INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 6, Number 2, Pages 256–273

ANOVA EXPANSIONS AND EFFICIENT SAMPLING METHODS FOR PARAMETER DEPENDENT NONLINEAR PDES

YANZHAO CAO, ZHENG CHEN, AND MAX GUNZBURGER

Abstract. The impact of parameter dependent boundary conditions on solutions of a class of nonlinear partial differential equations and on optimization problems constrained by such equations is considered. The tools used to gain insights about these issues are the Analysis of Variance (ANOVA) expansion of functions and the related notion of the effective dimension of a function; both concepts are reviewed. The effective dimension is then used to study the accuracy of truncated ANOVA expansions. Then, based on the ANOVA expansions of functionals of the solutions, the effects of different parameter sampling methods on the accuracy of surrogate optimization approaches to constrained optimization problems are considered. Demonstrations are given to show that whenever truncated ANOVA expansions of functionals provide accurate approximations, optimizers found through a simple surrogate optimization strategy are also relatively accurate. Although the results are presented and discussed in the context of surrogate optimization problems, most also apply to other settings such as stochastic ensemble methods and reduced-order modeling for nonlinear partial differential equations.

Key Words. ANOVA expansions, nonlinear partial differential equations, surrogate optimization, parameter sampling methods.

1. Introduction

The type of problems we consider requires the solutions of equations such as

$$F(u;\vec{\alpha}) = 0, \tag{1}$$

where $\vec{\alpha} \in A \subseteq \mathbb{R}^p$ is a vector of parameters and A is some admissibility set. In particular, we are interested in problems for which $F(\cdot;\vec{\alpha})$ represents a nonlinear partial differential equation or system. The specific situation that interests us is one in which approximate solutions of problems involving (1) are determined by using the solutions to the problems

$$F(u^{(j)};\vec{\alpha}^{(j)}) = 0 \quad j = 1,\dots,N, \tag{2}$$

where $\{\vec{\alpha}^{(j)}\}_{j=1}^{N}$ are a chosen set of parameter values. Settings in which such problems arise include ensemble approximations of solutions of (1) in case the components of the parameter vector $\vec{\alpha}$ are random variables with given probability distributions; building reduced-order models from solutions of (1) at sample values

Received by the editors September 11, 2008.

2000 *Mathematics Subject Classification.* 62H99, 65N30.

Y. Cao was supported in part by the United States Air Force Office of Scientific Research under grant number FA9550-07-1-0154 and the National Science Foundation under grant number DMS0609918. M. Gunzburger was supported in part by the United States Air Force Office of Scientific Research under grant number FA9550-08-1-0415.

of the parameter vector $\vec{\alpha}$; and the surrogate optimization of a functional. Here, we focus on the third setting; however, most of the discussions in this paper can be translated to the other settings.

For surrogate optimization problems, we are given a functional $\mathcal{J}(u)$ and are asked to find $\vec{\alpha}^* \in A$ and a corresponding u^* that solve the problem

$$\min_{\vec{\alpha}\in A} \mathcal{J}(u) \quad \text{subject to} \quad F(u;\vec{\alpha}) = 0, \tag{3}$$

where A is a bounded subset of $\mathbb{R}^p$. In this setting, u denotes the state variable, $\vec{\alpha}$ the vector of design parameters, and the constraint equation $F(u;\vec{\alpha}) = 0$ the state system. Note that through the constraint, the functional $\mathcal{J}(u)$ is implicitly a function of the components of the parameter vector $\vec{\alpha}$. A simple, derivative-free approach to finding approximate solutions of the problem (3) is to *first choose particular values* $\{\vec{\alpha}^{(j)}\}_{j=1}^N$ *for the parameters*, then solve the problems in (2), and then use those solutions to evaluate the functional so that one obtains, for $j = 1, \ldots, N$, the values $\mathcal{J}(u^{(j)})$ corresponding to the parameter vectors $\vec{\alpha}^{(j)}$. One would then use this information to build, e.g., by a least-squares or interpolatory method, a surrogate function $\mathcal{J}_{sur}(\vec{\alpha})$ defined over the parameter subset A that can be used as an approximation to $\mathcal{J}(u(\vec{\alpha}))$ over A. Finally, one would approximate the solution of the optimization problem (3) by the parameter values that minimize the simpler functional $\mathcal{J}_{sur}(\vec{\alpha})$, i.e.

$$\vec{\alpha}^* \approx \vec{\alpha}^*_{sur}, \quad \text{where } \vec{\alpha}^*_{sur} \text{ solves the problem} \quad \min_{\vec{\alpha}\in A} \mathcal{J}_{sur}(\vec{\alpha}). \tag{4}$$

Building the surrogate functional requires the evaluation of the functional $\mathcal{J}(\cdot)$ at the points $\{\vec{\alpha}^{(j)}\}_{j=1}^N$ sampled within the set A. In turn, evaluating the functional at the N parameter points requires N solves of the constraint equation as in (2). Since the latter step involves solving a nonlinear partial differential equation system, it dominates the overall computation; thus, the constraint equation should be solved as few times as possible. Thus, we want to sample only a "few" points in A, i.e., we want to sample sparsely. In addition, in practice, p, the number of control parameters, may be large so that, for the surrogate optimization problem, we are interested in intelligent, *sparse sampling in possibly high dimensions.*

In this paper, we treat a simple model problem, but the need for intelligent sampling would be even greater in more complicated settings. We even simplify things some more by assuming that the parameter vector is constrained to belong to a hypercube, that its components have no known bias or correlation so that we will sample them uniformly and independently, and that they appear linearly in the definition of the problem. Clearly, this work is only the beginning of what should be a much larger study that encompasses more general and more realistic situations.

The particular focus of this paper is to explore the connections that ANOVA (Analysis of Variance) expansions of multivariate functions (and the related notion of the effective dimension of those functions) have with the solution of parameter-dependent nonlinear partial differential equations. Specifically, we will study the general approximation properties of ANOVA expansions for functionals of solutions of nonlinear partial differential equations and the implications that particular features these expansions possess have on those solutions and on how one builds surrogate functionals.

Of course, the problem of solving partial differential equations with parameter-dependent input data is an active research area; see e.g., [1,4,12–15,17,19–22,26, 28,29]. However, these works are mostly focused at finding approximate solutions of partial differential equations problems while, in this paper, ANOVA expansions are mainly used to gain insight into the nature of solutions of parameter-dependent partial differential equations. In this sense, our study is similar to studies found in, e.g., [11].

The paper is organized as follows. In §2, we briefly discuss ANOVA expansions and some of their properties and the concept of effective dimensions. In §3, we define the model problem we use as a basis for our study of ANOVA expansions and their relation to solutions of partial differential equations, including a study the approximation property of the ANOVA expansion for solutions of the model problem in the case of small nonlinear perturbations. Then, in §4, we examine the ANOVA expansions and the effective dimension of several functionals of the solutions of the model problem. Finally, in §5, we consider several parameter sampling strategies and how they affect the accuracy of a simple surrogate optimization method.

We emphasize that we are not necessarily advocating ANOVA expansions as an approximation method, i.e., as a method for determining approximate solutions of nonlinear partial differential equations depending on parameters. Instead, this paper should be viewed as an effort towards acquainting the numerical analysis for partial differential equations community to the possibilities offered by ANOVA expansions and the notion of the effective dimension of functions so as to encourage further work in this direction.

2. ANOVA expansions and effective dimensions of multivariate functions

2.1. ANOVA expansions of multivariate functions. We review the ANOVA expansion of a function. ANOVA expansions are *exact* and contain a *finite* number of terms, although truncations of ANOVA expansions may provide good approximations with fewer terms. Our presentation is brief; further details may be obtained from [5, 8, 24].

Let $P = \{1, \dots, p\}$; for any subset of (ordered) coordinate indices $T \subseteq P$, let $|T|$ denote the cardinality of T, let $\vec{\alpha}_T \in \mathbb{R}^{|T|}$ denote the $|T|$-vector containing the components of the vector $\vec{\alpha} \in \mathbb{R}^p$ indexed by T, and let $A_T^{|T|}$ denote the $|T|$-dimensional unit hypercube which is the projection of the p-dimensional unit hypercube A^p onto the coordinates indexed by T. Any function $g \in L^2(A^p)$ may be written as the ANOVA expansion

$$g(\vec{\alpha}) = g_0 + \sum_{T \subseteq P} g_T(\vec{\alpha}_T), \tag{5}$$

where the terms in the expansion are determined recursively by

$$g_T(\vec{\alpha}_T) = \int_{A^{p-|T|}} g(\vec{\alpha})\, d\vec{\alpha}_{P \setminus T} - \sum_{V \subset T} g_V(\vec{\alpha}_V) - g_0 \tag{6}$$

starting with

$$g_0 = \int_{A^p} g(\vec{\alpha})\, d\vec{\alpha}$$

and where, by convention,

$$\int_{A^0} g(\vec{\alpha})\, d\vec{\alpha}_\emptyset = g(\vec{\alpha}).$$

Note that the integration in (6) is carried out over those coordinates having indices not included in the set T and that the sum is over strict subsets of T. The total number of terms in the expansion is 2^p. Each term is, in general, a nonlinear function of its arguments.

We explicitly write out the first few terms in the expansion. We have that

$$g_i(\alpha_i) = \int_{A^{p-1}} g(\vec{\alpha})\, d\vec{\alpha}' - g_0 \qquad \text{for } i = 1, \dots, p,$$

$$g_{ij}(\alpha_i, \alpha_j) = \int_{A^{p-2}} g(\vec{\alpha})\, d\vec{\alpha}'' - g_i(\alpha_i) - g_j(\alpha_j) - g_0 \qquad \text{for } i < j,\ i, j = 1, \dots, p,$$

$$g_{ijk}(\alpha_i, \alpha_j, \alpha_k) = \int_{A^{p-3}} g(\vec{\alpha})\, d\vec{\alpha}''' - g_{ij}(\alpha_i, \alpha_j) - g_{ik}(\alpha_i, \alpha_k) - g_{jk}(\alpha_j, \alpha_k) \\ - g_i(\alpha_i) - g_j(\alpha_j) - g_k(\alpha_k) - g_0 \qquad \text{for } i < j < k,\ i, j, k = 1, \dots, p$$

and so on, where $d\alpha'$ indicates integration over all coordinates except α_i, $d\alpha''$ indicates integration over all coordinates except α_i and α_j, and so on.[1]

The ANOVA expansion (5) has several remarkable and useful properties. A partial list includes (see, e.g., [24] for details):

(1) the expansion is exact and finite;

(2) the term $g_T(\vec{\alpha}_T)$ depends only on the coordinates with indices contained in T; g_0 is the average of $g(\cdot)$ and is a constant;

(3) the terms are mutually orthogonal, i.e.,

$$\int_{A^p} g_0 g_T(\vec{\alpha}_T)\, d\vec{\alpha} = 0$$

so that, since g_0 is constant, for all $T \subseteq S$, $g_T(\vec{\alpha}_T)$ has zero average, and

$$\int_{A^p} g_T(\vec{\alpha}_T) g_V(\vec{\alpha}_V)\, d\vec{\alpha} = 0 \quad \text{whenever one or more of the indices in } T \text{ and } V \text{ differ;}$$

note that this includes the cases for which the cardinality of the two index sets are the same;

(4) not only do the individual terms (other than g_0) have zero averages, but

$$\int_0^1 g_T(\vec{\alpha}_T)\, d\alpha_i = 0 \qquad \text{for every } i \in T;$$

(5) each term in the expansion is a projection, with respect to the $L^2(A^p)$ inner product, of $g(\vec{\alpha})$ onto a subspace of $L^2(A^p)$.

Since the nonlinear function $g_T(\vec{\alpha}_T)$ is the unique term in the ANOVA expansion (5) that depends on exactly the $|T|$ variables indexed by T, it provides the effect within $g(\vec{\alpha})$ of the interplay between those $|T|$ variables taken together.

The *order* of a term $g_T(\vec{\alpha}_T)$ appearing in (5) is the cardinality $|T|$ of the corresponding set T. A *truncated ANOVA expansion of order* r is defined by

$$g(\vec{\alpha}; r) = g_0 + \sum_{T \subseteq P,\, |T| \le r} g_T(\vec{\alpha}_T). \tag{7}$$

ANOVA expansions are of great interest because, in many practical settings, $g(\vec{\alpha}; r)$ with $r \ll p$ provides a good approximation to $g(\vec{\alpha})$. A truly remarkable feature of ANOVA expansions is that the degree of approximation of a truncated ANOVA expansion is independent of the measure used to define that expansion, i.e., if $\|g(\vec{\alpha}; r) - g(\vec{\alpha})\| = O(\epsilon)$, then the same is true for ANOVA expansions with respect to all other measures, where $\|\cdot\|$ represents the norm associated with the measure. Of course, the value of the constant in the $O(\cdot)$ relation may differ significantly for different measures.

[1] Instead of the Lebesgue measure on $L^2(A^p)$, an ANOVA expansion may be defined with respect to other measures. For example, if instead the Dirac measure with respect to a fixed point $\vec{\beta} \in A^p$ is used and if $g(\vec{\alpha}) \in C(A^p)$, one obtains the cut-HDMR expansion [24].

Let us examine some implications of being able to approximate well a multivariate function by a short truncated ANOVA expansion. Recall that the term $g_T(\vec{\alpha}_T)$ is only a function of the coordinates having indices in T. Thus, if $r \ll p$, the function $g(\vec{\alpha})$ that depends on p variables can be approximated well by a sum of functions each of which depends on at most r variables. The effect of the interplay between coordinate sets of more than r variables is thus negligible. Such a happenstance has serious implications on how one samples parameter space. For example, consider the simple quadrature rule

$$(8) \qquad \int_{A^p} g(\vec{\alpha})\, d\vec{\alpha} \approx \frac{1}{N} \sum_{j=1}^{N} g(\vec{\alpha}^{(j)}), \qquad \vec{\alpha}^{(j)} \in A^p \text{ for } j = 1, \dots, N,$$

that is in widespread use for high-dimensional integration. One wants to choose the integration points $\vec{\alpha}^{(j)} \in A^p$ so that the approximation is as good as possible. Now, suppose that $g(\vec{\alpha})$ can be approximated very well by a first-order truncated ANOVA expansion, i.e., by a sum of univariate functions, so that

$$g(\vec{\alpha}) = g_0 + \sum_{i=1}^{p} g_i(\alpha_i) + O(\epsilon)$$

with $\epsilon \ll 1$. Then, we have that (8) reduces to

$$(9) \qquad \int_{A^p} g(\vec{\alpha})\, d\vec{\alpha} \approx g_0 + \sum_{i=1}^{p} \Big(\frac{1}{N} \sum_{j=1}^{N} g_i(\alpha_i^{(j)}) \Big)$$

so that the p-dimensional quadrature rule (8) effectively reduces to a sum of p one-dimensional quadrature rules. Thus, the accuracy of the quadrature rule (8) is determined by how accurate one can do the implied one-dimensional quadratures in (9). As an example of how naive choices for the sampling points can have disastrous effects, choose $N = \widehat{N}^p$ for some positive integer $\widehat{N}$ and choose the quadrature points $\vec{\alpha}^{(j)} \in A^p$ to lie on a Cartesian grid in A^p. Then, if $\vec{\alpha}^{(\widehat{j})}$, $\widehat{j} = 1, \dots, \widehat{N}$ denote the points of the grid along a main diagonal on A^p, we have that

$$(10) \qquad \int_{A^p} g(\vec{\alpha})\, d\vec{\alpha} \approx g_0 + \sum_{i=1}^{p} \Big(\frac{1}{\widehat{N}} \sum_{j=1}^{\widehat{N}} g_i(\alpha_i^{(\widehat{j})}) \Big).$$

Thus, although we are using the N-point quadrature rule (8), we are effectively only getting the accuracy of a $\widehat{N} = N^{1/p}$ quadrature rule!

How can one make sure that the N-point quadrature rule (8) gives us what we think we are getting, i.e., the accuracy of an N-point rule, even if the integrand $g(\vec{\alpha})$ happens to be approximated well by a truncated ANOVA expansion of order one? In general, we want the points to have low *discrepancy* (see, e.g., [16]) which the Cartesian points do not. This need motivates the large body of work on the development of intelligent sampling strategies. In fact, high-dimensional integration is the prime example driving that work so that the greatest interest and most of the algorithms and analysis have been directed at sampling lots of points in high-dimensional hypercubes. Our own interest is more along the lines of sparse sampling in possibly high dimensions, a situation for which most of the results for dense sampling are not applicable. Thus, we will examine, though some computational experiments based on a simple model problem, intelligent sparse parameter sampling strategies.

2.2. The effective dimension of a function. Related to ANOVA expansions is the concept of the effective dimension of a multivariate function. Here, we introduce the concept and discuss some properties relevant to our interests; details may be found in [2, 18, 25, 27].

Let T be a subset of P and $\sigma^2(g)$ denote the variance of a function g. Then,

$$\sigma^2(g) = \sum_{|T|>0} \sigma_T^2(g), \qquad \text{where} \qquad \sigma_T^2(g) = \int_{A^p} \big(g_T(\vec{\alpha}_T)\big)^2 d\vec{\alpha}.$$

In the following definitions, the proportion $q > 0$ is chosen to be slightly less than 1; $q = 0.99$ is a common choice.

Definition 1. The effective dimension of g in the superposition sense or, in short, the *superposition dimension*, is the smallest integer p_s such that

$$\sum_{0<|T|\le p_s} \sigma_T^2(g) \ge q\sigma^2(g). \tag{11}$$

Definition 2. The effective dimension of g in the truncation sense or, in short, the *truncation dimension*, is the smallest integer p_t such that

$$\sum_{T\subseteq\{1,\dots,p_t\}} \sigma_T^2(g) \ge q\sigma^2(g). \tag{12}$$

Given a functions g and its putative approximation h, the *normalized approximation error* is defined as

$$E(g,h) = \frac{1}{\sigma^2(g)} \int_{A^p} \big(g(\vec{\alpha}) - h(\vec{\alpha})\big)^2 d\vec{\alpha}.$$

The following theorem (see [27]) is concerned with the approximation property of ANOVA expansions.

Theorem 2.1. *Assume that $g(\vec{\alpha})$ has superposition dimension p_s in proportion q and let $g(\vec{\alpha}; p_s) = \sum_{0<|T|\le p_s} g_T(\vec{\alpha}_T)$ denote its truncated ANOVA expansion of order p_s. Then,*

$$E\big(g(\vec{\alpha}), g(\vec{\alpha}; p_s)\big) \le (1-q). \tag{13}$$

Thus, if the superposition dimension of a function is small, it can be well approximated by short truncated ANOVA expansions. In §4, we will calculate the effective dimensions of some functionals of solutions of our model problem and demonstrate that, in some cases at least, truncated ANOVA expansions of order two provide excellent approximations.

3. The model problem and the approximation property of ANOVA expansions of its solution

In this section, we describe a model nonlinear, partial differential equation problem and study the ANOVA expansion of its solution. In particular, we examine the approximation property of the truncated ANOVA expansion of order one for the solution of a small nonlinear perturbation of the Laplace equation.

Let $\Omega \subset \mathbb{R}^2$ be a bounded open set with boundary $\partial\Omega$; we assume that $\partial\Omega \in C^1$ or is convex Lipschitz. Assume that Γ_0, Γ_1, and Γ_2 are subsets of $\partial\Omega$ such that $\Gamma_0 \cup \Gamma_1 \cup \Gamma_2 = \partial\Omega$ and $\Gamma_i \cap \Gamma_j = \emptyset$. We consider the following boundary value problem for the unknown function u:

$$\begin{cases} -\Delta u + \epsilon f(u) = 0 & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_0 \\ u = \displaystyle\sum_{i=1}^{m} \alpha_i \phi_i & \text{on } \Gamma_1 \\ \dfrac{\partial u}{\partial n} = \displaystyle\sum_{i=m+1}^{p} \alpha_i \phi_i & \text{on } \Gamma_2, \end{cases} \tag{14}$$

where $f(u)$ is a given function of u, ϕ_1 and ϕ_2 are given functions defined along the boundary segments Γ_1 and Γ_2 as appropriate, and α_i for $i = 1, \cdots p$ are paremeters. Obviously, the solution u depends on the parameters α_i.

We focus on optimization problems in which one seeks optimal states u and optimal parameter values $\{\alpha_i\}_{i=1}^p$ such that a cost functional such as

$$\mathcal{J}(\vec{\alpha}) = \int_\Omega w(u)\, d\Omega \tag{15}$$

is minimized, where $w(u)$ is a given function of u. In this setting, u is the state variable, the α_i's are the controls or design variables, and (14) is the constraint. We assume throughout that the α_i's are uncorrelated. Clearly, since for any set of control parameters $\{\alpha_i\}_{i=1}^p$, we may solve (14) for u, we may consider $\mathcal{J}(\vec{\alpha})$ to be a function of those parameters.

We assume that the parameters $\vec{\alpha} = \{\alpha_i\}_{i=1}^p$ are constrained to lie in a box which, without loss of generality, we take to be the unit hypercube A^p in $\mathbb{R}^p$. We will build a surrogate for $\mathcal{J}(\cdot)$ as described in §1: we sample points within the parameter hypercube; we solve the model problem (14) using the sampled parameters as inputs; we use the solutions of (14) to evaluate the functional $\mathcal{J}(\cdot)$; using that data, we build a response surface and then use the minimizer of that surrogate as the approximation of the minimizer of $\mathcal{J}(\cdot)$.

To complete the above surrogate optimization recipe, one must choose what kind of response surface to use, e.g., linear or quadratic polynomials or some other approximations and one must choose a method to build that surface, e.g., interpolation or least-squares approximation or something else. Even before either of these steps are effected, one must choose the points in parameter space that are used to evaluate the functional; this is the focus of the remainder of the paper. We shall see that the approximation property of the ANOVA expansion plays a key role in evaluating different methods for sampling points in the parameter hypercube. Here, however, we first examine the approximation property of the ANOVA expansion in terms of the perturbation parameter ϵ appearing in (14).

For simplicity, we assume that Γ_1 is empty so that (14) reduces to

$$-\Delta u + \epsilon f(u) = 0 \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \Gamma_0, \qquad \text{and} \qquad \frac{\partial u}{\partial n} = \sum_{i=1}^{p} \alpha_i \phi_i \quad \text{on } \Gamma_2. \tag{16}$$

For small $\epsilon > 0$, we view this problem as a perturbation of the linear problem

$$-\Delta v = 0 \quad \text{in } \Omega, \qquad v = 0 \quad \text{on } \Gamma_0, \qquad \text{and} \qquad \frac{\partial v}{\partial n} = \sum_{i=1}^{p} \alpha_i \phi_i \quad \text{on } \Gamma_2. \tag{17}$$

Note that (17) is linear with respect to both the unknown v and the parameters $\{\alpha_i\}_{i=1}^p$ so that we can use the principle of superposition to conclude that the solution is a linear combination of the α_i's. This implies that the truncated ANOVA expansion of order one for v coincides with the solution v itself.

In the following theorem, we assume that the solution of these two problems are "close" and then conclude that a truncated ANOVA expansion is a good approximation of the solution of the nonlinear problem (16). Note that $H^1(\Omega)$ denotes the usual Sobolev space of square integrable functions having one square integrable derivative.

Theorem 3.1. *Assume that the solutions u and v of* (16) *and* (17)*, respectively, satisfy*

$$\|u-v\|_{H^1(\Omega)} \le C\epsilon \tag{18}$$

for some positive constant C whose value is independent of α_j, $j = 1,\dots,p$. Let

$$u(\vec{\alpha};1) = u_0 + \sum_{j=1}^{p} u_j(\alpha_j) \tag{19}$$

denote the truncated ANOVA expansion of u of order 1*. Then, there exists a positive constant M whose value is also independent of α_j, $j = 1,\dots,p$, such that*

$$\|u-u(\vec{\alpha};1)\|_{H^1(\Omega)} \le \epsilon M. \tag{20}$$

Proof: Let $v(\vec{\alpha};1) = v_0 + \sum_{j=1}^{p} v_j(\vec{\alpha}_j)$ be the ANOVA expansion of v of order 1. Note that (17) is linear with respect to both the unknown v and the parameters $\{\alpha_i\}_{i=1}^{p}$ so that we can use the principle of superposition to conclude that the solution is a linear combination of the α_i's. This implies that the truncated ANOVA expansion of order one for v coincides with the solution v itself, i.e., $v = v(\vec{\alpha};1)$. Then,

$$\|u-u(\vec{\alpha};1)\|_{H^1(\Omega)} \le \|u-v\|_{H^1(\Omega)} + \|u(\vec{\alpha};1)-v(\vec{\alpha};1)\|_{H^1(\Omega)}.$$

By (18) and the definitions of u_0 and v_0, we have that

$$\|u_0-v_0\|_{H^1(\Omega)} = \left\| \int_{A^p} (u-v)\, d\vec{\alpha} \right\|_{H^1(\Omega)} \le \int_{A^p} \|u-v\|_{H^1(\Omega)}\, d\vec{\alpha} \le \epsilon C.$$

Similary, we can prove that there exist constants $C_1,\dots,C_p$ having values independent of $\{\alpha_j\}_{j=1}^{p}$ such that

$$\|u_j - v_j\|_{H^1(\Omega)} \le \epsilon C_j, \quad j = 1,\dots,p.$$

Therefore,

$$\|u-u(\vec{\alpha};1)\|_{H^1(\Omega)} \le \epsilon C + \sum_{j=1}^{p} \epsilon C_j = \epsilon M,$$

where $M = C + \sum_{j=1}^{p} C_j$. □

Define $H^1_{\Gamma_0}(\Omega) = \{w \in H^1(\Omega) \,|\, w = 0 \text{ on } \Gamma_0\}$; we use the semi-norm $\|\nabla(\cdot)\|_{L^2(\Omega)}$ as the norm on $H^1_{\Gamma_0}(\Omega)$. We also define $H^{-1}_{\Gamma_0}(\Omega)$ as the dual space of $H^1_{\Gamma_0}(\Omega)$.

The following proposition provides a simple example of a problem for which (18) holds.

Proposition 3.2. *In* (16) *and* (17)*, let $\phi_j \in H^{-1/2}(\Gamma_2), j = 1,\dots,p$, and $\|f(u)\|_{H^{-1}_{\Gamma_0}(\Omega)} \le C_1\|u\|_{H^1(\Omega)}$ for $u \in H^1_{\Gamma_0}(\Omega)$. Then, there exists positive constant C whose values are independent of u, v, and α_j, $j = 1,\dots,p$, such that, for $\epsilon < \epsilon_1$,* (18) *holds.*

Proof: By Green's formula, we have that

$$\begin{aligned}\|u\|^2_{H^1(\Omega)} = (\nabla u, \nabla u) &= -\epsilon\big(f(u), u\big) + \int_{\Gamma_2} \frac{\partial u}{\partial n} u \, d\Gamma \\ &\leq \epsilon C_1 \|u\|^2_{H^1(\Omega)} + \Big\|\frac{\partial u}{\partial n}\Big\|_{H^{-1/2}(\Gamma_2)} \|u\|_{H^1(\Omega)} \\ &\leq \epsilon C_1 \|u\|^2_{H^1(\Omega)} + \frac{p}{4\epsilon_0} \sum_{j=1}^{p} \|\phi_j\|^2_{H^{-1/2}(\Gamma_2)} + \epsilon_0 \|u\|^2_{H^1(\Omega)}.\end{aligned}$$

Choose ϵ_0 and ϵ_1 such that $\epsilon_1 C_1 + \epsilon_0 < 1$. Then, for $\epsilon < \epsilon_1$, we have that

$$\|u\|_{H^1(\Omega)} \leq \sqrt{\frac{p}{4(1-\epsilon_1 C - \epsilon_0)\epsilon_0} \sum_{j=1}^{p} \|\phi_j\|^2_{H^{-1/2}(\Gamma_2)}} := C_2. \tag{21}$$

Now, let $z = u - v$. We then have that

$$-\Delta z + \epsilon f(u) = 0 \quad \text{in } \Omega, \qquad z|_{\Gamma_0} = 0, \qquad \text{and} \qquad \frac{\partial z}{\partial n}\Big|_{\Gamma_2} = 0.$$

Thus,

$$\|z\|^2_{H^1(\Omega)} = -\epsilon\big(f(u), z\big) \leq \epsilon C_1 \|u\|_{H^1(\Omega)} \|z\|_{H^1(\Omega)}.$$

Then, (21) implies that

$$\|u - v\|_{H^1(\Omega)} = \|z\|_{H^1(\Omega)} \leq \epsilon C_1 \|u\|_{H^1(\Omega)} \leq \epsilon C_1 C_2 := \epsilon C. \qquad \square$$

Consider now the problem of minimizing the functional $\mathcal{J}(\vec{\alpha})$ given in (15) subject to (14) being satisfied. The following result about the approximation properties of optimization using the first-order ANOVA expansion.

Theorem 3.3. *Assume that ϕ_j, $j = 1, \ldots, p$, and f in* (16) *satisfy the conditions of Proposition* 3.2. *Let*

$$\mathcal{J}_1(\vec{\alpha}) = \int_\Omega w(u(\vec{\alpha}, 1)) \, d\Omega, \tag{22}$$

where w is the same as in (15) *and $u(\vec{\alpha}, 1)$ is the first-order ANOVA expansion of u* (*see* (19)). *If*

$$|w(x) - w(y)| \leq Ck(|x| + |y|)|x - y|, \tag{23}$$

where k is a monotone, non-negative function, then, there exists a constant C and $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$,

$$\Big|\min_{\alpha \in [0,1]^s} \mathcal{J}(\vec{\alpha}) - \min_{\alpha \in [0,1]^s} \mathcal{J}_1(\vec{\alpha})\Big| \leq C\epsilon. \tag{24}$$

Proof. From (15), (22), and (23), we have that

$$\begin{aligned}|\mathcal{J}(\vec{\alpha}) - \mathcal{J}_1(\vec{\alpha})| &\leq C \int_\Omega k\left(|u(\vec{\alpha})| + |u(\vec{\alpha}, 1)|\right) |u(\vec{\alpha})| - |u(\vec{\alpha}, 1)| \, d\Omega \\ &\leq C \|k(u(\vec{\alpha}) + u(\vec{\alpha}, 1))\|_{L^2(\Omega)} \|u(\vec{\alpha}) - u(\vec{\alpha}, 1)\|_{L^2(\Omega)}.\end{aligned}$$

From the proof of Proposition 3.2 we know that there exists a constant C such that $\|u\|_{L^2(\Omega)} + \|u(\vec{\alpha}, 1)\|_{L^2(\Omega)} \leq C$. Using Proposition 3.2, we obtain the desired result. $\square$

From (20) and (24), we see that the the error in the minimizer (i.e., the difference between the minimum of the exact cost functional and the minimum of the cost functional with u replaced by its first-order ANOVA expansion) and the error in the solution (i.e., the difference between u itself and its first-order ANOVA approximation) are of similar magnitude in terms of ϵ. From this observation and (13), it is reasonable to assume that when ANOVA effects of order higher than s are negligible, one can use polynomials of degree s to serve as surrogate cost functionals.

In §1, the building of surrogate functionals through interpolation was alluded to; see §5 for more details. Interpolatory error estimates may be combined with (24) to obtain estimates for the difference between surrogates and truncated ANOVA expansions of functionals. Examining these two contributions to the errors may shed some light on the choice of interpolation points, i.e., on different sampling strategies (see, e.g., §5) for selecting these points. However, as was noted in §1, for surrogate optimization problems, one is likely to sample very sparsely, i.e., use very few interpolation points, so that interpolatory error estimates may be of limited value when the parameter dimension is large.

4. ANOVA expansions and the effective dimension of cost functionals

The specific constraint system we consider for our computational examples is given by (14) with $\epsilon = 1$, Ω being the unit square, $m = 2$, $p = 4$, and

$$\begin{aligned}
&\phi_1(x,y) = \begin{cases} -16x(x-0.5) & \text{if } 0 \le x \le 0.5 \text{ and } y = 1, \\ 0 & \text{otherwise,} \end{cases} \\
&\phi_2(x,y) = \begin{cases} -16(x-0.5)(x-1.0) & \text{if } 0.5 \le x \le 1.0 \text{ and } y = 1, \\ 0 & \text{otherwise,} \end{cases} \\
&\phi_3(x,y) = \begin{cases} -16x(x-0.5) & \text{if } 0 \le x \le 0.5 \text{ and } y = 0, \\ 0 & \text{otherwise,} \end{cases} \\
&\phi_4(x,y) = \begin{cases} -16(x-0.5)(x-1.0) & \text{if } 0.5 \le x \le 1.0 \text{ and } y = 0, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{25}$$

so that the boundary segments Γ_i, $i = 0, 1, 2$, and the data in the boundary conditions in (14) have support as indicated in the sketch of Figure 1.

For the nonlinear term $f(u)$ in the partial differential equation we consider the three choices

$$f(u) = u^2 \qquad \text{or} \qquad e^u \qquad \text{or} \qquad \sqrt{|u|}. \tag{26}$$

For the integrand $w(u)$ of the functional $\mathcal{J}(u)$ we consider the three choices

$$w(u) = (u - \widehat{u})^2 \qquad \text{or} \qquad e^{(u-\widehat{u})} \qquad \text{or} \qquad \sqrt{|u - \widehat{u}|},$$

where we choose $\widehat{u}(x,y) = u(\{\alpha_i = 0.5\}_{i=1}^4)$, i.e., $\widehat{u}(x,y)$ is the solution of the problem (14) corresponding to the parameter values $\alpha_i = 0.5$, $i = 1, \ldots, 4$. Note that since $\epsilon = 1$ there is no a priori reason to believe that solutions of (14) or functionals that depend on those solutions can be accurately approximated by short truncated ANOVA expansions.

For a given set of parameters $\{\alpha_i\}_{i=1}^4$, the system (14) is discretized using standard, continuous, piecewise quadratic finite element functions defined with respect to a uniform triangulation of the spatial domain into 648 triangles. Approximate

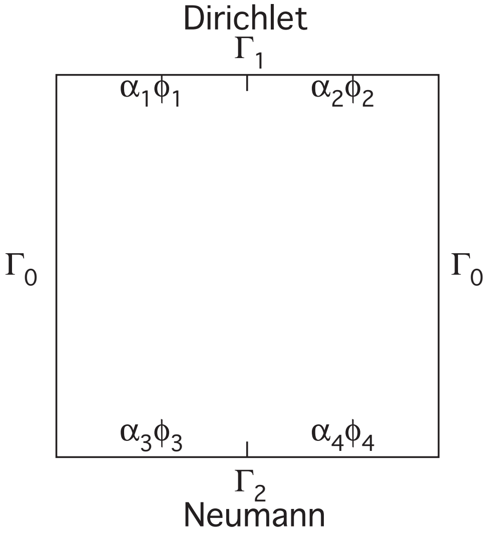

FIGURE 1. *The data configuration for the model problem used in the computational examples.*

solutions of (14) obtained in this manner are accurate enough so that the discretization error is not a factor in our consideration of ANOVA expansions. The resulting nonlinear discrete system is solved by Newton's method.

Two types of integrals enter into our determination of the ANOVA expansion of the cost functional $\mathcal{J}(u)$. We have to evaluate integrals such as

$$\int_{A^{4-j}} \mathcal{J}\big(u(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\big)\, d\vec{\alpha}_{4-j} \qquad \text{for } j=0,\ldots,3,$$

where $d\vec{\alpha}_{4-j}$ is the appropriate $(4-j)$-dimensional measure and here we view u as a function of the parameters. We approximate this type of integral by tensor products (in parameter space) of the classimathcal four-point Gauss rule. To apply that rule, we have to evaluate $\mathcal{J}(\cdot)$ at each of the Gauss quadrature points; this itself requires the evaluation of the spatial integral

$$\int_\Omega w(u(x,y))\, d\Omega,$$

where we now view u as the approximate solution of (14) with the input parameter vector $\vec{\alpha}$ chosen to correspond to a quadrature point in parameter space. This spatial integral is approximated using a simple nodal quadrature rule over the finite element grid.

In Table 1, the $L^2(A^p) = L^2(A^T)$ norms $\sigma_T^2\big(\mathcal{J}(\vec{\alpha})\big)$ of the individual terms $\mathcal{J}_T(\vec{\alpha}_T)$, $T \subseteq P = \{1,2,3,4\}$, in the ANOVA expansion of $\mathcal{J}(\vec{\alpha})$ are given for the various cases introduced above. The $L^2(A^p)$ norm of the constant term $\mathcal{J}_0$ is also provided. From that table, one sees the general trend that the terms in the ANOVA expansion become smaller as their order increases. In some cases, this trend is very dramatic. For example, for $w(u) = (|u-\widehat{u}|)^2$ and $w(u) = e^{(u-\widehat{u})}$, the norms of the third and fourth-order terms are less than 1% of that of the zeroth-order term. For $w(u) = \sqrt{|u-\widehat{u}|}$, the decay of the ANOVA terms is much less pronounced.

Table 2 provides information about the effective superposition and truncation dimensions of $\mathcal{J}(\vec{\alpha})$. For the superposition dimension, the values of q_{p_s}, $p_s = 1,2,3,$

	$w(u) = (u - \widehat{u})^2$			$w(u) = e^{u-\widehat{u}}$			$w(u) = \sqrt{\lvert u - \widehat{u}\rvert}$		
$f(u)$	u^2	e^u	$\sqrt{\lvert u\rvert}$	u^2	e^u	$\sqrt{\lvert u\rvert}$	u^2	e^u	$\sqrt{\lvert u\rvert}$
$\mathcal{J}_0$	3.8E-3	6.2E-3	6.4E-3	1.1E-1	1.1E-0	1.1E-0	1.7E-1	1.8E-1	1.9E-1
$\mathcal{J}_1$	1.5E-3	2.6E-3	2.7E-3	2.4E-2	2.4E-2	2.9E-2	1.5E-2	1.6E-2	1.7E-2
$\mathcal{J}_2$	1.5E-3	2.6E-3	2.7E-3	2.4E-2	2.4E-2	2.9E-2	1.5E-2	1.6E-2	1.7E-2
$\mathcal{J}_3$	1.7E-4	1.5E-4	1.4E-4	8.4E-3	3.2E-3	7.4E-3	5.7E-3	5.5E-3	5.5E-3
$\mathcal{J}_4$	1.7E-4	1.6E-4	1.4E-4	8.4E-3	3.3E-3	7.5E-3	5.8E-3	5.7E-3	5.6E-3
$\mathcal{J}_{12}$	1.1E-3	8.6E-4	9.9E-4	5.0E-4	3.9E-3	6.9E-4	2.2E-2	1.9E-2	2.0E-2
$\mathcal{J}_{13}$	2.3E-4	1.4E-4	1.7E-4	9.8E-5	2.1E-3	1.7E-4	7.7E-3	6.2E-3	6.8E-3
$\mathcal{J}_{14}$	2.2E-4	1.3E-4	1.6E-4	9.2E-5	1.9E-3	1.5E-4	6.6E-3	5.3E-3	5.7E-3
$\mathcal{J}_{23}$	2.2E-4	1.3E-4	1.6E-4	9.2E-5	1.9E-3	1.5E-4	6.6E-3	5.2E-3	5.6E-3
$\mathcal{J}_{24}$	2.3E-4	1.4E-4	1.7E-4	9.8E-5	2.1E-3	1.7E-4	7.7E-3	6.2E-3	6.8E-3
$\mathcal{J}_{34}$	2.4E-4	1.4E-4	1.9E-4	1.0E-4	2.5E-3	1.9E-4	9.3E-3	7.3E-3	8.6E-3
$\mathcal{J}_{123}$	2.4E-6	2.7E-5	8.9E-6	1.0E-6	4.2E-4	1.6E-5	2.8E-3	2.1E-3	2.4E-3
$\mathcal{J}_{124}$	2.4E-6	2.7E-5	8.9E-6	1.0E-6	4.2E-4	1.6E-5	2.8E-3	2.1E-3	2.4E-3
$\mathcal{J}_{134}$	1.6E-6	1.4E-5	6.6E-6	6.9E-7	4.1E-4	1.4E-5	2.3E-3	1.5E-3	1.9E-3
$\mathcal{J}_{234}$	1.6E-6	1.4E-5	6.6E-6	6.9E-7	4.1E-4	1.4E-5	2.3E-3	1.5E-3	1.9E-3
$\mathcal{J}_{1234}$	2.4E-8	6.7E-6	2.1E-6	1.9E-9	4.9E-4	6.4E-6	4.2E-3	3.2E-3	3.6E-3

TABLE 1. *The $L^2(A^p)$ norms of the terms in the ANOVA expansion of the functional $\mathcal{J}(\vec{\alpha})$.*

given in the table satistfy

$$\sum_{0<|T|\le p_s} \sigma_T^2(\mathcal{J}) = q_{p_s}\sigma^2(\mathcal{J}) \qquad \text{for } p_s = 1, 2, 3.$$

Let the superposition dimension p_s be defined by (11) with $q = 0.99$. Then, we see from Table 2, that for $\mathcal{J}(\vec{\alpha})$, p_s is given by

$$p_s = \begin{cases} 1 & \text{if } w(u) = e^{(u-\widehat{u})} \quad \text{and} \quad f(u) = u^2 \ \text{ or } \ \sqrt{|u|} \\ 3 & \text{if } w(u) = \sqrt{u - \widehat{u}} \quad \text{and} \quad f(u) = e^u \ \text{ or } \ \sqrt{|u|} \\ 4 & \text{if } w(u) = \sqrt{u - \widehat{u}} \quad \text{and} \quad f(u) = u^2 \\ 2 & \text{otherwise.} \end{cases} \tag{27}$$

We then conclude from (13) that if $\mathcal{J}(\vec{\alpha}; p_s)$ denotes the truncated ANOVA expansion of $\mathcal{J}(\vec{\alpha})$ of order p_s, with p_s given by (27), then

$$E\big(\mathcal{J}(\vec{\alpha}), \mathcal{J}(\vec{\alpha}; p_s)\big) \le 0.01.$$

Thus, in some cases, short truncations of the ANOVA expansion provide very good approximations of the functional $\mathcal{J}(\vec{\alpha})$. Remathcall that a truncated ANOVA expansion of order p_s is a sum of functions of at most p_s variables. Thus, for the cases for which $p_s = 1$ or $p_s = 2$, we conclude that the functional $\mathcal{J}(\vec{\alpha})$ can be approximated well by a sum of at most univariate and bivariate functionals, respectively. Note, however, that in one case, even the highest-order term contributes significantly to the ANOVA expansion.

	$w(u) = (u - \widehat{u})^2$			$w(u) = e^{u-\widehat{u}}$			$w(u) = \sqrt{\lvert u - \widehat{u}\rvert}$		
$f(u)$	u^2	e^u	$\sqrt{\lvert u\rvert}$	u^2	e^u	$\sqrt{\lvert u\rvert}$	u^2	e^u	$\sqrt{\lvert u\rvert}$
Superposition dimension									
q_1	0.757	0.943	0.928	1.000	0.968	1.000	0.388	0.523	0.489
q_2	1.000	1.000	1.000	1.000	0.999	1.000	0.967	0.978	0.976
q_3	1.000	1.000	1.000	1.000	1.000	1.000	0.986	0.991	0.990
Truncation dimension									
q_{12}	0.950	0.990	0.988	0.893	0.963	0.937	0.694	0.752	0.755
q_{13}	0.387	0.473	0.466	0.499	0.488	0.500	0.238	0.295	0.279
q_{14}	0.386	0.473	0.466	0.500	0.487	0.500	0.228	0.287	0.270
q_{23}	0.386	0.473	0.466	0.499	0.487	0.500	0.226	0.286	0.268
q_{24}	0.387	0.473	0.466	0.500	0.488	0.500	0.240	0.297	0.280
q_{34}	0.018	0.005	0.005	0.107	0.023	0.063	0.117	0.107	0.103
q_{123}	0.970	0.994	0.993	0.946	0.979	0.968	0.802	0.844	0.841
q_{124}	0.970	0.994	0.993	0.947	0.979	0.969	0.804	0.846	0.843
q_{134}	0.408	0.477	0.471	0.553	0.505	0.531	0.367	0.401	0.386
q_{234}	0.408	0.477	0.471	0.553	0.505	0.531	0.367	0.401	0.386

TABLE 2. *Proportionality constants for the effective dimensions of the functional $\mathcal{J}(\vec{\alpha})$.*

In Table 2, for the truncation dimension, the values $q_{1,\ldots,p_t}$ for $p_t > 1$ are such that

$$\sum_{T \subseteq \{1,\ldots,p_t\}} \sigma_T^2(\mathcal{J}) \geq q_{1,\ldots,p_t} \sigma^2(\mathcal{J}).$$

In the table, we have also included entries for ordered index sets that begin with integers greater than one. The reason for doing so is that the ordering of any variables as $\alpha_1, \alpha_2, \ldots, \alpha_p$ is usually arbitrary; certainly this is the case for the functional $\mathcal{J}(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ defined through (14) and (15). The entries included in Table 2 account for all possible orderings of the variables. The results in Table 2 relative to the three choices of $w(u)$ are consistent with the results in Table 1.

The proportions for the truncation dimension demonstrate that all variables have significant effects, but some more so than others. In fact, an unexpected trend is apparent from the truncation dimension proportions given in Table 2: the entries whose index set includes the integers 1 and 2 are considerably larger than those that do not contain those integers and the entries containing both the integers 1 and 2 are considerably larger than those that contain only one of those integers. From this we conclude that the variables α_1 and α_2 are of about equal importance to the ANOVA expansion of the functional $\mathcal{J}(\vec{\alpha})$ and that both are of greater importance compared to the variables α_3 and α_4. This type of information is useful in determining which parameters are more influential to a functional and which may perhaps be ignored. In the particular setting we are considering here, although none of the variables can be ignored, we see that the parameters α_1 and α_2 corresponding to the Dirichlet boundary conditions (see (14) and (25)) are more "important" than the parameters

α_3 and α_4 corresponding to the Neumann boundary conditions. We do not have a good explanation for this observation, but it seems that this issue deserves further study.

5. Sampling parameter space in the building of surrogate functionals

Consider the problem of minimizing the functional $\mathcal{J}(\vec{\alpha})$ given in (15) subject to (14) being satisfied. In the context of the concrete problem of §4, we find an approximate solution of the optimization problem by building a surrogate functional as described in §1.

The first step towards building a surrogate is to sample N points in A^p with $p = 4$ for the problem of §4. From the calculations of the last section, we see that the effective dimension is no more than 3 in most cases. Thus, we will choose quadratic functions as our surrogate functions. We note again that in the context of surrogate optimization problems, we are interested in the (very) sparse sampling of the hypercube so that the large body of literature related to sampling many points in hypercubes does not apply to our study.

The specific procedure we use to build the surrogate functional is as follows:

(1) choose 15 points in the parameter hypercube A^4;
(2) solve, using a finite element method, the nonlinear partial differential equation problem (14) for each of the parameter points chosen in step 1;
(3) use the solutions obtained in step 2 to evaluate the functional (15) at each of the parameter points obtained in step 1;
(4) determine the quadratic polynomial in parameter space that interpolates the functional values obtained in step 3 at the points obtained in step 1;
(5) determine the minimum value, within A^4, of the quadratic polynomial constructed in step 4.

For step 1, we will use several sampling methods: random or Monte Carlo sampling (MC), Latin hypercube sampling (LHS), Halton sampling (HAL), Hammersley sampling (HAM), and centroidal Voronoi tessellation sampling (CVT). Brief descriptions of these methods are given in the Appendix.

For our computational results, steps 2 and 3 are carried out in the same way as described in §4. We also use the concrete problem considered in that section, where we use the three choices given in (26) for the nonlinear term $f(u)$ in the partial differential equation and the two choices $w(u) = (u - \widehat{u})^2$ and $w(u) = \sqrt{|u - \widehat{u}|}$ for the integrand in the functional (15). For the target function $\widehat{u}$, we choose the solution of (14) for a specific choice $\vec{\alpha}^*$ of the parmeters; then, for the choices we make for $w(u)$, we know the functional $\mathcal{J}$ attains a minimum value 0 at $\vec{\alpha} = \vec{\alpha}^*$. In order to at least partially remove any statistical bias a single example might have, we use ten randomly chosen values of $\vec{\alpha}^*$ to determine the average and worst performance of each of the parameter sampling strategies.

For each sampling method, we respectively provide, in Tables 3 and 4, the absolute errors (averaged over ten realizations) in the location of the minimizer of the functional and in the value of the functional at its minimizing point. Specifically, in Table 3, we provide the average (over ten random choices for $\vec{\alpha}^*$) of $|\vec{\alpha}^*_{sur} - \vec{\alpha}^*|$, where $\vec{\alpha}^*$ denotes the minimizer of the functional $\mathcal{J}$ and $\vec{\alpha}^*_{sur}$ denotes the minimizer of the surrogate functional $\mathcal{J}_{sur}$. In Table 4, we provide the average (over ten random choices for $\vec{\alpha}^*$) of $|\mathcal{J}_{sur}(\vec{\alpha}^*_{sur}) - \mathcal{J}(\vec{\alpha}^*)| = |\mathcal{J}_{sur}(\vec{\alpha}^*_{sur})|$, where the last equality holds since, by construction, $\mathcal{J}(\vec{\alpha}^*) = 0$.

From Tables 1–4, we see a correlation between the accuracy of truncated ANOVA expansions of the functional and the accuracy of results obtained using surrogate

sampling method	$w(u) = (u - \widehat{u})^2$			$w(u) = \|u - \widehat{u}\|^{1/2}$		
	$f(u)$			$f(u)$		
	u^2	e^u	$\|u\|^{1/2}$	u^2	e^u	$\|u\|^{1/2}$
MC	0.036	0.018	0.086	0.556	0.566	0.574
LHS	0.021	0.012	0.035	0.378	0.427	0.380
HAL	0.032	0.022	0.081	0.601	0.634	0.670
HAM	0.064	0.040	0.172	0.447	0.447	0.441
CVT	0.067	0.026	0.069	0.436	0.433	0.436

Table 3. *Absolute error, averaged over 10 realizations, in the location of the minimizing parameter point for each of the point sampling methods used in the surrogate construction.*

sampling method	$w(u) = (u - \widehat{u})^2$			$w(u) = \|u - \widehat{u}\|^{1/2}$		
	$f(u)$			$f(u)$		
	u^2	e^u	$\|u\|^{1/2}$	u^2	e^u	$\|u\|^{1/2}$
MC	1.85E-5	7.87E-6	4.38E-5	0.052	0.051	0.052
LHS	9.75E-6	6.07E-6	2.12E-5	0.053	0.083	0.054
HAL	4.53E-5	2.35E-5	1.49E-4	0.457	0.448	0.447
HAM	6.36E-5	3.62E-5	1.84E-4	0.228	0.208	0.204
CVT	9.08E-5	3.39E-5	3.11E-4	0.289	0.256	0.234

Table 4. *Absolute error, averaged over 10 realizations, in the minimum value of the functionals for each of the point sampling methods used in the surrogate construction.*

optimization. Specifically, for the case $w(u) = |u - \widehat{u}|^{1/2}$, we see from Table 1 that short ANOVA expansions are inaccurate and from Tables 3 and 4 that minimizing points and minimum values of the surrogate functional are also inaccurate, compared to the results obtained for the integrand $w(u) = (u - \widehat{u})^2$. In fact, for the integrand $w(u) = |u - \widehat{u}|^{1/2}$, the minimizing points of the surrogate functionals are essentially useless as approximations of the minimizing points of the given functional. We also see from Tables 3 and 4 that, for $w(u) = (u - \widehat{u})^2$, results are generally worse for the nonlinearity $f(u) = |u|^{1/2}$ than for the other two nonlinearities; this is again consistent with the relative accuracy of truncated ANOVA expansions as indicated in Tables 1 and 2.

The results provided in Tables 1–4 are interesting for at least two reasons. First, one sees that parameter sampling strategies that are known to be better for some applications, e.g., multidimensional integration, that involve sampling a large number of points are not necessarily better in our setting which requires sparse sampling. For example, Halton and Hammersley sampling were developed as improvements over Monte Carlo sampling for multidimensional integration applications, but Monte Carlo sampling seems to be more effective in our specific example. Second, given that we cannot use results about sampling strategies that hold for a large number of sample points, we can use ANOVA expansions as a tool to provide guidance for choosing the form of the surrogate functional. For example, when short truncated ANOVA expansions provide good approximations to the functional, one sees that quadratic surrogate functionals based on sampling 15 points provide good approximations. On the other hand, for cases where the terms in the ANOVA expansion do not decay, poor approximations are obtained. It is likely that both the number of sampling points must be increased and the form of the surrogate functional must be changed in order to obtain better approximations in the latter case.

Of course, no definitive conclusion can be drawn from just the examples considered here. Further experiments on the various sampling strategies should be carried out as well as the testing of some of the many other sampling strategies available in the literature.

6. Concluding Remarks

We have studied the application of ANOVA expansions to optimization problems constrained by nonlinear elliptic partial differential equations. We demonstrate that when the the ANOVA expansion of the solution of the optimization problem has low effective dimension, we can use lower-order polynomials to build surrogate functionals for the optimization problem.

Our goal in this paper was to acquaint the reader of the possibilities offered by two notions that are well known in the statistics and multidimensional integration communities, namely ANOVA expansions of functions and the effective dimension of functions, for problems governed by nonlinear partial differential equations. We hope this paper encourages further work in this direction. Certainly, other sampling strategies can be tested and the dependence on the results on the number of sample points and on the grid size should be explored. Further mathematical analyses are also called for. From the computational standpoint, we note that, in §anovaex, all the numerical parameters, e.g., meshsize and number of quadrature points, are fixed. It would be interesting to vary them in order to assess the effects on the conclusions drawn.

Acknowledgment

The authors thank John Burkardt for allowing the use of his programs for determining Halton, Hammersley, and CVT sample points.

References

[1] M. Barrault, Y. Maday, N. Nguyen, and A. Patera, An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations, *C. R. Acad. Sci. Paris, Ser, I* **339**, 2004, 667-672.

[2] R. Caflisch, W. Morokoff, A. Owen, Valuation of Mortgage backed securities using Brownian bridges to reduce effective dimension, *J. Comp. Finan.* **1**, 1997, 27-46.

[3] Q. Du, V. Faber, and M. Gunzburger, Centroidal Voronoi tessellations: applications and algorithms, *SIAM Rev.* **41**, 1999, 637-676.

[4] M. Grepl and A. Patera, A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations, *ESAIM:M2AN* **39**, 2005, 157-181.

[5] C. Gu, *Smoothing Spline ANOVA Models*, Springer, Berlin, 2002.

[6] J. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, *Numer. Math.* **2** 1960, 84-90.

[7] J. Hammersley, Monte Carlo methods for solving multivariable problems, *Ann. New York Acad. Sci.* **86**, 1960, 844-874.

[8] F. Hickernell, Lattice rules: How well do they measure up?, in *Random and Quasi-Random Point Sets*, P. Hellekalek and G. Larcher, eds., Springer, Berlin, 1998, 109-166.

[9] G. Hooker, Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables, *J. Comput. Graph. Stat.* **16**, 2007, 709-732.

[10] L. Ju, Q. Du, and M. Gunzburger, Probablistic methods for centroidal Voronoi tessellations and their parallel implementations, *J. Parallel Comput.* **28**, 2002, 1477-1500.

[11] J. Kleijnen, *Design and Analysis of Simulation Experiments*, *International Series in Operations Research & Management Science* , Vol. **111**, 2008.

[12] K. Kunisch and S. Volkwein, Control of Burger's equation by a reduced order approach using proper orthogonal decomposition, *JOTA* **102**, 1999, 345-371.

[13] K. Kunisch and S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems, *Spezialforschungsbereich F003 Optimierung und Kontrolle,* Projektbereich Kontinuierliche Optimierung und Kontrolle, Bericht Nr. 153, Graz, 1999.

[14] O. Le Maitre, O. Knio, H. Najm, and R. Ghanem, A stochastic projection method for fluid flow. I. Basic formulation, *J. Comput. Phys.* **173**, 2001, 481-511.

[15] D. Nagy, Modal representation of geometrically nonlinear behavior by the finite element method, *Comput. Struct.* **10**, 1979, 693.

[16] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

[17] A. Noor, Recent advances in reduction methods for nonlinear problems, *Comput. Struc.* **13**, 1981, 31-44.

[18] A. Owen, The dimension distribution and quadrature test functions, *Stat. Sinica* **13**, 2003, 1-18.

[19] H. Park and W. Lee, An efficient method of solving the Navier-Stokes equations for flow control, *Inter. J. Numer. Meth. Engrg.* **41**, 1998, 1133-1151.

[20] J. Peterson, The reduced basis method for incompressible viscous flow calculations, *SIAM J. Sci. Stat. Comp.* **10**, 1989, 777-786.

[21] S. Ravindran, Proper orthogonal decomposition in optimal control of fluids, *Int. J. Numer. Meth. Fluids* **34**, 2000, 425-448.

[22] J. Rodríguez and L. Sirovich, Low-dimensional dynamics for the complex Ginzburg-Landau equations, *Physica D* **43**, 1990, 77-86.

[23] Y. Saka, M. Gunzburger, and J. Burkardt, Latinized, improved LHS, and CVT point sets in hypercubes, to appear.

[24] A. Saltelli, K. Chan, and E. Scott, *Sensitivity Analysis*, Wiley, Chichester, 2000.

[25] I. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comp. Simul.* **55** 2001, 271-280.

[26] S. Volkwein, Optimal control of a phase field model using the proper orthogonal decomposition, *ZAMM* **81**, 2001, 83-97.

[27] X. Wang and K. Fang, The effective dimension and quasi-Monte Carlo integration, *J. Complex.* **19** 2003, 101-124.

[28] D. Xiu and G. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, *Comp. Meth. Appl. Mech. Engrg.* **191**, 2002, 4927-4948.

[29] D. Xiu and G. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, *J. Comp. Phys.* **187**, 2003, 137-167.

Appendix A. Methods for uniform sampling in hypercubes

We describe the procedures used in §5 for sampling the unit hypercube in parameter space in order to obtain the values of the cost functional used in building surrogate functionals. In addition to the methods described below, we also used uniform, random sampling, i.e., Monte Carlo sampling, within the hypercube.
Latin hybercube sampling. To obtain a Latin Hybercube sample of N points in the hypercube A^p, one first subdivides the hypercube into N^p bins of equal size, then randomly places one point within N randomly chosen bins with the following restriction: if one projects the bins and points onto any one-dimensional face of the hypercube, then there will be exactly one projected point within each projected interval. See, e.g., [16] for details.
Halton and Hammersley sampling. Halton samples are an example of quasi-Monte Carlo sequences and are defined as follows. To generate a sample of N points in the unit hypercube $A^p \subset \mathbb{R}^p$, one first chooses p prime numbers $s_1, s_2, \ldots, s_p$. Then, each integer $j = 1, \ldots, N$ can be uniquely expressed in an expansion with respect to the base s_k of the form $j = \sum_{i\geq 0} b_{ki}(j) s_k^i$, where $b_{ki}(j)$ is the i-th coefficient of

j in the expansion. One then lets

$$\alpha_k^{(j)} = \sum_{i\geq 0} \frac{b_{ki}(j)}{s_k^{i+1}} \qquad \text{for} \quad k = 1,\dots,p \quad \text{and} \quad j = 1,\dots,N.$$

The N-point Halton sequence of points in the hypercube generated by $s_1,\dots,s_p$ is then defined by

$$\vec{\alpha}_j = (\alpha_1^{(j)} \ \alpha_2^{(j)} \ \cdots \ \alpha_p^{(j)})^T \qquad \text{for} \quad j = 1,\dots,N.$$

The corresponding N Hammersley sample points are given by

$$\vec{\alpha}_j = \big(\frac{j}{N} \ \ \alpha_1^{(j)} \ \alpha_2^{(j)} \ \cdots \ \alpha_{p-1}^{(j)}\big)^T \qquad \text{for} \quad j = 1,\dots,N.$$

See, e.g., [6, 7, 16] for details.

Centroidal Voronoi tessellation sampling. Given an open set $\Omega \subset \mathbb{R}^p$, the set of open subsets $\{\Omega_i\}_{i=1}^N$ is called a tessellation of Ω if $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^N \overline{\Omega}_i = \overline{\Omega}$. Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^p$. Given a set of points $\{\vec{\alpha}_i\}_{i=1}^N$ belonging to $\overline{\Omega}$, the Voronoi region Ω_i corresponding to the point $\vec{\alpha}_i$ is defined by

$$\widehat{\Omega}_i = \{\vec{\beta} \in \Omega \mid \|\vec{\beta} - \vec{\alpha}_i\| < \|\vec{\beta} - \vec{\alpha}_j\| \text{ for } j = 1,\dots,N, \, j \neq i\}.$$

The points $\{\vec{\alpha}_i\}_{i=1}^N$ are called the generators, the set $\{\widehat{\Omega}_i\}_{i=1}^N$ the Voronoi tessellation or Voronoi diagram of Ω corresponding to those generators, and each $\widehat{\Omega}_i$ the Voronoi region corresponding to $\vec{\alpha}_i$. The Voronoi regions are polyhedra.

Given a bounded region $D \subset \mathbb{R}^p$ and a density function $\rho(\cdot)$ defined on D, the center of mass or centroid $\overline{\alpha}$ of D is defined by

$$\overline{\alpha} = \frac{\displaystyle\int_D \rho(\vec{\beta})\vec{\beta}\, d\vec{\beta}}{\displaystyle\int_D \rho(\vec{\beta})\, d\vec{\beta}}.$$

Given N points $\vec{\alpha}_i$, $i = 1,\dots,N$, we can define their associated Voronoi regions $\widehat{\Omega}_i$, $i = 1,\dots,N$. Then, for each Voronoi region $\widehat{\Omega}_i$, we can define the corresponding centroid $\overline{\alpha}_i$. In general, $\overline{\alpha}_i \neq \vec{\alpha}_i$, i.e., the generators of a Voronoi tessellation do not coincide with the centers of mass of the Voronoi regions. The special situation for which $\overline{\alpha}_i = \vec{\alpha}_i$ for $i = 1,\dots,N$ is referred to as a centroidal Voronoi tessellation (CVT) of Ω. This situation is quite special, so that CVTs must be constructed. Details about CVTs, including methods for their construction, are given, e.g., in [3, 10, 23].

In the context of this paper, CVT uniform point sampling refers to choosing the sample points to be the generators of a centroidal Voronoi tessellation (with a constant density function) of the unit hypercube A^p.

Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849-5168, USA
E-mail: ycz0009@auburn.edu

Department of Mathematics, Southern University, New Orleans, LA 70126, USA
E-mail: zchen@suno.edu

Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA
E-mail: gunzburg@scs.fsu.edu