Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 326 (2017) 402–421
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

# A near-stationary subspace for ridge approximation

Paul G. Constantine[a,*], Armin Eftekhari[b], Jeffrey Hokanson[a], Rachel A. Ward[c]

[a] University of Colorado at Boulder, Department of Computer Science, 80309 Boulder, CO, United States
[b] Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, United Kingdom
[c] Department of Mathematics and Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, United States

Received 6 June 2016; received in revised form 7 June 2017; accepted 26 July 2017
Available online 24 August 2017

## Highlights

- We present ridge approximation as a method for constructing response surfaces of expensive computer simulations.
- We show that the gradient-based active subspace is near-stationary for optimal ridge approximation.
- We offer computational heuristics for fitting ridge approximations, and we demonstrate them on an airfoil design problem.

## Abstract

Response surfaces are common surrogates for expensive computer simulations in engineering analysis. However, the cost of fitting an accurate response surface increases exponentially as the number of model inputs increases, which leaves response surface construction intractable for high-dimensional, nonlinear models. We describe *ridge approximation* for fitting response surfaces in several variables. A ridge function is constant along several directions in its domain, so fitting occurs on the coordinates of a low-dimensional subspace of the input space. We review essential theory for ridge approximation – e.g., the best mean-squared approximation and an optimal low-dimensional subspace – and we prove that the gradient-based *active subspace* is near-stationary for the least-squares problem that defines an optimal subspace. Motivated by the theory, we propose a computational heuristic that uses an estimated active subspace as an initial guess for a ridge approximation fitting problem. We show a simple example where the heuristic fails, which reveals a type of function for which the proposed approach is inappropriate. We then propose a simple alternating heuristic for fitting a ridge function, and we demonstrate the effectiveness of the active subspace initial guess applied to an airfoil model of drag as a function of its 18 shape parameters.


*Keywords:* Active subspaces; Ridge functions; Projection pursuit

## 1. Introduction

Engineering computations often employ cheap response surfaces that mimic the input/output relationship between an expensive computer model's parameters and its predictions. The essential idea is to use a few expensive model runs

* Corresponding author.
*E-mail address:* paul.constantine@colorado.edu (P.G. Constantine).

http://dx.doi.org/10.1016/j.cma.2017.07.038
0045-7825/

at particular parameter values (i.e., a *design of experiments* [1]) to fit or train a response surface, where the surface may be a polynomial, spline, or radial basis approximation [2,3]. The same scenario motivates statistical tools for design and analysis of computer experiments [1,4,5], which use Gaussian processes to model uncertainty in the surrogate's predictions.

The cost of constructing an accurate response surface increases exponentially as the dimension of the input space increases; in approximation theory, this is the tractability problem [6,7], though it is colloquially referred to as the *curse of dimensionality* [8, Section 5.16]. Several techniques attempt to alleviate this curse—each with advantages and drawbacks for certain classes of problems; see [9] for an extensive survey. One idea is to identify unimportant input variables with global sensitivity metrics [10] and fix them at nominal values, which effectively reduces the dimension for response surface construction; Sobol' et al. studied the effects of such coordinate-based dimension reduction on the approximation [11]. A generalization of coordinate-based dimension reduction is to identify unimportant directions—not necessarily coordinate aligned. If the scientist can identify a few important linear combinations of inputs, then she may fit a response surface of only those linear combinations, which allows a higher degree of accuracy along important directions in the input space.

A *ridge function* [12] is a function of a few linear combinations of inputs that takes the form $g(\boldsymbol{U}^T\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^m$, $\boldsymbol{U} \in \mathbb{R}^{m\times n}$ with $n < m$, and $g : \mathbb{R}^n \to \mathbb{R}$. The term *ridge function* is more commonly used when $\boldsymbol{U}$ is a single vector ($n = 1$). Pinkus calls our definition a *generalized ridge function* [12, Chapter 1], though Keiper uses the qualifier *generalized* for a model where $\boldsymbol{U}$ depends on $\mathbf{x}$ [13]. A ridge function is constant along directions in its domain that are orthogonal to $\boldsymbol{U}$'s columns. To see this, let $\mathbf{v} \in \mathbb{R}^m$ be orthogonal to $\boldsymbol{U}$'s columns; then

$$g(\boldsymbol{U}^T(\mathbf{x}+\mathbf{v})) = g(\boldsymbol{U}^T\mathbf{x} + \underbrace{\boldsymbol{U}^T\mathbf{v}}_{=0}) = g(\boldsymbol{U}^T\mathbf{x}). \tag{1}$$

If $\boldsymbol{U}$ is known, then one need only construct $g$, which is a function of $n < m$ variables. Thus, constructing $g$ may require exponentially fewer model evaluations than constructing a comparably accurate response surface on all $m$ variables.

Let $f : \mathbb{R}^m \to \mathbb{R}$ represent the simulation model's input/output map to approximate, and let its domain be equipped with a probability function $\rho : \mathbb{R}^m \to \mathbb{R}_+$. The function $\rho$ may model uncertainty in the simulation's input parameters, which is a common modeling choice in *uncertainty quantification* [14,15]. The ridge approximation problem may be stated as: given $f$ and $\rho$, find $g$ and $\boldsymbol{U}$ that minimize the approximation error. After a brief survey of related concepts, we define a specific ridge approximation problem in Section 2. We then study a particular $\boldsymbol{U}$ derived from $f$'s gradient known as the *active subspace* [16]. We show that, under certain conditions, the active subspace is nearly stationary—i.e., that the gradient of the objective function defining the approximation problem is bounded; see Section 3. This result motivates a heuristic for the initial subspace when fitting a ridge approximation given pairs $\{(\mathbf{x}_i, f(\mathbf{x}_i))\}$. In Section 4, we show a simple bivariate example that exposes the limitations of the heuristic. We then study an 18-dimensional example from an airfoil shape optimization problem where the heuristic succeeds; in particular, we demonstrate a numerical procedure for estimating the active subspace using samples of the gradient, and we show how the estimated active subspace is a superior starting point for a numerical optimization heuristic for fitting the ridge approximation.

### 1.1. Related concepts

There are many concepts across subfields that relate to ridge approximation. In what follows, we briefly review three of these subfields with citations that point interested readers to representative works.

#### 1.1.1. Projection pursuit regression

In the context of statistical regression, Friedman and Stuetzle [17] proposed *projection pursuit regression* with a ridge function model:

$$y_i = \sum_{k=1}^{r} g_k(\mathbf{u}_k^T\mathbf{x}_i) + \varepsilon_i, \tag{2}$$

where $\mathbf{x}_i$'s are samples of the predictors, $y_i$'s are the associated responses, and $\varepsilon_i$'s model random noise—all standard elements of statistical regression [18]. The $g_k$'s are smooth univariate functions (e.g., splines), and the $\mathbf{u}_k$'s are