Gradient-based dimension reduction of multivariate vector-valued functions

Olivier Zahm*, Paul Constantine†, Clémentine Prieur‡, and Youssef Marzouk§

Abstract. We propose a gradient-based method for detecting and exploiting low-dimensional input parameter dependence of multivariate functions. The methodology consists in minimizing an upper bound, obtained by Poincaré-type inequalities, on the approximation error. The resulting method can be used to approximate vector-valued functions (e.g., functions taking values in $\mathbb{R}^n$ or functions taking values in function spaces) and generalizes the notion of active subspaces associated with scalar-valued functions. A comparison with the truncated Karhunen-Loève decomposition shows that using gradients of the function can yield more effective dimension reduction. Numerical examples reveal that the choice of norm on the codomain of the function can have a significant impact on the function’s low-dimensional approximation.

Key words. High-dimensional function approximation, dimension reduction, active subspace, ridge approximation, Karhunen-Loève decomposition, Poincaré inequality.

AMS subject classifications. 41A30, 41A63, 65D15

1. Introduction. Many problems that arise in uncertainty quantification—e.g., integrating or approximating multivariate functions—suffer from the curse of dimensionality: loosely, the cost of computing a sufficiently accurate approximation grows dramatically (typically exponentially) with the dimension of the input parameter space. One approach to alleviate this curse is to identify and exploit some notion of low-dimensional structure. For example, the function of interest might vary primarily along a few directions of the input parameter space. In this case, we can say that the problem has a low intrinsic dimension; algorithms for quantifying uncertainty can then focus on these important directions to reduce overall cost.

A common and rather simple approach to reducing a function’s input dimension is the truncated Karhunen-Loève decomposition [31], closely related to principal component analysis [19]. These techniques exploit the correlation structure of the function’s input space (specifically, decay in the spectrum of the covariance of the input measure). However more effective dimension reduction should be possible with techniques that exploit not only input correlations but also the structure of the input-output map itself. Global sensitivity analysis [29] seeks to rank the input coordinates’ importance relative to a particular output of interest. In practice, one may use this ranking to distinguish a subset of influential input variables; computations can proceed by fixing the uninfluential inputs and only varying the influential inputs. Limiting attention to subsets of input coordinates, however, may miss more general linear structure that is even lower-dimensional. Active subspaces [5, 7] are eigenspaces of the average outer product of the function’s gradient with itself. They capture directions along which the function “varies” the most—in the sense of its output responding most strongly to input perturbations, in expectation over the input measure. Active subspaces have been

*Massachusetts Institute of Technology, zahmo@mit.edu
†University of Colorado Boulder, paul.constantine@colorado.edu
‡Université Grenoble Alpes and Laboratoire Jean Kuntzmann, clementine.prieur@univ-grenoble-alpes.fr
§Massachusetts Institute of Technology, ymarz@mit.edu
discovered and exploited in a wide range of science and engineering models [22, 6, 17]. Active subspaces are not necessarily coordinate-aligned, and in this sense they generalize coordinate-based global sensitivity analysis. Yet global sensitivity analysis and active subspaces have primarily been focused on scalar-valued functions (as in models with a single output quantity of interest). In the presence of multiple outputs of interest, as is the case in many practical applications, new approaches are needed. Generalized Sobol’ indices for multivariate or functional outputs have recently been introduced in [20], and further studied in [14, 13]. In the context of active subspaces, one could try to identify important input parameter directions for each output and then “combine” all those directions, as in [18]. But it is not clear how to interpret or even best perform such a combination step, particularly when seeking rigorous error guarantees.

In this paper, we develop a methodology for detecting and exploiting the low intrinsic dimension of vector-valued functions—for instance, functions with multiple real-valued outputs or functions taking values in function spaces. Our analysis extends the theoretical development of active subspaces to vector-valued outputs. We formulate the problem as seeking a particular kind of structure-exploiting approximation (a ridge function; see below) whose error is below a prescribed tolerance. To solve this problem, we use a Poincaré-type inequality to derive an upper bound on the error. This bound, defined by means of gradients of the function, admits a simple expression and can be easily (analytically) minimized. Then we construct an approximation of the vector-valued function such that the upper bound is below the prescribed tolerance. We argue that this methodology can yield more effective dimension reduction than the truncated Karhunen-Loève decomposition. The choice of norm on the codomain (output space) of the function, used to measure the approximation error, is of central importance. In this paper, we assume that the domain (input space) of the function is equipped with a Gaussian probability measure, and we use the corresponding weighted norm on the codomain to define a Hilbert space within which we seek approximations. The approximation we obtain thus depends on the input measure, the function to be approximated, and the norm on the output space. When choosing the latter norm according to some particular objective, we naturally obtain a goal-oriented approximation.

We demonstrate our results on several analytical examples and on a PDE model with a correlated random field input. In the latter case, we show how the low-dimensional structure and its associated approximation change as various vector-valued outputs—from the PDE solution at two points in the spatial domain to the entire PDE solution field—are considered. When no low-dimensional structure is present in the input-output map, the only possible dimension reduction comes from the correlation structure of the inputs.

The low-dimensional model we use to approximate the map from inputs to outputs is a ridge function [27]. Ridge functions and their approximation properties were extensively studied in the 1980s because of their connection to both projection pursuit regression and early neural networks [9, 15]. Recent work has exploited compressed sensing to recover a ridge function from point queries [11, 4]. In contrast, we are not recovering a true but unknown ridge function; instead, we approximate a given map by a ridge function up to a prescribed precision.

Finally, note also that input space dimension reduction arises in the statistical regression literature under the name sufficient dimension reduction [1, 8]. Similar to the ridge function.
recovery problem, the goal is to identify linear combinations in the input space that are statistically sufficient to explain the regression response. Related ideas from regression include the well-known projection pursuit regression [12, 16], where the directions in the projection pursuit model are trained using given input/output pairs. In [30], gradient information is used to explore the underlying regression structure by means of average derivative functionals, estimated nonparametrically via kernels. Li et al. specifically studied vector-valued responses as groupwise dimension reduction [21]. Broadly, and in contrast with the approach proposed in this paper, these regression analyses are concerned with estimation from a given data set, and thus rely on statistical assessments of the error.

2. Dimension reduction of the input parameter space. Throughout the paper, the algebraic space $\mathbb{R}^d$ refers to a parameter space of dimension $d \gg 1$. The Borel set of $\mathbb{R}^d$ is denoted by $B(\mathbb{R}^d)$ and we let $\mu = N(m, \Sigma)$ be the Gaussian probability measure on $\mathbb{R}^d$ with mean $m \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d \times d}$, which is assumed to be non-singular. We let $V$ be a Hilbert space endowed with a scalar product $(\cdot, \cdot)_V$ and the associated norm $\| \cdot \|_V = (\cdot, \cdot)_V^{1/2}$. We denote by

$$H = L^2(\mathbb{R}^d, B(\mathbb{R}^d), \mu; V),$$

the Hilbert space which contains all the Borel-measurable functions $v : \mathbb{R}^d \to V$ such that $\|v\|_H < \infty$, where $\| \cdot \|_H$ is the norm associated with the scalar product $(\cdot, \cdot)_H$ defined by

$$(u, v)_H = \int (u(x), v(x))_V \mu(dx),$$

for any $u, v \in H$.

We consider the problem of finding a controlled approximation of a function $f \in H$ by a ridge function in $H$ of the form of

$$(1) \quad x \mapsto g(P_r x),$$

where $g : \mathbb{R}^d \to V$ is a Borel-measurable function and where $P_r \in \mathbb{R}^{d \times d}$ is a rank-$r$ projector. Notice that $g(P_r x) = g(P_r y)$ whenever $x - y \in \text{Ker}(P_r)$, which means that the function (1) is constant along the kernel of the projector. In other words, $x \mapsto g(P_r x)$ is a function of at most $r$ linear combinations of the input parameters. As in [24], the function $g$ is called the profile of the ridge function. Given a prescribed tolerance $\varepsilon \geq 0$, the problem consists in finding the function $g$ and the projector $P_r$, also seen as a (linear) function $P_r : \mathbb{R}^d \to \mathbb{R}^d$, such that

$$(2) \quad \| f - g \circ P_r \|_H \leq \varepsilon.$$
Remark 2.1. An equivalent formulation of the problem is the following. Given a tolerance \( \varepsilon > 0 \), we want to find a Borel function \( g : \mathbb{R}^d \to V \) and a low-rank projector \( P_r \in \mathbb{R}^{d \times d} \) such that
\[
\mathbb{E}(\| f(X) - g(P_r X) \|^2_V) \leq \varepsilon^2,
\]
where \( X \sim \mathcal{N}(m, \Sigma) \) is a random vector and where \( \mathbb{E}(\cdot) \) denotes the mathematical expectation. If \( \varepsilon^2 \ll \text{Var}(f(X)) = \mathbb{E}((f(X) - \mathbb{E}(f(X)))^2) \), the statistical interpretation is that the random variable \( X_r = P_r X \) is an explanatory variable for \( f(X) \), in the sense that most of the variance of \( f(X) \) can be explained by \( X_r \).

2.1. Optimal profile for the ridge function. In this section, we assume that the projector \( P_r \) is given. We denote by
\[
\mathcal{H}_{P_r} = L^2(\mathbb{R}^d, \sigma(P_r), \mu; V),
\]
the space which contains all the \( \sigma(P_r) \)-measurable functions \( v : \mathbb{R}^d \to V \) such that \( \| v \|_{\mathcal{H}} < \infty \). Here \( \sigma(P_r) \) is the \( \sigma \)-algebra generated by \( P_r \). By Doob-Dynkin’s lemma, see for example Lemma 1.13 in [25], the set of all \( \sigma(P_r) \)-measurable functions is exactly the set of the functions of the form \( x \mapsto g(P_r x) \) for some Borel function \( g \), so that
\[
\mathcal{H}_{P_r} = \{ g \circ P_r \mid g : \mathbb{R}^d \to V, \text{Borel function} \} \cap \mathcal{H}.
\]
Note that \( \mathcal{H}_{P_r} \) is a closed subspace in \( \mathcal{H} \). Then, for any \( f \in \mathcal{H} \), there exists a unique minimizer of \( f_r \mapsto \| f - f_r \|_{\mathcal{H}} \) over \( \mathcal{H}_{P_r} \). This minimizer corresponds to the orthogonal projection of \( f \in \mathcal{H} \) onto \( \mathcal{H}_{P_r} \) and is denoted by \( \mathbb{E}_\mu(f|\sigma(P_r)) \). We can write
\[
\| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\mathcal{H}} = \min_{f_r \in \mathcal{H}_{P_r}} \| f - f_r \|_{\mathcal{H}} = \min_{g : \mathbb{R}^d \to V, \text{Borel function}} \| f - g \circ P_r \|_{\mathcal{H}},
\]
which means that \( \mathbb{E}_\mu(f|\sigma(P_r)) \) yields an optimal profile \( g \). Note that \( \mathbb{E}_\mu(f|\sigma(P_r)) \in \mathcal{H}_{P_r} \) can be uniquely characterized by the variational equation
\[
\int (\mathbb{E}_\mu(f|\sigma(P_r)), h)_V \, d\mu = \int (f, h)_V \, d\mu,
\]
for all \( h \in \mathcal{H}_{P_r} \). In other words, \( \mathbb{E}_\mu(f|\sigma(P_r)) \) corresponds to the conditional expectation of \( f \) under the distribution \( \mu \) given the \( \sigma \)-algebra \( \sigma(P_r) \), which explains the choice of notation.

Proposition 2.2. Let \( P_r \) and \( Q_r \) be two projectors such that \( \text{Ker}(P_r) = \text{Ker}(Q_r) \). Then we have
\[
\mathcal{H}_{P_r} = \mathcal{H}_{Q_r}.
\]

Proof. Let \( h \in \mathcal{H}_{P_r} \). By (3) we can write \( h = g \circ P_r \) for some Borel function \( g \). Since \( \text{Ker}(Q_r) = \text{Ker}(P_r) \) we have \( P_r x = P_r Q_r x = 0 \) for all \( x \in \text{Ker}(Q_r) \). Also for any \( x \in \text{Im}(Q_r) \) we have \( Q_r x = x \) and then \( P_r x = P_r Q_r x \). Thus \( P_r x = P_r Q_r x \) holds for any \( x \in \mathbb{R}^d = \text{Ker}(Q_r) \oplus \text{Im}(Q_r) \) so that \( P_r = P_r Q_r \). Then \( h = g \circ P_r = (g \circ P_r) \circ Q_r \) which shows that \( h \in \mathcal{H}_{Q_r} \). Then the inclusion \( \mathcal{H}_{P_r} \subset \mathcal{H}_{Q_r} \) holds. By symmetry of the role of \( P_r \) and \( Q_r \) we obtain the result. \(\blacksquare\)
Let us recall that a projector is uniquely characterized by both its kernel and its image. Proposition 2.2 shows that $\mathcal{H}_{P_r}$ is invariant with respect to the image of $P_r$, and so is the conditional expectation $E_\mu(f|\sigma(P_r))$. In particular, the error $P_r \mapsto \|f - E_\mu(f|\sigma(P_r))\|_H$ depends only on the kernel of $P_r$. This means that, with regard to the initial dimension reduction problem (2), the goal is now to find a subspace where the function $f$ does not vary significantly.

By Proposition 2.2 and without loss of generality, we can assume that $P_r$ is an orthogonal projector with respect to an arbitrary norm on $\mathbb{R}^d$. In the present context, the natural norm to use is the one induced by the precision matrix $\Sigma$, meaning the norm $\|\cdot\|_{\Sigma^{-1}}$ defined by $\|x\|^2_{\Sigma^{-1}} = x^T \Sigma^{-1} x$ for any $x \in \mathbb{R}^d$. Thus we will say that $P_r$ is a $\Sigma^{-1}$-orthogonal projector iff

$$\|x\|^2_{\Sigma^{-1}} = \|P_r x\|^2_{\Sigma^{-1}} + \|(I_d - P_r)x\|^2_{\Sigma^{-1}},$$

holds for all $x \in \mathbb{R}^d$. The following proposition gives a simple expression for the conditional expectation $E_\mu(f|\sigma(P_r))$, provided $P_r$ satisfies (5).

**Proposition 2.3.** Let $\mu = \mathcal{N}(m, \Sigma)$ where $\Sigma \in \mathbb{R}^{d \times d}$ is a non-singular covariance matrix and $f \in \mathcal{H}$. Then for any $\Sigma^{-1}$-orthogonal projector $P_r$ we have

$$E_\mu(f|\sigma(P_r)) : x \mapsto E(f(P_r x + (I_d - P_r)Y)),$$

where the expectation is taken over the random vector $Y \sim \mu$.

**Proof.** Let $F : x \mapsto \int_{\mathbb{R}^d} f(P_r x + (I_d - P_r)y)\mu(dy)$ and $h \in \mathcal{H}_{P_r}$. By (3), $h$ can be written as a $g \circ P_r$ for some Borel function $g$ so that $h(x) = h(P_r x + (I_d - P_r)y)$ for all $x, y \in \mathbb{R}^d$. We can write

$$\int_{\mathbb{R}^d} (F(x), h(x))_V \mu(dx) = \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} f(P_r x + (I_d - P_r)y)\mu(dy), h(x) \right)_V \mu(dx)$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left( f(P_r x + (I_d - P_r)y), h(P_r x + (I_d - P_r)y) \right)_V \mu(dy) \mu(dx)$$

$$= \mathbb{E}((f(Z), h(Z))_V),$$

where the expectation is taken over the random vector $Z = P_r X + (I_d - P_r)Y$, where $X$ and $Y$ are two independent random vectors distributed as $\mu = \mathcal{N}(m, \Sigma)$. If $Z \sim \mu$ then the previous relation yields (4) for any $h \in \mathcal{H}_{P_r}$, which would conclude the proof.

It remains to show that $Z \sim \mu$. Note that $Z$ is Gaussian with mean $m$ and covariance

$$\text{Cov}(Z) = P_r \Sigma P_r^T + (I_d - P_r)\Sigma(I_d - P_r)^T = \Sigma - P_r \Sigma - \Sigma P_r^T + 2P_r \Sigma P_r^T.$$

Then $Z \sim \mu$ if and only if $P_r \Sigma + \Sigma P_r^T = 2P_r \Sigma P_r^T$. Since $P_r$ is $\Sigma^{-1}$-orthogonal, relation (5) holds for any $x \in \mathbb{R}^d$ which is equivalent to

$$P_r^T \Sigma^{-1} + \Sigma^{-1} P_r = 2P_r^T \Sigma^{-1} P_r.$$

---

1 Of course an orthogonal projector (orthogonal with respect to any given norm) is uniquely characterized either by its kernel or by its image, since the other subspace can be uniquely defined as the orthogonal complement.
Such that $\|\cdot\|$ is the norm on $\mathbb{R}^d$.

Therefore we have

$$P_r \Sigma + \Sigma P_r^T = 2P_r \Sigma P_r^T,$$

which concludes the proof.

### 2.2. Poincaré-based upper bound for the error.

In this section we show how Poincaré-type inequalities can be used to derive an upper bound for $\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H$. This upper bound holds for any projector and is quadratic in $P_r$ so that it can easily be minimized.

It is well known that the standard Gaussian distribution $\gamma = \mathcal{N}(0, I_d)$ satisfies the Poincaré inequality

$$\int (h - \mathbb{E}_\gamma(h))^2 \, d\gamma \leq \int \|\nabla h\|_2^2 \, d\gamma,$$

for any continuously differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, where $\nabla h$ denotes the gradient of $h$ (see for example Theorem 3.20 in [2]). Here $\mathbb{E}_\gamma(h) = \int h \, d\gamma$ and $\|\cdot\|_2 = \sqrt{\langle \cdot, \cdot \rangle}$ denotes the canonical norm of $\mathbb{R}^d$. As noticed in [3], non-standard Gaussian distributions also satisfy a Poincaré inequality. By replacing $h$ by $x \mapsto h(\Sigma^{1/2} x + m)$ in (6), where $\Sigma^{1/2}$ is a symmetric square root of $\Sigma$, we have that $\gamma = \mathcal{N}(m, \Sigma)$ satisfies

$$\int (h - \mathbb{E}_\mu(h))^2 \, d\mu \leq \int \|\nabla h\|_\Sigma^2 \, d\mu,$$

for any continuously differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, where $\|\cdot\|_\Sigma$ is the norm on $\mathbb{R}^d$ such that $\|x\|_\Sigma^2 = x^T \Sigma x$ for all $x \in \mathbb{R}^d$. The next proposition shows that $\mu$ satisfies another Poincaré-type inequality which we call the subspace Poincaré inequality.

**Proposition 2.4.** The probability distribution $\mu = \mathcal{N}(m, \Sigma)$ satisfies

$$\int (h - \mathbb{E}_\mu(h|\sigma(P_r)))^2 \, d\mu \leq \int \|(I_d - P_r)^T \nabla h\|_\Sigma^2 \, d\mu,$$

for any continuously differentiable function $h : \mathbb{R}^d \to \mathbb{R}$ and for any projector $P_r$.

**Proof.** First we assume that $P_r$ is a $\Sigma^{-1}$-orthogonal projector. Let $h : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function and define $g : x \mapsto h(P_r y + (I_d - P_r)x)$ for some $y \in \mathbb{R}^d$. For any $x \in \mathbb{R}^d$ we have $\nabla g(x) = (I_d - P_r)^T \nabla h(P_r y + (I_d - P_r)x)$. By Proposition 2.3 we have

$$\mathbb{E}_\mu(g) = \int_{\mathbb{R}^d} h(P_r y + (I_d - P_r)x') \mu(dx') = \mathbb{E}_\mu(h|\sigma(P_r))(y).$$

Notice that we can write $\mathbb{E}_\mu(h|\sigma(P_r))(y) = \mathbb{E}_\mu(h|\sigma(P_r))(P_r y + (I_d - P_r)x)$. Then the Poincaré inequality (7) applied with the function $g$ yields

$$\int_{\mathbb{R}^d} (h(P_r y + (I_d - P_r)x) - \mathbb{E}_\mu(h|\sigma(P_r))(P_r y + (I_d - P_r)x))^2 \mu(dx) \leq \int_{\mathbb{R}^d} \|(I_d - P_r)^T \nabla h(P_r y + (I_d - P_r)x\|_\Sigma^2 \, \mu(dx).$$
Recall that, since $P_r$ is $\Sigma^{-1}$-orthogonal, we have $P_r Y + (I_d - P_r) X \sim \mu$ whenever $X \sim \mu$ and $Y \sim \mu$ are independent; see the proof of Proposition 2.3. Thus, replacing $y$ by $Y$ in the previous inequality and taking the expectation over $Y$ yields (8).

It remains to show that (8) also holds for projectors that are not $\Sigma^{-1}$-orthogonal. Thus let $P_r$ be any projector and define $Q_r$ as the (unique) $\Sigma^{-1}$-orthogonal projector such that $\text{Ker}(Q_r) = \text{Ker}(P_r)$. Following the proof of Proposition 2.3, we have that $Q_r \Sigma + \Sigma Q_r^T = 2Q_r \Sigma Q_r^T$ which is equivalent to saying that the relation $\|x\|_{\Sigma}^2 = \|Q_r^T x\|_{\Sigma}^2 + \|(I_d - Q_r^T) x\|_{\Sigma}^2$ holds for any $x \in \mathbb{R}^d$. Then $\|x\|_{\Sigma}^2 \geq \|(I_d - Q_r^T) x\|_{\Sigma}^2$ for any $x \in \mathbb{R}^d$. Replacing $x$ by $(I_d - P_r^T) x$ we get

\[
\|(I_d - P_r^T) x\|_{\Sigma}^2 \geq \|(I_d - Q_r^T)(I_d - P_r^T) x\|_{\Sigma}^2 = \|(I_d - Q_r^T - P_r^T + Q_r^T P_r) x\|_{\Sigma}^2 = \|(I_d - Q_r^T) x\|_{\Sigma}^2.
\]

(9)

For the last equality we used relation $P_r = P_r Q_r$, which holds true since $\text{Ker}(P_r) = \text{Ker}(Q_r)$. Finally, Proposition 2.2 allows writing $\mathbb{E}_\mu(h|\sigma(P_r)) = \mathbb{E}_\mu(h|\sigma(Q_r))$ so that

\[
\int (h - \mathbb{E}_\mu(h|\sigma(P_r)))^2 d\mu = \int (h - \mathbb{E}_\mu(h|\sigma(Q_r)))^2 d\mu \leq \int \|(I_d - Q_r^T) \nabla h\|_{\Sigma}^2 d\mu \overset{(8)}{\leq} \int \|(I_d - P_r^T) \nabla h\|_{\Sigma}^2 d\mu,
\]

which shows that (8) holds for any projector $P_r$. 

The subspace Poincaré inequality stated in Proposition 2.4 allows us to derive an upper bound for the error $\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}}$, as shown by the following proposition.

**Proposition 2.5.** Let $\mu = N(m, \Sigma)$ where $\Sigma \in \mathbb{R}^{d \times d}$ is a non-singular covariance matrix and let $f \in \mathcal{H} = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu; V)$ where $V$ is a separable Hilbert space. Furthermore, assume that $f$ is continuously differentiable. Then for any projector $P_r$ we have

\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}} \leq \sqrt{\text{trace} \left( \Sigma(I_d - P_r^T) H(I_d - P_r) \right)},
\]

where $H \in \mathbb{R}^{d \times d}$ is the matrix defined by

\[
H = \int (\nabla f)^* (\nabla f) \ d\mu.
\]

Here, $\nabla f(x) : \mathbb{R}^d \to V$ denotes the Jacobian of $f$ at point $x$, which is the linear operator such that

\[
f(x + \delta x) = f(x) + \nabla f(x) \delta x + o(\|\delta x\|_2),
\]

(11)

for any $\delta x \in \mathbb{R}^d$. The linear operator $(\nabla f(x))^* : V \to \mathbb{R}^d$ is the adjoint of $\nabla f(x)$, defined by $y^T((\nabla f(x))^* v) = (\nabla f(x) y, v)_V$ for all $v \in V$ and for all $y \in \mathbb{R}^d$. 

\[
(\nabla f(x))^* v = \nabla f(x)^* v, \quad (\nabla f(x))^* v = \nabla f(x) v^*. 
\]

Note that $\nabla f(x)$ depends on $x$. The matrix $H$ is the Hessian of $f$ at $x$; i.e., $H(x) = \nabla^2 f(x)$. This means that $H$ is a linear operator that maps elements of $V$ to elements of $\mathbb{R}^d$. Furthermore, $H$ is a symmetric matrix because $\nabla f(x)$ is the Jacobian of $f$ at $x$, which is a linear function.

Finally, the trace of a matrix $A$ is defined as $\text{trace}(A) = \sum_{i=1}^d A_{ii}$, and the Frobenius norm of a matrix $A$ is defined as $\|A\|_F = \sqrt{\text{trace}(A^* A)}$. Thus, the right-hand side of the inequality in Proposition 2.5 can be interpreted as the square root of the sum of the squares of the eigenvalues of $H$.
Proof. Denote by \( \{e_i\}_{i \geq 1} \) an orthonormal basis of the separable Hilbert space \( V \). The function \( f \in \mathcal{H} \) can be represented as \( x \mapsto \sum_{i \geq 1} f_i(x) e_i \) where, for all \( i \geq 1 \), the function \( f_i : \mathbb{R}^d \to \mathbb{R} \) is defined by \( f_i(x) = (f(x), e_i)_V \). By linearity of the conditional expectation and by orthogonality of the basis, we can write

\[
\| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\tilde{\mathcal{H}}} = \sum_{i \geq 1} \int (f_i - \mathbb{E}_\mu(f_i|\sigma(P_r)))^2 \, d\mu.
\]

By assumption, \( f \) is continuously differentiable and so is \( f_i \) for all \( i \geq 1 \). Then the subspace Poincaré inequality (8) with \( h = f_i \) allows writing

\[
\int (f_i - \mathbb{E}_\mu(f_i|\sigma(P_r)))^2 \, d\mu \leq \int \| (I_d - P_r^T)\nabla f_i \|_2^2 \, d\mu
\]

\[
= \int \text{trace} (\Sigma(I_d - P_r^T)(\nabla f_i)(\nabla f_i)^T(I_d - P_r)) \, d\mu
\]

\[
= \text{trace} (\Sigma(I_d - P_r^T)\left( \int (\nabla f_i)(\nabla f_i)^T \, d\mu \right)(I_d - P_r)),
\]

where \( \nabla f_i(x) \in \mathbb{R}^d \) denotes the gradient of \( f_i \) at point \( x \in \mathbb{R}^d \). Together with (13), the above relation yields (10) with

\[
H = \int \sum_{i \geq 1} (\nabla f_i)(\nabla f_i)^T \, d\mu.
\]

To conclude the proof, it remains to show that this definition of \( H \) matches the one in (11). It is sufficient to show \( \sum_{i \geq 1} (\nabla f_i)(\nabla f_i)^T = (\nabla f(x))^* (\nabla f(x)) \) for any \( x \in \mathbb{R}^d \). Firstly, projecting (12) onto the basis vector \( e_i \) yields \( f_i(x + \delta x) = f_i(x) + (\nabla f(x)\delta x, e_i)_V + o(\|\delta x\|_2) \) which allows identifying the gradient of \( f_i \) by \( \nabla f_i(x)^T \delta x = (\nabla f(x)\delta x, e_i)_V \) for all \( \delta x \in \mathbb{R}^d \). Then

\[
\nabla f(x)^T \delta x = \sum_{i \geq 1} (\nabla f_i(x)^T \delta x) e_i.
\]

Secondly, recall that the adjoint operator \( \nabla f(x)^* : V \to \mathbb{R}^d \) is defined by \( \delta y^T (\nabla f(x)^* v) = (\nabla f(x)\delta y, v)_V \) for all \( v \in V \) and for all \( \delta y \in \mathbb{R}^d \). Replacing \( v \) by \( \nabla f(x)^T \delta x \) we get

\[
\delta y^T \nabla f(x)^* \nabla f(x) \delta x = (\nabla f(x)\delta y, \nabla f(x)\delta x)_V
\]

\[
= \sum_{i,j \geq 1} (\nabla f_i(x)^T \delta y)(\nabla f_j(x)^T \delta x)(e_i, e_j)_V
\]

\[
= \delta y^T \left( \sum_{i \geq 1} \nabla f_i(x) \nabla f_i(x)^T \right) \delta x,
\]

for any \( \delta x, \delta y \in \mathbb{R}^d \) so that \( \nabla f(x)^* \nabla f(x) = \sum_{i \geq 1} \nabla f_i(x) \nabla f_i(x)^T \). This concludes the proof. \( \blacksquare \)

We finish this section with an example. Assume \( V = \mathbb{R}^m \) is an algebraic space endowed with the norm \( \| \cdot \|_V \) defined by \( \| v \|_V^2 = v^T M v \) for all \( v \in V \), where \( M \in \mathbb{R}^{m \times m} \) is a symmetric positive-definite matrix. The Jacobian of \( f \) at point \( x \in \mathbb{R}^d \) is a \( m \)-by-\( d \) matrix defined by

\[
(\nabla f(x))_{i,j} = \frac{\partial f_i}{\partial x_j}(x).
\]
The adjoint of $\nabla f(x)$ is a $d$-by-$m$ matrix given by $(\nabla f(x))^* = (\nabla f(x))^T M$ so that we can write the matrix $H$ as follows

$$H = \int (\nabla f)^T M (\nabla f) \, d\mu. \tag{14}$$

Notice that $H$ depends on the choice the norm $\| \cdot \|_V$ via the matrix $M$.

### 2.3. Minimizing the upper bound

The following proposition enables minimization of the upper bound in Proposition 2.5.

**Proposition 2.6.** Let $\Sigma \in \mathbb{R}^{d \times d}$ be a symmetric positive-definite matrix and $H \in \mathbb{R}^{d \times d}$ a symmetric positive-semidefinite matrix. Denote by $(\lambda_i, v_i) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^d$ the $i$-th generalized eigenpair of the matrix pair $(H, \Sigma^{-1})$, meaning $H v_i = \lambda_i \Sigma^{-1} v_i$ with $\|v_i\|_{\Sigma^{-1}} = 1$. For any $r \leq d$ we have

$$\min_{\text{rank-}r \text{ projector}} \text{trace} \left( (\Sigma I_d - P_r^T) H (I_d - P_r) \right) = \sum_{i=r+1}^{d} \lambda_i. \tag{15}$$

Furthermore a solution to the above minimization problem is the $\Sigma^{-1}$-orthogonal projector defined by

$$P_r = \left( \sum_{i=1}^{r} v_i v_i^T \right) \Sigma^{-1}. \tag{16}$$

**Proof.** Let $H^{1/2}$ and $\Sigma^{1/2}$ be the symmetric positive square roots of $H$ and $\Sigma$ respectively. For any projector $P_r$ we have

$$\text{trace}(\Sigma (I_d - P_r^T) H (I_d - P_r)) = \|H^{1/2}(I_d - P_r)\Sigma^{1/2}\|_F^2 = \|A - X_r\|_F^2,$$

where $A = H^{1/2}\Sigma^{1/2}$ and $X_r = H^{1/2}P_r\Sigma^{1/2}$ and where $\| \cdot \|_F = \sqrt{\text{trace}(\cdot^T \cdot)}$ denotes the Frobenius norm. Consider the singular value decomposition of $A = UDV^T$ where $U, V \in \mathbb{R}^{d \times d}$ are two orthogonal matrices and $D = \text{diag}(a_1, \ldots, a_d)$ with $a_1 \geq a_2 \geq \ldots \geq 0$. The Eckart-Young theorem states that (i) the matrix $A_r = UD_rV^T$, with $D_r = \text{diag}(a_1, \ldots, a_r, 0, \ldots, 0)$, is a minimizer of $\|A - A_r\|_F^2$ over all matrices $A_r$ with rank($A_r$) $\leq r$ and (ii) that $\|A - A_r\|_F^2 = a_{r+1}^2 + \ldots + a_d^2$. We now show that $A_r$ can be written as $X_r = H^{1/2}P_r\Sigma^{1/2}$ for some rank-$r$ projector $P_r$. Let $V_r \in \mathbb{R}^{d \times r}$ be the matrix containing the $r$ first columns of $V$ and let $P_r = \Sigma^{1/2}V_rV_r^T\Sigma^{-1/2}$. Since $V_r^T V_r = I_r$ we have $P_r^2 = P_r$ so that $P_r$ is a rank-$r$ projector. Also we have $X_r = H^{1/2}P_r\Sigma^{1/2} = AV_rV_r^T = A_r$ then $\|A - X_r\|^2_2 = \|A - A_r\|^2_2 \leq \|A - \tilde{A}_r\|^2_2$ holds for any rank-$r$ matrix $A_r$, in particular for the ones of the form of $A_r = H^{1/2}P_r\Sigma^{1/2}$ for any rank-$r$ projector $P_r$. This shows that the minimum in (15) is reached by $P_r = \Sigma^{1/2}V_rV_r^T\Sigma^{-1/2}$. Furthermore it is easy to check that $P_r^T \Sigma^{-1} P_r = 2P_r^T \Sigma^{-1} P_r$ holds so that, as we saw in the proof of Proposition 2.3, $P_r$ is $\Sigma^{-1}$-orthogonal.

It remains to show that $P_r$ can be written as in (16). Notice that $A^T A = \Sigma^{1/2} H \Sigma^{1/2} = V D^2 V^T$ holds and yields $H \Sigma^{1/2} V = \Sigma^{-1/2} V D^2$. Denoting by $v_i$ the $i$-th column of $\Sigma^{1/2} V$ (which is such that $\|v_i\|_{\Sigma^{-1}} = 1$), the latter relation yields $H v_i = a_i^2 \Sigma^{-1} v_i$. This means that $v_i$
is the \(i\)-th generalized eigenvector of the matrix pair \((H, \Sigma^{-1})\) and the associated eigenvalue is \(\lambda_i = a_i^2\). Therefore \(P_r\) satisfies

\[
P_r = \Sigma^{1/2}V_r V_r^T \Sigma^{-1/2} = (\sum_{i=1}^{r} v_i v_i^T) \Sigma^{-1}
\]

as in (16) and

\[
\text{trace}(\Sigma(I_d - P_r^T H(I_d - P_r))) = \|A - A_r\|_{\mathcal{F}}^2 = \|U(D-D_r)V^T\|_{\mathcal{F}}^2 = \lambda_{r+1} + \ldots + \lambda_d
\]

as in (15).

Thanks to Propositions 2.5 and 2.6 we have that, for a sufficiently regular function \(f \in \mathcal{H}\), the error \(\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}}\) can be controlled by means of the generalized eigenvalues \(\lambda_1, \ldots, \lambda_d\) of the matrix pair \((H, \Sigma^{-1})\) as follows

\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}} \leq \left( \sum_{i=r+1}^{d} \lambda_i \right)^{1/2},
\]

where \(P_r\) is the projector defined as in (16) and

\[
H = \int (\nabla f)^\ast(\nabla f) d\mu.
\]

The matrix pair \((H, \Sigma^{-1})\) provides a test to reveal the low intrinsic dimension of the function \(f\). Indeed, a fast decay in the spectrum of \((H, \Sigma^{-1})\) ensures that \(\sum_{i=r+1}^{d} \lambda_i\) goes quickly to zero with \(r\). In that case, given \(\varepsilon > 0\), there exists \(r(\varepsilon) \ll d\) and a projector \(P_r\) with rank \(r(\varepsilon)\) such that

\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}} \leq \varepsilon.
\]

Notice, however, that a fast decay in the spectrum of \((H, \Sigma^{-1})\) is only a sufficient condition for the low intrinsic dimension: the absence of decay in the \((\lambda_i)\) does not mean that \(f\) cannot be well approximated by \(\mathbb{E}_\mu(f|\sigma(P_r))\) for some low-rank projector \(P_r\).

3. **Contrast with the truncated Karhunen-Loève decomposition.** A simple yet powerful dimension reduction method is the truncated Karhunen-Loève (K-L) decomposition. In the finite-dimensional setting, it consists in reducing the parameter space to the subspace spanned by the leading eigenvectors of the covariance matrix of \(\mu = N(m, \Sigma)\). This approach is based on the observation that

\[
\min_{P_r \in \mathbb{R}^{d \times d} \text{rank-} r \text{ projector}} \mathbb{E}(\|(X - m) - P_r(X - m)\|_2^2) = \sum_{i=r+1}^{d} \sigma_i^2,
\]

where \(X \sim \mu\) and where \(\sigma_i^2\) is the \(i\)-th eigenvalue of \(\Sigma\). If the left-hand side of (17) is small, then the random variable \(X\) can be well approximated (in the \(L^2\) sense) by \(m + P_r(X - m) = P_rX + (I_d - P_r)m\), where \(P_r\) is a solution\(^2\) to (17). In that case, given a function \(f \in \mathcal{H}\), we can hope that \(f(P_rX + (I_d - P_r)m)\) is a good approximation of \(f(X)\). In order to make a quantitative statement, we assume \(f\) is Lipschitz continuous, meaning that there exists a constant \(L \geq 0\) such that

\[
\|f(x) - f(y)\|_V \leq L \|x - y\|_2,
\]

for all \(x, y \in \mathbb{R}^d\). Letting \(g : x \mapsto f(P_rx + (I_d - P_r)m)\), we can write

\[
\|f - g \circ P_r\|_{\mathcal{H}} = \mathbb{E}(\|f(X) - f(P_rX + (I_d - P_r)m)\|_V^2)^{1/2}
\]

\[
\leq L \mathbb{E}(\|X - (P_rX + (I_d - P_r)m)\|_2^2)^{1/2} \overset{(18)}{\leq} L \left( \sum_{i=r+1}^{d} \sigma_i^2 \right)^{1/2}. \tag{19}
\]

\(^2\)Consider the eigendecomposition of \(\Sigma = \sum_{i=1}^{d} \sigma_i^2 u_i u_i^T\). Then the projector \(P_r = \sum_{i=1}^{r} u_i u_i^T\) is a solution to (17).
Thus we have for any $x,y$:

$$\|f - g \circ P_r\|_H \leq \varepsilon,$$

where $\text{rank}(P_r) = r(\varepsilon) \ll d$. In other words, the low intrinsic dimension of a Lipschitz continuous function can be revealed by the spectrum of $\Sigma$. Approximations that exploit this type of low-dimensional structure have been used extensively in forward and inverse uncertainty quantification; see, e.g., [23].

Notice that the function $g : x \mapsto f(P_r x + (I_d - P_r)m)$ considered here does not satisfy $g \circ P_r = \mathbb{E}_\mu(f|\sigma(P_r))$ in general, and therefore is not the optimal choice of profile; see Section 2.1.

**Proposition 3.1.** Let $f \in H = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu; V)$ be a continuously differentiable function and let $P_r$ be a minimizer of $P_r \mapsto \text{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r))$, where $H = f(\nabla f)^\top \nabla f \, d\mu$ and where $\Sigma$ is the covariance matrix of $\mu = \mathcal{N}(m, \Sigma)$. If $f$ is Lipschitz continuous such that (18) holds for some $L \geq 0$, we have

$$\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H \leq \left( \sum_{i=r+1}^d \lambda_i \right)^{1/2} \leq L \left( \sum_{i=r+1}^d \sigma_i^2 \right)^{1/2},$$

where $\sigma_i^2$ and $\lambda_i$ are the $i$-th eigenvalues of $\Sigma$ and of the matrix pair $(H, \Sigma^{-1})$ respectively.

**Proof.** The *trace duality* property allows writing $\text{trace}(AB) \leq \|A\| \text{trace}(B)$ for any symmetric positive-semidefinite matrices $A, B \in \mathbb{R}^{d \times d}$, where $\|A\| = \sup\{ |x^T Ax|, \ x \in \mathbb{R}^d \text{ s.t. } \|x\|_2 = 1 \}$ denotes the spectral norm of $A$. With the choice $A = H$ and $B = (I_d - Q_r)\Sigma(I_d - Q_r)^T$ we can write

$$\text{trace} \left( \Sigma(I_d - Q_r)^T H(I_d - Q_r) \right) = \text{trace} \left( H(I_d - Q_r)\Sigma(I_d - Q_r)^T \right) \leq \|H\| \text{trace} \left( (I_d - Q_r)\Sigma(I_d - Q_r)^T \right) = \|H\| \text{trace} \left( (X - m) - Q_r(X - m) \right)_2^2,$$

for any projector $Q_r$. Let $Q_r$ be a solution to (17) and $P_r$ be a minimizer of $P_r \mapsto \text{trace}(\Sigma(I_d - P_r)^T H(I_d - P_r))$. By Propositions 2.5 and 2.6 we can write

$$\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2 \leq \sum_{i=r+1}^d \lambda_i = \text{trace} \left( \Sigma(I_d - P_r)^T H(I_d - P_r) \right) \leq \|H\| \text{trace} \left( \Sigma(I_d - Q_r)^T H(I_d - Q_r) \right) \leq \|H\| \text{trace} \left( (X - m) - Q_r(X - m) \right)_2^2 = \|H\| \sum_{i=r+1}^d \sigma_i^2.$$

To conclude the proof, it remains to show that $\|H\| \leq L^2$. Since $f$ is Lipschitz we can write

$$\|\nabla f(x)(ty)\|_V = \|f(x + ty) - f(x) + o(\|ty\|_2)\|_V \leq \|f(x + ty) - f(x)\|_V + o(\|ty\|_2) \leq tL\|y\|_2 + o(\|ty\|_2),$$

for any $x, y \in \mathbb{R}^d$ and $t > 0$. Dividing by $t$ and letting $t \to 0$ we obtain $\|\nabla f(x)y\|_V \leq L\|y\|_2$. Thus we have

$$\|H\| = \sup_{y \in \mathbb{R}^d, \|y\|_2 = 1} |y^T H y| = \sup_{y \in \mathbb{R}^d, \|y\|_2 = 1} \int_{\mathbb{R}^d} \|\nabla f(x)y\|_V^2 \mu(dx) \leq \sup_{y \in \mathbb{R}^d, \|y\|_2 = 1} L^2 \|y\|_2^2 = L^2,$$
which concludes the proof.

Similar to the methodology proposed in this paper, the truncated K-L decomposition can be interpreted as a method that minimizes an upper bound of an approximation error; see equation (19). Proposition 3.1 shows that the minimum of the upper bound of the new method is always smaller or equal to that of the truncated K-L. Of course comparing upper bounds does not allow one to make any clear statement about which method performs better than the other. However, note that for the truncated K-L decomposition, the construction of the projector relies only on the covariance matrix \( \Sigma \), whereas the proposed method also takes into account the function \( f \) (through the matrix \( H \)) in the construction of \( P_r \). Thus we can expect the new approach to provide projectors that are better for the approximation of \( f \).

4. Illustrations.

4.1. Analytical examples. We give here three analytical examples for which we can compute a closed-form expression for the error \( \| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\mathcal{H}} \). This allows us to find the projector that minimizes the true error. We then compare this projector with the one that minimizes the upper bound of \( \| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\mathcal{H}} \).

4.1.1. Linear functions. Assume \( f \in \mathcal{H} \) is a linear function and let \( P_r \in \mathbb{R}^{d \times d} \) be a \( \Sigma^{-1} \)-orthogonal projector. By Proposition 2.3 and by linearity of \( f \) we have

\[
\mathbb{E}_\mu(f|\sigma(P_r))(x) = \mathbb{E}(f(P_r x + (I_d - P_r)Y)) = f(P_r x) + f((I_d - P_r)m),
\]

for any \( x \in \mathbb{R}^d \), where we recall that \( m \in \mathbb{R}^d \) is the mean of \( Y \sim \mu = \mathcal{N}(m, \Sigma) \). We can write

\[
\| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\mathcal{H}}^2 = \int_{\mathbb{R}^d} \| f(x) - f(P_r x) - f((I_d - P_r)m) \|_V^2 \mu(dx)
\]

\[
= \int_{\mathbb{R}^d} \| f((I_d - P_r)(x - m)) \|_V^2 \mu(dx)
\]

\[
= \int_{\mathbb{R}^d} (x - m)^T(I_d - P_r)^T f^*(I_d - P_r)(x - m) \mu(dx)
\]

\[
= \text{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r)),
\]

where, for the last equality, we used the relations \( \Sigma = \int_{\mathbb{R}^d} (x - m)(x - m)^T \mu(dx) \) and \( H = \int (\nabla f)^*(\nabla f) \mu = f^* f \), which holds true since \( f \) is linear. Thus we have that equality is attained in (10) for any linear functions \( f \in \mathcal{H} \) and for any \( \Sigma^{-1} \)-orthogonal projector \( P_r \). Therefore minimizing the upper bound is the same as minimizing the true error \( \| f - \mathbb{E}_\mu(f|\sigma(P_r)) \|_{\mathcal{H}} \).

4.1.2. Quadratic forms. Assume \( \mu = \mathcal{N}(0, I_d) \) is the standard normal distribution and let \( f \in \mathcal{H} \) be a quadratic form defined by \( f : x \mapsto \frac{1}{2} x^T A x \) for some symmetric matrix \( A \in \mathbb{R}^{d \times d} \). It is a real-valued function so that \( V = \mathbb{R} \) and \( \| \cdot \|_V = | \cdot | \), the absolute value. Let \( P_r \) be an orthogonal projector with rank \( r \) so that \( P_r^T = P_r \). One can easily check that the relation

\[
f(P_r x + (I_d - P_r)Y) = f(P_r x) + Y^T(I_d - P_r) A P_r x + f((I_d - P_r)Y),
\]
holds for all \( x \in \mathbb{R}^d \) where \( Y \sim \mu \). By taking the expectation with respect to \( Y \), Proposition 2.3 allows writing
\[
\mathbb{E}_\mu(f|\sigma(P_r))(x) = f(P_r x) + \mathbb{E}(f((I_d - P_r)Y)).
\]
The function \( f - \mathbb{E}_\mu(f|\sigma(P_r)) \) is quadratic and can be written as \( x \mapsto x^T \Lambda x + c \) where \( \Lambda = \frac{1}{2}(A - P_r AP_r) \) and \( c = -\mathbb{E}(Y^T \Lambda Y) \). We have
\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}}^2 = \mathbb{E}((Y^T \Lambda Y + c)^2) = \text{Var}(Y^T \Lambda Y).
\]
Consider the eigendecomposition of \( \Lambda = U \text{diag}(a_1, \ldots, a_d) U^T \) and let \( Z = U^T Y \sim \mathcal{N}(0, I_d) \). We have \( Y^T \Lambda Y = \sum_{i=1}^d a_i Z_i^2 \) so that
\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}}^2 = \sum_{i=1}^d a_i^2 \text{Var}(Z_i^2) = 2 \sum_{i=1}^d a_i^2 = 2 \text{trace}(\Lambda^2) = \frac{1}{2} \|A - P_r AP_r\|_F^2,
\]
where \( \| \cdot \|_F = \sqrt{\text{trace}(\cdot^T \cdot)} \) denotes the Frobenius norm. One can show that the rank-\( r \) projector which minimizes \( P_r \mapsto \|A - P_r AP_r\|_F \) is the projector onto the leading eigenspace of \( A^2 \). Denoting by \( \alpha_i^2 \) the \( i \)-th largest eigenvalue of \( A^2 \), we have
\[
\min_{P_r \in \mathbb{R}^{d \times d} \text{ rank-} r \text{ orth. projector}} \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}} = \frac{1}{\sqrt{2}} \left( \sum_{i=r+1}^d \alpha_i^2 \right)^{1/2}.
\]
Now we consider the projector that minimizes the upper bound given by Proposition 2.5. We can write \( \nabla f(x) = Ax \) so that \( H = \int (\nabla f)(\nabla f)^T d\mu = A^2 \). Therefore equation (10) yields
\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}}^2 \leq \text{trace}((I_r - P_r)A^2(I_r - P_r)) = \|A - P_r A\|_F^2,
\]
for any orthogonal projector \( P_r \) with rank \( r \). By Proposition 2.6, the rank-\( r \) orthogonal projector which minimizes the right-hand side in the above inequality is the projector onto the leading eigenspace of \( A^2 \), which is the same as the solution to (20). In other words the minimizer of the upper bound of the error is, for the considered example, the same as the minimizer of the error itself. In addition, the upper bound evaluated at the optimal projector allows controlling the error \( \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_{\mathcal{H}} \) by \( (\sum_{i>r}^d \alpha_i^2)^{1/2} \) which is, up to a factor of \( \sqrt{2} \), the same as the true error.

Remark 4.1. Notice that the same analysis can be carried out for centered normal distributions \( \mu = \mathcal{N}(0, \Sigma) \) with any non-singular covariance \( \Sigma \). To do so, it is sufficient to replace \( A \) by \( \Sigma^{1/2} A \Sigma^{1/2} \) and to consider \( \Sigma^{-1} \)-orthogonal projectors \( P_r \) in the previous analysis.

4.1.3. Sum of sines. Let \( \mu = \mathcal{N}(0, I_d) \) be a standard normal distribution. Consider the real-valued function \( f \in \mathcal{H} \) such that
\[
f : x \mapsto \sum_{i=1}^d a_i \sin(\omega_i x_i),
\]
It is readily seen that \( a \) tends to infinity. This shows that the upper bound is a poor estimator for the error, even if given by Proposition 2.5. Recall that \( H \) with the set \( \Lambda \) where \( \Lambda \) contains the indices of the \( r \) largest values of \( \omega \). For simplicity, we restrict our analysis to the case where \( \omega \) for any \( x \in \mathbb{R}^d \), \( a \in \mathbb{R}^d \) and \( \omega \in \mathbb{R}^d \) are two vectors. Let \( P_r \) be an orthogonal projector. For simplicity, we restrict our analysis to the case where \( P_r \) is a projector onto the span of \( r \) vectors from the canonical basis \( \{e_1, \ldots, e_d\} \) of \( \mathbb{R}^d \), meaning

\[
P_r = \sum_{i \in \Lambda_r} e_i e_i^T, \quad \text{where } \begin{cases} \Lambda_r \subset \{1, \ldots, d\} \\ \#\Lambda_r = r \end{cases}
\]

(21)

It is readily seen that \( \mathbb{E}_\mu(f|\sigma(P_r)) \) is the function \( x \mapsto \sum_{i \in \Lambda_r} a_i \sin(\omega_i x_i) \). We can show that

\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|^2_H = \mathbb{E}\left(\left(\sum_{i \in \Lambda_r^c} a_i \sin(\omega_i x_i)\right)^2\right) = \frac{1}{2} \sum_{i \in \Lambda_r^c} a_i^2 (1 - \exp(-2\omega_i^2)),
\]

where \( \Lambda_r^c \) is the complementary set of \( \Lambda_r \) in \( \{1, \ldots, d\} \) and \( X \sim \mu \). Therefore, the projector \( P_r \) of the form of (21) which minimizes the error \( \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|^2_H \) is the one associated with the set \( \Lambda_r \) containing the indices of the \( r \) largest values of \( a_i^2 (1 - \exp(-2\omega_i^2)) \).

Now we find the projector of the form (21) that minimizes the upper bound of the error given by Proposition 2.5. Recall that \( H = \int (\nabla f)(\nabla f)^T d\mu \), so we can write

\[
\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|^2_H \overset{(10)}{\leq} \text{trace}((I_d - P_r)^T H (I_d - P_r)) \overset{(21)}{=} \sum_{i \in \Lambda_r^c} e_i^T H e_i = \sum_{i \in \Lambda_r^c} \int \left(\frac{\partial f}{\partial x_i}\right)^2 d\mu = \sum_{i \in \Lambda_r^c} \mathbb{E}(a_i \omega_i \cos(\omega_i x_i))^2 = \frac{1}{2} \sum_{i \in \Lambda_r^c} a_i^2 \omega_i^2 (1 + \exp(-2\omega_i^2)).
\]

The projector (21) which minimizes the above upper bound is the one associated with the set \( \Lambda_r \) containing the indices of the \( r \) largest values of \( a_i^2 \omega_i^2 (1 + \exp(-2\omega_i^2)) \).

Let us consider the case where \( \omega_i = \omega \) for all \( i \leq d \). The index sets corresponding to the largest \( a_i^2 \omega^2 (1 + \exp(-2\omega^2)) \) and \( a_i^2 (1 - \exp(-2\omega^2)) \) are the same, and therefore the projector that minimizes the upper bound is the same as the minimizer of the true error. Notice, however, that when \( \omega \to \infty \) the true error tends to \( \frac{1}{2} \sum_{i \in \Lambda_r} a_i^2 \), whereas the upper bound tends to infinity. This shows that the upper bound is a poor estimator for the error, even if its minimization allows recovering the optimal projector.

Suppose now that \( \omega_i = a_i^{-2} \geq 1 \) for all \( i \leq d \). Then the index set corresponding to the largest \( a_i^2 \omega_i^2 (1 + \exp(-2\omega_i^2)) = \omega_i (1 + \exp(-2\omega_i^2)) \) is the same as the index set of the smallest \( a_i^2 (1 - \exp(-2\omega_i^2)) = \omega_i^{-1} (1 - \exp(-2\omega_i^2)) =: h_2(\omega_i) \). Indeed \( h_1 \) is increasing on \((1, \infty)\) whereas \( h_2 \) is decreasing. Hence, for this particular example, minimizing the upper bound yields the worst possible projector, i.e., the one that maximizes the true error.
4.2. Elliptic PDE. Consider the diffusion equation on the square-shaped domain \( \Omega = [0,1]^2 \), which consists in finding \( u \in H^1(\Omega) \) such that

\[
\begin{aligned}
\nabla_s (\kappa \nabla_s u) &= 0 & \text{in } \Omega, \\
u &= s_1 + s_2 & \text{on } \partial \Omega.
\end{aligned}
\]

Here \( s = (s_1, s_2) \in \Omega \) denotes the spatial coordinates and \( \nabla_s \) refers to the gradient in the spatial variable \( s \). The diffusion coefficient \( \kappa \) is a random field and follows a log-normal distribution such that \( \log(\kappa) \) is a Gaussian process on \( \Omega \) with zero mean and with a covariance function \( c: \Omega \times \Omega \to \mathbb{R} \) defined by \( c(s,t) = \exp(-\|s-t\|^2/(0.15)^2) \) for all \( s,t \in \Omega \). A numerical approximation of (22) is obtained with the finite element method (FEM); see, for example, [10]. The diffusion field \( \kappa \) is approximated by the piecewise constant random field

\[
\kappa(x) : s \mapsto \exp\left( \sum_{i=1}^d x_i \mathbf{1}_i(s) \right),
\]

where \( \mathbf{1}_i \) denotes the indicator function associated with the \( i \)th element of the mesh represented in Figure 1a. Here \( d = 3252 \) corresponds to the number of elements, and \( x \sim \mu = \mathcal{N}(0, \Sigma) \) with

\[
\Sigma_{i,j} = c(s_i, s_j), \quad 1 \leq i, j \leq d,
\]

and \( s_i \) being the center of the \( i \)th element. With a slight abuse of notation, we denote by \( u(x) \) the Galerkin projection of the solution to (22) onto the space of continuous piecewise affine functions associated with the mesh in Figure 1a. We consider the following scenarios, where the function \( f: \mathbb{R}^d \to V \) is defined by three different post-solution treatments of \( u(x) \):

1. \( f: x \mapsto u(x) \), which means that \( f \) is the solution map from the parameter \( x \) to the FEM solution to (22). In that case \( V \) is the FEM approximation space with dimension \( \dim(V) = 1691 \), the number of nodes in the mesh. Since \( V \subset H^1(\Omega) \), the natural choice for the norm \( \| \cdot \|_V \) is

\[
\|v\|_V^2 = \int_\Omega (v(s))^2 \, ds + \int_\Omega \|\nabla_s v(s)\|_2^2 \, ds.
\]

2. \( f: x \mapsto u|_{\Omega_s}(x) \), where \( \Omega_s = [0.35, 0.65]^2 \subset \Omega \). In other words, \( f(x) \) corresponds to the restriction of \( u(x) \) to a subdomain \( \Omega_s \) of \( \Omega \). For this scenario, \( V \subset H^1(\Omega_s) \) is of dimension 168 (the number of nodes in the mesh) and is endowed with the norm \( \| \cdot \|_V \) given by

\[
\|v\|_V^2 = \int_{\Omega_s} (v(s))^2 \, ds + \int_{\Omega_s} \|\nabla_s v(s)\|_2^2 \, ds.
\]

3. \( f: x \mapsto (u|_{s_a}(x), u|_{s_b}(x)) \), where \( s_a = (0.2, 0.8) \in \Omega \) and \( s_b = (0.8, 0.2) \in \Omega \). In this scenario, we are interested in the evaluation of the solution \( u(x) \) at two different spatial locations \( s_a \) and \( s_b \). There are two scalar-valued outputs so that \( V = \mathbb{R}^2 \) is an algebraic space. Consider the weighted norm \( \| \cdot \|_V \) defined by

\[
\|v\|_V^2 = \alpha v_1^2 + \beta v_2^2,
\]

where \( \alpha, \beta \geq 0 \) and \( \alpha + \beta = 1 \).
4.2.1. Computational aspects. We consider the problem of computing the matrix $H = f(\nabla f)^* (\nabla f) \, d\mu$. Since $H = \mathbb{E}((\nabla f(X))^* (\nabla f(X)))$, with $X \sim \mu$, $H$ can be approximated by the $K$-sample Monte-Carlo estimate

\begin{equation}
\hat{H} = \frac{1}{K} \sum_{i=1}^{K} (\nabla f(X_i))^* (\nabla f(X_i)),
\end{equation}

where $X_1, \ldots, X_K$ are independent copies of $X$. To numerically compute a realization of $\hat{H}$, one needs to evaluate the Jacobian of the function $f$ $K$ times. To do so, we employ the the adjoint method; see for example [28]. Then, to construct the projector, instead of minimizing

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Scenario 1 & Scenario 2 & Scenario 3 \\
\hline
$f(x) = \begin{array}{c}
\end{array}$ & $f(x) = \begin{array}{c}
\end{array}$ & $f(x) = \begin{pmatrix} 0.939 \\ 1.032 \end{pmatrix}$ \\
\hline
\end{tabular}
\end{table}

Figure 1: Illustration of the Elliptic PDE problem: geometry and mesh (Figure 1a), representation of the diffusion field associated with a parameter $x \in \mathbb{R}^d$ drawn randomly from $\mu$ (Figure 1b), corresponding solution (Figure 1c) and representation of $f(x)$ for the three different scenarios given this particular $x$ (Figure 1d).
trace(\(\Sigma(I_d - P_r^T)H(I_d - P_r)\)) we consider a projector \(\hat{P}_r\) such that

\[
\hat{P}_r \in \arg \min_{P_r \in \mathbb{R}^{d \times d} \ \text{rank}-r \ \text{projector}} \text{trace}\left(\Sigma(I_d - P_r^T)\hat{H}(I_d - P_r)\right).
\]

By construction, \(\hat{P}_r\) depends upon \(\hat{H}\), and thus it is random. Recall that such a projector can be obtained by computing the generalized eigendecomposition of the matrix pair \((\hat{H}, \Sigma^{-1})\); see Proposition 2.6.

To approximate the conditional expectation \(E_\mu(f|\sigma(\hat{P}_r))\), we consider the random function

\[
\hat{F}_r : x \mapsto \frac{1}{M} \sum_{i=1}^{M} f(\hat{P}_r x + (I_d - \hat{P}_r)Y_i),
\]

where \(Y_1, \ldots, Y_M\) are independent copies of \(Y \sim \mu\). Given a realization of the projector \(\hat{P}_r\), a realization of \(\hat{F}_r\) can be obtained by drawing \(M\) samples of \(Y\) and by using those samples to evaluate \(\hat{F}_r\) using (26). Notice that the samples are not redrawn for each new evaluation point \(x\) of \(\hat{F}_r\). By Proposition 2.3 and for any \(x \in \mathbb{R}^d\), \(\hat{F}_r(x)\) can be interpreted as an \(M\)-sample Monte Carlo approximation of \(E_\mu(f|\sigma(\hat{P}_r))(x)\). Finally, notice that if \(M = 1 \) and \(Y_1 = 0\) (i.e., the mean of \(Y\)), then our approximation of \(f\) reduces to the form used in Section 3 when truncating a K-L decomposition, albeit for a different projector; see relation (19) with \(m = 0\) and \(P_r = \hat{P}_r\).

### 4.2.2. Modes and influence of the norm \(\|\cdot\|_V\).

For each scenario, an approximation \(\hat{H}\) of \(H\) is computed with a large number of samples, \(K = 10^4\). This approximation is considered sufficiently accurate and will be used in place of \(H\) unless specified otherwise. Figure 2 illustrates the leading generalized eigenvectors of the matrix pair \((H, \Sigma^{-1})\) as well as the leading eigenvectors of \(\Sigma\), meaning the K-L modes; see Section 3. Since they do not depend upon \(f\), the K-L modes do not have any particular relation to the elliptic PDE solution other than some symmetry properties related to the shape of the domain \(\Omega\). In contrast, the modes associated with the three scenarios present specific features which depend on the function \(f\). For example with scenario 2, we observe that the modes in the parameter space somehow represent more information local to the region of interest \(\Omega_s\).

The choice of the norm \(\|\cdot\|_V\) also impacts the generalized eigenvectors of \((H, \Sigma^{-1})\) through the matrix \(H\). For instance, with scenario 3, relation (14) with \(M = \text{diag}(\alpha, \beta)\) allows us to write

\[
H = \alpha H_1 + \beta H_2, \quad \text{with} \quad H_i = \int (\nabla f_i)^T(\nabla f_i)dx.
\]

With the choice \(\alpha = \beta = 1\), the modes in Figure 2 suggest that the two points of interest \(s_a\) and \(s_b\) are considered equally important, whereas the choice \(\alpha = 10\) and \(\beta = 1\) leads to significantly more patterns around point \(s_a\) (on the top-left of \(\Omega\)) than around point \(s_b\).

### 4.2.3. Approximating the conditional expectation and comparison with K-L.

Assume the matrix \(H\) is known (again, an approximation \(\hat{H}\) with \(K = 10^4\) samples is used in place of \(H\)) and let \(P_r\) be the rank-\(r\) projector which minimizes \(\text{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r))\). We consider the approximation \(\hat{F}_r\) of the conditional expectation \(E_\mu(f|\sigma(P_r))\) given by (26) with
Figure 2: Parameter modes: each figure represents the function $s \mapsto \sum_{i=1}^{d} v_i 1_i(s)$ for different $v \in \mathbb{R}^d$, where $1_i$ is the indicator function of the $i$-th element of the mesh. In the first row (K-L) $v$ is the $i$-th eigenvector of $\Sigma$, which corresponds to the Karhunen-Loève modes. In the four other rows, $v$ is the $i$-th generalized eigenvector of the matrix pair $(H, \Sigma^{-1})$, for different $H$ depending on the scenario.

$\hat{P}_r = P_r$. Figure 3 shows the error $\|f - \hat{F}_r\|_H$ as a function of the rank $r$ of the projector. For each scenario, one realization of $\hat{F}_r$ is computed with either $M = 1$, $M = 5$, or $M = 20$ samples. We first note that, since we do not exactly compute the conditional expectation, the errors (dotted curves) are sometimes above the upper bound (solid red curves). In this inexact setting, $\text{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r))^{1/2}$ is no longer a certified upper bound for the error. However we observe it can still be used as a good error indicator.

The three scenarios do not have the same convergence rate with $r$: the first scenario has the slowest and the third the fastest. Even though they are different post-solution treatments of the same solution map $x \mapsto u(x)$, the functions $f$ associated with each scenario do not have the same complexity in terms of intrinsic dimension.
Interestingly, increasing $M$ does not lead to significant improvements of the approximation. This phenomenon can be explained by the following relation,

$$
\mathbb{E}(\|f - \hat{F}_r\|_H^2) \overset{\text{(4)}}{=} \mathbb{E}(\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2 + \|\hat{F}_r - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2) \\
\overset{\text{(20)}}{=} \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2 + \frac{1}{M} \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2 \\
= \left(1 + \frac{1}{M}\right) \|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2,
$$

where the expectation is taken over the samples $Y_1, \ldots, Y_M$ (the projector $P_r$ being fixed here). This result shows that even with small $M$, one can still hope to obtain a good approximation $\hat{F}_r$ of $f$ provided $P_r$ is chosen such that $\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H$ is sufficiently small. In other words a crude approximation of the conditional expectation yields at most a factor of two (when $M = 1$) in the expected error squared, so that it remains of the same order of magnitude as $\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H^2$; see also Theorem 3.2 from [7].

We now compare with the truncated Karhunen-Loève decomposition, for which $P_r$ is defined as the rank-$r$ orthogonal projector onto the leading eigenspace of the covariance matrix $\Sigma$. The black dash-dotted curves in Figure 3 represent the upper bound trace($\Sigma(I_d - P_r^T)H(I_d - P_r^T))^{1/2}$ for this choice of $P_r$, as a function of $r$. (The true error $\|f - \mathbb{E}_\mu(f|\sigma(P_r))\|_H$ is substantially the same as its upper bound, so we decided not to plot it.) It is interesting to see that in the first scenario, the K-L projector is essentially as effective as the projector obtained by minimizing the upper bound. As shown in Figure 4, the spectrum of $H$ is flat, which means that $H$ is close to a rescaled identity matrix. Then, minimizing trace($\Sigma(I_d - P_r^T)H(I_d - P_r^T))$ is nearly the same as minimizing trace($\Sigma(I_d - P_r)\Sigma(I_d - P_r^T) = \mathbb{E}(\|X - P_rX\|_2^2)$, where $X \sim \mathcal{N}(0, \Sigma)$, and yields the same projector as the truncated K-L method; see (17). However, this reasoning does not apply to scenarios 2 and 3, where the spectrum of $H$ decays rapidly. For these scenarios we observe in Figure 3 that the new method outperforms the truncated K-L method. For instance, in scenario 2 the new method reaches an error of $10^{-4}$ with only $r = 150$ whereas the truncated K-L method requires $r = 300$.

**4.2.4. Quality of the projector.** In this section we assess the quality of a projector $\hat{P}_r$ defined by (25), where $\hat{H}$ is the $K$-sample Monte Carlo approximation of $H$ given by (24). In the present context, an optimal projector would be a minimizer of $P_r \mapsto \text{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r^T))^{1/2}$ so that the only relevant criteria for the quality of $\hat{P}_r$ is how close $\text{trace}(\Sigma(I_d - \hat{P}_r^T)H(I_d - \hat{P}_r^T))^{1/2}$ is to the minimum of the upper bound. Figure 5 contains two sets of curves: the solid curves represent the error bound trace($\Sigma(I_d - \hat{P}_r^T)H(I_d - \hat{P}_r^T))^{1/2}$ as a function of the rank of $\hat{P}_r$, whereas the dotted curves correspond to the approximate error bound trace($\Sigma(I_d - \hat{P}_r^T)\hat{H}(I_d - \hat{P}_r^T))^{1/2}$. This approximate error bound is the quantity we would use in place of the error bound when the matrix $H$ is not known. For each scenario we observe that for small $K$, the approximate error bound underestimates the true error bound. This means that trace($\Sigma(I_d - \hat{P}_r^T)\hat{H}(I_d - \hat{P}_r^T))^{1/2}$ can be used as an error estimator only if $K$ is sufficient large.

Observe in Figure 5 that scenarios 1 and 2 need fewer samples to obtain a good projector (say around $K = 30$ samples) compared to the last scenario (at least $K = 400$ samples). To understand this result, let us note that if $r$ is larger than the rank of $\hat{H}$, the projector $\hat{P}_r$ is not
Figure 3: Error $\|f - \tilde{F}_r\|_H$ as a function of the rank of $P_r$. The error $\|f - \tilde{F}_r\|_H = \mathbb{E}(\|f(X) - \tilde{F}_r(X)\|^2_{V})^{1/2}$, $X \sim \mu$, is estimated via Monte Carlo with 300 samples for $X$. The red (solid) and black (dash-dot) lines represent the upper bound $\text{trace}(\Sigma(\tilde{I} - P_r H(I_d - P_r))^1/2$ with $P_r$ defined either as the minimizer of the upper bound (red lines) or as the projector onto the leading eigenspace of $\Sigma$ (black lines).

Figure 4: Spectrum of $H$ for the three scenarios (left) and spectrum of $\Sigma$ (right).

uniquely determined: any $\tilde{P}_r$ such that $\text{Im}(\tilde{H}) \subset \text{Im}(\tilde{P}_r)$ is a solution to (25). Therefore the rank of $\tilde{P}_r$ should not exceed that of $\tilde{H}$ which, thanks to (24), satisfies the following relation

$$\text{rank}(\tilde{H}) \leq K \text{rank}((\nabla f(X))^{*}(\nabla f(X))) \leq K \text{dim}(V).$$

With scenario 3 we have $\text{dim}(V) = 2$ so that the rank of $\tilde{P}_r$ should not exceed $2K$. This limitation is represented by the vertical lines on Figure 5c. With scenarios 1 and 2 we have
Figure 5: Error bound trace(Σ(I_d − \tilde{P}_r^T)H(I_d − \tilde{P}_r))^{1/2} (solid curves) and approximate error bound trace(\Sigma(I_d − \tilde{P}_r^T)\tilde{H}(I_d − \tilde{P}_r))^{1/2} (dotted curves) as a function of the rank of \tilde{P}_r. For each scenario, the curves correspond to one realization of \tilde{H} and \tilde{P}_r defined by (24) and (25) for different values of K. In Figure 5c, the vertical lines correspond to r = 2K.

5. Conclusions. We have addressed the problem of approximating multivariate functions taking values in a vector space. We approximate such functions by means of ridge functions that depend only on a few linear combinations of the input parameters. Our approach exploits structure in the original function, in the (Gaussian) measure on the domain, and in the chosen norm on the codomain. Rather than seeking an optimal approximation, we build a controlled approximation: we develop an upper bound on the approximation error and minimize this upper bound.

Our analytical and numerical examples demonstrate good performance of the method, and also illustrate conditions under which it might not work well. For example, we show cases where minimizing the upper bound leads to an optimal approximation, and contrasting cases where the error bound is not tight. Numerical demonstrations on an elliptic PDE also illustrate various computational issues: sampling to compute both the projector (yielding the important directions) and the conditional expectation (yielding the ridge profile). We also contrast our new approach with the K-L expansion: our approach is in general more efficient, but when the function itself provides little low-dimensional structure, we essentially revert to the dimension reduction provided by a low-rank approximation of the input covariance.

Future work may explore several natural extensions of the proposed methodology. First
is the extension to non-Gaussian input measures, e.g., uniform measure on bounded domains in \( \mathbb{R}^d \). Second is the extension to infinite-dimensional input spaces: for example, letting the domain of \( f \) be a function space endowed with Gaussian measure or Besov measure. Finally, it may be possible to develop sharper error bounds based on higher-order derivatives, e.g., Hessians of \( f \). For the last two points, we may be able to use recent results on higher-order Poincaré inequalities [26]. Another open issue is the existence of a best approximation, i.e., a ridge function that minimizes the \( L^2 \) error. To our knowledge, this problem has not been fully addressed.

Acknowledgments. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. O. Zahm and Y. Marzouk also gratefully acknowledge support from the DARPA EQUiPS program.

REFERENCES


