REGULAR ARTICLE

Sparse polynomial prediction

Hugo Maruri-Aguilar[1] [ID] · Henry Wynn[2]

Abstract

In numerical analysis, sparse grids are point configurations used in stochastic finite element approximation, numerical integration and interpolation. This paper is concerned with the construction of polynomial interpolator models in sparse grids. Our proposal stems from the fact that a sparse grid is an echelon design with a hierarchical structure that identifies a single model. We then formulate the model and show that it can be written using inclusion–exclusion formulæ. At this point, we deploy efficient methodologies from the algebraic literature that can simplify considerably the computations. The methodology uses Betti numbers to reduce the number of terms in the inclusion–exclusion while achieving the same result as with exhaustive formulæ.

Keywords Smolyak grids · Sparse designs · Inclusion–exclusion · Betti numbers

1 Introduction

The application of computational algebraic methods to experimental design has been extensively covered by the authors, co-workers and others following the seminal paper (Pistone and Wynn 1996), monograph (Pistone et al. 2001) and review by the authors (Maruri-Aguilar and Wynn 2015), with other developments such the study of duality between fractions of factorial design and models (Maruri-Aguilar et al. 2012). In this work a design is seen as a set of points and defines a natural polynomial model which can be used to interpolate data at design points. We concentrate on designs that are the union of product tensor structures, known as echelon designs. For such designs,

✉ Hugo Maruri-Aguilar
    H.Maruri-Aguilar@qmul.ac.uk

    Henry Wynn
    H.Wynn@lse.ac.uk

[1] School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

[2] Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK

the exponents of the interpolator have the same form as the design, giving a duality between designs and models, with both exhibiting hierarchical structure.

Another feature of this hierarchical design-model framework is the central role played by inclusion–exclusion (IE) formulæ. Inclusion–exclusion has been used extensively in reliability theory, where formulæ are derived from monomial ideals, see (Sáenz-de Cabezón Wynn 2009). We use the same principles to describe the echelon design points as IE of tensor (product) designs. The benefit of this method is the reduction in complexity achieved by using the so-called minimal free resolution. In short, the coefficients in the IE are Betti numbers of the resolution of the monomial ideal.

In this paper we apply the same machinery to study sparse grids. These grids are used widely in quadrature, particularly in the Finite Element solution of differential equations, and the difference of terminology should be noted. Sparse grids are sometimes referred to as Smolyak grids following the seminal paper (Smolyak 1963). Recent work has used sparse grids to the solution of differential equations (Nobile et al. 2008), for interpolation (Barthelmann et al. 2000) and for integration (Novak and Ritter 1996).

In our development, a set of quadrature points becomes an experimental design (design for short). This approach is in line with the recent history of the design of computer experiments, also termed as simulation experiments or simply simulation. We show that a Smolyak grid satisfies the same IE despite not having the same hierarchical form as discussed above.

The main result of this paper is that the IE formulæ which hold for interpolators over echelon designs extend to more general sparse grids. Polynomial interpolators based on the "echelon" designs inherit the IE decomposition that uses the natural product model interpolators. Remarkably, the same IE forms are inherited by sparse grid interpolators. Moreover the reduced forms use the multi-graded Betti numbers from the algebraic theory.

Betti numbers are at the foundation of algebraic topology, and describe features of the geometry of manifolds via topologically equivalent, and one might say more abstract entities, namely simplicial complexes. In statistics they have been used intensively within the relatively new area topological data analysis (TDA) (Edelsbrunner Harer 2010). However, here Betti numbers are used in a different way, namely as a way of reducing the complexity of the inclusion–exclusion lemma. It is quickly observed that in handling large inclusion–exclusion formulæ, there are redundancies; that is to say a large amount of cancellation of terms and many sets may be empty. This happens partly because the underlying dimension may be much smaller that the number of sets and partly because of the complexity of the inter-relation between the sets. Both in TDA and IE there are "minimal free resolutions" (a term from the underlying algebraic topology) which gives, in a well-defined sense the, least complex formulae in which the plus and minus signs of the formulae are replaced by (multigraded) Betti numbers (Naiman and Wynn 1997; Sáenz-de Cabezón Wynn 2009).

The "sets" in this paper are the tensor grids of exponents α in monomials x^α. Thus, in summary, the Betti numbers support less complex IE formulae, in the sense of number of terms, in which the complexity is, in a sense, buried in the Betti numbers. The main purpose of this paper is to point out this structure, that is to say complexity reduction

using the Betti numbers, is inherited by the natural interpolators based supported on the tensor grids and their extension to sparse grids.

The rest of the paper is ordered as follows. In Sect. 2 we introduce echelon designs and the duality between designs and models. We discuss the tensor product notation and the extreme corner points that define echelon designs. Section 3 contains the main results for interpolation. We define indicators which are the building blocks of a Lagrange interpolator over a single tensor design and give the result of inclusion–exclusion which we prove in Appendix 1. This result allows interpolation over general echelon designs which is achieved by IE formulæ. In Sect. 4 we apply our results to sparse grids and show this with an example. At the core of this paper are two approaches to response surface modelling, the numerical analysis and the statistical viewpoints. We briefly discuss these two approaches and suggest future work in Sect. 5.

2 Designs and models

We start with a general definition of a design that covers designs for response surface modelling and designs for computer experiments. The objective is to study the effect of k continuous factors $x_1, \ldots, x_k \in \mathbb{R}$ on a real valued output y, with information collected in an experimental design.

Definition 1 A design D is a finite set of n points in $\mathbb{R}^k$. For every factor, the values of the coordinates of points in D are the levels of the design.

In this paper we are interested in a special type of designs called *echelon designs*. The following two definitions form the core of this paper: a hierarchical grid in the integers and the echelon design built from it. Both objects below are designs as of Definition 1.

Definition 2 A reference grid G is a finite subset of $\mathbb{Z}_{\geq 0}^k$ with the order ideal property, that is for every point $\alpha = (\alpha_1, \ldots, \alpha_k) \in G$, all $\alpha' \in \mathbb{Z}_{\geq 0}^k$ such that $0 \leq \alpha' \leq \alpha$ (coordinatewise) are also in G.

Definition 3 An echelon design is a design obtained from a reference grid G by changing the levels of each factor to any finite real values using a one to one mapping.

Definition 3 starts from a reference grid G to build an echelon design D. The converse procedure is important, namely that any echelon design D maps to a unique reference grid $G(D)$, hence the use of the word reference. Using the identity map, a reference grid is trivially an echelon design.

Example 1 Full factorial experiments 2^k are echelon designs. The echelon property is not limited to designs with two levels, and if the design has all combinations of levels of factors, then the full factorial design $n_1 n_2 \cdots n_k$ is an echelon design, where n_i is the number of levels of the i-th factor.

In general, regular fractions of factorial designs are not echelon. A simple case of this is the fraction 2^{2-1} with point coordinates $(-1, -1)$, $(1, 1)$ that cannot be converted
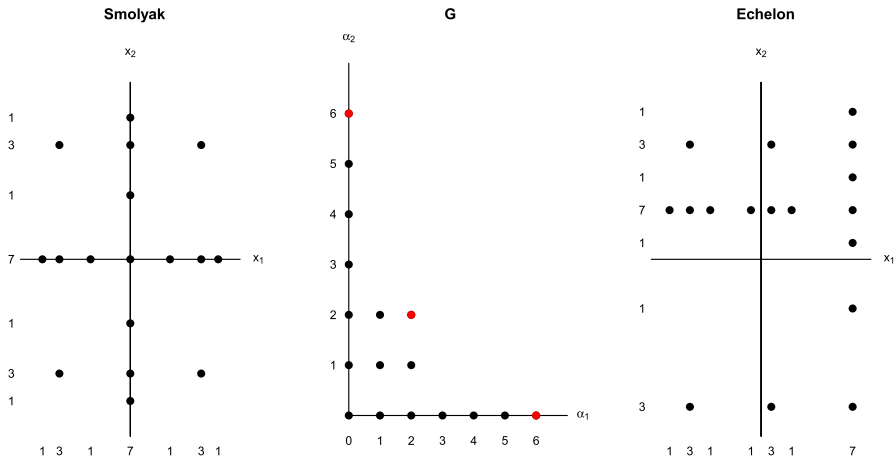
Fig. 1 From left to right, a Smolyak design, its reference grid G and an echelon design with the same reference grid. The points in red in G are the points $\mathcal{A}(G)$

Table 1 Map for generating the Smolyak design D from the grid G of Fig. 1

Coordinate of G	0	1	2	3	4	5	6
Coordinate of D	0.5	0.113	0.887	0.020	0.283	0.717	0.980
Frequency	7	3	3	1	1	1	1

The coordinates of D were obtained with the function `smolyak.quad` from the R library `gss` (Gu 2014). Given the symmetry of G, this map is valid for both coordinates x_1 and x_2

to a reference grid by relabeling levels. However, certain fractions of factorial designs are echelon designs because they can be converted to a reference grid, such as the designs used in the method of *elementary effects* (EE) for screening active factors in computer experiments, see (Morris 1991).

Example 2 The central panel of Fig. 1 shows a reference grid G with $n = 17$ points, which is transformed into a Smolyak design by the map shown in Table 1. For instance, all the 7 points of G with horizontal (vertical) coordinate equal to zero are relabeled to a x_1 (x_2) coordinate with value 0.5. The map is not unique, as points of G with coordinate value 1 (or 2) could be relabelled to be one of 0.113 or 0.887. The left panel in Fig. 1 shows the Smolyak design, while the right panel shows another echelon design with the same size and reference grid G. In both Smolyak and echelon, the numbers at the bottom (left) are numbers of points with the same x_1 (x_2) coordinate. The numbers shown in G give the actual coordinates of points.

A reference grid G is defined by a special set of corner points $\mathcal{A}(G)$. This set of corners $\mathcal{A}(G)$ is comprised with those points $\alpha \in G$ such that there are no other points $\alpha' \in G$ that satisfy $\alpha \leq \alpha'$ coordinatewise. Trivially, the reference grid defined by $\mathcal{A} = \mathcal{A}(G)$ equals G, i.e. $G(\mathcal{A}) = G$. In this last statement, we slightly abused notation using G for a reference grid and $G(\mathcal{A})$ as the reference grid built from the set of extreme corners $\mathcal{A}$.

The simplest reference grid is created with a single corner point $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{Z}_{\geq 0}^k$, that is $G(\alpha) = \{\alpha' \in \mathbb{Z}_{\geq 0}^k : 0 \leq \alpha' \leq \alpha\}$. The grid $G(\alpha)$ is a product design, written equivalently as

$$G(\alpha) = \{0, \ldots, \alpha_1\} \bigotimes \cdots \bigotimes \{0, \ldots, \alpha_k\},$$

and we refer to it as a *tensor* design. A tensor grid $G(\alpha)$ is already a full factorial design and any echelon created from it will also be a full factorial.

We now discuss the polynomial models for our designs. The building block for the model is a monomial whose exponents are entries of α, that is $x^\alpha = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$. We write the model over the tensor grid $G(\alpha)$ as

$$y_\alpha(x) = \sum_{\alpha' \in G(\alpha)} \theta_{\alpha'} x^{\alpha'},$$

and in general for a reference grid G with corners $\mathcal{A}(G)$ we have the model:

$$y_{\mathcal{A}(G)}(x) = \sum_{\alpha' \in G} \theta_{\alpha'} x^{\alpha'}.$$

This model has observations collected in a reference grid design G and uses points α in G as model exponents. This duality between design and model was referred to in the introduction, see also (Pistone et al. 2001). When we consider the elements of G as exponents of a model, then $\mathcal{A}(G)$ are the exponents of the set of *directing monomials*, see (Bates et al. 2003).

Although the model $y_{\mathcal{A}(G)}(x)$ uses all the points and exponents in G, we refer to it by its most extreme points $\mathcal{A}(G)$. This notation is necessary for later as our methods depend on inclusion exclusion. We also use the notation $G(\mathcal{A})$ when we wish to recapture the reference grid from its defining corner set $\mathcal{A}$.

Example 3 Consider the reference grid G shown in Fig. 1. If we start from G, we recover its extreme points as $\mathcal{A} = \mathcal{A}(G) = \{(0, 6), (2, 2), (6, 0)\}$. Taking a slightly different approach, we recover the full grid of $n = 17$ points from $\mathcal{A}$ as $G = G(\mathcal{A})$.

The following lemma links models built with the reference grid and an echelon design. This result is taken from Pistone et al. (2001).

Lemma 1 *If D is an echelon design, then it identifies a unique model with exponent vectors given by the elements of the reference grid with extreme corner $\mathcal{A}(G(D))$.*

In the following section we develop the models for the reference grids using indicator functions.

3 Interpolation

Consider a non-negative integer vector α. We are interested in interpolation over points in tensor design $G(\alpha)$. That is, assume that at every point $\alpha' = (\alpha_1', \ldots, \alpha_k')$ of $G(\alpha)$,

a real data value $y_{\alpha'}$ is available. We will construct a polynomial interpolator using the full set of data $\{y_{\alpha'} : \alpha' \in G(\alpha)\}$.

To build the interpolator, we start by defining the indicator function for the point α' in the grid $G(\alpha)$. The indicator is the product of k one dimensional indicators obtained by Lagrange interpolation:

$$I_{\alpha'}(x) = \prod_{i=1}^{k} \prod_{\substack{m=0 \\ m \neq \alpha_i'}}^{\alpha_i} \frac{x_i - m}{\alpha_i' - m}, \tag{1}$$

where $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ is the point at which the indicator is evaluated. For simplicity we write $I_{\alpha'}(x)$ instead of the notation $I_{\alpha'}^{(\alpha)}(x)$ that makes explicit that this is the indicator function for point α' in the grid $G(\alpha)$ for a given $\alpha \in \mathbb{Z}_{\geq 0}^k$. The notation $I_{\alpha'}^{(\alpha)}(x)$ is used in Appendix 1 to prove Lemma 2.

For a point x in $G(\alpha)$, the function $I_{\alpha'}(x)$ has the indicator property

$$I_{\alpha'}(x) = \begin{cases} 1 \text{ if } x = \alpha' \\ 0 \text{ if } x \neq \alpha' \end{cases}.$$

The interpolator function over the full tensor grid $G(\alpha)$ is built as a linear combination of the indicator functions, using the values y_α' as coefficients:

$$y_\alpha(x) = \sum_{0 \leq \alpha' \leq \alpha} y_{\alpha'} I_{\alpha'}(x). \tag{2}$$

Example 4 Consider the grid $\{(\alpha_1', \alpha_2') : \alpha_1' = 0, 1; \alpha_2' = 0, 1, 2, 3\}$ of $n = 8$ points in $\mathbb{Z}_{\geq 0}^2$. This is the reference grid $G(\alpha)$ for $\alpha = (1, 3)$, equivalently written as a tensor grid $G(\alpha) = \{0, 1\} \bigotimes \{0, 1, 2, 3\}$. The indicator function of point $\alpha' = (0, 3) \in G(\alpha)$ is

$$I_{\alpha'}(x) = \frac{x_1 - 1}{0 - 1} \cdot \frac{x_2}{3} \cdot \frac{x_2 - 1}{3 - 1} \cdot \frac{x_2 - 2}{3 - 2} = -\frac{1}{6}(x_1 - 1)x_2(x_2 - 1)(x_2 - 2).$$

3.1 Interpolation on echelon design

We consider interpolation over a reference grid $G(\mathcal{A})$, for a set of corner points $\mathcal{A}$. Suppose that this grid is the union of two tensor grids

$$G(\{\alpha, \gamma\}) = G(\alpha) \cup G(\gamma).$$

Here $\mathcal{A} = \{\alpha, \gamma\}$, where α, γ are distinct vectors with non-negative integer entries. To avoid trivial cancellations, in what follows the vectors α, γ do not dominate each other, that is neither $\alpha \leq \gamma$ nor $\alpha \geq \gamma$. By this requirement we avoid the inclusions $G(\alpha) \subset G(\gamma)$ and $G(\alpha) \supset G(\gamma)$.

The following is a simple inclusion–exclusion for the designs, using indicator functions in an obvious way:

$$I(G(\{\alpha, \gamma\})) = I(G(\alpha)) + I(G(\gamma)) - I(G(\alpha) \cap G(\gamma)),$$

where the indicator $I(G)$ takes the unit value on grid points in G and zero on integer vectors (grid points) outside G.

The greatest common divisor between monomials x^α and x^γ is the monomial $x^{\alpha \wedge \gamma}$, with $\alpha \wedge \gamma := (\min(\alpha_1, \gamma_1), \min(\alpha_2, \gamma_2), \ldots, \min(\alpha_k, \gamma_k))$. Note that $\alpha \wedge \gamma$ is the extreme corner point of the tensor grid $G(\alpha \wedge \gamma) = G(\alpha) \cap G(\gamma)$.

A main result of this paper is that the set inclusion–exclusion (IE) for the design points extends to the interpolators. The proof is given in Appendix 1.

Lemma 2 *Let $y_\alpha(x)$, $y_\gamma(x)$ and $y_{\alpha \wedge \gamma}(x)$ be the interpolator functions over grids $G(\alpha)$, $G(\gamma)$ and $G(\alpha \wedge \gamma)$, respectively; and let $y_{\{\alpha, \gamma\}}(x)$ be the interpolator function over the reference grid with corners $\mathcal{A} = \{\alpha, \gamma\}$, that is over $G(\alpha) \cup G(\gamma)$. Then the interpolator $y_{\{\alpha, \gamma\}}(x)$ satisfies the following inclusion–exclusion formula*

$$y_{\{\alpha, \gamma\}}(x) = y_\alpha(x) + y_\gamma(x) - y_{\alpha \wedge \gamma}(x). \tag{3}$$

Example 5 Consider $G(\gamma)$ with $\gamma = (2, 2)$ and $G(\alpha)$ of Example 4 with corner point $\alpha = (1, 3)$. These grids determine a design $D = G(\alpha) \cup G(\gamma) \subset \mathbb{Z}_{\geq 0}^2$ of $n = 11$ points.

We now show the interpolating property for point $x = (0, 3) \in G(\alpha)$, that is we show that $y_D(x)$ equals $y_{(0,3)}$ when evaluated at $x = (0, 3)$. There are three terms of Equation (3) to be evaluated at $x = (0, 3)$, and given the location of x in $G(\alpha)$ but not in $G(\gamma)$, this computation uses the second case in the proof of Lemma 2.

The first of the three terms in Equation (3) is $y_\alpha(x)$, to be evaluated at x. As y_α interpolates over $G(\alpha)$, this value is $y_{(0,3)}$, precisely the value that we wish to interpolate. The second evaluation is for $y_\gamma(x)$, composed of 9 summand terms. Of these, only the indicators $I_{(0,0)}^{(\gamma)}(x)$, $I_{(0,1)}^{(\gamma)}(x)$ and $I_{(0,2)}^{(\gamma)}(x)$ have non zero value when evaluated at x. For the first one, when evaluated at $x = (0, 3)$, the value is

$$I_{(0,0)}^{(\gamma)}(x) = \frac{1}{4}(x_1 - 1)(x_1 - 2)(x_2 - 1)(x_2 - 2) = 1,$$

and the other two values are -3 and 3 respectively so that

$$y_\gamma(x) = y_{(0,0)} - 3y_{(0,1)} + 3y_{(0,2)}$$

at $x = (0, 3)$. The third evaluation is for $y_{\alpha \wedge \gamma}(x)$ with 6 terms. The evaluation of indicators at $x = (0, 3)$ coincide with those of $y_\gamma(x)$. Specifically, when evaluated at $x = (0, 3)$, the indicators $I_{(0,0)}^{(\alpha \wedge \gamma)}(x)$, $I_{(0,1)}^{(\alpha \wedge \gamma)}(x)$ and $I_{(0,2)}^{(\alpha \wedge \gamma)}(x)$ give values 1, -3 and 3, respectively so that at $x = (0, 3)$ we have

$$y_{\alpha \wedge \gamma}(x) = y_{(0,0)} - 3y_{(0,1)} + 3y_{(0,2)} = y_\gamma(x).$$

G({α,γ,δ})   G(α)   G(γ)   G(δ)

G(α^γ)   G(α^δ)   G(γ^δ)   G(α^γ^δ)
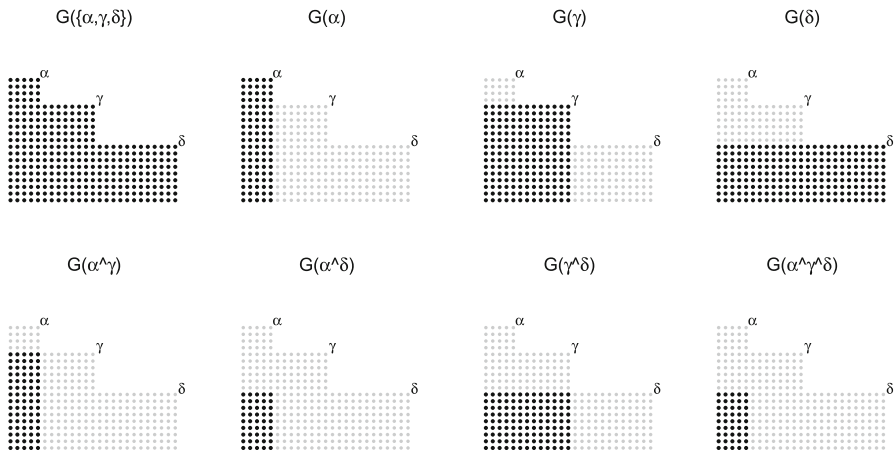

Fig. 2 Grids involved in Taylor IE

Indicators defined over grids $G(\alpha \wedge \gamma) = G(\alpha) \cap G(\gamma)$ and $G(\gamma)$ are different functions, for instance $I_{(0,0)}^{(\alpha \wedge \gamma)}(x) = -\frac{1}{2}(x_1 - 1)(x_2 - 1)(x_2 - 2)$ that coincides with $I_{(0,0)}^{(\gamma)}(x)$ at $x = (0, 3)$. We have thus shown the interpolating property at $x = (0, 3)$, that is

$$y_D(x) = y_{(0,3)} + \left(y_{(0,0)} - 3y_{(0,1)} + 3y_{(0,2)}\right) - \left(y_{(0,0)} - 3y_{(0,1)} + 3y_{(0,2)}\right) = y_{(0,3)}.$$

Example 6 In this example the grids $G(\alpha)$ and $G(\gamma)$ lie in different dimensions, consider $G(\alpha)$ with $\alpha = (2, 2, 2)$, a grid with 27 points in $\mathbb{Z}_{\geq 0}^3$; and grid $G(\gamma)$ with $\gamma = (0, 0, 3)$. Although $G(\gamma)$ is embedded in $\mathbb{Z}_{\geq 0}^3$, it is a set of 4 points lying along the x_3 coordinate only. The design is $D = G(\alpha) \cup G(\gamma)$, and we show the interpolating property for $y_D(x)$ at the point $x = (1, 1, 2) \in G(\alpha)$.

We begin with $y_\alpha(x) = y_{(1,1,2)}$ at $x = (1, 1, 2)$ because of interpolating property of y_α. For the inclusion–exclusion we require the grid $G(\alpha) \cap G(\gamma)$ that has extreme corner point $\alpha \wedge \gamma = (0, 0, 2)$. There is a single point $(0, 0, 2)$ in both $G(\gamma)$ and $G(\alpha) \cap G(\gamma)$ that shares the most coordinates with the point $(1, 1, 2)$. The indicators are $I_{(0,0,2)}^{(\gamma)}(x) = -x_3(x_3 - 1)(x_3 - 3)/2$ and $I_{(0,0,2)}^{(\alpha \wedge \gamma)}(x) = x_3(x_3 - 1)/2$ which when evaluated at $(1, 1, 2)$ take the value one. Completing Equation (3), at $x = (1, 1, 2)$ we have $y_D(x) = y_{(1,1,2)} + y_{(0,0,2)} - y_{(0,0,2)} = y_{(1,1,2)}$.

In general, reference grids we will have more than two corners and the inclusion–exclusion will require several layers. Figure 2 shows the grids involved in the IE with three generators α, γ, δ. The top left panel is the reference grid $G = G(\{\alpha, \gamma, \delta\})$; each of the other figures corresponds to a specific labeled grid; black dots are points in the grid while grey dots are points in G but outside the labeled grid.

The main result considers a reference grid defined by a set of corner points $\mathcal{A}$. We give the theorem below, whose proof by induction we omit.

Theorem 1 *Let $G(\mathcal{A})$ be a reference grid for a set of corner points $\mathcal{A}$. Then the following holds:*

(i) There is an IE for $G(\mathcal{A})$ in terms of indicators of tensor grids:

$$I(G(\mathcal{A})) = \sum_{\alpha \in \mathcal{A}} I(G(\alpha)) - \sum_{\alpha, \gamma \in \mathcal{A}, \alpha \neq \gamma} I(G(\alpha \wedge \gamma))$$

$$+ \sum_{\alpha, \gamma, \delta \in \mathcal{A}, \alpha \neq \gamma \neq \delta} I(G(\alpha \wedge \gamma \wedge \delta)) + \cdots$$

(ii) The unique interpolator $y_{\mathcal{A}}(x)$ for points in $G(\mathcal{A})$ satisfies the following

$$y_{\mathcal{A}}(x) = \sum_{\alpha \in \mathcal{A}} y_\alpha(x) - \sum_{\alpha, \gamma \in \mathcal{A}, \alpha \neq \gamma} y_{\alpha \wedge \gamma}(x) + \sum_{\alpha, \gamma, \delta \in \mathcal{A}, \alpha \neq \gamma \neq \delta} y_{\alpha \wedge \gamma \wedge \delta}(x) + \cdots$$

3.2 Betti numbers

In algebraic theory, the inclusion–exclusion of Theorem 1 is known as the Taylor resolution, which is the most complex case of IE, namely using all the singleton generators, then all possible pairs, triples and so on. The Taylor resolution is often highly redundant and it is possible to reduce the complexity of the IE by using other "resolutions" Sáenz-de Cabezón Wynn (2009). Here is a simple example to show reduction in IE complexity by cancelation.

Example 7 Consider the set of corner points $\mathcal{A} = \{(2, 0), (1, 1), (0, 2)\}$ that define a reference grid of six points in two dimensions. The full Taylor inclusion–exclusion of $I(G(\mathcal{A}))$ is:

$$\begin{aligned} I(G(\mathcal{A})) = &\ I\left(G((2, 0))\right) + I\left(G((1, 1))\right) + I\left(G((0, 2))\right) \\ &- I(G((2, 0)) \cap G((1, 1))) - I(G((1, 1)) \cap G((0, 2))) \\ &- I(\cancel{G((2, 0)) \cap G((0, 2))}) + I(\cancel{G((2, 0)) \cap G((1, 1)) \cap G((0, 2))}). \end{aligned}$$

The cancellation occurs between terms that are equal, despite coming from intersections involving different numbers of sets, that is $G((2, 0)) \cap G((0, 2)) = G((0, 0))$ which equals $G((2, 0)) \cap G((1, 1)) \cap G((0, 2))$. The interpolator has five terms as well $y_{\mathcal{A}}(x) = y_{(2,0)}(x) + y_{(1,1)}(x) + y_{(0,2)}(x) - y_{(2,0) \wedge (1,1)}(x) - y_{(1,1) \wedge (0,2)}(x)$.

The efficacy of cancellations to simplify IE depends on the specific design. IE in Example 7 needed only five terms. We give two more complex examples.

The first column of Table 2a gives the extreme corners $\mathcal{A}$ of a grid with $n = 112$ points in five factors. The rest of columns contain the layers of inclusion–exclusion for the Taylor IE, for example the column labelled "3-way" has the $\binom{5}{3} = 10$ results of the type $\alpha \wedge \gamma \wedge \delta$ for $\alpha, \gamma, \delta \in \mathcal{A}$. For this grid, the exhaustive Taylor computation is highly redundant with many cancellations so that out of the 31 terms, we only need

Table 2 Inclusion–exclusion for two grids each with five generators

| $\mathcal{A}$ | 2-way | 3-way | 4-way | 5-way |
+	−	+	−	+
(a)				
	11111	11111		
	11111	11111		
21111	11111	11111	11111	11111
12111	11111	11111	11111	
11211	11111	11111	11111	
11121	11111	11111	11111	
11112	11111	11111	11111	
	11111	11111		
	11111	11111		
	11111	11111		
(b)				
	11222	11122		
	12122	11212		
12222	12212	11221	11112	11111
21222	12221	12112	11121	
22122	21122	12121	11211	
22212	21212	12211	12111	
22221	21221	21112	21111	
	22112	21121		
	22121	21211		
	22211	22111		

The grid (a) has maximal redundancy, while (b) has maximal complexity with no redundancies. The extreme corners of each grid are given in the first column $\mathcal{A}$, and the multi-indices have simplified notation e.g., 11121 means the point with coordinates (1, 1, 1, 2, 1)

six for the IE. The indicator is

$$I(G(\mathcal{A})) = I(G((2, 1, 1, 1, 1))) + I(G((1, 2, 1, 1, 1))) + I(G((1, 1, 2, 1, 1)))$$
$$+ I(G((1, 1, 1, 2, 1))) + I(G((1, 1, 1, 1, 2))) - 4I(G((1, 1, 1, 1, 1))).$$

For this grid, or any echelon design built from it, the IE interpolator is

$$y_{\mathcal{A}}(x) = y_{(2,1,1,1,1)}(x) + y_{(1,2,1,1,1)}(x) + y_{(1,1,2,1,1)}(x)$$
$$+ y_{(1,1,1,2,1)}(x) + y_{(1,1,1,1,2)}(x) - 4y_{(1,1,1,1,1)}(x). \tag{4}$$

Table 2b has the results for a grid with $n = 242$ points and the same number of factors and generators as the grid of Table 2a. However, this case has maximal complexity and no simplifications can be made. In summary, for this grid, both IE indicator $I(G(\mathcal{A}))$ and interpolator $y_{\mathcal{A}}(x)$ need all 31 terms.
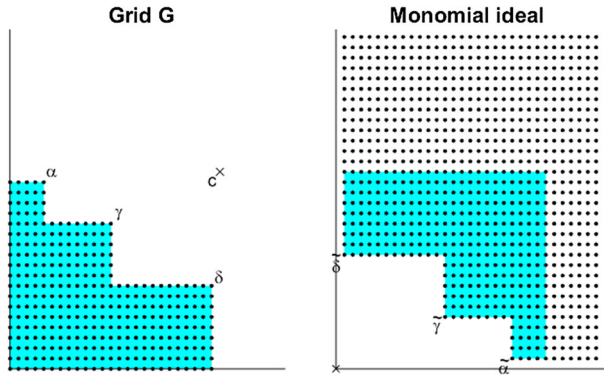
Fig. 3 The left panel shows a grid G, and the right panel shows exponents of the derived monomial ideal by flipping corners of G around point c. The shaded area is included to illustrate the flip operation

We require an efficient method to retrieve only the essential IE terms. There is no space to cover the theory of monomial ideals behind improved resolutions and simplified inclusion–exclusions for interpolators. We however we give a short explanation.

An efficient way of construction the sought inclusion–exclusion uses results from monomial ideals (Sáenz-de Cabezón Wynn 2009). Recall that a monomial ideal is a set of polynomials, and in our case, we build the ideal by flipping the grid G around a point. After flipping, the points become the exponents of monomials that generate an ideal. We use the shorthand that the monomial ideal obtained by this flipping the corners in $\mathcal{A}$ around a point c is the *derived* ideal. The objective of creating the derived ideal is to determine efficiently the IE decomposition of the grid G.

Figure 3 gives a graphical explanation of the process. We flip the corners in $\mathcal{A}$ around a point c as follows, every corner $\alpha \in \mathcal{A}$ is flipped to become $\tilde{\alpha} = c - \alpha$ and similarly for γ, δ in the same figure. When the whole grid is flipped, the point c becomes the origin as shown in the right panel. The choice of flip point c is not critical as long as it is at least equal to the maximum value of G for the coordinate in turn. In the figure, c was even beyond the maximum value per coordinate which has no influence in the result.

Example 8 We continue with the grid G of Example 7. To be specific, if the corner points $\mathcal{A}$ are flipped around the maximum point per coordinate $c = (2, 2)$, we retrieve the same $\mathcal{A}$ which when turned to exponents yields $x_1^2, x_1 x_2, x_2^2$ which are used to generate the derived ideal. For the generator corners of G in Example 2, we flip around $c = (6, 6)$ and convert to exponents so that we have ideal generators $x_1^6, x_1^4 x_2^4, x_2^6$. In a more complex case, the grid of Example 6 uses flip point $c = (2, 2, 3)$ which gives monomials x_3, x_1^2, x_2^2.

The result of homological computations with ideals is a collection of *Betti numbers* $\beta_{\alpha, j}$. These numbers are non-negative integers that provide the coefficients in a maximally reduced IE and thus are what we are looking for. Each number is indexed by a vector of non-negative integers α and an integer j that controls the role of the number in the inclusion–exclusion. We only use positive $\beta_{\alpha, j}$, as those zeroes do not

contribute to the result and for simplicity, we collect the Betti numbers for the same j in a set that we label B_j. The computations are carried out using, for example, the function `HilbertSeriesMultiDeg` of the CoCoA computer package (Abbott et al.). The results of computations over the derived ideal have to be reversed around c so that we can use them for the grid.

The following theorem is a simplification of the exhaustive series of Theorem 1 and it is given without a proof.

Theorem 2 *Let $G(\mathcal{A})$ be an echelon for a set of corner points $\mathcal{A}$. Let $\beta_{\alpha,j}$ be the multigraded Betti numbers associated with the minimal free resolution of the derived monomial ideal, and let the non-zero Betti numbers with the same value of j be collected in B_j. Then*

(i) There is an IE for $I(G(\mathcal{A}))$:

$$I(G(\mathcal{A})) = \sum_{j=0}^{k-1}(-1)^j \sum_{\alpha \in B_j} \beta_{\alpha,j} I(G(\alpha)),$$

(ii) The unique interpolator over $G(\mathcal{A})$ can be decomposed into tensor grid interpolator of the same form:

$$y_{\mathcal{A}}(x) = \sum_{j=0}^{k-1}(-1)^j \sum_{\alpha \in B_j} \beta_{\alpha,j} y_{\alpha}(x).$$

Theorems 1 and 2 yield exactly the same result concerning either the composition of $I(G(\mathcal{A}))$, or the interpolation formulæ for $y_{\mathcal{A}}(x)$ over $G(\mathcal{A})$. However the former is much more complex, involving $2^{|\mathcal{A}|} - 1$ terms, while the latter has many of these redundancies removed. As shown by example earlier, the reduction in complexity achieved with Theorem 2 depends on the generators of the derived ideal. The following example continues an earlier grid with maximal redundancy.

Example 9 Consider the grid of Table 2a, which will be flipped around the point $c = 22222$, written using the notation of Table 2. The derived ideal only requires the extreme corners, that once flipped are 01111, 10111, 11011, 11101 and 11110 so that the ideal is generated by monomials $x_2x_3x_4x_5$, $x_1x_3x_4x_5$, $x_1x_2x_4x_5$, $x_1x_2x_3x_5$ and $x_1x_2x_3x_4$. The non-zero Betti numbers are $\tilde{\beta}_{01111,0} = 1$, $\tilde{\beta}_{10111,0} = 1$, $\tilde{\beta}_{11011,0} = 1$, $\tilde{\beta}_{11101,0} = 1$, $\tilde{\beta}_{11110,0} = 1$ and $\tilde{\beta}_{11111,1} = 4$, where the tilde indicates that results refer to the derived ideal. To use these numbers in the grid, we need to reverse the multi-indices around c, dropping the tilde notation in the process. The Betti numbers are $\beta_{21111,0} = 1$, $\beta_{12111,0} = 1$, $\beta_{11211,0} = 1$, $\beta_{11121,0} = 1$, $\beta_{11112,0} = 1$ and $\beta_{11111,1} = 4$. With these numbers we retrieve the simplified IE interpolator of Equation (4). By contrast, the grid of Table 2b yields all Betti numbers for multi-indices in Table 2b equal to one, with no simplification possible.

4 Extension to sparse grids

Sparse grids are designs used in stochastic finite element approximation, numerical integration and interpolation Nobile et al. (2008) and Plumlee (2014). They have a reduced number of runs while preserving the accuracy of approximations.

To construct a sparse design, we use the construction of Definition 3. That is, we start from a reference grid $G(\mathcal{A})$. Recall that although the reference grid lies in $\mathbb{Z}_{\geq 0}^k$, it is already an echelon design, and each factor x_i has n_i levels $0, 1, \ldots, n_i - 1$. Next, define new levels for the design D, in which the levels of variable x_i are replaced by new levels

$$z_{i,1}, \ldots, z_{i,n_i},$$

for $i = 1, \ldots, k$. The replacement of levels is arbitrary and only requires a one to one mapping, hence the design D may not resemble at all its reference grid G. The operations described above can be reversed and design D in the class defined by the operation has a unique reference grid $G(D)$.

The reversibility of operations implies that the design D still identifies a single polynomial model with exponents given by points in the reference grid G. The main result of the paper is that the inclusion–exclusion results for the tensor designs and interpolators that hold for the echelon designs extend to the sparse designs obtained by a set of transformation of level of each factor. We shall now call such a design a Smolyak grid.

It is clear that any component tensor grid of the reference grid is transformed into a tensor grid in the design D. Thus a Smolyak grid defined from a reference grid $G(D)$ has a set of defining corner points $\mathcal{A}$ which then indexes an IE, for D.

We use the simple notation $y_{\tilde{\alpha}}$ to indicate the interpolator over the block defined by the "corner" α, and $\tilde{\mathcal{A}}$ for the full Smolyak grid. Over the sparse grid $\tilde{\mathcal{A}}$, the (Betti) reduced IE polynomial interpolator is

$$y_{\tilde{\mathcal{A}}}(x) = \sum_{j=0}^{k-1} (-1)^j \sum_{\alpha \in B_j} \beta_{\alpha,j} y_{\tilde{\alpha}}(x).$$

Example 10 Consider a Smolyak design D with $k = 2$ factors and $n = 161$ points and with 15 levels per factor. The reference grid G has generators $\{(14, 6), (6, 14)\}$ so that the inclusion–exclusion involves three tensor grids and corresponding design points as shown in Fig. 4.

5 Conclusions and future research

The sparse grids described are based on a reference grid and despite the apparently predominant one-factor-at-a-time structure, the relabelling of coordinates allows an efficient exploration of the design space while keeping the number of runs low. These
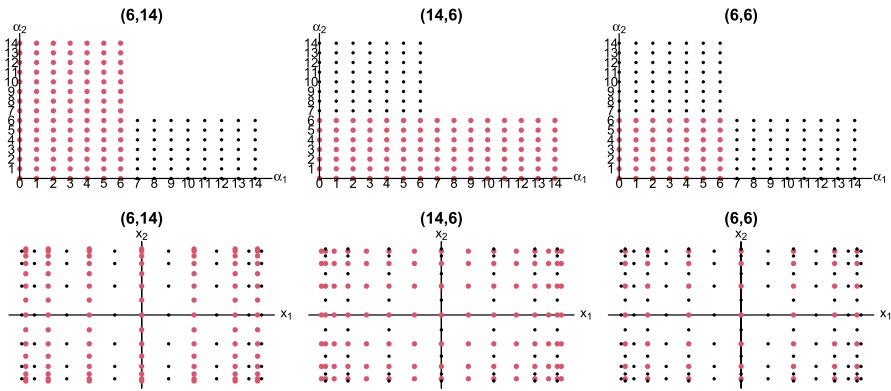
Fig. 4 The plots in the top row show in red the points of grid $G(\alpha)$ and in black the points in $G(\mathcal{A}) \setminus G(\alpha)$. Each plot is headed by α. The plots in the bottom row are the corresponding points in the Smolyak design

designs are an important subclass of experimental designs both from a theoretical point of view and because of their widespread use in applications.

Sparse grids have been mainly used in two disciplines: numerical analysis (NA) and experimental design (DE). A motivation for the current paper was to rediscover and clarify intricate formulæ used in the numerical analysis literature by using the Betti number decomposition of the last section.

For a fixed reference grid $G(\mathcal{A})$, the construction of a design for which $G(D) = G(\mathcal{A})$ still requires the specification of the numerical values of levels of each factor $z_{i,1}, \ldots, z_{i,n_i}$. Thus if the design is to optimised, it should be optimised with respect both the reference grid $G(D)$ and the spacings.

By analogy with factorial design and response surface design, one may outline a design first without specifying the spacing. A "star-composite design" is an example, being made up of a factorial design and a star without specifying the spacing of either. The spacing may be decided later based on the shape of the whole design region, preset candidate points, optimum design or some structural property e.g., rotatability. We could say, informally, that we shown one method of producing a generalised star design, but have left till later issues of spacing. This is partly because the numerical analysis definitions of optimality and the statistical ones are different so that more research is needed. We also note that although we worked on saturated models, actions like adding replicate points to improve statistical capabilities do not change the model identified.

In NA one may consider an approximation of a function by $y_{\tilde{\mathcal{A}}}(x)$ and select the spare grid $\tilde{\mathcal{A}}$ making use of the inequality

$$\left\| f - y_{\tilde{\mathcal{A}}}(x) \right\| \le \left\| f - y^* \right\| \times \Delta,$$

where y^* is some best approximator to f, in some class, Δ is the so-called Lebesgue constant, and $\| \cdot \|$ a chosen metric. But sparse grids are beginning to attract interest in the statistical literature; an example is Plumlee (2014). There is an opportunity to draw the extensive literature on optimal design for application to sparse grids, in the choice

of reference grid and spacing, whether for physical response surface experiments or computer experiments.

Whether in NA or DE, the use of algebra reduces complexity by tensor decomposition and we believe, this will help us to understand in future research the important trade off between sparsity and optimality.

Appendix A—Proof of IE theorem

Proof The proof is the verification of the interpolating property of $y_{\{\alpha,\gamma\}}(x)$ at grid points α' in $D = G(\mathcal{A}) = G(\alpha) \cup G(\gamma)$. There are three cases to consider, based upon the partition of D into disjoint sets $G(\alpha) \cap G(\gamma)$, $G(\alpha)\backslash(G(\alpha) \cap G(\gamma))$ and $G(\gamma) \backslash (G(\alpha) \cap G(\gamma))$. In what follows we compute $y_{\{\alpha,\gamma\}}(\alpha')$, for α' taking values in each of the three disjoint sets.

Case 1. The first case we consider is for the intersection of grids, that is when $\alpha' \in G(\alpha) \cap G(\gamma)$. Concerning $y_\alpha(x)$, this function interpolates over the intersection of grids $G(\alpha) \cap G(\gamma)$ as $G(\alpha) \cap G(\gamma) \subset G(\alpha)$ so that $\alpha' \in G(\alpha) \cap G(\gamma)$ implies $\alpha' \in G(\alpha)$ and finally $y_\alpha(\alpha') = y_{\alpha'}$. A similar argument shows that $\alpha' \in G(\gamma)$ and thus $y_\gamma(\alpha') = y_{\alpha'}$. Finally, the function $y_{\alpha \wedge \gamma}(x)$ also has interpolating property in the intersection $G(\alpha) \cap G(\gamma)$, that is $y_{\alpha \wedge \gamma}(\alpha') = y_{\alpha'}$ and

$$y_{\{\alpha,\gamma\}}(\alpha') = y_\alpha(\alpha') + y_\gamma(\alpha') - y_{\alpha \wedge \gamma}(\alpha')$$
$$= y_{\alpha'} + y_{\alpha'} - y_{\alpha'} = y_{\alpha'}.$$

Case 2. The second case is when $\alpha' \in G(\alpha)\backslash(G(\alpha) \cap G(\gamma))$. The function $y_\alpha(x)$ has interpolating property because $\alpha' \in G(\alpha)\backslash(G(\alpha) \cap G(\gamma)) \subset G(\alpha)$ and thus $y_\alpha(\alpha') = y_{\alpha'}$. Neither of the functions $y_\gamma(x)$ and $y_{\alpha \wedge \gamma}(x)$, have interpolating properties for $\alpha' \in G(\alpha)\backslash(G(\alpha) \cap G(\gamma))$. To show why this is the case, recall Equation (2) for interpolators over $G(\gamma)$ and over $G(\alpha) \cap G(\gamma) = G(\alpha \wedge \gamma)$:

$$y_\gamma(x) = \sum_{0 \leq \alpha'' \leq \gamma} y_{\alpha''} I_{\alpha''}^{(\gamma)}(x)$$

and

$$y_{\alpha \wedge \gamma}(x) = \sum_{0 \le \alpha'' \le \alpha \wedge \gamma} y_{\alpha''} I_{\alpha''}^{(\alpha \wedge \gamma)}(x),$$

where we used the notation $I_{\alpha''}^{(\gamma)}(x)$, $I_{\alpha''}^{(\alpha \wedge \gamma)}(x)$ emphasizing that these are indicators over different reference grids, namely $G(\alpha)$ and $G(\alpha \wedge \gamma)$. The ranges of summation above are $G(\alpha)$ and $G(\alpha) \cap G(\gamma)$ respectively and neither of these include the current case $\alpha' \in G(\alpha) \setminus (G(\alpha) \cap G(\gamma))$ hence $y_\gamma(x)$ and $y_{\alpha \wedge \gamma}(x)$ do not interpolate for α'. However, the core part of this case is that when evaluated at $x = \alpha'$, these two functions take the same value

$$y_\gamma(\alpha') = y_{\alpha \wedge \gamma}(\alpha'),$$

which we show next. By the construction of $y_\gamma(x)$, the only non zero terms $I_{\alpha''}^{(\gamma)}(\alpha')$ of $y_\gamma(\alpha')$ are for those points in $G(\gamma)$ that share the largest number of coordinates with the given α', as no single point shares all coordinates with α' because $\alpha' \notin G(\gamma)$. Without lack of generality, assume that a point $\alpha'' \in G(\gamma)$ shares the first $k-1$ coordinates with $\alpha' \in G(\alpha) \setminus (G(\alpha) \cap G(\gamma))$, and that the coordinate that is not equal between this generic $\alpha'' \in G(\gamma)$ and α'_k is the k-th coordinate, that is $\alpha' = (\alpha''_1, \alpha''_2, \ldots, \alpha''_{k-1}, \alpha'_k)$. The inner product in the i-th coordinate in Equation (1) is

$$\prod_{\substack{m=0 \\ m \ne \alpha''_i}}^{\gamma_i} \frac{x_i - m}{\alpha''_i - m} = \frac{x_i}{\alpha''_i} \frac{x_i - 1}{\alpha''_i - 1} \cdots \frac{x_i - \alpha''_i + 1}{\alpha''_i - \alpha''_i + 1} \frac{x_i - \alpha''_i - 1}{\alpha''_i - \alpha''_i - 1} \cdots \frac{x_i - \gamma_i}{\alpha'_i - \gamma_i}, \quad \text{(A1)}$$

and this product is equal to one because it is evaluated at $x_i = \alpha''_i$. This result holds for coordinates $i = 1, \ldots, k-1$.

To complete our development for $I_{\alpha''}^{(\gamma)}(\alpha')$, we compute the product of Equation (A1) evaluated at $x_k = \alpha'_k$ that becomes the indicator, that is

$$I_{\alpha''}^{(\gamma)}(\alpha') = \frac{\alpha'_k}{\alpha''_k} \frac{\alpha'_k - 1}{\alpha''_k - 1} \cdots \frac{\alpha'_k - \alpha''_k + 1}{\alpha''_k - \alpha''_k + 1} \frac{\alpha'_k - \alpha''_k - 1}{\alpha''_k - \alpha''_k - 1} \cdots \frac{\alpha'_k - \gamma_k}{\alpha''_k - \gamma_k}.$$

This non-zero quantity is the coefficient of $y_{\alpha''}$ in the evaluation of $y_\gamma(\alpha')$.

For the indicator $I_{\alpha \wedge \gamma, \alpha'}(\alpha')$ the results above hold, namely that the indicator terms are zero except for when α'' shares all but one coordinate with α' and that the product of terms equals one. This is true despite the fact that the product of Equation (A1) in $I_{\alpha \wedge \gamma, \alpha'}(x)$ runs for values of m in the range $0, 1, \alpha''_i - 1, \alpha''_i + 1, \ldots, \min(\alpha_i, \gamma_i)$ with $\min(\alpha_i, \gamma_i) \le \gamma_i$. In other words, indicators for different reference grids involve a different number of products.

The only fact needed to complete the equality between $y_\gamma(\alpha')$ and $y_{\alpha \wedge \gamma}(\alpha')$ is to show that $I_{\alpha''}^{(\gamma)}(\alpha') = I_{\alpha''}^{(\alpha \wedge \gamma)}(\alpha')$, namely that the indices of the last coordinate run

over the same range for $I_{\alpha''}^{(\alpha \wedge \gamma)}(\alpha')$. This is true because for the k-th coordinate we have $(\alpha \wedge \gamma)_k = \min(\alpha_k, \gamma_k) = \gamma_k$ and this is the case because the $G(\alpha)$ differs (exceeds) at least on the $k - th$ coordinate and for this to happen $\alpha_k > \gamma_k$. This completes this second case and we have shown that $y_\gamma(\alpha') = y_{\alpha \wedge \gamma}(\alpha')$ so that

$$
\begin{aligned}
y_{\{\alpha, \gamma\}}(\alpha') &= y_{\alpha'} + y_\gamma(\alpha') - y_{\alpha \wedge \gamma}(\alpha') \\
&= y_{\alpha'} + y_\gamma(\alpha') - y_\gamma(\alpha') = y_{\alpha'}.
\end{aligned}
$$

Case 3. The third case $\alpha' \in G(\gamma) \backslash (G(\alpha) \cap G(\gamma))$ is shown exactly as the second case, only reversing with the role of $G(\alpha)$ by taking $G(\gamma)$.

Combining the three cases, we have shown the interpolation property of $y_{\{\alpha, \gamma\}}(x)$ of Equation (3) over $G(\mathcal{A}) = G(\alpha) \cup G(\gamma)$, that is $y_{\{\alpha, \gamma\}}(\alpha') = y_{\alpha'}$ for $\alpha' \in G(\mathcal{A})$. □

References

Abbott J, Bigatti AM, Robbiano L.: CoCoA: a system for doing computations in commutative algebra. http://cocoa.dima.unige.it

Barthelmann V, Novak E, Ritter K (2000) High dimensional polynomial interpolation on sparse grids. Adv Comp Math 12(4):273–288

Bates RA, Giglio B, Wynn HP (2003) A global selection procedure for polynomial interpolators. Technometrics 45(3):246–255

Edelsbrunner H, Harer JL (2010) Computational topology. American Mathematical Society, Providence

Gu C (2014) Smoothing Spline ANOVA Models: R Package GSS. J Stat Softw 58(5):1–25

Maruri-Aguilar H, Wynn HP (2015) The algebraic method in experimental design. In: Dean A, Morris M, Stufken J, Bingham D (eds) Handbook of design and analysis of experiments. Chapman & Hall/CRC, pp 415–456

Maruri-Aguilar H, Sáenz-de Cabezón E, Wynn HP (2012) Alexander duality in experimental designs. Ann Inst Stat Math 65:667–686

Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33:161–74

Naiman DQ, Wynn HP (1997) Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. Ann Stat 25(5):1954–1983

Nobile F, Tempone R, Webster CG (2008) A sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J Numer Anal 46(5):2309–2345

Novak E, Ritter K (1996) High dimensional integration of smooth functions over cubes. Numer Math 75(1):79–97

Pistone G, Wynn HP (1996) Generalised confounding with Gröbner bases. Biometrika 83(3):653–666

Pistone G, Riccomagno E, Wynn HP (2001) Algebraic statistics, vol 89. Monographs on statistics and applied probability. Chapman & Hall CRC, Boca Raton

Plumlee M (2014) Fast prediction of deterministic functions using sparse grid experimental designs. J Am Stat Assoc 109(508):1581–1591

Sáenz-de Cabezón E, Wynn HP (2009) Betti numbers and minimal free resolutions for multi-state system reliability bounds. J Symb Comput 44(9):1311–1325

Smolyak SA (1963) Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl Akad Nauk SSSR 148:1042–1045