# Active Subspaces in Infinite Dimension

**Poorbita Kundu**
Public Health Sciences Division
Fred Hutchinson Cancer Center

**Nathan Wycoff**
Department of Mathematics and Statistics
University of Massachusetts Amherst

## Abstract

Active subspace analysis uses the leading eigenspace of the gradient's second moment to conduct supervised dimension reduction. In this article, we extend this methodology to real-valued functionals on Hilbert space. We define an operator which coincides with the active subspace matrix when applied to a Euclidean space. We show that many of the desirable properties of Active Subspace analysis extend directly to the infinite dimensional setting. We also propose a Monte Carlo procedure and discuss its convergence properties. Finally, we deploy this methodology to create visualizations and improve modeling and optimization on complex test problems.

## 1 INTRODUCTION

The increasing availability of computational resources has allowed for significantly more sophisticated models in science and engineering. But this increase in complexity is not always accompanied by a commensurate increase in understanding. Often, these models are parameterized by a set of variables governing the behavior of the simulation. When there are many such parameters, a *sensitivity analysis* (Razavi et al., 2021) can be key to unlocking knowledge of which are most important in terms of driving the system. A linear sensitivity analysis in particular looks for important linear combinations of variables; this can advance understanding of even complex systems where all variables play an important role.

*Gradient-based sensitivity analysis* (Peter and Dwight, 2010) uses the derivative of a target system's behavior with respect to its parameters to determine importance. Active subspace analysis (Constantine, 2015) is the most prominent linear gradient-based method, which given a differentiable function $f : \mathcal{X} \subseteq \mathbb{R}^D \to \mathbb{R}$, proceeds by defining an *active subspace matrix*:

$$\mathbf{C} = \mathbb{E}_\rho \left[ \nabla f(X) \nabla f(X)^\top \right] \in \mathbb{R}^{D \times D}, \qquad (1)$$

where $X$ is a $\mathcal{X}$-valued random variable with probability law $\rho$. In conducting an eigenanalysis on $\mathbf{C}$, we have the opportunity to discover some subspace of $\mathbb{R}^D$ which captures most or even all of $f$'s variation. For many practical functions, the gradient is restricted primarily to some lower dimensional subspace (Constantine et al., 2016). The strongest such circumstance is that of a *ridge function*, which has the form $f(x) = \tilde{f}(\mathbf{A}x)$ for some $\mathbf{A} \in \mathbb{R}^{R \times D}$ with $R < D$.

The active subspace method has found many applications across science and engineering (Kim et al., 2024; Grey and Constantine, 2018; Khatamsaz et al., 2021; Ji et al., 2019; Constantine et al., 2015; Bittner et al., 2021). In part, this is due to the proliferation of computing environments which allow for automatic calculation of gradients with respect to model parameters, including deep learning frameworks (e.g. Tensorflow (Abadi et al., 2015), Pytorch (Paszke et al., 2019), JAX (Bradbury et al., 2018)) and scientific computing packages (e.g. FEniCS (Alnæs et al., 2015), custEM (Rochlitz et al., 2019)). These allows for the computation of gradients with respect to parameters in arbitrarily high dimension, and even infinite dimensional spaces in the case of e.g. FEniCS.

The computational capabilities of software systems therefore appear to have run ahead of the statistical methodology and mathematical understanding of the active subspace method. The purpose of this article is to rectify this. In particular, our primary contributions are as follows:

1. We define an extended Active Subspace analysis for real-valued functionals on Hilbert spaces.
2. We show that many celebrated properties of the active subspace extend to the functional context.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

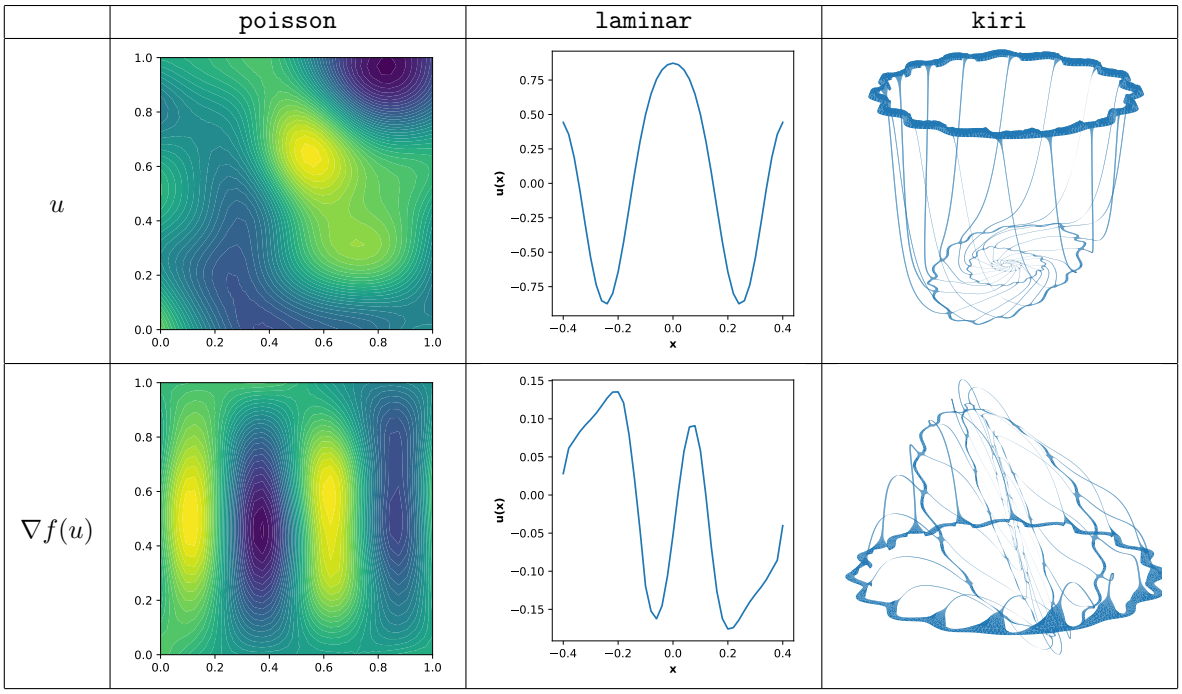


Figure 1: Random input functions (top row) and corresponding gradient functions (bottom row).

3. We propose a computable Monte-Carlo estimator and establish its convergence.

4. Sophisticated case studies show how to use this active subspace to improve practical analyses.

5. We release an open-source code-base building on FEniCS that implements this methodology [1].

## 2 MOTIVATION AND BACKGROUND

We begin this section by reviewing related existing work in infinite dimensional sensitivity analysis. Next, we provide this article's *raison d'être*: since we never represent functions exactly on a computer, but must instead discretize them, a natural objection to this article's project is that we can always first discretize the function and then apply the active subspace method to the discretization. We will demonstrate in a moment why this is insufficient. Finally, we provide an overview of the class of test functions, optimization of PDE parameters, which motivated our methodology.

### 2.1 SENSITIVITY ANALYSIS IN INFINITE DIMENSION

Though active subspaces themselves have not previously been applied to infinite dimensional problems to the best of our knowledge, we now review existing work on sensitivity analysis more broadly in the infinite dimensional setting. The procedure in this article will be closely related to functional PCA (Ramsay and Silverman, 2005), by analogy to the close connection between PCA and the active subspace method in the finite dimensional setting. In the context of finance, sensitivity analysis is commonly used under the name "the Greeks"; Benth et al. (2021) extended these to the infinite dimensional setting. Sobol indices, an alternative to gradient-based sensitivity, have been studied in infinite dimension (e.g. Iooss and Ribatet, 2009; Gamboa et al., 2014). These are most useful in the observational context, when gradient information is not available. We will in this article rather be concerned with the analysis of a computational model for which gradients may be computed. Wu et al. (2023) proposed an interesting technique for optimal Bayesian experimental design over PDE parameters and Beskos et al. (2017) propose using gradients to accelerate Bayesian computation.

### 2.2 THE NONCOMMUTATIVITY OF DISCRETIZATION

Diez and Serani (2024) review sensitivity analysis with respect to functions using the perfectly reasonable idea to first discretize the functions using a finite parameter vector $\theta \in \mathbb{R}^P$ and subsequently to view the objective function as being defined with respect to $\theta$, upon which standard finite dimensional methods

[1] `https://github.com/NathanWycoff/inf_active`

can be applied. Though useful, this is distinct from performing sensitivity analysis directly in the function space, which we now illustrate by example. Let $u(\theta) := \theta_1 h_1 + (1-\theta_1)h_2 + \theta_2 h_3$, where $h_1, h_2, h_3$ is a dictionary of three functions; maybe they're sin, cos, and exp. Now say that our objective functional is: $f = \langle \cdot, h_1 \rangle + \langle \cdot, h_2 \rangle$. According to a standard active subspace analysis with respect to $\theta$, there is only one important dimension: $\theta_1$. By contrast, the functional active subspace analysis with respect to $u$ as we will introduce in this article would give that $\text{span}(\{h_1, h_2\})$ is the two dimensional active subspace (assuming $h_1 \neq a h_2 \ \forall a \in \mathbb{R}$). Since the vector space dimension of these two active subspaces disagree, there cannot be an isomorphism between them: they are truly different objects. A similar inequivalence is present in the infinite dimensional optimization literature, where it is well known that discretizing then optimizing may lead to different results than optimizing then discretizing (e.g. Liu and Wang, 2019; Gholami et al., 2019; Onken and Ruthotto, 2020).

There are also some nice practical advantages to working with the active subspace directly in the function space. For one, we obtain eigenfunctions rather than eigenvectors. If the functions residing in the Hilbert space are defined on a low dimensional domain, they can be directly visualized in a plot; by contrast, interpreting an eigenvector beyond simply noting the magnitude of its largest elements can be difficult. Additionally, staying in the function space means we can make comparisons across different mesh sizes or otherwise across discretization schemes, which is important in adaptive meshes (Nochetto et al., 2009), or in any other circumstance where the number and meaning of the parameters in the discretization can vary.

## 2.3 PDE-CONSTRAINED OPTIMIZATION

Though the active subspace methodology we develop is generally applicable, we will in this article primarily consider applications in applied math, and especially in modeling with differential equations. Function-valued parameters can be of interest in such contexts in various ways. Sometimes, they are equated to a differential operator in a differential state equation of the form $\mathbb{D}v = l$, where $\mathbb{D}$ is some differential operator. When $l$ does not depend on $v$, it is called a *forcing function*. A function may also specify an inhomogeneous Dirichlet or Neumann boundary condition. These may be of interest to develop an optimal controller (Evans, 2024) or to solve an inverse problem (Vogel, 2002). We next introduce three such problems from the applied engineering literature which serve as our example real-valued functionals, the sensitivity of which we want to compute. We give additional details in Appendix L.2.

### 2.3.1 DISTRIBUTED POISSON CONTROL

As an illustration, consider a Poisson Distributed Control problem (see Figure 1, left). $\mathcal{X} = (0,1) \times (0,1)$ with homogeneous Dirichlet boundary conditions. Given a desired state $v_d(x,y) = \sin(4\pi x)\sin(\pi y)$ and regularization parameter $\alpha > 0$, the suitability of a candidate control function $m$ is is given by:

$$J(m) = \frac{1}{2}\int_{\mathcal{D}} (v - v_d)^2 \, \mathrm{d}x + \frac{\alpha}{2}\int_{\mathcal{D}} \mathrm{d}x, \qquad (2)$$

subject to the state equation:

$$-\Delta v \;=\; m \quad \text{in } \mathcal{X}, \qquad v \;=\; 0 \quad \text{on } \partial\mathcal{X}\,. \qquad (3)$$

Given a candidate setting $m$, the function $J$ measures its suitability for driving $v$ toward $v_d$ subject to $L^2$ regularization. Modern computational platforms like FEniCS (Alnæs et al., 2015; Logg et al., 2012) allow for nearly exact calculation of functional gradients of $J$ at any value of $m$ under a variety of discretization schemes.

### 2.3.2 LAMINAR JET

Following Klein et al. (2003); Beskos et al. (2017), we study the problem of reconstructing the flow of a non-reacting laminar jet at the inlet boundary given 41 measurements on the outflow boundary. This involves solving a Navier-Stokes equation over a 2D rectangular region to evaluate the resultant pressure and velocity fields. In this problem the input function gives the 1D inflow profile of the jet, the forward model is run, and predictions at the boundaries are compared with ground truth data using mean squared error (see Figure 1, middle).

### 2.3.3 KIRIGAMI ELECTRONICS

Yang et al. (2024) propose the idea of *Kirigami Electronics*, where Kirigami is, like Origami, a Japanese art form involving folded paper, but which also places cuts in the paper. By shaping electronics in this manner, they unfurl into complex shapes based on their environment. To develop their technology, Yang et al. (2024) have developed a sophisticated computational model to determine what shape their intricate geometry will obtain in the steady-state balancing elastic energy and the load. This problem involves specifying a 3D function over an irregular domain, and the cost function to be minimized is the energy of the configuration (see Figure 1, right).

# 3 ACTIVE SUBSPACES IN HILBERT SPACE

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real, separable Hilbert space and let $f : \Omega \subset \mathcal{H} \to \mathbb{R}$ be a real-valued functional defined on an open set $\Omega$. We use extensively the notion of a derivative on a Hilbert space in this article by way of Fréchet and Gâteaux differentiation. Intuition is most readily gained when $\mathcal{H}$ contains real-valued functions $u$ defined over some domain $\mathcal{D}$. Then, the functional derivative $Df(u)$ may be identified with a member of $\mathcal{H}$, i.e., viewed as a real-valued function on $\mathcal{D}$. As with a classical derivative, evaluating it at some element $x \in \mathcal{D}$ gives in some sense the infinitesimal change of $f(u)$ as $u(x)$ is varied. We now make this precise.

**Definition 1.** *Fix $u \in \Omega$. If $f$ is Fréchet differentiable at $u$, its derivative $Df(u) : \mathcal{H} \to \mathbb{R}$ is a bounded (continuous) linear functional. By the Riesz representation theorem (Fréchet, 1907; Riesz, 1907; Rudin, 1991), there exists a* unique *$\nabla f : \mathcal{H} \to \mathcal{H}$ such that for all $u \in \mathcal{H}$, $\nabla f(u) \in \mathcal{H}$ and*

$$Df(u)[h] \;=\; \langle h,\, \nabla f(u)\rangle \qquad \text{for all } h \in \mathcal{H}. \tag{4}$$

*We call $\nabla f(u)$ the (Fréchet) gradient of $f$ at $u$.*

Basic properties of $\nabla f$ are reviewed in Appendix A.

The integrand in the active subspace matrix definition is typically given as a vector times its transpose: $\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top$. Taking the linear map perspective on matrices, we see that, for any vector $\mathbf{a}$: $\nabla f(\mathbf{x})\nabla f(\mathbf{x})^\top \mathbf{a} = \langle \nabla f(\mathbf{x}), \mathbf{a}\rangle \nabla f(\mathbf{x})$. This perspective immediately suggests an extension to the general Hilbert space setting. We will denote this tensor product by $(a \otimes b)\, h := \langle h, b\rangle\, a$.

**Definition 2.** *Let $U$ be an $\mathcal{H}$-valued random variable with law $\rho$ on $\mathcal{H}$, and assume $\mathbb{E}\left[\|\nabla f(U)\|^2\right] < \infty$. Define the* active subspace operator *as:*

$$\mathcal{C} \;:=\; \mathbb{E}\big[\nabla f(U) \otimes \nabla f(U)\big] \;\;:\;\; \mathcal{H} \to \mathcal{H}. \tag{5}$$

In Appendix B, we show that $\mathcal{C}$ is well defined as a Bochner expectation and is trace-class (and thus compact), self-adjoint, and positive semidefinite. It therefore is governed by the Spectral Theorem.

**Proposition 1** (Spectral decomposition of $\mathcal{C}$)**.** *There exists orthonormal eigenfunctions $\{w_i\}_{i\geq 1} \subset \mathcal{H}$ and a nonincreasing sequence $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ such that*

1. $\mathcal{C}w_i = \lambda_i w_i$,
2. $\operatorname{trace}(\mathcal{C}) = \sum_{i\geq 1} \lambda_i$.
3. $\mathcal{C} = \sum_{i\geq 1} \lambda_i\, w_i \otimes w_i$
4. $\mathbb{E}\Big[\big(Df(U)[w_i]\big)^2\Big] = \lambda_i$

*Proof.* Appendix C □

Practically, the spectrum of $\mathcal{C}$ quantifies how strongly the functional varies along different directions in $\mathcal{H}$. In particular, each $\lambda_i$ equals the average squared directional derivative of $f$ along the corresponding eigenfunction $w_i$, so larger $\lambda_i$ indicate directions with greater average variability of $f$. In other words, large eigenvalues correspond to directions along which perturbations in the input function produce substantial average change in $f$, whereas small eigenvalues indicate directions with relatively weaker influence.

Thus we have provided an extension of the active subspace matrix. Next, we give our definition of the active subspace itself.

**Definition 3.** *Let the $n$-dimensional active subspace be given by $\mathcal{A}_n := \operatorname{span}\{w_1, \ldots, w_n\}$.*

The value of $n$ is so chosen that the leading eigenvalues $\lambda_1, \ldots, \lambda_n$ account for the dominant share of the spectrum. With the orthogonal decomposition $\mathcal{H} = \mathcal{A}_n \oplus \mathcal{A}_n^\perp$, projecting $u \in \mathcal{H}$ onto $\mathcal{A}_n$ retains the most informative coordinates (on average) and discards the insensitive ones.

With the active subspace in hand, it is very useful to define a surrogate function which depends only on the active coordinates. To this end, for every $u \in \mathcal{H}$,

$$f(u) \approx \tilde{f}\big(P_{\mathcal{A}_n} u\big) = \tilde{f}\Big(\sum_{i=1}^{n} \langle u, w_i\rangle\, w_i\Big), \tag{6}$$

where $P_{\mathcal{A}_n}$ be the orthogonal projector onto $\mathcal{A}_n$.

We can extend the notion of a ridge function to infinite dimension. In this case, $f(u) = \tilde{f}(\mathcal{M}u)$, where $\mathcal{M} : \mathcal{H} \to \mathbb{R}^R$ is a linear operator. In this case, $\mathcal{C}$ has rank at most $R$, its nonzero eigenfunctions span the range of $\mathcal{M}$, and $\mathcal{A}_n = \operatorname{Range}(\mathcal{M})$ captures all gradient energy.

The following result is a Hilbert-space analogue of Proposition 2.3 in Constantine et al. (2014), which describes how ridge structure is reflected in the active subspace operator.

**Theorem 1.** *Fix $n \in \mathbb{N}$. Let $\mathcal{A}_n := \operatorname{span}\{w_1, \ldots, w_n\}$ and let $P_{\mathcal{A}_n}$ denote the corresponding orthogonal projector. Assume the trailing spectrum vanishes:*

$$\lambda_{n+1} = \lambda_{n+2} = \cdots = 0. \tag{7}$$

*Then the following hold.*

*(i) (*Gradient lies in the active subspace*) For all $u \in \mathcal{H}$, $P_{\mathcal{A}_n^\perp} \nabla f(u) = 0$ (equivalently, $\nabla f(u) \in \mathcal{A}_n$ everywhere).*

*(ii) (*Level-set invariance*) If $u_1, u_2 \in \mathcal{H}$ satisfy $P_{\mathcal{A}_n} u_1 = P_{\mathcal{A}_n} u_2$, then $f(u_1) = f(u_2)$.*

*(iii)* (Gradient equality) *Under the same hypothesis,* $\nabla f(u_1) = \nabla f(u_2)$.

*(iv)* (Factorization through $P_{\mathcal{A}_n}$) *There exists a* $C^1$ *map* $\tilde{f} : P_{\mathcal{A}_n}(\mathcal{H}) \to \mathbb{R}$ *such that* $f(u) = \tilde{f}(P_{\mathcal{A}_n} u)$ *for all* $u \in \mathcal{H}$, *and* $\nabla \tilde{f}(P_{\mathcal{A}_n} u) = \nabla f(u) \in \mathcal{A}_n$.

*Proof.* Appendix D □

Next, we broaden our discussion to functions which do not enjoy ridge structure. We show how we can nevertheless construct low-dimensional approximations which enjoy low MSE under certain assumptions. For a random element $U$ with law $\rho$ on $\Omega \subset \mathcal{H}$, define $Y := P_{\mathcal{A}_n} U \in \mathcal{A}_n$, $Z := P_{\mathcal{A}_n^\perp} U \in \mathcal{A}_n^\perp$, so that $U = Y + Z$. Denote by $\rho_Y$ the law of $Y$. We begin by defining an Active-surrogate extending (6) by conditional expectation:

$$F(y) \;:=\; \mathbb{E}\big[f(U) \mid Y = y\big], \qquad y \in \mathcal{A}_n\,. \tag{8}$$

Then $F$ is the $L^2(\rho)$-best predictor of $f$ among functions of $Y$. We now demonstrate a bound on the error introduced by ignoring the remaining directions, extending Theorem 3.1 of Constantine et al. (2014).

**Theorem 2** (Hilbert-space MSE bound for the active-subspace surrogate)**.** *Under Assumption* 2 *and* $\mathbb{E}\|\nabla f(U)\|^2 < \infty$, *the mean squared error of the active-surrogate* $F$ *defined in* (8) *satisfies*

$$\mathbb{E}\big(f(U) - F(Y)\big)^2 \;\le\; C_\star \sum_{i>n} \lambda_i \;=\; C_\star\, \mathbb{E}\big\|P_{\mathcal{A}_n^\perp} \nabla f(U)\big\|^2. \tag{9}$$

*Proof.* Appendix E. □

Wycoff et al. (Section 3.1.2; 2022) showed how to establish a transformation which equalizes the active subspace importance of input dimensions with nonzero impact. In particular, they showed that if $\mathcal{C}^{1/2}$ is positive matrix square root of $\mathcal{C}$, then the mapping between the "warped" input $V$ transformed by $\mathcal{C}^{\frac{1}{2}}$ and the output $f(U)$ is equally sensitive to the input directions in the range of $\mathcal{C}$. Our next result extends this to the Hilbert space setting.

**Theorem 3** (Active directions equalized by $\mathcal{C}^{1/2}$)**.** *Assume* $\mathbb{E}\,\|\nabla f(U)\|^2 < \infty$ *and Assumption 3 holds. Let* $\mathcal{B} := \overline{\operatorname{ran}(\mathcal{C}^{1/2})} = (\ker \mathcal{C})^\perp$ *and let* $P_{\mathcal{B}}$ *denote the orthogonal projector onto* $\mathcal{B}$. *Then:*

$$\mathbb{E}_{\rho_V}\left[\nabla_v \tilde{f}(V) \otimes \nabla_v \tilde{f}(V)\right] = P_{\mathcal{B}}\,, \tag{10}$$

*where* $V = \mathcal{C}^{\frac{1}{2}} U$ *with distribution* $\rho_V$ *implied by* $\rho$.

*Proof.* Appendix K □

# 4 MONTE CARLO ESTIMATION

We have thus shown that with $\mathcal{C}$ in hand, we can perform dimension reduction. However, we will typically not have an analytic expression for $\mathcal{C}$. Following the typical workflow in the finite dimensional setting (Constantine and Gleich, 2014), we will in this section develop and analyze a Monte-Carlo estimator $\hat{\mathcal{C}}$ for $\mathcal{C}$.

**Definition 4.** *Let* $U_1, \ldots, U_B \overset{i.i.d.}{\sim} \rho$. *Set* $g_b := \nabla f(U_b) \in \mathcal{H}$, $b = 1, \ldots, B$, *and let* $h \in \mathcal{H}$. *Define the empirical covariance operator*

$$\widehat{\mathcal{C}}_B \;:=\; \frac{1}{B}\sum_{b=1}^{B} g_b \otimes g_b \in \mathcal{L}(\mathcal{H});\; \widehat{\mathcal{C}}_B h = \frac{1}{B}\sum_{b=1}^{B} \langle h, g_b\rangle\, g_b.$$

Though this is an operator on $\mathcal{H}$, it is of finite rank, and its action is computable via $B$ many $\mathcal{H}$ inner products. This is precisely what is straightforwardly approximated in modern numerical environments.

We next show how to compute its eigendecomposition. To this end, we define the semi-infinite matrix $G_B : \mathbb{R}^B \to \mathcal{H}$ by $G_B e_b = g_b/\sqrt{B}$ so that $\widehat{\mathcal{C}}_B = G_B G_B^*$, and let $\Gamma_B := G_B^* G_B \in \mathbb{R}^{B\times B}$. Perform the eigendecomposition $\Gamma_B = \sum_{i=1}^{B} \sigma_i v_i v_i^\top$.

**Lemma 1** (Eigenanalysis for $\hat{\mathcal{C}}$)**.** *For all* $i \in \{1, \ldots, B\}$, *define* $\hat{w}_i := \frac{G_B v_i}{\sqrt{\sigma_i}}$. *Then the* $\hat{w}_i$ *are* $\mathcal{H}$-*orthogonal,* $\hat{\mathcal{C}}\hat{w}_i = \sigma_i w_i$, *and:* $\widehat{\mathcal{C}}_B = \sum_{i=1}^{r} \sigma_i\, \widehat{w}_i \otimes \widehat{w}_i$.

*Proof.* Appendix F. □

We see that an eigenanalysis of the matrix $\Gamma_B$ unlocks an eigenanalysis for the operator $\hat{\mathcal{C}}$. Their nonzero eigenvalues are shared, and the eigenfunctions of $\hat{\mathcal{C}}$ are linear combinations of the sampled gradients $g_b$ with coefficients given by $[\frac{v_{i,1}}{\sigma_i}, \ldots, \frac{v_{i,B}}{\sigma_i}]$. See Appendix F for a more detailed statement of Lemma 1. From a computational standpoint, the primary cost will often come from evaluating the objective functional and its functional gradient, and this cost can vary substantially based on the kind of problem being addressed. Once the gradients have been obtained, the most expensive operation that remains is the eigen decomposition of the functional Gram matrix $\Gamma_B \in \mathbb{R}^{B\times B}$, which is of cubic computational complexity $B$, in the Monte Carlo sample size.

Our next result establishes convergence of this Monte Carlo estimator.

**Theorem 4** (Consistency)**.** *Assume* $\mathbb{E}\|\nabla f(U)\|^2 < \infty$. *Then* $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \to 0$ *almost surely.*

*Proof.* Appendix G. □

Since the active subspace operator itself is frequently used as a means for obtaining the active subspace, we next investigate the convergence behavior of an eigenanalysis on $\hat{\mathcal{C}}$ to that on $\mathcal{C}$. Recall that $(\lambda_i, w_i)_{i\geq 1}$ are the eigenpairs of $\mathcal{C}$ and $(\sigma_i, \hat{w}_i)_{i\in 1,\ldots,B}$ are those of $\hat{\mathcal{C}}$.

**Corollary 1.** *Assume $\mathbb{E}\|\nabla f(U)\|^2 < \infty$. Then:*

*(i)* Eigenvalues. *For every $i \geq 1$,*

$$|\sigma_i - \lambda_i| \xrightarrow[B\to\infty]{\text{a.s.}} 0.$$

*(ii)* Eigenfunctions. *Fix $i \geq 1$ and suppose $\lambda_i$ is simple and isolated with spectral gap*

$$\gamma_i := \min\{\lambda_{i-1}-\lambda_i,\ \lambda_i-\lambda_{i+1}\} > 0 \quad (\lambda_0 := +\infty).$$

*Then $\|s_i\hat{w}_i - w_i\| \xrightarrow[B\to\infty]{\text{a.s.}} 0$ for appropriately chosen signs $s_i \in \{-1, 1\}$.*

*Proof.* Appendix H. □

We next determine the convergence rate under an additional moment assumption.

**Theorem 5** ($\sqrt{B}$-rate of convergence)**.** *Assume $\mathbb{E}\|\nabla f(U)\|^4 < \infty$. Then*

$$\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{op} = O(B^{-1/2}). \tag{11}$$

*Proof.* Appendix I, which also presents Chebyshev bounds. □

While the $\sqrt{B}$ convergence rate is typical, it also means we get diminishing marginal returns. This makes a uncertainty quantification about the $\mathcal{C}$ estimate itself practically useful. Our final result concerns the asymptotic distribution of $\hat{\mathcal{C}}$, which we establish under the following assumption:

**Assumption 1** (Lindeberg/Lyapunov condition in $\mathcal{S}_2$)**.** Let $Z := Y - \mathbb{E}Y \in \mathcal{S}_2(\mathcal{H})$ with $Y = \nabla f(U) \otimes \nabla f(U)$. Assume $\mathbb{E}\|Z\|^2_{\mathcal{S}_2} < \infty$ and either

- (*Lindeberg*) for every $\varepsilon > 0$, $\mathbb{E}\big[\|Z\|^2_{\mathcal{S}_2}\mathbf{1}\{\|Z\|_{\mathcal{S}_2} > \varepsilon\sqrt{B}\}\big] \to 0$, or
- (*Lyapunov*) there exists $\eta > 0$ such that $\mathbb{E}\|Z\|^{2+\eta}_{\mathcal{S}_2} < \infty$.

A sufficient condition is $\mathbb{E}\|\nabla f(U)\|^{4+\eta} < \infty$ for some $\eta > 0$.

**Theorem 6** (Hilbert–space CLT in $\mathcal{S}_2(\mathcal{H})$)**.** *Under Assumption 1, in the separable Hilbert space $\mathcal{S}_2(\mathcal{H})$,*

$$\sqrt{B}\left(\widehat{\mathcal{C}}_B - \mathcal{C}\right) \Rightarrow Z_\infty,$$

*where $Z_\infty$ is a mean-zero Gaussian element in $\mathcal{S}_2(\mathcal{H})$ with covariance operator given in the Appendix.*

*Proof.* Appendix J □

# 5 APPLICATIONS

We now give deploy our methodology on our motivating test functions and discuss the results. For all experiments, we use FEniCS (Alnæs et al., 2015) to generate the meshes, solve the differential equations, and determine the functional gradients.

## 5.1 EXPLORATORY ANALYSIS AND VISUALIZATION

The first question an analyst will have is whether an active subspace exists and if so what its dimension is. Constantine (2015) suggests to look for a significant gap in the log-eigenvalues, while traditionally for PCA, cumulative variance retained being above some threshold is used to determine the reduced subspace dimension. As in finite dimension, this may in either case be achieved by visually inspecting the decay of the spectrum of the operator. We show the estimated spectra in the top row of Figure 2.

A key application of the active subspace method is visualization of high dimensional functions. Just as in finite dimension, we can project some functions $u_1, \ldots, u_N$ onto the two leading eigenfunctions $w_1, w_2$. Coloration can be used to indicate the value $f(u_i)$ of each point. However, if there is significant variation in $f$ beyond the first two eigenfunctions, this may lead to a difficult to interpret plot. Therefore, we can enrich the plot by fitting a statistical model to the reduced dataset using $\langle u_i, w_1\rangle$, $\langle u_i, w_2\rangle$ as predictor variables being used to predict $f(u_i)$, and treat the remaining variation in the function as noise. A plot of the model's predictions can be used as an estimate of the optimal $L^2$ active subspace surrogate. We show these surfaces in the bottom row of Figure 2. For the `poisson` and `laminar` functions, we see that most of the variation is captured by the low dimensional plot. Two dimensions is not sufficient to capture all variation for the `kiri` function, but we nevertheless observe some low-dimensional, bowl-shaped structure.

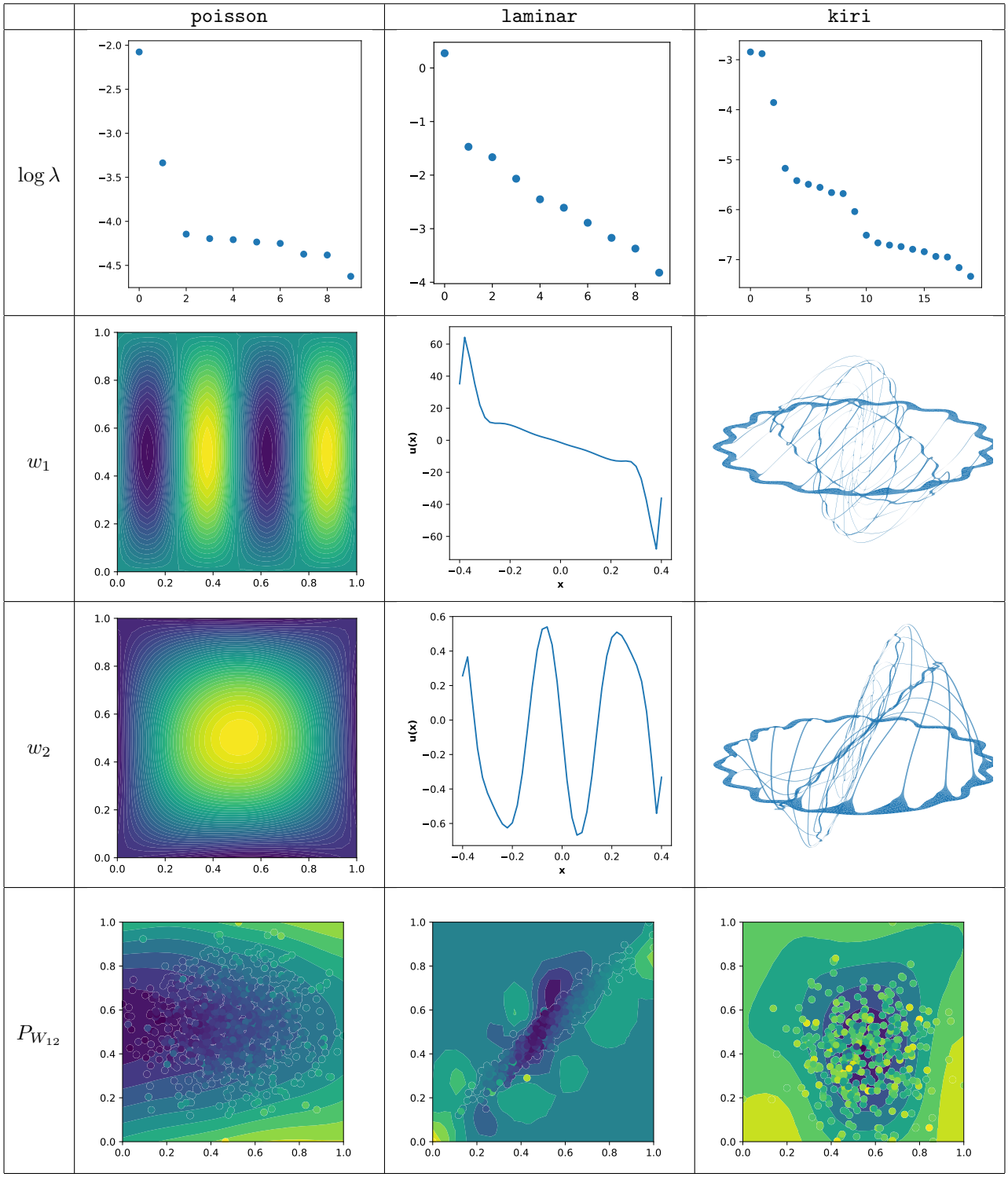


Figure 2: **Visualization with Active Subspaces.** Each row corresponds to a function. *Top row:* Monte-Carlo estimates of first 10 eigenvalues of active subspace operator; log-eigenvalues are plotted to help determine jumps indicating an active subspace. *Middle rows:* Estimate for first and second eigenfunctions. *Bottom row:* Projection of samples along first two eigenfunctions (colored points); surface gives a Gaussian process conditional mean estimate of the $L^2$-optimal surrogate fit using `hetGPy` (O'Gara et al., 2025).

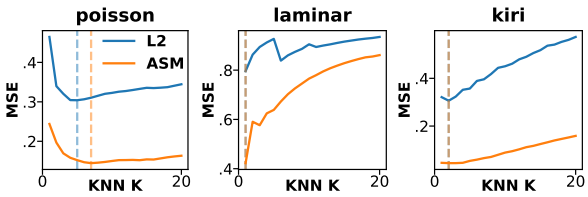


Figure 3: **Active Subspace for functional KNN Regression**. Cross Validation MSE comparison for KNN with standard $L^2$ distance versus Euclidean distance along active subspace projection.

### 5.2 SURROGATE MODELING

Active subspaces can also serve to increase the predictive accuracy of local surrogate models. Following Fukumizu and Leng (2014), we test our dimension reduction method by evaluating its ability to improve the performance of a K Nearest Neighbors (KNN) regression. KNN regression is most straightforwardly applied to our functional context by predicting the output of $f$ at some unobserved point $u^* \in \mathcal{H}$ by finding the $K$ training points with the least $L^2$ distance to $u^*$. This will serve as our baseline method.

To deploy our active subspace operator in this context, we first project functions at which predictions are desired onto the active subspace $\mathcal{A}$, and then compute standard $L_2$ distance. In particular:

$$\|u_1 - u_2\|_{\mathcal{A}} := \|P_{\mathcal{A}}(u_1 - u_2)\|_{L^2} \,. \tag{12}$$

This can be calculated by computing the coordinates of the projection of the functions $u_1, u_2$ in an orthonormal basis for $\mathcal{A}$, the $b$th coordinate of which is denoted $\omega_{1,b}$ and $\omega_{2,b}$, and then evaluating the Euclidean distance:

$$\|u_1 - u_2\|_{\hat{\mathcal{C}}} = \sqrt{\sum_{b=1}^{B} (\omega_{1,b} - \omega_{2,b})^2} \,. \tag{13}$$

Using leave-one-out Cross Validation, we compare the predictive error of KNN using the standard $L^2$ distance versus the distance along the active subspace for various $K \in \{1, \ldots, 20\}$ on our test functions. Figure 3 gives the results. We see that on all three functions, the active subspace distance is better from a predictive perspective at the optimal value of $K$. For the `poisson` and `kiri` problems, the worst performing value of $K$ for the active subspace method is superior to the best performing value of $K$ for the $L^2$ method.

### 5.3 BAYESIAN OPTIMIZATION

Bayesian Optimization (BO; Jones et al., 1998; Garnett, 2023) has been classically viewed as a method appropriate only for mild dimensional input spaces (Binois and Wycoff, 2022), and thus it may be unsurprising that its application to function spaces has been somewhat limited in the academic literature. However, building on Wang et al. (2016) working in the finite dimensional case, Vellanki et al. (2019); Shilton et al. (2020) propose to methods to conduct Bayesian optimization on spaces of functions by searching within random finite dimensional subspaces. These subspaces are generated as the span of a randomly drawn set of functions $m_1, \ldots, m_R \sim \rho$ from some measure on $\mathcal{H}$. This choice of $\rho$, together with the target function, implicitly defines an active subspace operator. We will propose to use as a basis for the search space not $\{m_1, \ldots, m_R\}$ as in previous approaches, but rather the leading eigenbasis of $\hat{\mathcal{C}} = \sum_{r=1}^{R} \nabla f(m_r) \otimes \nabla f(m_r)$.

We define a new function $g : [-1, 1]^R \to \mathbb{R}$ as:

$$g(\mathbf{c}) = \sum_{i=1} \frac{c_i}{\ell} q_i \,, \tag{14}$$

where $q_i$ are either the randomly generated functions from $\rho$ (the `Rand` method) or the leading $R$ eigenfunctions of $\hat{\mathcal{C}}$ (the `ASM` method). This gives us a function defined on a finite dimensional cube which makes it compatible with standard BO machinery. We use in particular the BO framework Ax (Olson et al., 2025). We sampled $N_{\text{init}} = 10$ random initial functions $m_1, \ldots, m_{N_{\text{init}}}$ from $\rho$ and then projected them onto $\{q_1, \ldots, q_R\}$. The coefficients can be computed by $(Q^*Q)^{-1}Q^*M$, where $Q^*Q$ is the gram matrix of $\{q_1, \ldots, q_R\}$ and $Q^*M$ is the matrix of inner products between the two sets of functions. The scalar $\ell$ is 1.5 times the maximum absolute value of any coefficient.

We then evaluated the objective function $g$ at the $N_{\text{init}}$ projected initial functions. Although for each repetition we use the same initial points in function space, the initial design for the finite dimensional BO subroutine are different because the projection of this initial set onto the respective bases differs. After the $N_{\text{init}}$ initial random functions, we used an additional 40 sequential evaluations chosen by the BO subroutine using expected improvement. We conducted 100 repetitions with different random seeds for each function and method.

We present the $10^{th}, 50^{th}$ and $90^{th}$ percentiles of the optimization progress in Figure 4. On the `poisson` problem, the `Rand` method is barely able to make progress, while the `ASM` method quickly finds the global optimum (we show only the first 20 iterations to better illustrate the difference). On the `laminar` problem, the initial set of functions generated by the random projection actually generates better starting output values, but in the sequential design phase the `ASM` method rapidly surpasses the `Rand`. On the `kiri` prob-

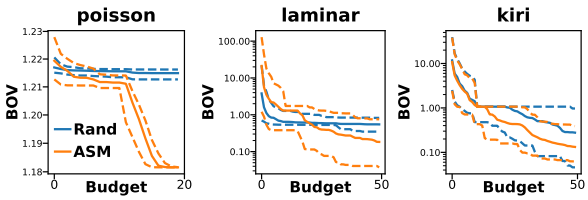


Figure 4: **Functional Optimization in Active Subspace.** Compares searching a random vs active subspace; y-axis gives Best Observed Value. Solid line gives median and dotted $10^{th}$ and $90^{th}$ percentile.

lem, both methods quickly find that the zero function gives a decent solution (as the midpoint of the space, Ax specifically checks this point at initialization), but the `ASM` method much more quickly discovers improved solutions. Overall, the active subspace approach surpasses the random subspace method.

## 6 DISCUSSION

**Summary and Conclusions.** In this article, we introduced an extension of the active subspace matrix to infinite dimensions, established its theoretical properties, and developed a computable approximation. We found that the active subspace method can be a powerful tool for developing finite dimensional approximations of functionals on a Hilbert space and dominated existing approaches on our complex test functions.

**Future Work.** In the decade since the active subspace method has swept computational engineering, authors have proposed a number of extensions and improvements. For instance, several authors have proposed active subspaces for vector-valued functions (Zahm et al., 2020; Tripathy and Bilionis, 2019; Edeling, 2023; Musayeva and Binois, 2024; Rumsey et al., 2025) which could be ported to our infinite dimensional input setting. Lam et al. (2020) consider active subspaces for multifidelity models. Several authors have proposed strategies for estimating active subspaces without gradients using statistical models (e.g. Wycoff et al., 2021; Rumsey et al., 2024), and extending these to the infinite dimensional setting would expand the applicability of this method. Finally, Lee (2019) propose a modification of the active subspace matrix which emphasizes the mean gradient, and it would be interesting to study a functional analog.

## Acknowledgements

Author NW gratefully acknowledges support from the NSF and AFOSR under award #2529277 as well as from the NSF and NGA under award #2428033, and computing resources from the Unity Research Computing Platform at the Massachusetts Green High Performance Computing Center.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N. (2015). The fenics project version 1.5. *Archive of numerical software*, 3(100).

Araujo, A. and Giné, E. (1980). *The central limit theorem for real and Banach valued random variables*. John Wiley & Sons.

Benth, F. E., Di Nunno, G., and Simonsen, I. C. (2021). Sensitivity analysis in the infinite dimensional heston model. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 24(02):2150014.

Beskos, A., Girolami, M., Lan, S., Farrell, P. E., and Stuart, A. M. (2017). Geometric mcmc for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351.

Binois, M. and Wycoff, N. (2022). A survey on high-dimensional gaussian process modeling with application to bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization*, 2(2):1–26.

Bittner, D., Engel, M., Wohlmuth, B., Labat, D., and Chiogna, G. (2021). Temporal scale-dependent sensitivity analysis for hydrological model parameters using the discrete wavelet transform and active subspaces. *Water Resources Research*, 57(10):e2020WR028511.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs.

Chen, Y.-X. and Zhu, W.-J. (2011). Note on the strong law of large numbers in a Hilbert space. *Gen. Math*, 19(3):11–18.

Constantine, P. and Gleich, D. (2014). Computing active subspaces with monte carlo. *arXiv preprint arXiv:1408.0545*.

Constantine, P. G. (2015). *Active subspaces: Emerging ideas for dimension reduction in parameter studies*. SIAM.

Constantine, P. G., del Rosario, Z., and Iaccarino, G. (2016). Many physical laws are ridge functions. *arXiv preprint arXiv:1605.07974*.

Constantine, P. G., Dow, E., and Wang, Q. (2014). Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524.

Constantine, P. G., Zaharatos, B., and Campanelli, M. (2015). Discovering an active subspace in a single-diode solar cell model. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8(5-6):264–273.

Courant, R. and Hilbert, D. (1954). *Methods of Mathematical Physics, Vol. I*. John Wiley & Sons.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46.

Diez, M. and Serani, A. (2024). Design-space dimensionality reduction in global optimization of functional surfaces: recent developments and way forward. *Ship Technology Research*, 71(2):141–152.

Edeling, W. (2023). On the deep active-subspace method. *SIAM/ASA Journal on Uncertainty Quantification*, 11(1):62–90.

Evans, L. C. (2024). An introduction to mathematical optimal control theory spring, 2024 version. *Accessed on June*, 28:2024.

Fréchet, M. (1907). Sur les ensembles de fonctions et les opérations linéaires. *CR Acad. Sci. Paris*, 144:1414–1416.

Fukumizu, K. and Leng, C. (2014). Gradient-based kernel dimension reduction for regression. *Journal of the American Statistical Association*, 109(505):359–370.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs.

Garnett, R. (2023). *Bayesian optimization*. Cambridge University Press.

Gholami, A., Keutzer, K., and Biros, G. (2019). Anode: unconditionally accurate memory-efficient gradients for neural odes. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, IJCAI'19, page 730–736. AAAI Press.

Grey, Z. J. and Constantine, P. G. (2018). Active subspaces of airfoil shape parameterizations. *AIAA Journal*, 56(5):2003–2017.

Iooss, B. and Ribatet, M. (2009). Global sensitivity analysis of computer models with functional inputs. *Reliability Engineering & System Safety*, 94(7):1194–1204.

Ji, W., Ren, Z., Marzouk, Y., and Law, C. K. (2019). Quantifying kinetic uncertainty in turbulent combustion simulations using active subspaces. *Proceedings of the Combustion Institute*, 37(2):2175–2182.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492.

Khatamsaz, D., Molkeri, A., Couperthwaite, R., James, J., Arróyave, R., Srivastava, A., and Allaire, D. (2021). Adaptive active subspace-based efficient multifidelity materials design. *Materials & Design*, 209:110001.

Kim, J., Wang, Z., and Song, J. (2024). Adaptive active subspace-based metamodeling for high-dimensional reliability analysis. *Structural Safety*, 106:102404.

Klein, M., Sadiki, A., and Janicka, J. (2003). Investigation of the influence of the reynolds number on a plane jet using direct numerical simulation. *International journal of heat and fluid flow*, 24(6):785–794.

Lam, R. R., Zahm, O., Marzouk, Y. M., and Willcox, K. E. (2020). Multifidelity dimension reduction via active subspaces. *SIAM Journal on Scientific Computing*, 42(2):A929–A956.

Ledoux, M. and Talagrand, M. (2013). *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.

Lee, M. R. (2019). Modified active subspaces using the average of gradients. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):53–66.

Liu, J. and Wang, Z. (2019). Non-commutative discretize-then-optimize algorithms for elliptic pde-constrained optimal control problems. *Journal of Computational and Applied Mathematics*, 362:596–613.

Logg, A., Mardal, K.-A., and Wells, G. (2012). *Automated solution of differential equations by the finite element method: The FEniCS book*, volume 84. Springer Science & Business Media.

Musayeva, K. and Binois, M. (2024). Shared active subspace for multivariate vector-valued functions. *arXiv preprint arXiv:2401.02735*.

Nochetto, R. H., Siebert, K. G., and Veeser, A. (2009). Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation: Dedicated to Wolfgang Dahmen on the occasion of his 60th birthday*, pages 409–542. Springer.

O'Gara, D., Binois, M., Garnett, R., and Hammond, R. A. (2025). hetgpy: Heteroskedastic gaussian process modeling in python. *Journal of Open Source Software*, 10(106):7518.

Olson, M., Santorella, E., Tiao, L. C., Cakmak, S., Garrard, M., Daulton, S., Lin, Z. J., Ament, S., Beckerman, B., Onofrey, E., Igusti, P., Lara, C., Letham, B., Cardoso, C., Shen, S. S., Lin, A. C., Grange, M., Kashtelyan, E., Eriksson, D., Balandat, M., and Bakshy, E. (2025). Ax: A Platform for Adaptive Experimentation. In *AutoML 2025 ABCD Track*.

Onken, D. and Ruthotto, L. (2020). Discretize-optimize vs. optimize-discretize for time-series regression and continuous normalizing flows. *arXiv preprint arXiv:2005.13420*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peter, J. E. and Dwight, R. P. (2010). Numerical sensitivity analysis for aerodynamic optimization: A survey of approaches. *Computers & Fluids*, 39(3):373–391.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Piano, S. L., Iwanaga, T., Becker, W., et al. (2021). The future of sensitivity analysis: an essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137:104954.

Riesz, F. (1907). *Sur une espèce de géométrie analytique des systèmes de fonctions sommables*. Gauthier-Villars.

Rochlitz, R., Skibbe, N., and Günther, T. (2019). custem: Customizable finite-element simulation of complex controlled-source electromagnetic data. *Geophysics*, 84(2):F17–F33.

Rudin, W. (1991). Functional analysis. *Inc, New York*, 45(46):4.

Rumsey, K., Francom, D., and Vander Wiel, S. (2024). Discovering active subspaces for high-dimensional computer models. *Journal of Computational and Graphical Statistics*, 33(3):896–908.

Rumsey, K. N., Hardy, Z. K., Ahrens, C., and Vander Wiel, S. (2025). Co-active subspace methods for the joint analysis of adjacent computer models. *Technometrics*, 67(1):133–146.

Shilton, A., Gupta, S., Rana, S., and Venkatesh, S. (2020). Sequential subspace search for functional bayesian optimization incorporating experimenter intuition. *arXiv preprint arXiv:2009.03543*.

Stewart, G. W. and Sun, J.-G. (1990). *Matrix perturbation theory*. San Diego: Academic Press.

Tripathy, R. and Bilionis, I. (2019). Deep active subspaces: A scalable method for high-dimensional uncertainty propagation. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 59179, page V001T02A074. American Society of Mechanical Engineers.

Vellanki, P., Rana, S., Gupta, S., de Celis Leal, D. R., Sutti, A., Height, M., and Venkatesh, S. (2019). Bayesian functional optimisation with shape prior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1617–1624.

Vogel, C. R. (2002). *Computational methods for inverse problems*. SIAM.

Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and De Feitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55:361–387.

Wu, K., O'Leary-Roseberry, T., Chen, P., and Ghattas, O. (2023). Large-scale bayesian optimal experimental design with derivative-informed projected neural network. *Journal of Scientific Computing*, 95(1):30.

Wycoff, N., Binois, M., and Gramacy, R. B. (2022). Sensitivity prewarping for local surrogate modeling. *Technometrics*, 64(4):535–547.

Wycoff, N., Binois, M., and Wild, S. M. (2021). Sequential learning of active subspaces. *Journal of Computational and Graphical Statistics*, 30(4):1224–1237.

Yang, X., Forró, C., Li, T. L., Miura, Y., Zaluska, T. J., Tsai, C.-T., Kanton, S., McQueen, J. P., Chen, X., Mollo, V., et al. (2024). Kirigami electronics for long-term electrophysiological recording of human neural organoids and assembloids. *Nature biotechnology*, 42(12):1836–1843.

Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323.

Zahm, O., Constantine, P. G., Prieur, C., and Marzouk, Y. M. (2020). Gradient-based dimension reduction of multivariate vector-valued func-

tions. *SIAM Journal on Scientific Computing*, 42(1):A534–A558.

## Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes, all assumptions explicitly given in supplementary.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Only algorithms with well-known complexities (e.g. matrix decompositions) are used.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Yes.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes we explicitly give all assumptions.
   (b) Complete proofs of all theoretical results. Complete, detailed proofs are in the supplementary materials.
   (c) Clear explanations of any assumptions. We give remarks in the supplementary explaining assumptions.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). All code is in supplementary.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Yes we give training details.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). We describe how we use quantiles to measure uncertainty.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). No, we don't give this information, because this approach is not as computationally intensive as, for instance a neural network. It just relies on matrix decompositions.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. Yes, we cite all assets used.
   (b) The license information of the assets, if applicable. All licenses used are open source.
   (c) New assets either in the supplemental material or as a URL, if applicable. Yes, all code is in supplementary material.
   (d) Information about consent from data providers/curators. Not Applicable.
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not applicable.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. Not Applicable.
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable.
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable.

# Active Subspaces in Infinite Dimension Supplementary Materials

## A Background: Differentiation in function space

**Properties of $\nabla f$ (see Definition 1)**

- Uniqueness: The Fréchet gradient $\nabla f(u)$ is uniquely determined by $Df(u)$.
- Norm identity: $\|\nabla f(u)\| = \|Df(u)\|$, where $\|Df(u)\| = \sup_{\|h\|=1} |Df(u)[h]|$.
- Linearity in $h$: $Df(u)[\cdot]$ is linear and continuous, hence $h \mapsto \langle h, \nabla f(u)\rangle$ is the correct representation.
- (Gateaux variant) If all directional limits $D_G f(u)[h] = \lim_{t\to 0} \frac{f(u+th)-f(u)}{t}$ exist and the map $h \mapsto D_G f(u)[h]$ is *continuous and linear* on $\mathcal{H}$, then $Df(u) = D_G f(u)$ and the same $\nabla f(u)$ satisfies $D_G f(u)[h] = \langle h, \nabla f(u)\rangle$ for all $h$.

## B Basic properties of $\mathcal{C}$

Since $\|a \otimes b\|_1 = \|a\|\,\|b\|$ for the trace norm, we have $\|\nabla f(U) \otimes \nabla f(U)\|_1 = \|\nabla f(U)\|^2$. From $\mathbb{E}\|\nabla f(U)\|^2 < \infty$, we obtain that the *operator-valued* random variable $T := \nabla f(U) \otimes \nabla f(U)$ satisfies $\mathbb{E}\|T\|_1 = \mathbb{E}\|\nabla f(U) \otimes \nabla f(U)\|_1 = \mathbb{E}\|\nabla f(U)\|^2 < \infty$, so $T$ is Bochner integrable in the Banach space of trace-class operators $\mathcal{L}_1(\mathcal{H})$. Therefore $\mathcal{C} := \mathbb{E}[\,\nabla f(U) \otimes \nabla f(U)\,] \in \mathcal{L}_1(\mathcal{H})$ as a Bochner expectation is well-defined.

We also have the following properties for $\mathcal{C}$.

- **Trace-class.** By definition, $\mathcal{L}_1(\mathcal{H})$ is the space of trace-class operators with norm $\|A\|_1 = \mathrm{trace}\big((A^*A)^{1/2}\big)$. Since $C \in \mathcal{L}_1(\mathcal{H})$, it is trace-class. Equivalently, $\|\mathcal{C}\|_1 \le \mathbb{E}\|\nabla f(U) \otimes \nabla f(U)\|_1 = \mathbb{E}\|\nabla f(U)\|^2 < \infty$.
- **Compact.** Every trace-class operator on a Hilbert space is compact (in fact, trace-class $\subset$ Hilbert–Schmidt $\subset$ compact). Hence $\mathcal{C}$ is compact.
- **Self-adjoint.** For any $a, b \in \mathcal{H}$, $(a \otimes b)^* = b \otimes a$. In particular, $(a \otimes a)^* = a \otimes a$, so $\nabla f(U) \otimes \nabla f(U)$ is self-adjoint almost surely. The adjoint is continuous with respect to the trace norm, and the Bochner integral preserves linear operations. Hence

$$\mathcal{C}^* = \big(\mathbb{E}[\,\nabla f(U) \otimes \nabla f(U)\,]\big)^* = \mathbb{E}\big[(\nabla f(U) \otimes \nabla f(U))^*\big] = \mathbb{E}[\,\nabla f(U) \otimes \nabla f(U)\,] = \mathcal{C}.$$

- **Positive semidefinite.** By definition, $A \succeq 0$ iff $\langle Ah, h\rangle \ge 0$ for all $h \in \mathcal{H}$. For rank-one tensors with $a = b$,

$$\langle (a \otimes a)h,\, h\rangle = \langle h, a\rangle\,\langle a, h\rangle = |\langle h, a\rangle|^2 \ge 0.$$

  Thus $\nabla f(U) \otimes \nabla f(U) \succeq 0$ almost surely. Taking expectations, for any $h \in \mathcal{H}$,

$$\langle \hat{\mathcal{C}}, h\rangle = \mathbb{E}\,\langle (\nabla f(U) \otimes \nabla f(U))h,\, h\rangle = \mathbb{E}\,|\langle h,\, \nabla f(U)\rangle|^2 \;\ge\; 0,$$

  so $\mathcal{C} \succeq 0$.

**Remark 1** (Interpretation)**.** *For any fixed unit vector $w \in \mathcal{H}$, $\langle \mathcal{C}w, w\rangle = \mathbb{E}\big[\langle \nabla f(U), w\rangle^2\big] = \mathbb{E}\big[(Df(U)[w])^2\big]$ equals the average squared directional derivative of $f$ along direction $w$. In finite dimensions, with $\nabla f(U) = \nabla f(X)$ and $w \in \mathbb{R}^P$, this reads $\langle \mathcal{C}w, w\rangle = \mathbb{E}\big[((\nabla f(X))^\top w)^2\big]$. Note that, $\mathcal{C}$ acts on $\mathcal{H}$, so it is in general* infinite-dimensional. *However, if $\nabla f(U)$ almost surely lies in an $n$-dimensional subspace, then $\mathcal{C}$ has finite rank $n$.*

**Expansion of $\nabla f(U)$ in the eigenbasis.** We can expand $\nabla f(U)$ in the eigenbasis. Write $\nabla f(U) = \sum_{i\geq 1} \xi_i\, w_i$, where $\xi_i := \langle \nabla f(U), w_i\rangle$. Then, using $C = \mathbb{E}[\nabla f(U) \otimes \nabla f(U)]$, $\mathbb{E}[\xi_i \xi_j] = \mathbb{E}\big[\langle \nabla f(U), w_i\rangle \langle \nabla f(U), w_j\rangle\big] = \langle C w_j,\, w_i\rangle = \lambda_j \langle w_j, w_i\rangle = \lambda_i\, \delta_{ij}$, so the coordinates are uncorrelated and

$$\mathbb{E}[\xi_i^2] = \lambda_i \quad \text{for all } i. \tag{15}$$

**Gradient energy in the active subspace.** Since $P_{\mathcal{A}_n} w_i = w_i$ for $i \leq n$ and $P_{\mathcal{A}_n} w_i = 0$ for $i > n$,

$$P_{\mathcal{A}_n} \nabla f(U) \;=\; \sum_{i=1}^{n} \xi_i\, w_i, \qquad P_{\mathcal{A}_n^\perp} \nabla f(U) \;=\; \sum_{i>n} \xi_i\, w_i.$$

Hence,

$$\|P_{\mathcal{A}_n} \nabla f(U)\|^2 = \sum_{i=1}^{n} \xi_i^2, \qquad \|P_{\mathcal{A}_n^\perp} \nabla f(U)\|^2 = \sum_{i>n} \xi_i^2,$$

and taking expectations with (15) gives

$$\mathbb{E}\big[\|P_{\mathcal{A}_n} \nabla f(U)\|^2\big] = \sum_{i=1}^{n} \lambda_i, \quad \mathbb{E}\big[\|P_{\mathcal{A}_n^\perp} \nabla f(U)\|^2\big] = \sum_{i>n} \lambda_i. \tag{16}$$

Using $\|P_V \nabla f\|^2 = \langle P_V \nabla f,\, P_V \nabla f\rangle = \langle \nabla f,\, P_V \nabla f\rangle$ and linearity of expectation, $\mathbb{E}\|P_V \nabla f(U)\|^2 = \mathbb{E}\, \langle \nabla f(U), P_V \nabla f(U)\rangle$
$= \mathbb{E}\, \mathrm{trace}\big(P_V (\nabla f(U) \otimes \nabla f(U))\big) = \mathrm{trace}\big(P_V\, \mathbb{E}[\nabla f(U) \otimes \nabla f(U)]\big) = \mathrm{trace}(P_V C)$. Therefore, $\mathrm{trace}(P_{\mathcal{A}_n} C) = \mathbb{E}\big[\|P_{\mathcal{A}_n} \nabla f(U)\|^2\big] = \sum_{i=1}^{n} \lambda_i$. The mean squared gradient contribution from the discarded complement $P_{\mathcal{A}_n^\perp}$ equals the *tail sum* $\sum_{i>n} \lambda_i$.

This is what motivates the active-subspace surrogates defined in the main text, as in this case the mean squared *discarded* gradient energy is the above tail sum.

## C Proposition 1: Detailed statement and derivation

**Spectral decomposition of $\mathcal{C}$.** There exists orthonormal eigenfunctions $\{w_i\}_{i\geq 1} \subset \mathcal{H}$ and a non-increasing sequence $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ such that

$$\mathcal{C} \;=\; \sum_{i\geq 1} \lambda_i\, w_i \otimes w_i, \qquad \mathcal{C} w_i = \lambda_i w_i, \qquad \mathrm{trace}(\mathcal{C}) = \mathbb{E}\|\nabla f(U)\|^2 = \sum_{i\geq 1} \lambda_i. \tag{17}$$

Moreover, for each $i$,

$$\mathbb{E}\Big[\big(Df(U)[w_i]\big)^2\Big] = \lambda_i. \tag{18}$$

**Proof.** By definition, for each $i$,

$$\mathbb{E}\Big[\big(Df(U)[w_i]\big)^2\Big] \;=\; \mathbb{E}\big[\langle \nabla f(U), w_i\rangle^2\big] \;=\; \langle \mathcal{C} w_i, w_i\rangle \;=\; \lambda_i\, \|w_i\|^2 \;=\; \lambda_i. \tag{19}$$

## D Proof of Theorem 1

*Proof. (i) To show:* $P_{\mathcal{A}_n^\perp} \nabla f \equiv 0$.
For any $h \in \mathcal{H}$, by definition of $\mathcal{C}$

$$\langle \mathcal{C} h, h\rangle = \mathbb{E}_\rho\big[\langle \nabla f(U), h\rangle^2\big].$$

If $h \in \mathcal{A}_n^\perp$, then using Theorem 1,

$$\langle \mathcal{C} h, h\rangle = \sum_{i>n} \lambda_i\, \langle h, w_i\rangle^2 = 0 \quad \Longrightarrow \quad \mathbb{E}_\rho\big[\langle \nabla f(U), h\rangle^2\big] = 0.$$

Let $\{e_j\}_{j\geq 1}$ be an orthonormal basis set of $\mathcal{A}_n^\perp$, then the above implies $\langle \nabla f(U), e_j\rangle = 0$ for $\rho$-a.e. $U$, each $j$. Therefore,

$$P_{\mathcal{A}_n^\perp}\nabla f(U) = 0 \quad \text{for } \rho\text{-a.e. } U.$$

Assume $g$ is continuous (required assumption: this follows, for example, if $f$ is $C^1$ as in (4)). If there were $u_0$ with $\|P_{\mathcal{A}_n^\perp}\nabla f(u_0)\| > 0$, continuity would result in an open ball $B \ni u_0$ with $\|P_{\mathcal{A}_n^\perp}\nabla f(u)\| \geq \varepsilon > 0$ for all $u \in B$, and since $\rho$ has full support on open sets (required assumption), $\rho(B) > 0$, contradicting $\rho$-a.s. vanishing. Hence $P_{\mathcal{A}_n^\perp}\nabla f(u) = 0$ for all $u \in \mathcal{H}$.

*(ii) Level-set invariance.*
Let $u_1, u_2 \in \mathcal{H}$ with $P_{\mathcal{A}_n}u_1 = P_{\mathcal{A}_n}u_2$. Then, by definition of complementary space, $z := u_2 - u_1 \in \mathcal{A}_n^\perp$. Define $u_t := u_1 + tz$ for $t \in [0,1]$ and $\phi(t) := f(u_t)$. By the chain rule and (4),

$$\phi'(t) = Df(u_t)[z] = \langle \nabla f(u_t), z\rangle = \langle P_{\mathcal{A}_n}\nabla f(u_t), z\rangle + \langle P_{\mathcal{A}_n^\perp}\nabla f(u_t), z\rangle = 0,$$

since $z \perp \mathcal{A}_n$ and $P_{\mathcal{A}_n^\perp}\nabla f \equiv 0$ by (i). Thus $\phi$ is constant and $f(u_1) = f(u_2)$.

*(iii) Gradient equality.*
Let $u_1, u_2 \in \mathcal{H}$ with $P_{\mathcal{A}_n}u_1 = P_{\mathcal{A}_n}u_2$. Fix $h \in \mathcal{H}$ and $\varepsilon > 0$. Then, by linearity of the orthogonal projector $P_{\mathcal{A}_n}$, $P_{\mathcal{A}_n}(u_1+\varepsilon h) = P_{\mathcal{A}_n}(u_2+\varepsilon h)$. So by (ii), $f(u_1+\varepsilon h) = f(u_2+\varepsilon h)$. Therefore, assuming Gateaux differentiability,

$$\langle \nabla f(u_1), h\rangle - \langle \nabla f(u_2), h\rangle = \lim_{\varepsilon\to 0}\frac{f(u_1+\varepsilon h) - f(u_1)}{\varepsilon} - \lim_{\varepsilon\to 0}\frac{f(u_2+\varepsilon h) - f(u_2)}{\varepsilon} = 0.$$

Since $h$ was arbitrary, take $h = \nabla f(u_1) - \nabla f(u_2)$. Then we have $\langle (\nabla f(u_1) - \nabla f(u_2)), (\nabla f(u_1) - \nabla f(u_2))\rangle = 0$, which in turn implies $\nabla f(u_1) = \nabla f(u_2)$.

*(iv) Factorization.*
Define $\tilde{f} : P_{\mathcal{A}_n}(\mathcal{H}) \to \mathbb{R}$ by $\tilde{f}(y) := f(u)$ for any $u$ with $P_{\mathcal{A}_n}u = y$. This is well-defined by (ii). For $y = P_{\mathcal{A}_n}u$ and any $h \in \mathcal{A}_n$,

$$D\tilde{f}(y)[h] = \lim_{\varepsilon\to 0}\frac{\tilde{f}(y+\varepsilon h) - \tilde{f}(y)}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{f(u+\varepsilon h) - f(u)}{\varepsilon} = \langle \nabla f(u), h\rangle.$$

Hence $\nabla\tilde{f}(y) = \nabla f(u) \in \mathcal{A}_n$, and $f(u) = \tilde{f}(P_{\mathcal{A}_n}u)$.

□

## E Additional discussion and proof of Theorem 2

**Disintegration and a fiberwise Poincaré inequality.** Let $\mathcal{H} = \mathcal{A}_n \oplus \mathcal{A}_n^\perp$ with orthogonal projectors $P_{\mathcal{A}_n}$ and $P_{\mathcal{A}_n^\perp}$. For a random element $U$ with law $\rho$ on $\Omega \subset \mathcal{H}$, define

$$Y := P_{\mathcal{A}_n}U \in \mathcal{A}_n, \qquad Z := P_{\mathcal{A}_n^\perp}U \in \mathcal{A}_n^\perp, \quad \text{so that } U = Y + Z.$$

Denote by $\rho_Y$ the law of $Y$. By disintegration, there exist regular conditional probabilities $\{\rho_{Z|Y}(\cdot \mid y)\}_{y\in\mathcal{A}_n}$ on $\mathcal{A}_n^\perp$ such that for every bounded measurable $\psi : \mathcal{H} \to \mathbb{R}$,

$$\int_\Omega \psi(u)\,\rho(du) \;=\; \int_{\mathcal{A}_n}\left(\int_{\mathcal{A}_n^\perp}\psi(y+z)\,\rho_{Z|Y}(dz \mid y)\right)\rho_Y(dy).$$

Fix $y \in \mathcal{A}_n$. We view the inactive space $(\mathcal{A}_n^\perp, \langle\cdot,\cdot\rangle)$ has a Hilbert space in its own right. For a measurable $\varphi : \mathcal{A}_n^\perp \to \mathbb{R}$, write $\varphi \in W^{1,2}(\rho_{Z|Y=y})$ if $\varphi \in L^2(\rho_{Z|Y=y})$ and there exists a (Riesz) gradient $\nabla_z\varphi \in L^2(\rho_{Z|Y=y}; \mathcal{A}_n^\perp)$ such that for each $z$ the directional derivative satisfies

$$D\varphi(z)[v] = \langle \nabla_z\varphi(z),\, v\rangle \qquad \text{for all } v \in \mathcal{A}_n^\perp.$$

**Assumption 2** (Uniform conditional Poincaré inequality)**.** There exists a constant $C_\star \in (0,\infty)$ such that for $\rho_Y$-a.e. $y$ and for every $\varphi \in W^{1,2}(\rho_{Z|Y=y})$,

$$\mathrm{Var}(\varphi(Z)\,|\,Y=y) \;\le\; C_\star\,\mathbb{E}\big[\,\|\nabla_z\varphi(Z)\|^2 \;\big|\; Y=y\big]\,. \tag{20}$$

Here $\mathrm{Var}(\varphi(Z)\mid Y=y)$ denotes the variance of $\varphi(Z)$ under the conditional law $\rho_{Z|Y=y}$, computed as

$$\mathrm{Var}(\varphi(Z)\mid Y=y) = \mathbb{E}\Big[\big(\varphi(Z)-\mathbb{E}[\varphi(Z)\mid Y=y]\big)^2 \;\Big|\; Y=y\Big]\,.$$

The gradient $\nabla_z$ acts only along the inactive directions $\mathcal{A}_n^\perp$.

**Remark 2** (Special cases where the fiberwise Poincaré (20) holds automatically)**.** *Assume* $\dim(\mathcal{A}_n^\perp) = m-n < \infty$ *and* $\rho_{Z|Y=y}(dz) = \pi_y(z)\,dz$ *with a* $C^1$ *positive density. The uniform constant in* (20) *is guaranteed in some special cases. As for example, it holds in the* ***Gaussian case:*** *If the conditional law* $\rho_{Z|Y=y}$ *is Gaussian on* $\mathcal{A}_n^\perp$ *with covariance* $Q_y$*, then* (20) *holds with* $C_\star = \|Q_y\|_{\mathrm{op}}$*; in particular, for an identity covariance along* $\mathcal{A}_n^\perp$ *one has* $C_\star = 1$*, independent of* $y$*. Mathematically, if* $\pi_y(z) \propto \exp(-\frac{1}{2}z^\top Q_y^{-1}z)$ *with* $Q_y \succeq 0$*, then*

$$\mathrm{Var}(\varphi\mid y) \;\le\; \sup_y \|Q_y\|_{\mathrm{op}}\,\mathbb{E}\big[\|\nabla_z\varphi\|^2 \mid y\big].$$

*If* $Q_y \equiv I$*, then* $C_\star = 1$*. Absent such conditions, a uniform spectral gap may fail, so* (20) *must be assumed.*

We now proceed with the actual proof of the theorem.

*Proof.* By tower property,

$$\mathbb{E}\big(f-F(Y)\big)^2 \;=\; \mathbb{E}\Big[\;\mathrm{Var}\big(f(U)\mid Y\big)\;\Big] \;=\; \mathbb{E}\Big[\;\mathbb{E}\big((f-\mathbb{E}[f\mid Y])^2\mid Y\big)\;\Big]. \tag{21}$$

Fix $y$ and consider the function of the inactive variable

$$\varphi_y(z) \;:=\; f(y+z), \qquad z\in\mathcal{A}_n^\perp.$$

Then by (20) (with $\varphi=\varphi_y$) under Assumption 2,

$$\mathrm{Var}\big(f(U)\mid Y=y\big) = \mathrm{Var}\big(\varphi_y(Z)\mid Y=y\big) \;\le\; C_\star\,\mathbb{E}\big[\,\|\nabla_z\varphi_y(Z)\|^2 \;\big|\; Y=y\big]. \tag{22}$$

Further, by the Riesz formula $Df(u)[h] = \langle\nabla f(u),h\rangle$, the derivative of $z\mapsto f(y+z)$ in a direction $v\in\mathcal{A}_n^\perp$ is given by

$$D_z\varphi_y(z)[v] \;=\; Df(y+z)[v] \;=\; \langle\nabla f(y+z),v\rangle.$$

Hence, the Riesz gradient *along* $\mathcal{A}_n^\perp$ is the orthogonal projection of $\nabla f(y+z)$ onto $\mathcal{A}_n^\perp$:

$$\nabla_z\varphi_y(z) \;=\; P_{\mathcal{A}_n^\perp}\nabla f(y+z). \tag{23}$$

Therefore, combing (22) and (23)

$$\mathrm{Var}\big(f(U)\mid Y=y\big) \;\le\; C_\star\,\mathbb{E}\Big[\;\big\|P_{\mathcal{A}_n^\perp}\nabla f(U)\big\|^2 \;\Big|\; Y=y\Big]. \tag{24}$$

Taking $\mathbb{E}[\cdot]$ of both sides and using (21) yields

$$\mathbb{E}\big(f-F(Y)\big)^2 \;\le\; C_\star\,\mathbb{E}\big\|P_{\mathcal{A}_n^\perp}\nabla f(U)\big\|^2.$$

Furthermore, with $C = \mathbb{E}[\nabla f(U)\otimes\nabla f(U)] = \sum_{i\ge1}\lambda_i w_i\otimes w_i$,

$$\mathbb{E}\big\|P_{\mathcal{A}_n^\perp}\nabla f(U)\big\|^2 = \mathbb{E}\sum_{i>n}\langle\nabla f(U),w_i\rangle^2 = \sum_{i>n}\mathbb{E}\langle\nabla f(U),w_i\rangle^2 = \sum_{i>n}\lambda_i,$$

which proves (9). If the tail spectrum vanishes, the bound is 0, so $f=F(Y)$ a.s. □

### E.1 Additional result: Monte Carlo estimation of active-surrogate

Theorem 2 gives us confidence that the optimal $L^2$ approximant can do well using only information from the active subspace. But this function will typically not be available to us, and thus must be estimated.

Recall the active/inactive split $\mathcal{H} = \mathcal{A}_n \oplus \mathcal{A}_n^\perp$, the random variables

$$Y := P_{\mathcal{A}_n} U \in \mathcal{A}_n, \qquad Z := P_{\mathcal{A}_n^\perp} U \in \mathcal{A}_n^\perp, \qquad U = Y + Z,$$

and the conditional laws $\{\rho_{Z|Y=y}\}_{y\in\mathcal{A}_n}$ from disintegration. Define

$$G(y) \;:=\; \mathbb{E}\big[f(Y+Z) \mid Y = y\big] \;=\; \int_{\mathcal{A}_n^\perp} f(y+z)\,\rho_{Z|Y}(dz \mid y).$$

The $L^2(\rho)$–best $Y$–measurable surrogate is then given by,

$$F(Y) \;:=\; G(Y).$$

This is equivalent to the definition in (8).

For fixed $y \in \mathcal{A}_n$ and let $(Z_i^{(y)})_{i=1}^N$ be i.i.d. samples from $\rho_{Z|Y=y}$. Define

$$\widehat{G}_N(y) \;:=\; \frac{1}{N}\sum_{i=1}^N f\big(y + Z_i^{(y)}\big), \qquad \widehat{F}_N(Y) \;:=\; \widehat{G}_N(Y). \tag{25}$$

Then $\mathbb{E}[\widehat{G}_N(y) \mid Y = y] = G(y)$ and $\mathrm{Var}(\widehat{G}_N(y) \mid Y = y) = \mathrm{Var}(f(y+Z) \mid Y = y)/N$.

**Theorem 7** (Hilbert-space Monte Carlo MSE bound)**.** *Under Assumption* 2 *and* $\mathbb{E}\|\nabla f(U)\|^2 < \infty$*, for* $N \geq 1$

$$\mathbb{E}\big(f(U) - \widehat{F}_N(Y)\big)^2 \;\leq\; C_\star\Big(1 + \frac{1}{N}\Big)\sum_{i>n}\lambda_i \tag{26}$$

$$=\; C_\star\Big(1 + \frac{1}{N}\Big)\,\mathbb{E}\big\|P_{\mathcal{A}_n^\perp}\nabla f(U)\big\|^2. \tag{27}$$

*In particular, if* $\lambda_{n+1} = \lambda_{n+2} = \cdots = 0$*, then* $f(U) = \widehat{F}_N(Y)$ *almost surely for every* $N$*.*

*Proof.* Let $\sigma^2(y) := \mathrm{Var}\big(f(y+Z) \mid Y = y\big)$. By the tower property and conditional unbiasedness,

$$\mathbb{E}\big(f - \widehat{F}_N\big)^2 = \mathbb{E}\Big[\,\mathbb{E}\big((f-F) + (F - \widehat{F}_N)\big)^2 \;\big|\; Y\Big] \leq \mathbb{E}\big[(f-F)^2\big] + \mathbb{E}\big[(F - \widehat{F}_N)^2\big] = \mathbb{E}\big[(f-F)^2\big] + \mathbb{E}\Big[\frac{\sigma^2(Y)}{N}\Big].$$

Under Assumption 2, following the same line if logic as in (22),

$$\sigma^2(y) \;\leq\; C_\star\,\mathbb{E}\big[\;\|\nabla_z\varphi_y(Z)\|^2 \;\big|\; Y = y\big]\,.$$

The fiber gradient is the inactive Riesz gradient: $\nabla_z\varphi_y(z) = P_{\mathcal{A}_n^\perp}\nabla f(y+z)$. By tower property,

$$\mathbb{E}\big[(F - \widehat{F}_N)^2\big] = \frac{1}{N}\mathbb{E}[\sigma^2(Y)] \;\leq\; \frac{C_\star}{N}\,\mathbb{E}\big\|P_{\mathcal{A}_n^\perp}\nabla f(U)\big\|^2. \tag{28}$$

Furthermore, the Hilbert-space MSE bound for $F$ (Theorem 2) gives

$$\mathbb{E}[(f-F)^2] \leq C_\star\,\mathbb{E}\|P_{\mathcal{A}_n^\perp}\nabla f(U)\|^2.$$

Combining the two bounds and using $\mathbb{E}\|P_{\mathcal{A}_n^\perp}\nabla f(U)\|^2 = \sum_{i>n}\lambda_i$ proves (26).

If the tail eigenvalues vanish, then $P_{\mathcal{A}_n^\perp}\nabla f \equiv 0$, so both terms are zero and $f$ is $z$–invariant, implying $f = \widehat{F}_N(Y)$ a.s. □

**Remark 3** (Interpretation)**.** *The quantity $\sum_{i>n} \lambda_i$ is the mean squared "inactive" gradient energy of $f$. So* (20) *bounds the fiberwise variance of $f$ by the mean squared* inactive *gradient. The constant $C_\star$ depends only on the geometry and the conditional law on the inactive fibers.*

**Remark 4** (On assumptions and variants)**.**

- *The proof requires $f$ to be (weakly) differentiable along inactive directions with $P_{\mathcal{A}_n^\perp}\nabla f \in L^2(\rho;\mathcal{H})$ so that the right-hand side in* (9) *is finite.*
- *The choice $F(Y) = \mathbb{E}[f \mid Y]$ is optimal in $L^2$: it minimizes $\mathbb{E}(f - h(Y))^2$ over all measurable $h : \mathcal{A}_n \to \mathbb{R}$. The bound in Theorem 2 shows how the approximation error is controlled by the inactive gradient energy.*

## F Proof of Lemma 1

**Lemma 1 (Extended version)** Let $U_1, \dots, U_B \overset{\text{i.i.d.}}{\sim} \rho$. Set $g_b := \nabla f(U_b) \in \mathcal{H}$, $b = 1, \dots, B$. Define the empirical covariance operator

$$\widehat{\mathcal{C}}_B \;:=\; \frac{1}{B}\sum_{b=1}^{B} g_b \otimes g_b \in \mathcal{L}(\mathcal{H}), \text{ such that } \widehat{\mathcal{C}}_B h = \frac{1}{B}\sum_{b=1}^{B} \langle h, g_b\rangle\, g_b.$$

Let $G_B : \mathbb{R}^B \to \mathcal{H}$ be the linear map $G_B e_b := g_b/\sqrt{B}$, so that $\widehat{\mathcal{C}}_B = G_B G_B^*$. Its (finite) Gram matrix is

$$\Gamma_B := G_B^* G_B \in \mathbb{R}^{B\times B}, \qquad (\Gamma_B)_{ij} = \frac{1}{B}\langle g_i, g_j\rangle.$$

Let $\Gamma_B = V\Sigma V^\top$ with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_r > 0$ ($r \leq B$) and orthonormal eigenvectors $v_1, \dots, v_r$ (columns of $V$). Define, for $i = 1, \dots, r$.

$$\widehat{w}_i \;:=\; \frac{1}{\sqrt{\sigma_i}}\, G_B v_i \;=\; \frac{1}{\sqrt{\sigma_i B}} \sum_{b=1}^{B} v_{i,b}\, g_b \;\in\; \mathcal{H}, \tag{29}$$

Then $\{\widehat{w}_i\}_{i=1}^r$ is an orthonormal system in $\mathcal{H}$. Further, each $\widehat{w}_i$ is an eigenvector of $\widehat{\mathcal{C}}_B$ with eigenvalue $\sigma_i$, and

$$\widehat{\mathcal{C}}_B \;=\; \sum_{i=1}^{r} \sigma_i\, \widehat{w}_i \otimes \widehat{w}_i. \tag{30}$$

*Proof.* **Orthonormality.** Using $G_B^* G_B = \Gamma_B$ and $v_i^\top v_j = \delta_{ij}$,

$$\langle \widehat{w}_i, \widehat{w}_j\rangle = \frac{1}{\sqrt{\sigma_i\sigma_j}}\, v_i^\top G_B^* G_B v_j = \frac{1}{\sqrt{\sigma_i\sigma_j}}\, v_i^\top \Gamma_B v_j.$$

Further, using $\Gamma_B v_j = \sigma_j v_j$,

$$\langle \widehat{w}_i, \widehat{w}_j\rangle = \frac{1}{\sqrt{\sigma_i\sigma_j}}\, \sigma_j\, v_i^\top v_j = \sqrt{\frac{\sigma_j}{\sigma_i}}\, \delta_{ij} = \delta_{ij} \tag{31}$$

Hence, $\{\widehat{w}_i\}_{i=1}^r$ is an orthonormal system in $\mathcal{H}$.

**Eigen components.** Moreover,

$$G_B^* \widehat{w}_i = \frac{1}{\sqrt{\sigma_i}}\, G_B^* G_B v_i = \frac{1}{\sqrt{\sigma_i}}\, \Gamma_B v_i = \sqrt{\sigma_i}\, v_i.$$

Therefore,

$$\widehat{\mathcal{C}}_B \widehat{w}_i = G_B G_B^* \widehat{w}_i = G_B\big(\sqrt{\sigma_i}\, v_i\big) = \sqrt{\sigma_i}\, G_B v_i = \sigma_i\, \widehat{w}_i \tag{32}$$

**Spectral decomposition.** Observe that

$$\mathrm{range}(\widehat{\mathcal{C}}_B) = \mathrm{range}(G_B G_B^*) \subset \mathrm{range}(G_B) = \mathrm{span}\{g_1, \dots, g_B\},$$

and $\mathrm{range}(G_B) = \mathrm{span}\{\widehat{w}_1, \dots, \widehat{w}_r\}$ since the $\widehat{w}_i$ are obtained by applying $G_B$ to the eigenvectors $v_i$ (see (29)). Moreover, if $h \perp \mathrm{span}\{g_b\}$ then $G_B^* h = 0$ and thus $\widehat{\mathcal{C}}_B h = G_B G_B^* h = 0$.

Now, take any $h \in \mathcal{H}$ and decompose it orthogonally as

$$h \;=\; h_\parallel + h_\perp, \qquad h_\parallel \in \mathrm{span}\{\widehat{w}_1, \dots, \widehat{w}_r\}, \;\; h_\perp \perp \mathrm{span}\{\widehat{w}_1, \dots, \widehat{w}_r\}.$$

Write $h_\parallel = \sum_{i=1}^r \alpha_i \, \widehat{w}_i$. Using (32),

$$\widehat{\mathcal{C}}_B h = \widehat{\mathcal{C}}_B h_\parallel + \widehat{\mathcal{C}}_B h_\perp = \sum_{i=1}^r \alpha_i \, \widehat{\mathcal{C}}_B \widehat{w}_i + 0 = \sum_{i=1}^r \alpha_i \, \sigma_i \, \widehat{w}_i.$$

On the other hand, the finite-rank operator $\sum_{i=1}^r \sigma_i \, \widehat{w}_i \otimes \widehat{w}_i$ acts as

$$\Big(\sum_{i=1}^r \sigma_i \, \widehat{w}_i \otimes \widehat{w}_i\Big) h = \sum_{i=1}^r \sigma_i \, \langle h, \widehat{w}_i \rangle \, \widehat{w}_i = \sum_{i=1}^r \sigma_i \, \langle h_\parallel, \widehat{w}_i \rangle \, \widehat{w}_i = \sum_{i=1}^r \sigma_i \, \alpha_i \, \widehat{w}_i,$$

and vanishes on $h_\perp$ since $\langle h_\perp, \widehat{w}_i \rangle = 0$ for all $i$.

Hence, both the operators agree on every $h \in \mathcal{H}$, proving (30). □

**Remark 5** (Comments on the Gram construction)**.**

- *The spectral form in Lemma 1 uses the eigenpairs $\{(\sigma_i, \widehat{w}_i)\}$ of $\widehat{\mathcal{C}}_B$. These* empirical *eigenfunctions lie in* $\mathrm{span}\{g_1, \dots, g_B\}$ *and provide a rank-$r$ approximation of $\mathcal{C}$.*
- *If one builds the Gram matrix without the $1/B$ factor, i.e. $\widetilde{\Gamma}_{ij} := \langle g_i, g_j \rangle$, then the corresponding eigenvalues $\widetilde{\sigma}_i$ satisfy $\sigma_i = \widetilde{\sigma}_i / B$, and* (29) *becomes $\widehat{w}_i = \widetilde{\sigma}_i^{-1/2} \sum_b v_{i,b} \, g_b$ (see detailed statement of Lemma 1 in Appendix F). The operator $\widehat{\mathcal{C}}_B$ is unchanged; the normalization only moves between $\sigma_i$ and $\widehat{w}_i$.*

## G Detailed statement and proof of Theorem 4

**Theorem 4 (Expanded version of pointwise and operator-norm convergence).**

Assume $\mathbb{E}\|\nabla f(U)\|^2 < \infty$. Then, for every $h \in \mathcal{H}$,

$$\widehat{\mathcal{C}}_B h \;\xrightarrow[B\to\infty]{\text{a.s.}}\; \mathcal{C} h. \tag{33}$$

Equivalently, $\widehat{\mathcal{C}}_B \to \mathcal{C}$ in the strong operator topology, almost surely. Furthermore, regard $\nabla f(U) \otimes \nabla f(U)$ as a random element of the Hilbert–Schmidt class $\mathcal{S}_2(\mathcal{H})$ (since $\|\nabla f(U) \otimes \nabla f(U)\|_{\mathcal{S}_2} = \|\nabla f(U)\|^2$). Then

$$\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2} \;\xrightarrow[B\to\infty]{\text{a.s.}}\; 0. \tag{34}$$

Consequently, $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \to 0$ almost surely.

**Pointwise convergence** (33)

*Proof.* Fix $h \in \mathcal{H}$ and define the $\mathcal{H}$-valued i.i.d. random variables

$$X_b(h) := \langle h, \nabla f(U_b) \rangle \, \nabla f(U_b) \in \mathcal{H}.$$

Then by definition,

$$\widehat{\mathcal{C}}_B h = \frac{1}{B}\sum_{b=1}^{B} X_b(h),$$

and for each $b = 1, 2, \ldots, B$

$$\mathbb{E}\big[X_b(h)\big] = \mathbb{E}\big[\langle h, \nabla f(U)\rangle \nabla f(U)\big] = \Big(\mathbb{E}\big[\nabla f(U) \otimes \nabla f(U)\big]\Big)h = \mathcal{C}h.$$

Moreover, $\mathbb{E}\|X_b(h)\| \le \mathbb{E}\big(|\langle h, \nabla f(U)\rangle|\, \|\nabla f(U)\|\big) \le \|h\|\, \mathbb{E}\|\nabla f(U)\|^2 < \infty$ by Cauchy–Schwarz inequality.

Thus the strong law of large numbers in separable Hilbert spaces (Araujo and Giné, 1980; Chen and Zhu, 2011; Ledoux and Talagrand, 2013) applies to the i.i.d. sequence $\{X_b(h)\}_{b=1}^{B}$, yielding

$$\widehat{\mathcal{C}}_B h = \frac{1}{B}\sum_{b=1}^{B} X_b(h) \to \mathbb{E}[X_1(h)] = \mathcal{C}h, \quad \text{almost surely.}$$

□

**Operator-norm convergence** (34)

*Proof.* Let $Y_b := \nabla f(U_b) \otimes \nabla f(U_b) \in \mathcal{S}_2(\mathcal{H})$. Then $\widehat{\mathcal{C}}_B = \frac{1}{B}\sum_{b=1}^{B} Y_b$ and $\mathbb{E}\|Y_b\|_{\mathcal{S}_2} = \mathbb{E}\|\nabla f(U)\|^2 < \infty$. Therefore, $Y_b$ is Bochner integrable in the separable Hilbert space $\mathcal{S}_2(\mathcal{H})$, with $\mathbb{E}Y_b = \mathcal{C}$, viewed as an $\mathcal{S}_2$–valued expectation.

By the strong law of large numbers in separable Hilbert spaces (Araujo and Giné, 1980; Chen and Zhu, 2011; Ledoux and Talagrand, 2013),

$$\|\frac{1}{B}\sum_{b=1}^{B} Y_b - \mathbb{E}Y_1\|_{\mathcal{S}_2} \xrightarrow{\text{a.s.}} 0,$$

giving the stated $\mathcal{S}_2$–convergence.

Finally, $\|\cdot\|_{\text{op}} \le \|\cdot\|_{\mathcal{S}_2}$ implies $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\text{op}} \to 0$ a.s. □

**Remark 6** (Operator norm convergence stronger than pointwise convergence)**.** *The Hilbert–Schmidt norm dominates the operator norm, so* (34) *in Theorem 4 implies* $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\text{op}} \to 0$ *a.s., which in turn implies* $\widehat{\mathcal{C}}_B h \to \mathcal{C}h$ *for every fixed* $h \in \mathcal{H}$*. Hence,* (34) *strictly strengthens* (33)*.*

## H Proof of Corollary 1

**Corollary 1 (Extended version).** Assume $\mathbb{E}\|\nabla f(U)\|^2 < \infty$ so that $\widehat{\mathcal{C}}_B \to \mathcal{C}$ in Hilbert–Schmidt (hence operator) norm almost surely. Then

(i) *Eigenvalues.* For every $i \ge 1$,

$$|\sigma_i - \lambda_i| \xrightarrow[B\to\infty]{\text{a.s.}} 0.$$

(ii) *Eigenfunctions.* Fix $i \ge 1$ and suppose $\lambda_i$ is simple and isolated with spectral gap

$$\gamma_i \;:=\; \min\{\lambda_{i-1} - \lambda_i,\ \lambda_i - \lambda_{i+1}\} \;>\; 0 \quad (\lambda_0 := +\infty).$$

Choose the sign of $\widehat{w}_i$ so that $\langle \widehat{w}_i, w_i\rangle \ge 0$. Then $\|\widehat{w}_i - w_i\| \xrightarrow[B\to\infty]{\text{a.s.}} 0$, and quantitatively $\sin\angle(\widehat{w}_i, w_i) \le \frac{\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\text{op}}}{\gamma_i}$, so in particular $\|\widehat{w}_i - w_i\| \le \frac{2}{\gamma_i}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\text{op}} \xrightarrow{\text{a.s.}} 0$.

*Proof.* Since $\mathbb{E}\|\nabla f(U)\|^2 < \infty$ and $\|\nabla f \otimes \nabla f\|_{\mathcal{S}_2} = \|\nabla f\|^2$, the strong law in the separable Hilbert space $\mathcal{S}_2(\mathcal{H})$ yields $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2} \to 0$ a.s., hence $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \to 0$ a.s.

**(i) Eigenvalues.** By Courant–Fischer characterization (Courant and Hilbert, 1954): For a compact self-adjoint $T$ on $\mathcal{H}$ and $i \geq 1$,

$$\lambda_i(T) = \min_{\substack{S \subset \mathcal{H} \\ \dim S = i}} \max_{\substack{x \in S \\ \|x\|=1}} \langle Tx, x \rangle$$

Let $A = \widehat{\mathcal{C}}_B$, $B = \mathcal{C}$, and $E = A - B$. For any fixed $i$ and any $i$-dimensional subspace $S$,

$$\max_{\|x\|=1,\, x \in S} \langle Ax, x \rangle = \max_{\|x\|=1,\, x \in S} \big( \langle Bx, x \rangle + \langle Ex, x \rangle \big) \leq \max_{\|x\|=1,\, x \in S} \langle Bx, x \rangle \; + \; \|E\|_{\mathrm{op}}.$$

Minimize over $S$,

$$\lambda_i(A) = \min_{\dim S = i} \max_{\|x\|=1,\, x \in S} \langle Ax, x \rangle \leq \min_{\dim S = i} \max_{\|x\|=1,\, x \in S} \langle Bx, x \rangle + \|E\|_{\mathrm{op}} = \lambda_i(B) + \|E\|_{\mathrm{op}}.$$

Swapping $A$ and $B$ gives $\lambda_i(B) \leq \lambda_i(A) + \|E\|_{\mathrm{op}}$. Combining,

$$|\lambda_i(A) - \lambda_i(B)| \; \leq \; \|E\|_{\mathrm{op}}.$$

Substituting $A = \widehat{\mathcal{C}}_B$, $B = \mathcal{C}$, and $E = \widehat{\mathcal{C}}_B - \mathcal{C}$,

$$|\sigma_i - \lambda_i| \; \leq \; \|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}},$$

which converges to 0 almost surely by the operator-norm consistency of $\widehat{\mathcal{C}}_B$ (equation (34)).

**(ii) Eigenfunctions.** Fix $i \geq 1$. Suppose $\lambda_i$ is simple and isolated with spectral gap

$$\gamma_i \; := \; \min\{\lambda_{i-1} - \lambda_i, \; \lambda_i - \lambda_{i+1}\} \; > \; 0 \quad (\lambda_0 := +\infty).$$

In simpler terms, this assumption ensures that the i-th mode is a single, well-separated direction in the data. The direction is now identifiable and stable under small perturbations.

Let $P_i := w_i \otimes w_i$ be the spectral projector of $\mathcal{C}$ at $\lambda_i$, and analogously, $\widehat{P}_i := \widehat{w}_i \otimes \widehat{w}_i$ for $\widehat{\mathcal{C}}_B$ at $\sigma_i$. For $B$ large, part (i) implies $\sigma_i \in (\lambda_i - \gamma_i/2, \lambda_i + \gamma_i/2)$ and $\sigma_i$ is the unique eigenvalue of $\widehat{\mathcal{C}}_B$ in this interval. Then by the Davis–Kahan $\sin\Theta$ theorem (Davis and Kahan, 1970; Stewart and Sun, 1990; Yu et al., 2015) for self-adjoint compact perturbations (spectral projectors),

$$\sin \angle(\widehat{w}_i, w_i) \; \leq \; \frac{\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}}}{\gamma_i}.$$

Since $\|\widehat{w}_i - w_i\| \leq 2 \sin \angle(\widehat{w}_i, w_i)$ for aligned unit vectors, the claim follows from $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \to 0$ a.s. (equation (34)). □

**Remark 7.** *To prove convergence of a* single *empirical eigenfunction $\widehat{w}_i$ to the population eigenfunction $w_i$, one typically assumes that $\lambda_i$ is* simple *(multiplicity 1) and* isolated *(there is a spectral gap $\gamma_i > 0$). Under this gap, Davis–Kahan yields* $\sin \angle(\widehat{w}_i, w_i) \; \leq \; \frac{\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}}}{\gamma_i} \xrightarrow[B \to \infty]{a.s.} 0$, *so $\widehat{w}_i \to w_i$ up to sign.*

**Remark 8.** *Without simplicity (i.e., if $\lambda_i = \cdots = \lambda_{i+m-1}$ for some $m > 1$), the eigenvector is* not unique*: any orthonormal basis of the $m$-dimensional eigenspace is valid, and sampling noise can rotate $\widehat{w}_i$ within that space. In this case, individual eigenfunctions need not converge. However, the subspace consistency still holds true, provided there is a* block gap *$\delta > 0$ separating this $m$-fold block from the rest of the spectrum (e.g., $\lambda_{i-1} - \lambda_i \geq \delta$ and $\lambda_{i+m-1} - \lambda_{i+m} \geq \delta$). That is, if $P$ (resp. $\widehat{P}$) denotes the spectral projector onto the population (resp. empirical) $m$-dimensional eigenspace, then*

$$\|\widehat{P} - P\|_{\mathrm{op}} \; \leq \; \frac{\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}}}{\delta} \xrightarrow[B \to \infty]{a.s.} 0.$$

*Thus, with a gap, we can still recover the whole eigenspace consistently.*

# I Proof of Theorem 5 and Corollary 2

**Theorem 5 ($\sqrt{B}$-rate of convergence).** Let $Y := \nabla f(U) \otimes \nabla f(U) \in \mathcal{S}_2(\mathcal{H})$ and assume $\mathbb{E}\|\nabla f(U)\|^4 < \infty$. Then

$$\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2}^2 = \frac{1}{B}\Big(\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2\Big), \tag{35}$$

hence

$$\sqrt{\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2}^2} = O(B^{-1/2}) \quad \text{and} \tag{36}$$

$$\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2} \le \sqrt{\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2}^2} = O(B^{-1/2}). \tag{37}$$

*Proof.* Write $Y_b := g_b \otimes g_b$ and $\Delta_B := \widehat{\mathcal{C}}_B - \mathcal{C} = \frac{1}{B}\sum_{b=1}^{B}(Y_b - \mathbb{E}Y_1)$. Independence and zero mean of $\{Y_b - \mathbb{E}Y_1\}_{b=1}^{B}$ imply

$$\mathbb{E}\|\Delta_B\|_{\mathcal{S}_2}^2 = \frac{1}{B^2}\sum_{b=1}^{B}\mathbb{E}\|Y_b - \mathbb{E}Y_1\|_{\mathcal{S}_2}^2 = \frac{1}{B}\mathbb{E}\|Y_1 - \mathbb{E}Y_1\|_{\mathcal{S}_2}^2.$$

Expanding the variance in $\mathcal{S}_2(\mathcal{H})$,

$$\mathbb{E}\|Y_1 - \mathbb{E}Y_1\|_{\mathcal{S}_2}^2 = \mathbb{E}\|Y_1\|_{\mathcal{S}_2}^2 - 2\,\mathbb{E}\langle Y_1, \mathbb{E}Y_1\rangle_{\mathcal{S}_2} + \|\mathbb{E}Y_1\|_{\mathcal{S}_2}^2.$$

Since $\mathbb{E}\langle Y_1, \mathbb{E}Y_1\rangle_{\mathcal{S}_2} = \langle \mathbb{E}Y_1, \mathbb{E}Y_1\rangle_{\mathcal{S}_2} = \|\mathbb{E}Y_1\|_{\mathcal{S}_2}^2$ and $\mathbb{E}Y_1 = C$, we get

$$\mathbb{E}\|Y_1 - \mathbb{E}Y_1\|_{\mathcal{S}_2}^2 = \mathbb{E}\|Y_1\|_{\mathcal{S}_2}^2 - \|C\|_{\mathcal{S}_2}^2.$$

Finally,

$$\|Y_1\|_{\mathcal{S}_2}^2 = \|\nabla f(U_1) \otimes \nabla f(U_1)\|_{\mathcal{S}_2}^2 = \sum_k \|(\nabla f(U_1) \otimes \nabla f(U_1))e_k\|^2 = \|\nabla f(U)\|^2 \sum_k |\langle e_k, \nabla f(U)\rangle|^2 = \|\nabla f(U)\|^4,$$

where $\{e_k\}_{k\ge1}$ represents any orthonormal basis of $\mathcal{H}$. The second and fourth equalities use Parseval's identity $\sum_{k=1}^{\infty}|\langle e_k, x\rangle|^2 = \|x\|^2$ for $x \in \mathcal{H}$. Therefore, $\|Y_1\|_{\mathcal{S}_2}^2 = \|\nabla f(U)\|^4 < \infty$ yields (35) and hence (36). □

**Corollary 2.** *Assume $\mathbb{E}\|\nabla f(U)\|^4 < \infty$. Then,*

$$\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathrm{op}} \;\le\; \mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2} = O(B^{-1/2}),$$

*and by Chebyshev,*

$$\mathbb{P}\Big(\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2} \ge t\Big) \le \frac{\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2}{B\,t^2}, \tag{38}$$

$$\mathbb{P}\Big(\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathrm{op}} \ge t\Big) \le \frac{\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2}{B\,t^2}. \tag{39}$$

*Proof.* Since $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_{\mathcal{S}_2}$, we have

$$\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \;\le\; \mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2} \;\le\; \sqrt{\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2}^2}\,,$$

By the mean–square identity from Theorem 5,

$$\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2}^2 = \frac{\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2}{B},$$

hence $\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2} = O(B^{-1/2})$ and therefore also $\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} = O(B^{-1/2})$.

For the tail bounds, Chebyshev's inequality gives

$$\mathbb{P}\Big(\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2} \geq t\Big) \;\leq\; \frac{\mathbb{E}\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2}^2}{t^2} \;=\; \frac{\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2}{B\,t^2}.$$

Finally, since $\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \leq \|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2}$,

$$\mathbb{P}\Big(\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathrm{op}} \geq t\Big) \;\leq\; \mathbb{P}\Big(\|\widehat{\mathcal{C}}_B - \mathcal{C}\|_{\mathcal{S}_2} \geq t\Big) \;\leq\; \frac{\mathbb{E}\|\nabla f(U)\|^4 - \|C\|_{\mathcal{S}_2}^2}{B\,t^2}.$$

□

**Remark 9.**

*(i)* ***Finite second moment assumption for a.s. consistency.*** *With $Y := \nabla f(U) \otimes \nabla f(U) \in \mathcal{S}_2(\mathcal{H})$, $\|Y\|_{\mathcal{S}_2} = \|\nabla f(U)\|^2$, so $\mathbb{E}\|Y\|_{\mathcal{S}_2} < \infty$ whenever $\mathbb{E}\|\nabla f(U)\|^2 < \infty$. Thus to apply strong law of large numbers for separable Hilbert spaces, we assume $E\|\nabla f(U)\|^2 < \infty$.*

*(ii)* ***Finite fourth moment assumption for convergence rate.*** *Note that $\mathbb{E}\big\|\widehat{\mathcal{C}}_B - \mathcal{C}\big\|_{\mathcal{S}_2}^2 = \frac{1}{B}\Big(\mathbb{E}\|Y\|_{\mathcal{S}_2}^2 - \|C\|_{\mathcal{S}_2}^2\Big)$, where $\|Y\|_{\mathcal{S}_2}^2 = \|\nabla f(U)\|^4$. Thus we assume $\mathbb{E}\|\nabla f(U)\|^4 < \infty$ to ensure finite variance and hence, for the $B^{-1/2}$ rate in $L^2$ to hold.*

## J Proof of Theorem 6

*Proof.* Let $Y_b := \nabla f(U_b) \otimes \nabla f(U_b) \in \mathcal{S}_2(\mathcal{H})$ and $Z_b := Y_b - \mathbb{E}Y_1$, so that $\mathbb{E}Z_b = 0$, the $Z_b$ are i.i.d. in the separable Hilbert space $\mathcal{S}_2(\mathcal{H})$, and

$$\sqrt{B}\,\big(\widehat{\mathcal{C}}_B - \mathcal{C}\big) = \frac{1}{\sqrt{B}} \sum_{b=1}^{B} Z_b =: S_B.$$

Define the covariance operator $\mathcal{K} : \mathcal{S}_2(\mathcal{H}) \to \mathcal{S}_2(\mathcal{H})$ by

$$\mathcal{K} \;:=\; \mathbb{E}\big[\, Z \otimes Z \,\big], \qquad \langle \mathcal{K}(W_1), W_2 \rangle_{\mathcal{S}_2} = \mathrm{Cov}\big(\langle Z, W_1 \rangle_{\mathcal{S}_2},\, \langle Z, W_2 \rangle_{\mathcal{S}_2}\big).$$

**CLT for scalar projections.** Fix $W \in \mathcal{S}_2(\mathcal{H})$ and set $\xi_b(W) := \langle Z_b, W \rangle_{\mathcal{S}_2} \in \mathbb{R}$. Then $(\xi_b(W))_{b \geq 1}$ are i.i.d., mean zero, with $\mathrm{Var}(\xi_1(W)) = \langle \mathcal{K}W, W \rangle_{\mathcal{S}_2} < \infty$ since $\mathbb{E}\|Z\|_{\mathcal{S}_2}^2 < \infty$. By Cauchy–Schwarz, $|\xi_b(W)| \leq \|Z_b\|_{\mathcal{S}_2}\,\|W\|_{\mathcal{S}_2}$. Hence under Lindeberg's conditon in Assumption 1, for every $\varepsilon > 0$,

$$\mathbb{E}\Big[\xi_1(W)^2\,\mathbf{1}\{|\xi_1(W)| > \varepsilon\sqrt{B}\}\Big] \;\leq\; \|W\|_{\mathcal{S}_2}^2\,\mathbb{E}\Big[\|Z\|_{\mathcal{S}_2}^2\,\mathbf{1}\Big\{\|Z\|_{\mathcal{S}_2} > \frac{\varepsilon}{\|W\|_{\mathcal{S}_2}}\sqrt{B}\Big\}\Big] \;\xrightarrow[B\to\infty]{}\; 0. \tag{40}$$

Note that (40) also holds true under the Lyapunov condition. Using $|\xi| \leq \|Z\|_{\mathcal{S}_2}\,\|W\|_{\mathcal{S}_2}$, we have $|\xi|^{2+\eta} \leq \|W\|_{\mathcal{S}_2}^{2+\eta}\,\|Z\|_{\mathcal{S}_2}^{2+\eta}$. Also, for any $t > 0$ and $\eta > 0$, $x^2\mathbf{1}\{x > t\} \;\leq\; \frac{x^{2+\eta}}{t^\eta}$, for all $x \geq 0$. Substituting $t = \varepsilon\sqrt{B}$, this gives

$$\mathbb{E}\Big[\xi^2\,\mathbf{1}\{|\xi| > \varepsilon\sqrt{B}\}\Big] \;\leq\; \frac{\mathbb{E}|\xi|^{2+\eta}}{(\varepsilon\sqrt{B})^\eta} \;\leq\; \frac{\|W\|_{\mathcal{S}_2}^{2+\eta}\,\mathbb{E}\|Z\|_{\mathcal{S}_2}^{2+\eta}}{\varepsilon^\eta\,B^{\eta/2}}.$$

Finally, by the Lyapunov assumption $\mathbb{E}\|Z\|_{\mathcal{S}_2}^{2+\eta} < \infty$, $O(B^{-\eta/2}) \to 0$ as $B \to \infty$.

Hence, by (40), the (scalar) Lindeberg condition also holds for $\xi_b(W)$. Then the Lindeberg–Feller CLT yields

$$\frac{1}{\sqrt{B}} \sum_{b=1}^{B} \xi_b(W) \;\Rightarrow\; \mathcal{N}\big(0,\; \langle \mathcal{K}W, W \rangle_{\mathcal{S}_2}\big), \quad \text{i.e.} \quad \langle S_B, W \rangle_{\mathcal{S}_2} \;\Rightarrow\; \mathcal{N}\big(0,\; \langle \mathcal{K}W, W \rangle_{\mathcal{S}_2}\big). \tag{41}$$

**Tightness in $\mathcal{S}_2(\mathcal{H})$.** Using independence and $\mathbb{E}Z_b = 0$,

$$\mathbb{E}\|S_B\|_{\mathcal{S}_2}^2 = \mathbb{E}\left\langle \frac{1}{\sqrt{B}}\sum_{b=1}^{B} Z_b,\ \frac{1}{\sqrt{B}}\sum_{b'=1}^{B} Z_{b'} \right\rangle_{\mathcal{S}_2} = \mathbb{E}\|Z\|_{\mathcal{S}_2}^2 \quad \text{(independent of } B\text{)}.$$

Hence, for any $R > 0$,

$$\sup_{B\geq 1} \mathbb{P}\big(\|S_B\|_{\mathcal{S}_2} > R\big) \ \leq\ \frac{\sup_B \mathbb{E}\|S_B\|_{\mathcal{S}_2}^2}{R^2} \ =\ \frac{\mathbb{E}\|Z\|_{\mathcal{S}_2}^2}{R^2} \ \xrightarrow[R\to\infty]{} \ 0, \tag{42}$$

so $\{S_B\}$ is tight in the Hilbert space $\mathcal{S}_2(\mathcal{H})$.

**Hilbert–space CLT.** By (41), all one-dimensional projections $\langle S_B, W\rangle_{\mathcal{S}_2}$ converge to centered normals with the correct variances, and by Step (42), $\{S_B\}$ is tight. In a separable Hilbert space this implies

$$S_B \ \Rightarrow\ Z_\infty \ \text{ in } \mathcal{S}_2(\mathcal{H}), \tag{43}$$

where $Z_\infty$ is a mean-zero Gaussian element with covariance operator $\mathcal{K}$, characterized by $\langle \mathcal{K}(W_1), W_2\rangle_{\mathcal{S}_2} = \mathrm{Cov}(\langle Z, W_1\rangle_{\mathcal{S}_2}, \langle Z, W_2\rangle_{\mathcal{S}_2})$. Recalling $S_B = \sqrt{B}\,(\widehat{\mathcal{C}}_B - \mathcal{C})$ completes the proof. □

## K Additional discussion and proof of Theorem 3

**Assumption 3** (Inactive–invariance)**.** The gradient depends only on the active component:

$$\nabla f(P_{\mathcal{B}} u) = \nabla f(u) \qquad \text{for } \rho\text{-a.e. } u \in \mathcal{H}.$$

Equivalently, $f$ is (Gateaux) flat along $\mathcal{B}^\perp$, i.e. $Df(u)[v] = 0$ for all $v \in \mathcal{B}^\perp$ and $\rho$-a.e. $u$.

**Theorem 3 (Extended version).** Assume $\mathbb{E}\,\|\nabla f(U)\|^2 < \infty$. Let $\mathcal{C}^{1/2}$ denote the unique positive square root of the gradient covariance (trace-class) operator $\mathcal{C} := \mathbb{E}\big[\nabla f(U) \otimes \nabla f(U)\big] \in \mathcal{L}_1(\mathcal{H})$. Let $\mathcal{B} := \overline{\mathrm{ran}(\mathcal{C}^{1/2})} = (\ker \mathcal{C})^\perp$ and let $P_{\mathcal{B}}$ denote the orthogonal projector onto $\mathcal{B}$.

Further, let $L := \mathcal{C}^{1/2} : \mathcal{H} \to \mathcal{B}$ and let $L^\dagger$ denote the Moore–Penrose pseudoinverse of $L$, i.e. $L^\dagger = \mathcal{C}^{-1/2}$ on $\mathcal{B}$ and $L^\dagger = 0$ on $\mathcal{B}^\perp$. Define the *warped input* $V := LU \in \mathcal{B}$ and the *warped functional*

$$\tilde{f} : \mathcal{B} \to \mathbb{R}, \qquad \tilde{f}(v) := f(L^\dagger v).$$

Let $\rho_V$ be the pushforward of $\rho$ under $L$ (the law of $V$). Then:

(i) $\tilde{f}$ is Fréchet differentiable on $\mathcal{B}$, and its gradient (Riesz representer) satisfies

$$\nabla_v \tilde{f}(v) \ =\ L^\dagger\, \nabla f(L^\dagger v) \ \in\ \mathcal{B} \qquad (v \in \mathcal{B}).$$

(ii) Under Assumption 3, the warped gradient covariance operator

$$\mathcal{C}_{\text{warp}} \ :=\ \mathbb{E}_{\rho_V}\big[\nabla_v \tilde{f}(V) \otimes \nabla_v \tilde{f}(V)\big] \ :\ \mathcal{H} \to \mathcal{H}$$

equals the orthogonal projector onto the active subspace

$$\mathcal{C}_{\text{warp}} \ =\ P_{\mathcal{B}}.$$

Equivalently, in any eigenbasis $\{w_i\}_{i\geq 1}$ of $\mathcal{C}$ with $\mathcal{C}w_i = \lambda_i w_i$,

$$\mathbb{E}_{\rho_V}\big[\langle \nabla_v \tilde{f}(V), w_i\rangle^2\big] = \begin{cases} 1, & \lambda_i > 0, \\ 0, & \lambda_i = 0, \end{cases}$$

so all *active* directions are equalized to unit sensitivity and all *inactive* directions carry zero sensitivity.

*Proof.* Note that $L := \mathcal{C}^{1/2}$ is (bounded) self-adjoint and positive, and so its Moore–Penrose pseudoinverse $L^\dagger$. Using the spectral decomposition in Theorem 1 and definition of $\mathcal{B} := (\ker \mathcal{C})^\perp$, it is easy to see

$$\mathcal{B} \;:=\; \operatorname{span}\{\, w_i : \; \lambda_i > 0 \,\} \qquad P_{\mathcal{B}} \;:=\; \sum_{\lambda_j > 0} w_j \otimes w_j$$

Let $T : \mathcal{H} \to \mathcal{H}$ be bounded linear and set $H(v) := f(Tv)$. Then

$$DH(v)[h] = Df(Tv)[Th] = \langle \nabla f(Tv), Th \rangle = \langle T^* \nabla f(Tv), h \rangle,$$

so $\nabla H(v) = T^* \nabla f(Tv)$. With $T = L^\dagger = (L^\dagger)^*$ and $H = \tilde{f}$ we obtain

$$\nabla_v \tilde{f}(v) = L^\dagger \nabla f(L^\dagger v) \in \mathcal{B}.$$

Since $V := LU$ and $\rho_V = \rho \circ L^{-1}$ (pushforward of $\rho$ by $L$), for any integrable $\Phi$ we have $\mathbb{E}_{\rho_V}[\Phi(V)] = \mathbb{E}_\rho[\Phi(LU)]$. Applying this for $\Phi(v) \;:=\; \nabla_v \tilde{f}(v) \otimes \nabla_v \tilde{f}(v) \;=\; \big(L^\dagger \nabla f(L^\dagger v)\big) \;\otimes\; \big(L^\dagger \nabla f(L^\dagger v)\big)$, we obtain

$$\begin{aligned}
\mathcal{C}_{\text{warp}} &= \mathbb{E}_{\rho_V}\big[\nabla_v \tilde{f}(V) \otimes \nabla_v \tilde{f}(V)\big] \\
&= \mathbb{E}_\rho\big[\big(L^\dagger g\big(L^\dagger(LU)\big)\big) \otimes \big(L^\dagger g\big(L^\dagger(LU)\big)\big)\big] \\
&= \mathbb{E}_\rho\big[\big(L^\dagger \nabla f(P_{\mathcal{B}} U)\big) \otimes \big(L^\dagger \nabla f(P_{\mathcal{B}} U)\big)\big] \qquad (\text{since } L^\dagger L = P_{\mathcal{B}}) \\
&= L^\dagger \, \mathbb{E}_\rho\big[\nabla f(P_{\mathcal{B}} U) \otimes \nabla f(P_{\mathcal{B}} U)\big] \, (L^\dagger)^*.
\end{aligned}$$

By Assumption 3,

$$\mathcal{C}_{\text{warp}} = L^\dagger \, \mathbb{E}_\rho\big[\nabla f(U) \otimes \nabla f(U)\big] \, (L^\dagger)^* = L^\dagger \mathcal{C} (L^\dagger)^* = L^\dagger \mathcal{C} L^\dagger,$$

since $(Sa) \otimes (Sb) = S\,(a \otimes b)\,S^*$ and $(L^\dagger)^* = L^\dagger$.

By definition, $L^\dagger = \mathcal{C}^{-1/2}$ on $\mathcal{B}$ and $L^\dagger = 0$ on $\mathcal{B}^\perp$. Hence

$$L^\dagger \mathcal{C} L^\dagger = \begin{cases} I & \text{on } \mathcal{B}, \\ 0 & \text{on } \mathcal{B}^\perp, \end{cases}$$

i.e. $L^\dagger \mathcal{C} L^\dagger = P_{\mathcal{B}}$. Therefore $\mathcal{C}_{\text{warp}} = P_{\mathcal{B}}$. The directional statement follows since $\langle P_{\mathcal{B}} w_i, w_i \rangle = \mathbf{1}_{\{\lambda_i > 0\}}$. □

**Remark 10.** *Assumption 3 is invoked only in Theorem 3 and serves as an additional constraint specific to this result. All core theoretical developments, including the definitions, existence, spectral properties, and consistency/convergence of $\mathcal{C}$, hold independently of this assumption. Hence, our framework genuinely extends active-subspace methodology to arbitrary functions on infinite-dimensional Hilbert spaces, beyond the classical ridge-function setting.*

# L Details of Numerical Experiments

## L.1 Practical Aspects of Gradient Computation in Function Space

While, conceptually speaking, calculation of the gradient functional can be achieved automatically using the adjoint method, in practice, numerical issues mean that special attention is required to ensure the gradients are reasonable in practice. For example, extreme oscillations can be observed in gradient functionals. This can be mitigated in several ways. For one, choosing to a Riesz representer for the gradient in a space which encourages smoothness, such as a Sobolev space instead of $L^2$, can lead to improved gradients. Additionally, the actual approximation space itself can impact the smoothness of the gradient. Finally, the gradients can also be explicitly smoothed with a filter. In our Laminar jet problem, we found that using an linear CG space lead to much smoother gradients than using the quadratic CG space used by the authors.

In the Kirigami problem, we actually used an H1 inner product for the Riesz representer, while maintaining the sophisticated mixed approximation space of the authors. We subsequently provided further smoothing using a Helmholtz filter with lengthscale 0.5.

### L.2 Application Problem Details

We now describe in additional details the Laminar Jet and Kirigami Electronics test problems from our article, which are based on existing articles in the academic literature.

#### L.2.1 Laminar Jet Problem Details

On this problem, the domain of the PDE is a two dimensional rectangle given by $\mathcal{D} = (0, L_x) \times (-L_y/2, L_y/2)$. The boundary is broken up into three parts, the "left" part $\Gamma_{\text{in}} = \{x = 0\}$, which is called the *inlet*, the "right" part $\Gamma_{\text{out}} = \{x = L_x\}$ called the *outlet*, and the remaining part $\Gamma_w = \{y = \pm L_y/2\}$, which may be thought of as representing the walls of the test chamber.

The velocity $v : \mathcal{D} \to \mathbb{R}^2$ and pressure $p : \mathcal{D} \to \mathbb{R}$ solve the Navier-Stokes equations:

$$-\nu \operatorname{div}\left(\nabla v + \nabla v^\top\right) + v \cdot \nabla v + \nabla p = 0 \quad \text{in } \mathcal{D}, \tag{44}$$

$$\operatorname{div} v = 0 \quad \text{in } \mathcal{D}, \tag{45}$$

where $\nu > 0$ is the viscosity.

Let $n$ denote the outward normal to the mesh boundary. Then we have the following boundary conditions:

$$v \cdot n + u = 0 \quad \text{on } \Gamma_{\text{in}}. \tag{46}$$

$$\sigma_n + \beta(v \cdot n)^- v = 0 \text{ on } \Gamma_w \tag{47}$$

$$v \cdot n = 0 \text{ on } \Gamma_{\text{out}} \tag{48}$$

The second equation implements the backflow stabilization while the third equation prevents anything from going through walls. The first boundary condition is perhaps most interesting, as it contains the function $u$ which we seek to optimize, giving the profile of the flow at the inlet. This is a one-dimensional function on $-L_y/2, L_y/2$. This problem was obtained from Beskos et al. (2017), and more details may be found within. The code used in this project was adapted from their codebase[2].

#### L.2.2 Kirigami Electronics Problem Details

This problem involves a five dimensional state-space, consisting of the 3 dimensional location of the irregularly shape electrode in space, as well as 2 dimensions giving the orientation of the electrode. The state equations involve balancing elastic energy, which keeps the form of the material, with the force from gravity interacting with the medium that the electrodes are in. The boundary conditions are that the circular outer rim of the electrode must be at zero displacement.

The task is to find a configuration which minimizes the energy. To simplify the problem, we accept as input only the 3 dimensional location of the lattice, and numerically optimize the orientation conditional on that configuration, which is much simpler than solving the entire problem. We set the load parameter $P = 5$, which is a difficult problem to solve and involves solving a sequence of problems for increasing $P$ starting at $P = 0$.

This problem was obtained from Yang et al. (2024), and more details may be found within. We used the authors' codebase[3]. as a starting point for our code.

[2] `https://bitbucket.org/lanzithinking/geom-infmcmc/src/master/`
[3] `https://github.com/cforro/KiriE_2023_NatureBiotech`