{i+1}, 0, \bar{h}_{i+1,B_i}) \leq -\max\{H_i, H_{i+1}\}, \]

\[ Q_3(\alpha'_{D_i}, \beta'_{D_i}, \gamma'_{D_i}, \delta'_{D_i}, 0, \bar{h}_{D_i,B_i}) \leq \bar{H}_{D_i}, Q_3(-\alpha'_{D_i}, -\beta'_{D_i}, -\gamma'_{D_i}, -\delta'_{D_i}, 0, \bar{h}_{D_i,B_i}) \leq -\bar{H}_{D_i}, \]

\[ H_0 \leq \delta'_0 \leq H_0, H_{n+1} \leq \alpha'_n h_n^3 + \beta'_n h_n^2 + \gamma'_n h_n + \delta'_n \leq \bar{H}_{n+1}. \]

The position continuous constraints are

\[ \delta_k = \alpha'_{k-1} h_{k-1}^3 + \beta'_{k-1} h_{k-1}^2 + \gamma'_{k-1} h_{k-1} + \delta'_{k-1}, k = 1, 2, ... , i, i + 2, ... , n, \]

\[ \alpha(h_{i+1})^3 + \beta(h_{i+1})^2 + \gamma h_{i+1} + \delta = \delta_{i+1}, \]

\[ \alpha_k h_k^3 + \beta_k h_k^2 + \gamma_k h_k + \delta_k = \delta'_k, k = 1, 2, ... , i - 1, i + 2, ... , n, \]

\[ \alpha_{i+1}(\bar{h}_{i+1,B_i})^3 + \beta_{i+1}(\bar{h}_{i+1,B_i})^2 + \gamma_{i+1}(\bar{h}_{i+1,B_i}) + \delta_{i+1} = \delta'_{D_i}, \]

\[ \alpha_{D_i}(\bar{h}_{D_i,B_i})^3 + \beta_{D_i}(\bar{h}_{D_i,B_i})^2 + \gamma_{D_i}(\bar{h}_{D_i,B_i}) + \delta_{D_i} = \delta_{D_i}, \]

and the first-order continuous constraints are

\[ \gamma_k = 3 \alpha_k h_{k-1}^2 + 2 \beta_k h_{k-1} + \gamma_k, k = 1, 2, ... , i, i + 2, ... , n, \]

\[ 3\alpha(h_{i+1})^2 + 2\beta h_{i+1} + \gamma = \gamma_{i+1}, \]

\[ 3\alpha_k h_k^2 + 2\beta_k h_k + \gamma_k = \gamma'_k, k = 1, 2, ... , i - 1, i + 2, ... , n, \]

\[ 3\alpha_{i+1}(\bar{h}_{i+1,B_i})^2 + 2\beta_{i+1}(\bar{h}_{i+1,B_i}) + \gamma_{i+1} = \gamma'_{D_i}, \]

\[ 3\alpha_{D_i}(\bar{h}_{D_i,B_i})^2 + 2\beta_{D_i}(\bar{h}_{D_i,B_i}) + \gamma_{D_i} = \gamma'_{D_i}, \]

with the conditions of \( \alpha_k \neq 0, \beta_k^2 - 3\alpha_k \gamma_k \geq 0 \) and \( \alpha'_{D_i} \neq 0, \beta'_{D_i}^2 - 3\alpha'_{D_i} \gamma'_{D_i} \geq 0 \). Note that Eq. 27, together with Eq. 29-Eq. 42, is a finite optimization model called PHOM-S-3 with the coefficients of the cubic polynomial pieces as the decision variables.

6. A Numerical Experiment

Base on the fact that PIO is the overlapping type we usually meet in engineering, an example of PHOM-S-3 with PIO, solving by the optimization function 'fmincon' in MATLAB, is presented in this section.
Consider a set of interval data \( I_{0}, I_{1}, I_{2}, I_{3}, I_{4}, I_{5} \) as position data with \( I_{6} \cap I_{7} = \emptyset \) and the corresponding height intervals \( H_{0}, H_{1}, H_{2}, H_{3}, H_{4}, H_{5} \) as height data defined on \( I_{3} \). We start at an initial value of the constant vector \((1.9, 1.9, \ldots, 1.9)\) for decision variables to make a contrast. A non-optimal interpolation curve is shown in Fig. 5(a) and the red blocks show the variation ranges \([I_{k}, I_{k+1}]\) for data \((I_{k}, H_{k})\) \((k = 1, 2, 3, 4)\). We can see that although the curve basically interpolates every block we intend to, some parts of the interpolation curve have undesired fluctuation. Obviously, it is not a very good interpolating result. Also, the calculation report shows that it does not reach the optimum value because of the function evaluation limit. Then, we continue to change the initial value to the stop value and repeat this process until we obtain the optimal result \((0.9302, -0.0008, -3.9320, 24.0018, -0.3558, 0.9805, -1.1428, 20.999, 0.2099, -0.0128, -1.4907, 19.7891, -0.1881, 0.6373, -0.8867, 18.4955, -0.0046, 0.1583, -0.1762, 18.0581, 0.0958, -0.0143, 0.1268, 18.0356, -0.7525, 0.5375, 1.2187, 18.9980, -0.0177, 0.0173, 0.0363, 20.0018, 0.1753, -1.0498, -0.1071, 20.0018)\). The corresponding curve is shown in Fig. 5(b). We can see that this curve is very smooth with less fluctuation. Moreover, the interpolation works much better than that in Fig. 5(a). In brief, it has a very good shape as an interpolation curve. Another fact is that whatever the initial value we take, the result always converges to a very small range of this stable one.

7. Conclusion
In this paper, we discussed a general interpolation optimization model with overlapping interval data and the corresponding computation. We analyzed the similarities and differences of different overlapping situations, proposed the definitions for two overlapping types--PIO and WIO--which are able to be used to compose all the simple overlapping situations. A PHOM-S was formulated as a semi-infinite convex optimization problem for solving both PIO and WIO with a single overlapping interval. A point worth mentioning is that this model can degenerate into PHOM with non-overlapping intervals by the setting of the function notation in this paper. The discussion for building PHOM-M based on PHOM-S was also briefly presented. We furthermore simplified PHOM-S using the cubic polynomial spline to a finite optimization model called PHOM-S-3 to make this simplified model compatible with easy computation. By using the proposed model, we can solve such interval data interpolation problems without reordering the knot sequence to a sorted one and make it possible to share one interpolation optimization model between overlapping and non-overlapping cases.

Acknowledgement
This research was financially supported by the National Natural Science Foundation of China (No.11771453) and the Graduate Student Innovation Project of Central South University (No.2016zzts013).
References
[1] Francisco de A.T. de Carvalho, Fuzzy c-means clustering methods for symbolic interval data, Pattern Recognition Letters. 28 (2007) 423-437.
[2] Tatiana G. and Michael K. and Andreas N., Graph clusterings with overlaps: Adapted quality indices and a generation model, Neurocomputing. 123 (2014) 13-22.
[3] Amit K. S. and Pranab K. M., Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets, Engineering Applications of Artificial Intelligence. 77 (2019) 268-282.
[4] Yi Chen and L. Billard, A study of divisive clustering with Hausdorff distances for interval data, Pattern Recognition. 96 (2019) 106969.
[5] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, T. C. Havens, From interval-valued data to general type-2 fuzzy sets, IEEE Transactions on Fuzzy Systems. 23 (2014) 248-269.
[6] D. Bryce, T. S. Dübón, D. Bratzke, How are overlapping time intervals perceived? Evidence for a weighted sum of segments model, Acta Psychologica. 156 (2015) 83-95.
[7] H. Cheng, S. C. Fang, J. E. Lavery, Shape-preserving properties of univariate cubic $L_1$ spline, Journal of Computational and Applied Mathematics. 174 (2005) 361-382.
[8] M. Huard, R. T. Farouki, N. Sprynski, L. Biard, $C^2$ interpolation of spatial data subject to arc-length constraints using Pythagorean-hodograph quintic splines, Graphical Models. 76 (2014) 30-42.
[9] L. Zhou, Y. Wei, Y. Yao, Optimal multi-degree reduction of Bézier curves with geometric constraints, Computer-Aided Design. 49 (2014) 18-27.
[10] X. Han, J. Yang, Multi-degree reduction of Bézier curves with distance and energy optimization, Journal of Applied Mathematics and Physics. 4 (2016) 8-15.
[11] M. Amirfakhrian, H. Mafikandi, Approximation of parametric curves by Moving Least Squares method, Applied Mathematics and Computation. 283 (2016) 290-298.
[12] F. I. Utreras, On generalized cross-validation for multivariate smoothing spline functions, Siam Journal on Scientific Computing. 8 (1987) 630-643.
[13] I. Averbakh, S. Fang, Y. Zhao, Robust univariate cubic $L_2$ splines: interpolating data with uncertain positions of measurements, Journal of Industrial and Management Optimization. 5 (2009) 351-361.
[14] I. Averbakh, Y. Zhao, Robust univariate spline models for interpolating interval data, Operations Research Letters. 39 (2011) 62-66.