Generalized bounds for active subspaces[*]

Mario Teixeira Parente

Chair for Numerical Mathematics
Technical University of Munich
e-mail: parente@ma.tum.de

and

Jonas Wallin

Department of Statistics
Lund University
e-mail: jonas.wallin@stat.lu.se

and

Barbara Wohlmuth

Chair for Numerical Mathematics
Technical University of Munich
e-mail: wohlmuth@ma.tum.de

Abstract: In this article, we consider scenarios in which traditional estimates for the active subspace method based on probabilistic Poincaré inequalities are not valid due to unbounded Poincaré constants. Consequently, we propose a framework that allows to derive generalized estimates in the sense that it enables to control the trade-off between the size of the Poincaré constant and a weaker order of the final error bound. In particular, we investigate independently exponentially distributed random variables in dimension two or larger and give explicit expressions for corresponding Poincaré constants showing their dependence on the dimension of the problem. Finally, we suggest possibilities for future work that aim for extending the class of distributions applicable to the active subspace method as we regard this as an opportunity to enlarge its usability.

Contents

1. Introduction

Many modern computational problems, having a large number of input variables or parameters, suffer from the "curse of dimensionality" in that their solution becomes computationally expensive or even intractable as the dimension of the problem grows. The *active subspace method* (ASM), or shorter, *active subspaces* [17, 18], is a set of tools for dimension reduction which reduce the effects caused by the curse of dimensionality. ASM splits an Euclidean input space into a so-called active and inactive subspace based on average sensitivities of a real-valued function of interest. The sensitivities are found by an eigendecomposition of weighted outer products of the function's gradient with itself. That is, eigenvalues indicate average sensitivities of a function of interest in the direction of the corresponding eigenvector. Eigenvectors and eigenvalues belonging to the active subspace are then considered as dominant for the global behavior of the function of interest, whereas the inactive subspace is regarded as negligible.

The usefulness of ASM has already been demonstrated for several real case studies in various applied disciplines; see, e.g., [22, 31, 36, 37, 39]. It has also motivated other methodological advances, e.g., in the solution of Bayesian inverse problems [35] by an accelerated Markov chain Monte Carlo algorithm [20], in uncertainty quantification and propagation [15, 40], and in the theory of ridge approximation; see, e.g., [19, 24, 25].

However, ASM is only one dimension reduction technique among others. For example, likelihood-informed dimension reduction for the solution of Bayesian inverse problems [21] is based on a similar idea. This approach, however, analyzes the Hessian matrix of the function of interest instead of the gradient. An extension to vector-valued functions in gradient-based dimension reduction is given by [45]. Dimension reduction for nonlinear Bayesian inverse problems based on the Kullback-Leibler (KL) divergence of approximate posteriors and (subspace) logarithmic Sobolev inequalities, including a comprehensive comparison of several other techniques, was provided by the authors of [46]. Furthermore, Active

Manifolds [11], as a nonlinear analogue to ASM, and PTU [13], as an extension to a framework for nonlinear dimension reduction called Isomap [38], both have demonstrated a lot of promise.

A main result in ASM theory is an upper bound on the mean squared error between the original function of interest and its low-dimensional approximation on the active subspace. The corresponding proof is based on an inequality of Poincaré type which is probabilistic in nature since ASM involves a probability distribution that weights sensitivities of the function of interest at different locations in the input space. The upper bound consists of the product of a Poincaré type constant and the sum of eigenvalues corresponding to the inactive subspace, called inactive trace in the following. The constant derived in [18] is claimed to depend only on the original distribution which is generally incorrect. Also, to the knowledge of the authors, existing theory for dimension reduction techniques based on Poincaré or logarithmic Sobolev inequalities are subject to quite restrictive assumptions on the involved probability distribution. These assumptions comprise either the distribution having compact support or its density ρ being of uniformly log-concave form, i.e., $\rho(\boldsymbol{x}) = \exp(-V(\boldsymbol{x}))$, where V is such that its Hessian matrix $V''(\boldsymbol{x}) \succeq \alpha I$ for each $\boldsymbol{x}$ and some $\alpha > 0$. By the famous Bakry-Émery criterion, the latter assumption implies a logarithmic Sobolev inequality and Poincaré inequality with universal Poincaré constant $1/\alpha$; see, e.g., [3, 41]. Note that the case $\alpha = 0$, i.e., V being only convex, is not covered. However, Bobkov [8] showed that a Poincaré inequality is still satisfied in this case and gave lower and upper bounds on the corresponding Poincaré constant. Distributions with heavier tails, i.e., for $\alpha = 0$, as, e.g., exponential or Laplace distributions, do not satisfy the assumptions above, but are, however, of practical relevance.

In ASM theory, it is not the original distribution that must satisfy a Poincaré inequality, but a conditional distribution on the inactive subspace, which depends on a variable defined on the active subspace. Both assumptions on the original distribution from above are in fact passed on to the conditional distribution. However, the case $\alpha = 0$ is cumbersome. We shall give an example for this case regarding a distribution that itself satisfies a Poincaré inequality, but might not be applicable at all or only with care due to an arbitrary large Poincaré constant in the final bound for the mentioned mean squared error. Our arguments are based on the bounds for corresponding Poincaré constants given by Bobkov in [8]. We also describe a way to still get upper bounds in this situation, however with a weaker, reduced order in the inactive trace. This order reduction is controllable in the sense that the practitioner can decide for the actual trade-off between the order of the inactive trace and the size of the corresponding Poincaré constant. The mentioned general problem and its solution is exemplified on independently exponentially distributed random variables in dimension two and larger. Also, it is shown that the final constant is very much depending on the dimension of the problem. However, since this example is rather special, we eventually propose opportunities for future work that aim for extending the class of distributions for which the bounds and the involved constants are explicitly available in order to expand the applicability of ASM

to more scenarios of practical interest. In particular, the class of multivariate generalized hyperbolic distributions is a rich class that is, in our opinion, worthwhile to get investigated. Details on arising difficulties with this class are also provided.

The outline of the manuscript is as follows. Section 2 gives an introduction to ASM and its formal context. In Section 3, we recall results involving compactly supported and normal distributions. The main results consisting of a motivation and discussion of the mentioned problems, with independently exponentially distributed random variables as an extreme example, are presented in Section 4. In Section 5, we propose possibilities for future work. Finally, a summary is given in Section 6.

2. Active subspaces

The active subspace method is a set of tools for gradient-based dimension reduction [17, 18]. Its aim is to find directions in the domain of a function f along which the function changes dominantly, on average. For illustration, consider a function of the form $f(\boldsymbol{x}) = g(A^\top \boldsymbol{x})$ with a so-called profile function g and a matrix $A \in \mathbf{R}^{n \times k}$, $1 \leq k \leq n$, $n \geq 2$. Functions of this type are called *ridge functions* [34]. Note that f is constant along the null space of $A^\top$. Indeed, for $\boldsymbol{x} \in \mathrm{dom}(f) \subseteq \mathbf{R}^n$ and $\boldsymbol{v} \in \mathcal{N}(A^\top)$ such that $\boldsymbol{x} + \boldsymbol{v} \in \mathrm{dom}(f)$, it holds that

$$f(\boldsymbol{x} + \boldsymbol{v}) = g(A^\top(\boldsymbol{x} + \boldsymbol{v})) = g(A^\top \boldsymbol{x}) = f(\boldsymbol{x}). \tag{2.1}$$

That is, f is intrinsically at most k-dimensional. For arbitrary f, the general task is to find a suitable dimension k, a function $g : \mathrm{dom}(g) \to \mathbf{R}$, $\mathrm{dom}(g) \subseteq \mathbf{R}^k$, and a matrix $A \in \mathbf{R}^{n \times k}$ such that $f(\boldsymbol{x}) \approx g(A^\top \boldsymbol{x})$.

For this, the active subspace method assumes that the function of interest $f : \mathcal{X} \to \mathbf{R}$ is continuously differentiable with partial derivatives that are square-integrable w.r.t. a probability density function $\rho_{\boldsymbol{X}}$. We define $\mathcal{X} := \mathrm{dom}(f) \subseteq \mathbf{R}^n$ to be the support of $\rho_{\boldsymbol{X}}$, i.e., the closure of the set $\mathcal{X}^+ := \{\boldsymbol{x} \in \mathbf{R}^n \,|\, \rho_{\boldsymbol{X}}(\boldsymbol{x}) > 0\}$. We assume that $\mathcal{X}$ is a continuity set, that is, its boundary is a Lebesgue null set. The central object of interest is a matrix constructed by outer products of the gradient of f, $\nabla f = \nabla^{\boldsymbol{x}} f$, with itself weighted by $\rho_{\boldsymbol{X}}$,

$$C := \int_{\mathbf{R}^n} \nabla f(\boldsymbol{x}) \, \nabla f(\boldsymbol{x})^\top \rho_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{2.2}$$

Since C is real symmetric, there exists an eigendecomposition $C = W \Lambda W^\top$ with an orthogonal matrix $W \in \mathbf{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbf{R}^{n \times n}$ with descending eigenvalues $\lambda_1, \ldots, \lambda_n$ on its diagonal. The positive semidefiniteness of C additionally ensures that $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Note that the matrices C, W, and Λ all depend on f and $\rho_{\boldsymbol{X}}$.

The behavior of the function f and the eigendecomposition of C have an interesting, exploitable relation, i.e.,

$$\lambda_i = \boldsymbol{w}_i^\top C \boldsymbol{w}_i = \int_{\mathbf{R}^n} (\boldsymbol{w}_i^\top \nabla f(\boldsymbol{x}))^2 \rho_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad i = 1, \ldots, n. \tag{2.3}$$

If, for example, $\lambda_i = 0$ for some i, then we can conclude that f does not change in the direction of the corresponding eigenvector $\boldsymbol{w}_i$. That is, if eigenvalues λ_i, $i = k+1, \ldots, n$, are sufficiently small for a suitable $k \leq n-1$, or even zero as in the case of ridge functions, then f can be approximated by a lower-dimensional function. Formally, this corresponds to a split of Λ and W, i.e.,

$$\Lambda = \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} W_1 & W_2 \end{pmatrix}, \tag{2.4}$$

where $\Lambda_1 \in \mathbf{R}^{k \times k}$, $\Lambda_2 \in \mathbf{R}^{n-k \times n-k}$ and $W_1 \in \mathbf{R}^{n \times k}$, $W_2 \in \mathbf{R}^{n \times n-k}$.

Since

$$\boldsymbol{x} = WW^\top \boldsymbol{x} = W_1 W_1^\top \boldsymbol{x} + W_2 W_2^\top \boldsymbol{x} = W_1 \boldsymbol{y} + W_2 \boldsymbol{z}, \tag{2.5}$$

the split of W suggests a new coordinate system $(\boldsymbol{y}, \boldsymbol{z})$ for the *active variable* $\boldsymbol{y} := W_1^\top \boldsymbol{x} \in \mathbf{R}^k$ and the *inactive variable* $\boldsymbol{z} := W_2^\top \boldsymbol{x} \in \mathbf{R}^{n-k}$. The range of W_1, $\mathcal{R}(W_1) := \{W_1 \boldsymbol{y} \mid \boldsymbol{y} \in \mathbf{R}^k\} \subseteq \mathbf{R}^n$, is called the *active subspace of f*. Note that the new variable $\boldsymbol{y}$ is aligned to directions on which f changes much more, on average, than on directions the variable $\boldsymbol{z}$ is aligned to.

For the remainder, we define

$$\mathcal{Y} := W_1^\top \mathcal{X} = \{W_1^\top \boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{X}\} \quad \text{and} \quad \mathcal{Z} := W_2^\top \mathcal{X} = \{W_2^\top \boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{X}\}. \tag{2.6}$$

Also, for $\boldsymbol{y} \in \mathcal{Y}$ and $\boldsymbol{z} \in \mathcal{Z}$, let

$$[\![\boldsymbol{y}, \boldsymbol{z}]\!] := [\![\boldsymbol{y}, \boldsymbol{z}]\!]_W := W_1 \boldsymbol{y} + W_2 \boldsymbol{z} \tag{2.7}$$

to concisely denote changes of the coordinate system.

Variables $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ can also be regarded as random variables $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, respectively, that are defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The orthogonal variable transformation $\boldsymbol{x} \mapsto (\boldsymbol{y}, \boldsymbol{z})$ induces probability density functions for the random variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$. That is, the joint distribution of $(\boldsymbol{Y}, \boldsymbol{Z})$ is

$$\rho_{\boldsymbol{Y}, \boldsymbol{Z}}(\boldsymbol{y}, \boldsymbol{z}) = \rho_{\boldsymbol{X}}([\![\boldsymbol{y}, \boldsymbol{z}]\!]) \tag{2.8}$$

for $\boldsymbol{y} \in \mathcal{Y}$ and $\boldsymbol{z} \in \mathcal{Z}$. Corresponding marginal and conditional densities are defined as usual. Additionally, set

$$\mathcal{Y}^+ := \{\boldsymbol{y} \in \mathbf{R}^k \mid \rho_{\boldsymbol{Y}}(\boldsymbol{y}) > 0\} \tag{2.9}$$

to denote the set of all values for the active variable $\boldsymbol{y}$ with a strictly positive density value. We frequently use that for a $\rho_{\boldsymbol{X}}$-integrable function $h : \mathcal{X} \to \mathbf{R}$, it holds that

$$\mathbf{E}[h(\boldsymbol{X})] = \mathbf{E}[\mathbf{E}[h([\![\boldsymbol{Y}, \boldsymbol{Z}]\!]) \mid \boldsymbol{Y}]]. \tag{2.10}$$

Given the eigenvectors in W, we still need to define a lower-dimensional function g approximating f. For $\boldsymbol{y} \in \mathcal{Y}^+$, a natural way is to define $g(\boldsymbol{y})$ as the conditional expectation of f given $\boldsymbol{y}$, i.e., as an integral over the inactive subspace

weighted with the conditional density $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$. Recall that this approximation is the best in an L^2 sense [28, Corollary 8.17]. Hence, we set

$$
\begin{aligned}
g(\boldsymbol{y}) &:= \mathbf{E}[f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!]) \,|\, \boldsymbol{Y} = \boldsymbol{y}] \\
&= \int_{\mathbf{R}^{n-k}} f([\![\boldsymbol{y}, \boldsymbol{z}]\!]) \, \rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{z}.
\end{aligned}
\tag{2.11}
$$

Additionally, we define

$$
f_g(\boldsymbol{x}) := g(W_1^\top \boldsymbol{x})
\tag{2.12}
$$

for $\boldsymbol{x} \in \mathcal{X}^\circ$, where $\mathcal{X}^\circ$ denotes the interior of $\mathcal{X}$. Note that $W_1^\top \boldsymbol{x} \in \mathcal{Y}^\circ \subseteq \mathcal{Y}^+$ for $\boldsymbol{x} \in \mathcal{X}^\circ$.

Remark. In practice, both the matrix C from (2.2) and the low-dimensional function g from (2.11) are often not exactly available. Our results can, however, also be adapted to a corresponding perturbation analysis, provided in [18], which we do not perform since it would require additional notation and complexity but not contribute to the central aspects of this manuscript.

One of the main results in ASM theory is a theorem that gives an upper bound on the mean squared error of f_g approximating f. The upper bound is the product of a Poincaré constant $C_{\mathrm{P},W} > 0$ and the sum of $n - k$ eigenvalues corresponding to the inactive subspace, called inactive trace. That is, if the inactive trace is small, then the mean squared error of f_g approximating f is also small. Mathematically, for a given probability density function $\rho_{\boldsymbol{X}}$, the theorem states that [18, Theorem 3.1]

$$
\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] \leq C_{\mathrm{P},W}(\lambda_{k+1} + \cdots + \lambda_n)
\tag{2.13}
$$

for a Poincaré constant $C_{\mathrm{P},W} = C_{\mathrm{P},W}(W, \rho_{\boldsymbol{X}}) > 0$. Note that $C_{\mathrm{P},W}$ depends on $W = W(f)$ and thus also indirectly on f. If desired, we could remove this dependence by considering the supremum of $C_{\mathrm{P},W}$ over all orthogonal matrices, i.e.,

$$
C_{\mathrm{P}} := \sup_{W \mathrm{orth.}} C_{\mathrm{P},W},
\tag{2.14}
$$

and get

$$
\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] \leq C_{\mathrm{P}}(\lambda_{k+1} + \cdots + \lambda_n),
\tag{2.15}
$$

provided the constant $C_{\mathrm{P}} = C_{\mathrm{P}}(\rho_{\boldsymbol{X}})$ exists. Deriving such an upper bound for a certain class of distributions would allow to choose $\rho_{\boldsymbol{X}}$ independently of f. Note that [45, 46] also control the Poincaré constant for any orthogonal matrix W.

The derivation of (2.13) starts with

$$
\begin{aligned}
\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] &= \mathbf{E}[\mathbf{E}[(f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!]) - g(\boldsymbol{Y}))^2 \,|\, \boldsymbol{Y}]] \\
&\leq \mathbf{E}[C_{\boldsymbol{Y}} \, \mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]],
\end{aligned}
\tag{2.16}
\tag{2.17}
$$

where we used a probabilistic Poincaré inequality w.r.t. $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ for a given $\boldsymbol{y} \in \mathcal{Y}^+$. Note that the Poincaré constant $C_{\boldsymbol{y}}$ of $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ depends on $\boldsymbol{y}$. In [18, Theorem 3.1], it was indirectly assumed that this constant does not depend on $\boldsymbol{y}$.

Under the assumption that $C_{\mathrm{P},W} := \operatorname{ess\,sup} C_{\boldsymbol{Y}} < \infty$, i.e., the distribution of $C_{\boldsymbol{Y}}$ has compact support, we can continue with

$$\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] \leq C_{\mathrm{P},W}\,\mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]]. \tag{2.18}$$

However, as we see in Subsection 4.3, this assumption on $C_{\boldsymbol{Y}}$ is not always fulfilled.

The continuation of (2.18) follows [18, Lemma 2.2 and Theorem 3.1]. We repeat the steps here for the sake of completeness. So, first, note that $\nabla^{\boldsymbol{z}} f([\![\boldsymbol{y},\boldsymbol{z}]\!]) = W_2^\top \nabla^{\boldsymbol{x}} f([\![\boldsymbol{y},\boldsymbol{z}]\!])$ for $\boldsymbol{y} \in \mathcal{Y}$ and $\boldsymbol{z} \in \mathcal{Z}$. Then, we write

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} &f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]] \\
&= \operatorname{trace}\left(\mathbf{E}[\mathbf{E}[\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])\,\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])^\top \,|\, \boldsymbol{Y}]]\right) \\
&= \operatorname{trace}\left(W_2^\top\,\mathbf{E}[\mathbf{E}[\nabla^{\boldsymbol{x}} f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])\,\nabla^{\boldsymbol{x}} f([\![\boldsymbol{Y},\boldsymbol{Z}]\!])^\top \,|\, \boldsymbol{Y}]]\,W_2\right) \\
&= \operatorname{trace}\left(W_2^\top\,\mathbf{E}[\nabla^{\boldsymbol{x}} f(\boldsymbol{X})\,\nabla^{\boldsymbol{x}} f(\boldsymbol{X})^\top]\,W_2\right) \\
&= \operatorname{trace}\left(W_2^\top C W_2\right) = \operatorname{trace}\left(W_2^\top W \Lambda W^\top W_2\right) \\
&= \operatorname{trace}(\Lambda_2) = \lambda_{k+1} + \cdots + \lambda_n.
\end{aligned} \tag{2.19}$$

The next section gives two examples for types of densities $\rho_{\boldsymbol{X}}$ that are well-known to imply a probabilistic Poincaré inequality for $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ and allow for a bound on its constant $C_{\boldsymbol{y}}$ that is uniform in $\boldsymbol{y}$ and W. Again, we emphasize that it is not $\rho_{\boldsymbol{X}}$ that should satisfy a probabilistic Poincaré inequality but $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$.

3. Compactly supported and normal distributions

The uniform distribution, as a canonical example of a distribution with compact support $\mathcal{X}$, is well-known to satisfy a probabilistic Poincaré inequality on its own and to imply the same for densities $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ which are also uniform. Note that a probabilistic Poincaré inequality involving a uniform distribution is actually equivalent to a regular Poincaré inequality w.r.t. the Lebesgue measure. The following theorem is a slightly more general result. We add a convexity assumption on $\mathcal{X}^\circ$ since it makes Poincaré constants explicit. Recall that the Poincaré constant for a convex domain with diameter $d > 0$ is d/π; see, e.g., [7].

Theorem 3.1. *Assume that $\mathcal{X}^\circ$ is a bounded and convex domain. If $0 < \delta \leq \rho_{\boldsymbol{X}}(\boldsymbol{x}) \leq D < \infty$ for all $\boldsymbol{x} \in \mathcal{X}^\circ$, then*

$$\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] \leq C_{\mathrm{P}}(\lambda_{k+1} + \cdots + \lambda_n) \tag{3.1}$$

with

$$C_{\mathrm{P}} = C_{\mathrm{P}}(\delta, D, \mathcal{X}) := \frac{\operatorname{diam}(\mathcal{X})}{\pi} \cdot \frac{D}{\delta} > 0. \tag{3.2}$$

Proof. Define

$$\mathcal{Z}_{\boldsymbol{y}}^\circ = \{\boldsymbol{z} \in \mathbf{R}^{n-k} \,|\, [\![\boldsymbol{y},\boldsymbol{z}]\!] \in \mathcal{X}^\circ\} \subseteq \mathcal{Z} \tag{3.3}$$

and note that it is convex for $\boldsymbol{y} \in \mathcal{Y}^+$. It holds that $\mathrm{diam}(\mathcal{Z}_{\boldsymbol{y}}^\circ) \leq \mathrm{diam}(\mathcal{Z}) \leq \mathrm{diam}(\mathcal{X})$. Note that

$$\frac{\delta}{\rho_{\boldsymbol{Y}}(\boldsymbol{y})} \leq \rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) \leq \frac{D}{\rho_{\boldsymbol{Y}}(\boldsymbol{y})} \tag{3.4}$$

for $\boldsymbol{y} \in \mathcal{Y}^+$ and $\boldsymbol{z} \in \mathcal{Z}_{\boldsymbol{y}}^\circ$. This justifies the following lines of computation for $\boldsymbol{y} \in \mathcal{Y}^+$,

$$\mathbf{E}[(f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!]) - g(\boldsymbol{Y}))^2 \,|\, \boldsymbol{Y}] \tag{3.5}$$

$$= \int_{\mathcal{Z}_{\boldsymbol{y}}^\circ} (f([\![\boldsymbol{y}, \boldsymbol{z}]\!]) - g(\boldsymbol{y}))^2 \, \rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{z} \tag{3.6}$$

$$\leq \frac{D}{\rho_{\boldsymbol{Y}}(\boldsymbol{y})} \int_{\mathcal{Z}_{\boldsymbol{y}}^\circ} (f([\![\boldsymbol{y}, \boldsymbol{z}]\!]) - g(\boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{z} \tag{3.7}$$

$$\leq \frac{\mathrm{diam}(\mathcal{Z}_{\boldsymbol{y}}^\circ)}{\pi} \frac{D}{\rho_{\boldsymbol{Y}}(\boldsymbol{y})} \int_{\mathcal{Z}_{\boldsymbol{y}}^\circ} \|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{y}, \boldsymbol{z}]\!])\|_2^2 \, \mathrm{d}\boldsymbol{z} \tag{3.8}$$

$$\leq \frac{\mathrm{diam}(\mathcal{X})}{\pi} \frac{D}{\delta} \int_{\mathcal{Z}_{\boldsymbol{y}}^\circ} \|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{y}, \boldsymbol{z}]\!])\|_2^2 \, \rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{z} \tag{3.9}$$

$$= \frac{\mathrm{diam}(\mathcal{X})}{\pi} \frac{D}{\delta} \mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]. \tag{3.10}$$

Then, combining (2.19) with (3.10) yields the result in (3.1). $\qquad\square$

Also, it is well-known that the Poincaré constant is one for the multivariate standard normal distribution $\mathcal{N}(\boldsymbol{0}, I)$ [14]. Since its density is rotationally symmetric, random variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are independent and each follow again a standard normal distribution. Hence, it holds that $C_{\mathrm{P}} = 1$. For general multivariate normal distributions $\mathcal{N}(\boldsymbol{m}, \Sigma)$ with mean $\boldsymbol{m}$ and non-degenerate covariance matrix Σ, shifting and scaling arguments give that $C_{\mathrm{P}} = \lambda_{\max}(\Sigma)$.

Remark. Note that the constant C_{P} in the previous two examples is independent of W.

4. Main results

This section contains the main contribution of the manuscript which lies in an investigation of general log-concave probability measures w.r.t. their applicability for ASM. Log-concave distributions have Lebesgue densities of the form $\rho_{\boldsymbol{X}}(\boldsymbol{x}) = \exp(-V(\boldsymbol{x}))$ for a convex function $V : \mathbf{R}^n \to (-\infty, +\infty]$. Note that $+\infty$ is included in the codomain of V. The conditional density $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ for a given $\boldsymbol{y} \in \mathcal{Y}^+$ is then given by

$$\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) = \frac{\exp(-V([\![\boldsymbol{y}, \boldsymbol{z}]\!]))}{\rho_{\boldsymbol{Y}}(\boldsymbol{y})} = \exp(-\tilde{V}_{\boldsymbol{y}}(\boldsymbol{z})), \tag{4.1}$$

where $\tilde{V}_{\boldsymbol{y}}(\boldsymbol{z}) := V([\![\boldsymbol{y}, \boldsymbol{z}]\!]) + \log(\rho_{\boldsymbol{y}}(\boldsymbol{y}))$. Note that $\tilde{V}_{\boldsymbol{y}}$ inherits convexity (in $\boldsymbol{z}$) from V. Bobkov [8] shows that general log-concave densities satisfy a Poincaré

inequality and gives lower and upper bounds on the corresponding Poincaré constant.

First, we discuss the special case of α-uniformly convex functions V for which the corresponding density $\rho_{\boldsymbol{X}}$ is known to satisfy a Poincaré inequality with universal Poincaré constant $1/\alpha$. However, the assumption of the density $\rho_{\boldsymbol{X}}$ being of uniformly log-concave type is somewhat restrictive since it excludes distributions with heavier tails as, for example, exponential or Laplace distributions. For this reason, we secondly investigate general log-concave densities and show that there might arise problems with this class of probability distributions due to arbitrary large Poincaré constants $C_{\boldsymbol{Y}}$. In particular, the problems and their proposed solution are exemplified on an extreme case example involving independently exponentially distributed random variables in $n \geq 2$ dimensions.

4.1. α-uniformly convex functions V

Definition 4.1 (α-uniformly convex function)**.** A function $V \in \mathcal{C}^2$ is said to be α-uniformly convex, if there is an $\alpha > 0$ such that for all $\boldsymbol{x} \in \mathbf{R}^n$ it holds that

$$\boldsymbol{u}^\top V''(\boldsymbol{x})\boldsymbol{u} \geq \alpha\|\boldsymbol{u}\|_2^2 \tag{4.2}$$

for all $\boldsymbol{u} \in \mathbf{R}^n$, where V'' denotes the Hessian matrix of V.

In [41, pp. 43–44], it was shown that there is a dimension-free Poincaré constant $1/\alpha$ for α-uniformly log-concave $\rho_{\boldsymbol{X}}$. Note that this says nothing about the special case $\alpha = 0$. The existence of a dimension-free Poincaré constant for this special case is actually a consequence of the famous Kannan-Lovász-Simonovits conjecture; see, e.g., [1, 30]. However, since we need a Poincaré inequality for $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{Y}^+$, we have to prove the following lemma similar to [46, Subsection 7.2]

Lemma 4.2. *If $\rho_{\boldsymbol{X}}$ is α-uniformly log-concave, then $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ is α-uniformly log-concave for each $\boldsymbol{y} \in \mathcal{Y}^+$.*

Proof. Let $\boldsymbol{y} \in \mathcal{Y}^+$. Recall that $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) = \exp(-\tilde{V}_{\boldsymbol{y}}(\boldsymbol{z}))$ for a convex function $\tilde{V}_{\boldsymbol{y}}(\boldsymbol{z}) := V(\llbracket\boldsymbol{y}, \boldsymbol{z}\rrbracket) + \log(\rho_{\boldsymbol{y}}(\boldsymbol{y}))$. The Hessian matrix $\tilde{V}_{\boldsymbol{y}}''(\boldsymbol{z})$ (w.r.t. $\boldsymbol{z}$) computes to

$$\tilde{V}_{\boldsymbol{y}}''(\boldsymbol{z}) = W_2^\top \, V''(\llbracket\boldsymbol{y}, \boldsymbol{z}\rrbracket) \, W_2. \tag{4.3}$$

Choose $\boldsymbol{w} \in \mathbf{R}^{n-k}$ arbitrarily. Then, for every $\boldsymbol{z} \in \mathbf{R}^{n-k}$, it holds that

$$\boldsymbol{w}^\top \tilde{V}_{\boldsymbol{y}}''(\boldsymbol{z})\boldsymbol{w} = (W_2\boldsymbol{w})^\top V''(\llbracket\boldsymbol{y}, \boldsymbol{z}\rrbracket)(W_2\boldsymbol{w}) \tag{4.4}$$

$$\geq \alpha\|W_2\boldsymbol{w}\|_2^2 = \alpha\|\boldsymbol{w}\|_2^2. \tag{4.5}$$

$\square$

Since $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ inherits the universal Poincaré constant $1/\alpha$ from $\rho_{\boldsymbol{X}}$, the result in (2.15) also holds for α-uniformly log-concave densities with $C_{\mathrm{P}} = 1/\alpha$ (independent of W) which is similar to [46, Corollary 2].

For example, α-uniformly log-concave densities comprise multivariate normal distributions $\mathcal{N}(\boldsymbol{m}, \Sigma)$ with mean $\mathbf{m}$ and covariance matrix Σ ($\alpha = 1/\lambda_{\max}(\Sigma)$). However, distributions that satisfy the assumption only for $\alpha = 0$ as, e.g., Weibull distributions with the exponential distribution as a special case or Gamma distributions with shape parameter $\beta \geq 1$, only belong to the class of general log-concave distributions.

4.2. General convex functions V

Since we cannot make use of a universal dimension-free Poincaré constant involving general convex functions $V : \mathbf{R}^n \to (-\infty, +\infty]$, we look at them more closely in this subsection. Recall that $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) = \exp(-\tilde{V}_{\boldsymbol{y}}(\boldsymbol{z}))$, $\boldsymbol{y} \in \mathcal{Y}^+$, for a convex function $\tilde{V}_{\boldsymbol{y}}$. We have to deal with the fact that the essential supremum of the random Poincaré constant $C_{\boldsymbol{Y}}$ of $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{Y})$ does possibly not exist. A corresponding example is given in Subsection 4.3.1. In the step from (2.17) to (2.18), we have applied Hölder's inequality with Hölder conjugates $(p, q) = (+\infty, 1)$. Since this is not possible for unbounded random variables $C_{\boldsymbol{Y}}$, we can only show a weaker result.

Lemma 4.3. *If* $\|\nabla f\|_2^2 \leq L$ *for some constant* $L > 0$, *then*

$$\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] \leq C_{\mathrm{P},\varepsilon,W}(\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)}, \tag{4.6}$$

where

$$C_{\mathrm{P},\varepsilon,W} = C_{\mathrm{P},\varepsilon,W}(\varepsilon, n, k, L, W, \rho_{\boldsymbol{X}}) := L^{\varepsilon/(1+\varepsilon)}\mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)}. \tag{4.7}$$

Proof. The boundedness of ∇f implies that also $\|\nabla^{\boldsymbol{z}} f\|_2^2 \leq L$. Choosing a weaker pair of conjugates $(p, q) = ((1+\varepsilon)/\varepsilon, 1+\varepsilon)$, $\varepsilon > 0$, we compute

$$\mathbf{E}[C_{\boldsymbol{Y}} \, \mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]] \tag{4.8}$$

$$\leq \mathbf{E}[C_{\boldsymbol{Y}}^p]^{1/p} \, \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]^q]^{1/q} \tag{4.9}$$

$$= \mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \, \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]^{1+\varepsilon}]^{1/(1+\varepsilon)} \tag{4.10}$$

$$\leq L^{\varepsilon/(1+\varepsilon)} \, \mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \, \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]]^{1/(1+\varepsilon)} \tag{4.11}$$

$$\leq L^{\varepsilon/(1+\varepsilon)} \, \mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \, (\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)} \tag{4.12}$$

$$= C_{\mathrm{P},\varepsilon,W} \, (\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)}. \tag{4.13}$$

The step in (4.12) uses (2.19). The result follows by (2.16) and (2.17). □

Remark. The previous lemma requires the gradient of f to be uniformly bounded, an assumption that is not needed in [18] and [46].

However, first, applying ASM, in the sense that the matrix C from (2.2) is estimated by a finite Monte Carlo sum, requires the same assumption to prove results on corresponding approximations of eigenvalues λ_i and eigenvectors $\boldsymbol{w}_i$; see [16] and [17, Section 3.3].

Secondly, this assumption can be weakened by applying another Hölder's inequality analogous to (4.9). Indeed, for $\varepsilon \in (0,1)$, we would get

$$\mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]^{1+\varepsilon}]^{1/(1+\varepsilon)} \tag{4.14}$$

$$\leq \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]^{1/(1-\varepsilon)}]^{(1-\varepsilon)/(1+\varepsilon)} \tag{4.15}$$

$$\cdot \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]]^{\varepsilon/(1+\varepsilon)}. \tag{4.16}$$

Since

$$\mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^2 \,|\, \boldsymbol{Y}]^{1/(1-\varepsilon)}] \tag{4.17}$$

$$\leq \mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}} f([\![\boldsymbol{Y}, \boldsymbol{Z}]\!])\|_2^{2/(1-\varepsilon)} \,|\, \boldsymbol{Y}]] \tag{4.18}$$

$$\leq \mathbf{E}[\|\nabla^{\boldsymbol{x}} f(\boldsymbol{X})\|_2^{2/(1-\varepsilon)}], \tag{4.19}$$

we would only require $\|\nabla^{\boldsymbol{x}} f(\boldsymbol{X})\|_2^{2/(1-\varepsilon)}$ to be integrable. What we, however, would have to accept in this case, is the resulting weaker order $\varepsilon/(1+\varepsilon)$ in the inactive trace.

The L- and $\rho_{\boldsymbol{X}}$-dependence of $C_{\mathrm{P},\varepsilon,W}$ is notationally neglected in the following. If possible, we can choose a suitable $\varepsilon > 0$ to get $\mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}] < \infty$ and thus a finite constant $C_{\mathrm{P},\varepsilon,W}$. Note that we lose first order in the eigenvalues from the inactive subspace, but have instead order $1/(1+\varepsilon) < 1$. Of course, the constant $C_{\mathrm{P},\varepsilon,W}$ could get arbitrarily large as $\varepsilon \to 0$, but this strongly depends on W and the moments of $C_{\boldsymbol{Y}}$; see the example given in Subsection 4.3.1.

It is known by Bobkov [8, Eqs. (1.3), (1.8) and p. 1906] that there exists a (dimensionally dependent) Poincaré constant $C_{\boldsymbol{y}}$ for a general log-concave density $\rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\cdot|\boldsymbol{y})$ that is bounded from below and above by

$$\begin{aligned}
\mathbf{E}[(\|\boldsymbol{Z} - \boldsymbol{z}_0\|_2 - \mathbf{E}[\|\boldsymbol{Z} - \boldsymbol{z}_0\|_2 \,|\, \boldsymbol{Y} = \boldsymbol{y}])^2 \,|\, \boldsymbol{Y} = \boldsymbol{y}] &\leq C_{\boldsymbol{y}} \\
&\leq K \, \mathbf{E}[\|\boldsymbol{Z} - \boldsymbol{z}_0\|_2^2 \,|\, \boldsymbol{Y} = \boldsymbol{y}] \\
&= K \sum_{i=1}^{n-k} \mathbf{V}\mathrm{ar}(Z_i \,|\, \boldsymbol{Y} = \boldsymbol{y}),
\end{aligned} \tag{4.20}$$

where $\boldsymbol{z}_0 := \mathbf{E}[\boldsymbol{Z} \,|\, \boldsymbol{Y} = \boldsymbol{y}]$ and $K = 432$ [8, Eqs. (1.8) and (3.4)] is a universal constant. To the authors' knowledge, the constant $C_{\boldsymbol{y}}$ is the best available. We provide a scenario in Subsection 4.3.1 ("Rotation by $\theta = \pi/4$") in which the lower bound viewed as a random variable has no finite essential supremum implying the same for $C_{\boldsymbol{Y}}$.

However, to make use of Lemma 4.3, we need to investigate the involved constant $C_{\mathrm{P},\varepsilon,W}(\varepsilon, n, k)$.

Lemma 4.4. *It holds that*

$$\mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \leq K(n-k)^{1/(1+\varepsilon)} C_{\mathbf{V}\mathrm{ar},W}, \tag{4.21}$$

where

$$C_{\mathbf{V}\mathrm{ar},W} = C_{\mathbf{V}\mathrm{ar},W}(\varepsilon, n, k, W) := \left(\sum_{i=1}^{n-k} \mathbf{E}[\mathbf{V}\mathrm{ar}(Z_i|\mathbf{Y})^{(1+\varepsilon)/\varepsilon}] \right)^{\varepsilon/(1+\varepsilon)}. \quad (4.22)$$

Proof. Using Jensen's inequality for weighted sums, it follows that

$$\mathbf{E}[C_{\mathbf{Y}}^{(1+\varepsilon)/\varepsilon}] \le K^{(1+\varepsilon)/\varepsilon}(n-k)^{(1+\varepsilon)/\varepsilon}\frac{1}{n-k}\sum_{i=1}^{n-k}\mathbf{E}[\mathbf{V}\mathrm{ar}(Z_i|\mathbf{Y})^{(1+\varepsilon)/\varepsilon}] \quad (4.23)$$

$$= K^{(1+\varepsilon)/\varepsilon}(n-k)^{1/\varepsilon}\sum_{i=1}^{n-k}\mathbf{E}[\mathbf{V}\mathrm{ar}(Z_i|\mathbf{Y})^{(1+\varepsilon)/\varepsilon}]. \quad (4.24)$$

The result follows.   □

Eventually, we get

$$C_{\mathrm{P},\varepsilon,W}(\varepsilon, n, k) \le L^{\varepsilon/(1+\varepsilon)}K(n-k)^{1/(1+\varepsilon)}C_{\mathbf{V}\mathrm{ar},W}(\varepsilon, n, k). \quad (4.25)$$

As before, we can remove the dependence of $C_{\mathbf{V}\mathrm{ar},W}$ on $W = W(f)$ by considering the supremum over all orthogonal matrices. That is, we define

$$C_{\mathrm{P},\varepsilon} := \sup_{W \text{ orth.}} C_{\mathrm{P},\varepsilon,W} \quad (4.26)$$

and

$$C_{\mathbf{V}\mathrm{ar}} := \sup_{W \text{ orth.}} C_{\mathbf{V}\mathrm{ar},W}, \quad (4.27)$$

and get

$$\mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] \le C_{\mathrm{P},\varepsilon}(\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)}, \quad (4.28)$$

provided the constant $C_{\mathrm{P},\varepsilon} = C_{\mathrm{P},\varepsilon}(\varepsilon, n, k, L, \rho_{\mathbf{X}})$ exists.

For $C_{\mathbf{V}\mathrm{ar}}$, we argue that it is actually enough to take the supremum only over the set of rotation matrices. Indeed, any orthogonal matrix W is either a *proper* ($\det W = 1$) or an *improper* ($\det W = -1$) rotation which is the combination of a proper rotation and an inversion of the axes; see, e.g., [27, 33]. However, since the constant $C_{\mathbf{V}\mathrm{ar},W}$ from (4.22) is invariant to inversions of the axes, it holds that

$$\sup_{W \text{ orth.}} C_{\mathbf{V}\mathrm{ar},W} = \sup_{R \text{ rot.}} C_{\mathbf{V}\mathrm{ar},R}. \quad (4.29)$$

This equality is exploited in the next subsection.

4.3. Independently exponentially distributed random variables as an extreme case

In this subsection, we take a closer look at independently exponentially distributed random variables in $n \ge 2$ dimensions as an example for a general log-concave distribution. In particular, we use the lower bound of Bobkov

from (4.20) in Subsection 4.3.1 to show that there exists a scenario in which the random Poincaré constant $C_{\mathbf{Y}}$ does not have an essential supremum implying that C_{P} from (2.14) does not exist. Therefore, the quantity $C_{\mathbf{Var}}$ from (4.27) is investigated in Subsections 4.3.1 and 4.3.2 to derive a (finite) upper bound for $C_{\mathrm{P},\varepsilon}$ from (4.26) in this special case.

We regard a random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top$ whose components are independently exponentially distributed with unit rates $\nu_i = 1$, $i = 1, \ldots, n$ and will see that investigations with unit rates are sufficient to derive statements also involving other rates. The distribution of $\mathbf{X}$ has the density

$$\rho_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \exp(-x_1 - \cdots - x_n) & \text{if } \mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbf{R}_{\geq 0}^n, \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

That is, in this case $\mathcal{X} = \mathbf{R}_{\geq 0}^n$ and

$$V(\mathbf{x}) = \begin{cases} x_1 + \cdots + x_n & \text{if } \mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbf{R}_{\geq 0}^n, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.31}$$

Note that V is convex.

Since we are interested in $C_{\mathbf{Var}}$ as a supremum over all orthogonal matrices, we assume that, in this subsection, $W = \begin{pmatrix} W_1 & W_2 \end{pmatrix}$ is an *arbitrary* orthogonal matrix not depending on f and $\rho_{\mathbf{X}}$. Indeed, as the equality in (4.29) motivates, we can further assume that W is a rotation matrix.

4.3.1. 2 dimensions

The joint density of two independently exponentially distributed random variables X_1 and X_2 both with unit rate is

$$\rho_{\mathbf{X}}(x_1, x_2) = \begin{cases} \exp(-x_1 - x_2) & \text{if } x_1, x_2 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.32}$$

First, let us regard a rotation of the two-dimensional Cartesian coordinate system by a general angle $\theta \in [-\pi, \pi)$ to a coordinate system for (y, z), i.e.,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = R_\theta \begin{pmatrix} y \\ z \end{pmatrix} \tag{4.33}$$

for a rotation matrix

$$W = R_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{4.34}$$

That is, in two dimensions, it holds that

$$C_{\mathbf{Var}} = \sup_{\theta \in [-\pi, \pi)} C_{\mathbf{Var}, R_\theta}. \tag{4.35}$$
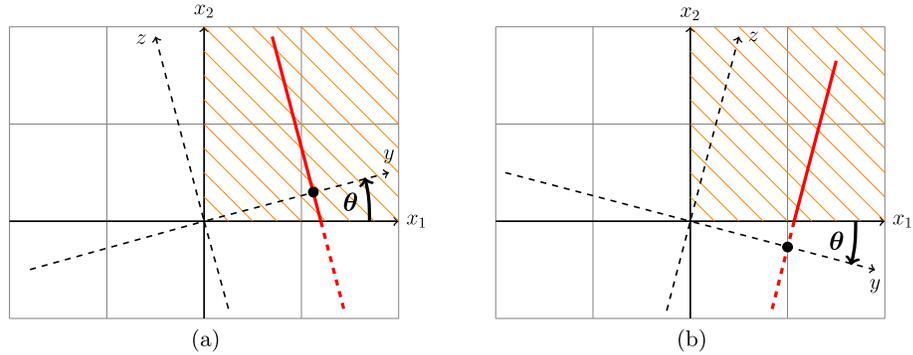
FIG 1. *Rotations of the coordinate system with a positive (a) and a negative (b) angle. The orange lines depict contour lines in the support of $\rho_{\mathbf{X}}$. The red lines show the values of (y, z) for a given y. Their solid parts mark regions within the support of $\rho_{\mathbf{X}}$, whereas the dashed parts identify values with density zero.*

Subsequently, we look at the special case $\theta = \pi/4$ as an example for an unbounded Poincaré constant C_y of $\rho_{z|y}(\cdot|y)$. Variables are written in thin letters in this subsection since they denote real values and not multidimensional vectors.

Note that the bound from (4.21) in this two-dimensional setting becomes

$$\mathbf{E}[C_Y^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \leq K C_{\mathbf{V}\mathrm{ar},W}(\varepsilon, 2, 1) \tag{4.36}$$

with

$$C_{\mathbf{V}\mathrm{ar},W}(\varepsilon, 2, 1) = \mathbf{E}[\mathbf{V}\mathrm{ar}(Z|Y)^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)}. \tag{4.37}$$

Rotation by general θ

Let $\theta \in [-\pi, \pi)$. Then, the joint density of (Y, Z) is

$$\rho_{Y,Z}(y, z) = \exp(-(y\cos\theta - z\sin\theta) - (y\sin\theta + z\cos\theta)) \tag{4.38}$$

$$= \exp(-(\cos\theta + \sin\theta)y - (\cos\theta - \sin\theta)z), \tag{4.39}$$

for (y, z) with $[\![y, z]\!] \in \mathbf{R}_{\geq 0}^2$ and zero otherwise. If we define $a_\theta^+ := \cos\theta + \sin\theta$ and $a_\theta^- := \cos\theta - \sin\theta$, we have

$$\rho_{Y,Z}(y, z) = \begin{cases} \exp(-a_\theta^+ y - a_\theta^- z) & \text{if } [\![y, z]\!] \in \mathbf{R}_{\geq 0}^2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.40}$$

Fig. 1 illustrates the situation for a positive (Fig. 1a) and a negative (Fig. 1b) angle θ.

The interval of investigation for $\theta \in [-\pi, \pi)$ can be reduced by reasons of periodicity and symmetry. First, note that the map

$$Q_\varepsilon(\theta) := C_{\mathbf{V}\mathrm{ar},R_\theta}(\varepsilon, 2, 1), \tag{4.41}$$
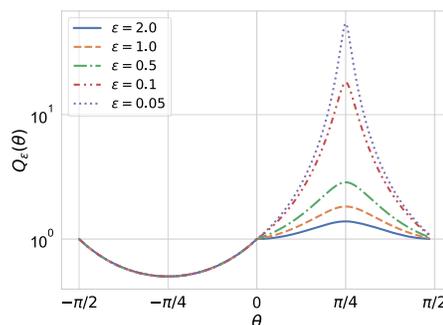
FIG 2. *Illustration of symmetries in θ of the map $Q_\varepsilon(\theta)$ for several $\varepsilon > 0$.*

is π-periodic in θ since an additional rotation by π corresponds to changing signs of y and z which is not important for integrals in Q_ε. Hence, it suffices to consider $\theta \in [-\pi/2, \pi/2]$. Secondly, from Fig. 1 it can be deduced that Q_ε, as a map of θ, is symmetric around $-\pi/4$ in $[-\pi/2, 0]$ and symmetric around $\pi/4$ in $[0, \pi/2)$. This fact is also shown in Fig. 2. That is, it is enough to investigate angles $\theta \in [-\pi/4, \pi/4]$.

For the computation of integrals in $Q_\varepsilon(\theta)$, $\theta \in [-\pi/4, \pi/4]$, it is necessary, for a given y, to determine boundaries $\ell_0(y)$ and $\ell_1(y)$ of intervals for z that lie in the support of the joint density $\rho_{Y,Z}(y, z)$ (see the thick solid lines in Fig. 1). The integrals in $Q_\varepsilon(\theta)$ are computed using the computer algebra system *Wolfram Mathematica* [43]. The computation requires to treat the cases $\theta \in [-\pi/4, 0)$ and $\theta \in [0, \pi/4]$ differently (see Fig. 1).

For negative $\theta \in [-\pi/4, 0)$ and arbitrary $y \in \mathbf{R}$, we have that

$$\ell_0(y) = \begin{cases} |y| \cot(|\theta|) & \text{if } y < 0 \\ y \tan(|\theta|) & \text{if } y \geq 0 \end{cases} = |y| \tan(|\theta|)^{\text{sgn}(y)} \tag{4.42}$$

and $\ell_1(y) = \infty$, i.e.,

$$\rho_{Y,Z}(y, z) = \exp(-a_\theta^+ y - a_\theta^- z) \cdot \mathbf{1}_{[\ell_0(y), \ell_1(y)]}(z). \tag{4.43}$$

We compute that

$$\mathbf{Var}(Z \,|\, Y = y) = (\cos(|\theta|) + \sin(|\theta|))^{-2} \tag{4.44}$$

which is constant in y and yields

$$Q_\varepsilon(\theta) = C_{\mathbf{Var}, R_\theta}(\varepsilon, 2, 1) = (\cos(|\theta|) + \sin(|\theta|))^{-2}. \tag{4.45}$$

Note that this explains the left part of the graph of $Q_\varepsilon(\theta)$ in Fig. 2 which shows that $Q_\varepsilon(\theta)$ does not depend on ε for $\theta \in [-\pi/2, 0)$.

For non-negative $\theta \in [0, \pi/4]$ and a given $y \geq 0$, the boundaries are computed to $\ell_0(y) = -y \tan(\theta)$ and $\ell_1(y) = y \cot(\theta)$, i.e.,

$$\rho_{Y,Z}(y, z) = \exp(-a_\theta^+ y - a_\theta^- z) \cdot \mathbf{1}_{[0,\infty)}(y) \cdot \mathbf{1}_{[\ell_0(y), \ell_1(y)]}(z). \tag{4.46}$$
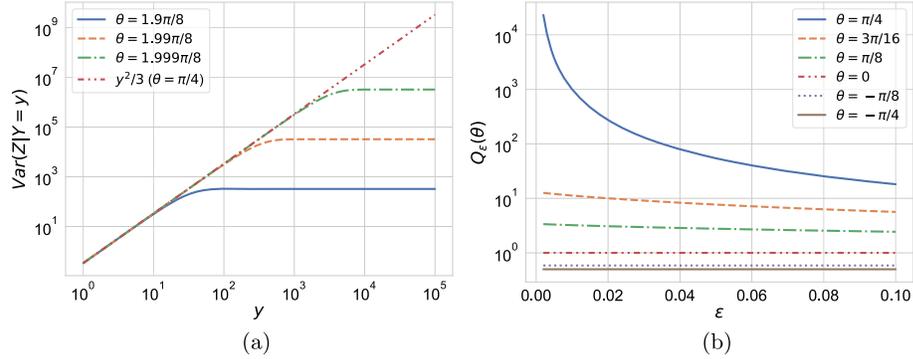
FIG 3. *(a) The log-log plot of the map $y \mapsto \mathbf{V}\mathrm{ar}(Z \mid Y = y)$ shows that it is bounded for angles $\theta \in [0, \pi/4)$, but approaching the unbounded function $y^2/3$, which corresponds to $\theta = \pi/4$, as $\theta \to \pi/4$. (b) The plot shows the map $\varepsilon \mapsto Q_\varepsilon(\theta)$ for several angles θ. Also, it illustrates the fact that $\theta = \pi/4$ is a special case for which $Q_\varepsilon(\theta)$ can get arbitrarily large.*

We compute that

$$
\begin{aligned}
&\mathbf{V}\mathrm{ar}(Z \mid Y = y) \\
&= \frac{a_\theta}{8b_\theta^2}\left(\frac{1 - 2\exp(b_\theta y) + \exp(2b_\theta y) - 8\exp(b_\theta y)y^2(1 - d_\theta)}{(\exp(b_\theta y) - 1)^2} - c_\theta \right)
\end{aligned} \quad (4.47)
$$

for $a_\theta := \csc(\theta)^4 \sec(\theta)^4$, $b_\theta := \sec(\theta) - \csc(\theta)$, $c_\theta := \cos(4\theta)$, and $d_\theta := \sin(2\theta)$. $\mathbf{V}\mathrm{ar}(Z \mid Y = y)$ can actually be bounded in y for $\theta \in [0, \pi/4)$. Indeed, since $d_\theta \in [0, 1)$, it holds that $1 - d_\theta \in (0, 1]$ implying that $8\exp(b_\theta y)y^2(1 - d_\theta) > 0$. It follows that

$$
\mathbf{V}\mathrm{ar}(Z \mid Y) \leq \frac{a_\theta}{8b_\theta^2}\left(\frac{1 - 2\exp(b_\theta y) + \exp(2b_\theta y)}{(\exp(b_\theta y) - 1)^2} - c_\theta \right) \quad (4.48)
$$

$$
= \frac{a_\theta}{8b_\theta^2}\left(\frac{(\exp(b_\theta y) - 1)^2}{(\exp(b_\theta y) - 1)^2} - c_\theta \right) \quad (4.49)
$$

$$
= \frac{a_\theta(1 - c_\theta)}{8b_\theta^2}. \quad (4.50)
$$

Fig. 3a illustrates the boundedness of $\mathbf{V}\mathrm{ar}(Z \mid Y = y)$ and additionally shows that it approaches the unbounded function $y \mapsto y^2/3$ as $\theta \to \pi/4$. Hence, for $\theta \in [0, \pi/4)$, it holds that

$$
Q_\varepsilon(\theta) = C_{\mathbf{V}\mathrm{ar}, R_\theta}(\varepsilon, 2, 1) \leq \frac{a_\theta(1 - c_\theta)}{8b_\theta^2}. \quad (4.51)
$$

This bound is itself unbounded in θ since $b_\theta \to 0$ and $a_\theta(1 - c_\theta) \to 32$ as $\theta \to \pi/4$ implying that we can see $\theta = \pi/4$ as a special case. This assessment is also supported by Fig. 3b. In particular, note that

$$
C_{\mathbf{V}\mathrm{ar}} = C_{\mathbf{V}\mathrm{ar}, R_{\pi/4}}. \quad (4.52)
$$
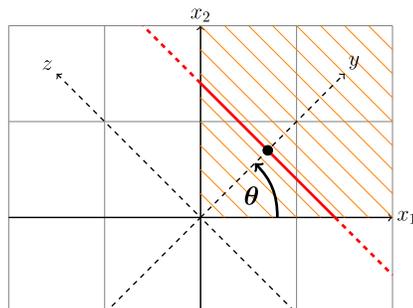
FIG 4. *Exponential distribution in 2 dimensions with a coordinate system rotated by 45°. The orange lines depict the contour levels of the distribution in the support of $\rho_{\boldsymbol{X}}$. The solid red line marks the interval of the uniform distribution of $Z \mid Y = y$ for $y > 0$.*

Rotation by $\theta = \pi/4$

A rotation of 45°, i.e., $\theta = \pi/4$ and $W = R_{\pi/4}$, is a limit case since $a_{\pi/4}^-$ from (4.40) becomes zero. The joint density for Y and Z is then

$$\rho_{Y,Z}(y,z) = \exp(-\sqrt{2}y) \cdot \mathbf{1}_{[0,\infty)}(y) \cdot \mathbf{1}_{[-y,y]}(z). \tag{4.53}$$

A graphical illustration of this case is given in Fig. 4. Consequently, the marginal distribution of Y is

$$\rho_Y(y) = \int_{-\infty}^{\infty} \rho_{Y,Z}(y,z)\,\mathrm{d}z = 2y \exp(-\sqrt{2}y) \cdot \mathbf{1}_{[0,\infty)}(y) \tag{4.54}$$

and the conditional density $\rho_{Z|Y}(\cdot|y)$ computes to

$$\rho_{Z|Y}(z|y) = \frac{\mathbf{1}_{[-y,y]}(z)}{2y} \tag{4.55}$$

for $y > 0$. Note that $\rho_{Z|Y}(\cdot|y)$ is the density of a uniform distribution on the interval $[-y,y]$. For $Y > 0$, it follows that

$$\mathbf{V}\mathrm{ar}(Z|Y) = (2Y)^{-1} \int_{-Y}^{Y} z^2\,\mathrm{d}z = Y^2/3, \tag{4.56}$$

which is the expression that variances of $Z|Y$ for other angles θ^* approach to as $\theta^* \to \pi/4$ (see Fig. 3a).

Note that the lower bound from (4.20) for C_Y in this case becomes

$$\mathbf{E}[(|Z| - \mathbf{E}[|Z|\,|\,Y])^2 \,|\, Y] = \mathbf{V}\mathrm{ar}(|Z|\,|\,Y) = Y^2/12, \tag{4.57}$$

since $|Z|\,|\,Y \sim \mathcal{U}([0,Y])$ and, hence, its distribution is not compactly supported implying the same for the distribution of C_Y. Therefore, we found a scenario in which the constants $C_{\mathrm{P},W}$ and C_{P} indeed do not exist.

However, there is still a chance that the constants $C_{\mathrm{P},\varepsilon,W}$ and $C_{\mathrm{P},\varepsilon}$ from (4.7) and, respectively, (4.26) exist. It holds that

$$C_{\mathbf{Var}}(\varepsilon, 2, 1) = C_{\mathbf{Var},R_{\pi/4}}(\varepsilon, 2, 1) = \frac{1}{3}\mathbf{E}[Y^{2+2/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \qquad (4.58)$$

implying that the constant $C_{\mathrm{P},\varepsilon}(\varepsilon, 2, 1)$ can be bounded from above by

$$C_{\mathrm{P},\varepsilon}(\varepsilon, 2, 1) \le L^{\varepsilon/(1+\varepsilon)}\frac{K}{3}\mathbf{E}[Y^{2+2/\varepsilon}]^{\varepsilon/(1+\varepsilon)}. \qquad (4.59)$$

For example, choosing $\varepsilon = 2$ would give

$$C_{\mathrm{P},\varepsilon}(2, 2, 1) \le 2K\left(\frac{L^2}{3}\right)^{1/3}. \qquad (4.60)$$

4.3.2. n dimensions

This subsection aims to generalize the results of the previous subsection, i.e., we investigate the constant $C_{\mathrm{P},\varepsilon}$ from (4.26) for n independently exponentially distributed random variables.

Motivated by the two-dimensional case, we regard the rotation of the coordinate system by a matrix $W = R^*$ that rotates the vector $(1, 0, \ldots, 0)^\top \in \mathbf{R}^n$ to $(1/\sqrt{n}, \ldots, 1/\sqrt{n})^\top \in \mathbf{R}^n$. Note that in the two-dimensional case, a rotation by $\theta = \pi/4$ corresponds to a matrix rotating $(1, 0)^\top$ to $(1/\sqrt{2}, 1/\sqrt{2})^\top$. This is the worst case in the sense that $Z_i|Y$ is uniformly distributed for each component Z_i in $\mathbf{Z} = (Z_1, \ldots, Z_{n-k})^\top$ and hence, similar to the two-dimensional case, the conditional variance of $Z_i|Y$ has no finite essential supremum. In the context from above, it holds that

$$C_{\mathbf{Var}}(\varepsilon, n, k) = C_{\mathbf{Var},R^*}(\varepsilon, n, k). \qquad (4.61)$$

The following theorem studies this case and investigates the dimensional dependence of the involved constant.

Theorem 4.5. *For $\rho_{\mathbf{X}}$ as in* (4.30), *it holds that*

$$\mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] \le C_{\exp^n}(\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)} \qquad (4.62)$$

for a constant

$$C_{\exp^n} = C_{\exp^n}(\varepsilon, n, k, L, \rho_{\mathbf{X}}) \ge C_{\mathrm{P},\varepsilon} \qquad (4.63)$$

Proof. In the support of $\rho_{\mathbf{X}}$, i.e., in $\mathcal{X} = \mathbf{R}_{\ge 0}^n$, $\rho_{\mathbf{X}}$ is greater than zero and constant on the intersection of $\mathbf{R}_{\ge 0}^n$ and planes

$$P_a := \{\mathbf{x} \mid x_1 + \cdots + x_n = a\} = \{\mathbf{x} \mid (1, \ldots, 1)^\top \mathbf{x} = a\} \subset \mathbf{R}^n, \quad a > 0, \quad (4.64)$$

i.e., on hypersurfaces $T_a := P_a \cap \mathbf{R}_{\ge 0}^n$. The situation is illustrated by Fig. 5 for $n = 3$ dimensions.
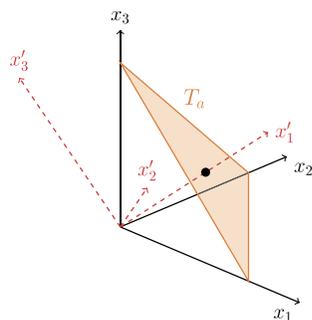
FIG 5. *Exponential distribution in 3D with a rotated coordinate system.*

For $\boldsymbol{x} = [\![\boldsymbol{y}, \boldsymbol{z}]\!] \in \mathbf{R}^n_{\geq 0}$, the value of $\rho_{\boldsymbol{Y}, \boldsymbol{Z}}(\boldsymbol{y}, \boldsymbol{z})$ is only determined by $y_1 \geq 0$. Reversely, if $y_1 < 0$, then $\rho_{\boldsymbol{Y}, \boldsymbol{Z}}(\boldsymbol{y}, \boldsymbol{z}) = 0$. We know that the point at $\boldsymbol{x}_0 := \beta(1, \ldots, 1)^\top \in \mathbf{R}^n$ with $\|\boldsymbol{x}_0\|_2 = y_1$ is supposed to lie on P_a for some $\beta > 0$. It follows immediately that $\beta = y_1/\sqrt{n}$. Also, we determine a with

$$a = (1, \ldots, 1)^\top \boldsymbol{x}_0 = \frac{y_1}{\sqrt{n}} n = \sqrt{n} y_1. \tag{4.65}$$

Let us define $T(y_1) := T_{\sqrt{n}y_1}$. That is,

$$\rho_{\boldsymbol{Y}, \boldsymbol{Z}}(\boldsymbol{y}, \boldsymbol{z}) = \exp(-\sqrt{n} y_1) \cdot \mathbf{1}_{[0,\infty)}(y_1) \cdot \mathbf{1}_{T(y_1)}(\boldsymbol{y}, \boldsymbol{z}). \tag{4.66}$$

$T(y_1)$, as a geometric figure, is a *regular $(n-1)$-simplex* in n dimensions. $T(y_1)$ is intrinsically $(n-1)$-dimensional and has n corners which are

$$(\sqrt{n}y_1, 0, \ldots, 0), \ldots, (0, \ldots, 0, \sqrt{n}y_1) \in \mathbf{R}^n. \tag{4.67}$$

It follows that the side length of $T(y_1)$ is $\sqrt{2n}y_1$. Note that the coordinates $\check{\boldsymbol{y}} = (y_2, \ldots, y_k)^\top$ and $\boldsymbol{z} = (z_1, \ldots, z_{n-k})^\top$ all move on $T(y_1)$.

We can rewrite $T(y_1)$ as

$$T(y_1) = \{\boldsymbol{x} \in \mathbf{R}^n_{\geq 0} \,|\, (W^\top \boldsymbol{x})_1 = y_1\} \tag{4.68}$$

$$= \{[\![\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{z}}]\!] \,|\, [\![\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{z}}]\!] \in \mathbf{R}^n_{\geq 0}, \tilde{y}_1 = y_1\}. \tag{4.69}$$

This motivates to view $T(y_1)$ as an $(n-1)$-dimensional set in the rotated coordinate system, i.e., we define

$$\check{T}(y_1) := \{(\check{\boldsymbol{y}}, \boldsymbol{z}) \in \mathbf{R}^{k-1} \times \mathbf{R}^{n-k} \,|\, [\![(y_1, \check{\boldsymbol{y}}), \boldsymbol{z}]\!] \in T(y_1)\} \subset \mathbf{R}^{n-1}. \tag{4.70}$$

We observe that the conditioned random variable $(\check{\boldsymbol{Y}}, \boldsymbol{Z})|Y_1$ is uniformly distributed on the regular $(n-1)$-simplex $\check{T}(y_1)$. The basic idea to get a bound for $\mathbf{E}[\mathbf{Var}(Z_i|\boldsymbol{Y})^{(1+\varepsilon)/\varepsilon}]$ is based on the fact that z_i, moving as the $(k+i-1)$-th coordinate inside $\check{T}(y_1)$, takes values in $[0, h_i(y_1)]$, where $h_i(y_1)$ is the height of a regular $(k+i-1)$-simplex with side length $\sqrt{2n}y_1$ and is thus bounded.

In general, the height of a regular n-simplex is the distance of a vertex to the circumcentre of its opposite regular $(n-1)$-simplex. By [12, p. 367], it holds that

$$h_i(y_1) = \sqrt{\frac{n(k+i)}{k+i-1}} y_1. \tag{4.71}$$

We start the computation by noting that

$$\rho_{\boldsymbol{Y},\boldsymbol{Z}}(\boldsymbol{y},\boldsymbol{z}) = \exp(-\sqrt{n}y_1) \cdot \mathbf{1}_{[0,\infty)}(y_1) \cdot \mathbf{1}_{\check{T}(y_1)}(\check{\boldsymbol{y}},\boldsymbol{z}). \tag{4.72}$$

The marginal distribution of $Z_i|\boldsymbol{Y}$ is given by

$$\rho_{Z_i|\boldsymbol{Y}}(z_i|\boldsymbol{y}) = \int \cdots \int \rho_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{z}|\boldsymbol{y}) \, \mathrm{d}z_1 \ldots \mathrm{d}z_{i-1} \, \mathrm{d}z_{i+1} \ldots \mathrm{d}z_{n-k} \tag{4.73}$$

and so we get

$$\rho_{\boldsymbol{Y}}(\boldsymbol{y})\rho_{Z_i|\boldsymbol{Y}}(z_i|\boldsymbol{y}) \tag{4.74}$$

$$= \int \cdots \int \rho_{\boldsymbol{Y},\boldsymbol{Z}}(\boldsymbol{y},\boldsymbol{z}) \, \mathrm{d}z_1 \ldots \mathrm{d}z_{i-1} \, \mathrm{d}z_{i+1} \ldots \mathrm{d}z_{n-k} \tag{4.75}$$

$$= \exp(-\sqrt{n}y_1) \cdot \mathbf{1}_{[0,\infty)}(y_1) \tag{4.76}$$

$$\cdot \int \cdots \int \mathbf{1}_{T(y_1)}(\check{\boldsymbol{y}},\boldsymbol{z}) \, \mathrm{d}z_1 \ldots \mathrm{d}z_{i-1} \, \mathrm{d}z_{i+1} \ldots \mathrm{d}z_{n-k}. \tag{4.77}$$

Using Jensen's inequality in a first step, we can continue with

$$\mathbf{E}[\mathbf{V}\mathrm{ar}(Z_i|\boldsymbol{Y})^{(1+\varepsilon)/\varepsilon}] \le \mathbf{E}[\mathbf{E}[Z_i^{2(1+\varepsilon)/\varepsilon} \mid \boldsymbol{Y}]] \tag{4.78}$$

$$= \int \left(\int z_i^{2(1+\varepsilon)/\varepsilon} \, \rho_{Z_i|\boldsymbol{Y}}(z_i|\boldsymbol{y}) \, \mathrm{d}z_i \right) \rho_{\boldsymbol{y}}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \tag{4.79}$$

$$= \int_0^\infty \exp(-\sqrt{n}y_1) \left(\int \int z_i^{2(1+\varepsilon)/\varepsilon} \cdot \mathbf{1}_{\check{T}(y_1)}(\check{\boldsymbol{y}},\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \, \mathrm{d}\check{\boldsymbol{y}} \right) \mathrm{d}y_1 \tag{4.80}$$

$$\le \int_0^\infty \exp(-\sqrt{n}y_1) \, h_i(y_1)^{2(1+\varepsilon)/\varepsilon} \left(\int \int \mathbf{1}_{\check{T}(y_1)}(\check{\boldsymbol{y}},\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \, \mathrm{d}\check{\boldsymbol{y}} \right) \mathrm{d}y_1 \tag{4.81}$$

$$= \int_0^\infty \exp(-\sqrt{n}y_1) \left(\sqrt{\frac{n(k+i)}{k+i-1}} y_1 \right)^{2(1+\varepsilon)/\varepsilon} \frac{\sqrt{n^n}}{(n-1)!} y_1^{n-1} \, \mathrm{d}y_1 \tag{4.82}$$

$$= \left(\frac{n(k+i)}{k+i-1} \right)^{(1+\varepsilon)/\varepsilon} \frac{\sqrt{n^n}}{(n-1)!} \int_0^\infty y_1^{n+1+2/\varepsilon} \exp(-\sqrt{n}y_1) \, \mathrm{d}y_1 \tag{4.83}$$

$$= \left(\frac{n(k+i)}{k+i-1} \right)^{(1+\varepsilon)/\varepsilon} \frac{\sqrt{n^n}}{(n-1)!} \frac{\Gamma(n+2+2/\varepsilon)}{n^{(1+\varepsilon)/\varepsilon}\sqrt{n^n}} \tag{4.84}$$

$$= \left(\frac{k+i}{k+i-1} \right)^{(1+\varepsilon)/\varepsilon} \frac{\Gamma(n+2+2/\varepsilon)}{(n-1)!}. \tag{4.85}$$

Note that an intermediate step of the previous calculation uses the fact that the volume of the regular $(n-1)$-simplex $\check{T}(y_1)$ with side length $\sqrt{2n}y_1$ is (see [12,

p. 367])

$$\int \int \mathbf{1}_{\check{T}(y_1)}(\check{\boldsymbol{y}}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \, \mathrm{d}\check{\boldsymbol{y}} = \frac{\sqrt{n}^n}{(n-1)!} y_1^{n-1}. \tag{4.86}$$

Remember from (4.21) and (4.22) that

$$\mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \le K(n-k)^{1/(1+\varepsilon)} C_{\mathbf{Var}}(\varepsilon, n, k) \tag{4.87}$$

with

$$C_{\mathbf{Var}}(\varepsilon, n, k) = \left(\sum_{i=1}^{n-k} \mathbf{E}[\mathbf{Var}(Z_i | \boldsymbol{Y})^{(1+\varepsilon)/\varepsilon}] \right)^{\varepsilon/(1+\varepsilon)} \tag{4.88}$$

$$\le \left(\frac{\Gamma(n+2+2/\varepsilon)}{(n-1)!} \sum_{i=1}^{n-k} \left(\frac{k+i}{k+i-1} \right)^{(1+\varepsilon)/\varepsilon} \right)^{\varepsilon/(1+\varepsilon)}. \tag{4.89}$$

Defining

$$C_\varepsilon(n, k) := (n-k)^{1/(1+\varepsilon)} \left(\frac{\Gamma(n+2+2/\varepsilon)}{(n-1)!} \sum_{i=1}^{n-k} \left(\frac{k+i}{k+i-1} \right)^{(1+\varepsilon)/\varepsilon} \right)^{\varepsilon/(1+\varepsilon)} \tag{4.90}$$

then yields

$$\mathbf{E}[C_{\boldsymbol{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} \le K C_\varepsilon(n, k). \tag{4.91}$$

Combining all bounds, we get that

$$C_{\mathrm{P},\varepsilon}(\varepsilon, n, k) \le K \cdot L^{\varepsilon/(1+\varepsilon)} \cdot C_\varepsilon(n, k) =: C_{\exp^n}(\varepsilon, n, k, L), \tag{4.92}$$

where $C_{\mathrm{P},\varepsilon}(\varepsilon, n, k)$ was defined in (4.26). We recall that n denotes the dimension of the problem, k the dimension of the active subspace, L is the upper bound on $\|\nabla f\|_2^2$, and K the universal constant from (4.21).

The result follows by Lemma 4.3. □

Fig. 6 depicts the quantity $C_\varepsilon(n, k = 1)$ from (4.90) as a function of $\varepsilon > 0$ for some $n \in \mathbf{N}$ (left plot) and as a function of $n \ge 2$ for several $\varepsilon > 0$ (right plot). We set $k = 1$ since this gives the maximum value for C_ε over all $k \ge 1$. As expected, the curves increase quickly as ε approaches zero or, respectively, n becomes large.

Remark. In the previous theorem, the exponentially distributed random variables are assumed to have unit rates. The computations can also be made for arbitrary rates ν_i, $i = 1, \ldots, n$. However, some modifications are necessary. Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^\top$ denote the vector of rates. To get again the worst case scenario as in the previous subsection (uniform distribution on a simplex structure), the coordinate system has to be rotated in such a way that the vector $(1, 0, \ldots, 0)^\top$ rotates to $\boldsymbol{\nu}/\|\boldsymbol{\nu}\|_2$. The structure of a regular simplex that is used in the estimates above is not present in this more general case. Instead, we get a general simplex whose heights are not as easy to compute as in the regular case. However, rough estimates can be achieved by enclosing the general simplex with a larger regular one.
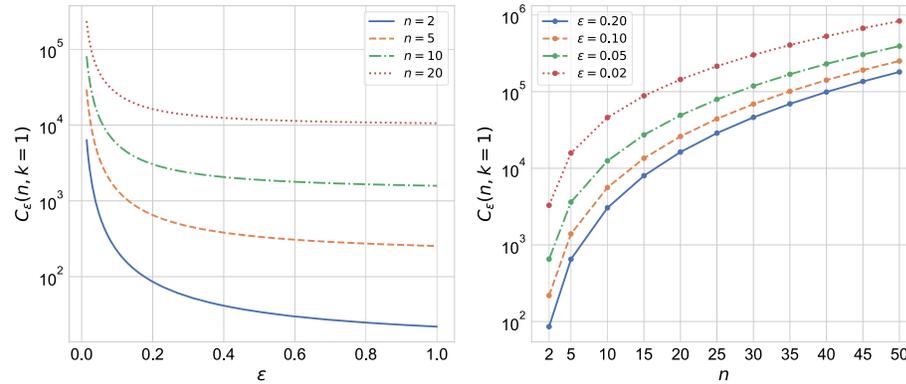
FIG 6. *The left plot shows curves of the map $\varepsilon \mapsto C_\varepsilon(n, k = 1)$ for $n \in \{2, 5, 10, 20\}$. Curves of $n \mapsto C_\varepsilon(n, k = 1)$ for $\varepsilon \in \{0.02, 0.05, 0.1, 0.2\}$ are displayed on the right.*

5. Future work with MGH distributions

The generalized bound from Lemma 4.3 and the study of corresponding Poincaré type constants $C_{\mathrm{P},\varepsilon,W}$ and $C_{\mathrm{P},\varepsilon}$ for independently exponentially distributed random variables in Subsection 4.3 motivate further similar investigations of more general distributions. From a statistical perspective, a study of the class of *multivariate generalized hyperbolic distributions* (MGH) (see e.g., [4]) can be considered as a next step since it allows for distributions with both non-zero skewness and heavier tails. An MGH is a distribution of the random vector

$$\boldsymbol{X} = \mu + \beta A + \sqrt{A} M \boldsymbol{V} \tag{5.1}$$

with location parameter $\mu \in \mathbf{R}^n$, skewness parameter $\beta \in \mathbf{R}^n$, and a symmetric positive definite matrix $M \in \mathbf{R}^{n \times n}$. The scalar random variable A, called the mixing variable, follows a generalized inverse Gaussian distribution (GIG) [26], and $\boldsymbol{V} \sim \mathcal{N}(0, I)$ is independent of A. As a particular example, for $\boldsymbol{X}$ to be Laplace distributed, we set $\beta = 0$ and let A be exponentially distributed [29]. Note that, however, the example from Subsection 4.3, assuming independently exponentially distributed random variables, is not an MGH. In order to include this case, we would need to introduce a mixing random matrix as scaling for $\boldsymbol{V}$.

Nevertheless, MGH is a large class containing classical distributions like the normal-inverse Gaussian, generalized Laplace, and Student's t-distribution. In particular, these distributions are interesting since they have been used in areas like, for instance, economics and financial markets [5, 6, 23], spatial and Geostatistics [9, 10, 42], and linear mixed-effects [2, 32, 47] which are used, e.g., for linear non-Gaussian time series models in medical longitudinal studies [2].

We mention that, under an assumption on a parameter, MGH distributions are log-concave [44], i.e., we can use the estimates on Poincaré constants $C_{\boldsymbol{Y}}$ of Bobkov from (4.20).

In our opinion, it is preferable to start the investigation with the subclass of symmetric MGH distributions, i.e., $\beta = 0$ in (5.1). The following lines demonstrate particular difficulties that we already encounter in this smaller subclass. Let us choose $\mu = \beta = 0$ and $M = I$ in (5.1) such that

$$\boldsymbol{X} = \sqrt{A}\boldsymbol{V} \tag{5.2}$$

with $\boldsymbol{V} \sim \mathcal{N}(0, I)$. A common first step is to study $\boldsymbol{X}$ conditioned on A, i.e., $\boldsymbol{X}|A \sim \mathcal{N}(0, AI)$, and to use the tower property of conditional expectations. That is, analogously to (2.2), we define

$$C := \mathbf{E}[C_A] = W\Lambda W^\top \tag{5.3}$$

with

$$C_A := \mathbf{E}[\nabla f(\boldsymbol{X})\nabla f(\boldsymbol{X})^\top \mid A] = W_A\Lambda_A W_A^\top. \tag{5.4}$$

Choosing $k \leq n - 1$ independent of A, we further set

$$\boldsymbol{Y}_A := W_{A,1}^\top \boldsymbol{X} \quad \text{and} \quad \boldsymbol{Z}_A := W_{A,2}^\top \boldsymbol{X}. \tag{5.5}$$

The computation starts, similar to (2.16), with

$$\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2 \mid A] \tag{5.6}$$

$$= \mathbf{E}[\mathbf{E}[(f([\![\boldsymbol{Y}_A, \boldsymbol{Z}_A]\!]_{W_A}) - g(\boldsymbol{Y}_A))^2 \mid \boldsymbol{Y}_A] \mid A] \tag{5.7}$$

$$\leq \mathbf{E}[C_{\boldsymbol{Y}_A}\mathbf{E}[\|\nabla^{\boldsymbol{z}_A} f([\![\boldsymbol{Y}_A, \boldsymbol{Z}_A]\!]_{W_A})\|_2^2 \mid \boldsymbol{Y}_A] \mid A] \tag{5.8}$$

$$= A\,\mathbf{E}[\mathbf{E}[\|\nabla^{\boldsymbol{z}_A} f([\![\boldsymbol{Y}_A, \boldsymbol{Z}_A]\!]_{W_A})\|_2^2 \mid \boldsymbol{Y}_A] \mid A] \tag{5.9}$$

$$= A\,\mathrm{trace}\,(\Lambda_{A,2})\,. \tag{5.10}$$

In (5.9), we use the fact that the Poincaré constant of a normal distribution $\mathcal{N}(0, AI)$ is $\lambda_{\max}(AI) = A$; see Section 3. The last step to (5.10) is equal to (2.19). This yields

$$\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2] = \mathbf{E}[\mathbf{E}[(f(\boldsymbol{X}) - f_g(\boldsymbol{X}))^2 \mid A]] \tag{5.11}$$

$$\leq \mathbf{E}[A \cdot \mathrm{trace}\,(\Lambda_{A,2})], \tag{5.12}$$

where the random variable $A\cdot\mathrm{trace}\,(\Lambda_{A,2})$ is assumed to have finite first moment.

At this point, as long as A is not compactly supported, we can only continue by applying another Hölder's inequality similar to the proof of Lemma 4.3. However, in any case, we have to face the problem that $\mathbf{E}[\mathrm{trace}\,(\Lambda_{A,2})]$ is, in general, not equal to $\mathrm{trace}\,(\Lambda_2)$ which denotes the inactive trace of C that we actually aim for. Nevertheless, we know that

$$\mathbf{E}[\mathrm{trace}\,(\Lambda_A)] = \mathrm{trace}\,(\Lambda)\,, \tag{5.13}$$

but it is unclear whether, and how, this equality can be exploited for our purposes.

6. Summary

This manuscript discusses bounds for the mean squared error of a given function of interest and a low-dimensional approximation of it which is found by the active subspace method. These bounds, consisting of the product of a Poincaré constant and a sum of eigenvalues belonging to a non-dominant subspace, are based on a probabilistic Poincaré inequality. Existing literature applies this Poincaré inequality with indirect non-explicit assumptions that, as a consequence, limit the class of distributions applicable for the active subspace method. For example, these assumptions exclude distributions with exponential tails as, e.g., exponential distributions. In this respect, the main results of this manuscript give details on the problem that arises when applying the active subspace method with log-concave distributions (which include exponential distributions). We are able to provide a scenario, involving independently exponentially distributed random variables, in which the usual estimates are not achievable due to an unbounded Poincaré constant. However, using Hölder's inequality with conjugates (p, q) $(p, q \in (1, \infty))$ instead of $(\infty, 1)$, we show that it is possible to derive a generalized result in a way that enables to balance the size of the Poincaré constant and the remaining order of the error. We exemplify this trade-off on the mentioned scenario and show that the size of the involved constant is very much depending on the dimension of the problem. Finally, we propose directions for future work related to the applicability of active subspaces to the large class of multivariate generalized hyperbolic distributions. Also, details are provided for particular difficulties that already arise with a smaller subclass of these.

Source code

Wolfram Mathematica notebooks and code for generating the plots in this manuscript are available in a repository at

<https://bitbucket.org/m-parente/asm-poincare-pub/>.

Acknowledgments

References

[1] Alonso-Gutiérrez, D. and Bastero, J. (2015). *Approaching the Kannan-Lovász-Simonovits and Variance Conjectures* **2131**. Springer. MR3308635

[2] ASAR, Ö., BOLIN, D., DIGGLE, P. J. and WALLIN, J. (2018). Linear Mixed-Effects Models for Non-Gaussian Repeated Measurement Data. *accepted for publication in Journal of the Royal Statistical Society: Series C.*

[3] BAKRY, D., GENTIL, I. and LEDOUX, M. (2013). *Analysis and Geometry of Markov Diffusion Operators* **348**. Springer Science & Business Media. MR3155209

[4] BARNDORFF-NIELSEN, O. (1977). Exponentially Decreasing Distributions for the Logarithm of Particle Size. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **353** 401–419.

[5] BARNDORFF-NIELSEN, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of statistics* **24** 1–13. MR1436619

[6] BARNDORFF-NIELSEN, O. E. (1997). Processes of Normal Inverse Gaussian Type. *Finance and stochastics* **2** 41–68. MR1804664

[7] BEBENDORF, M. (2003). A Note on the Poincaré Inequality for Convex Domains. *Zeitschrift für Analysis und ihre Anwendungen* **22** 751–756. MR2036927

[8] BOBKOV, S. G. (1999). Isoperimetric and Analytic Inequalities for Log-Concave Probability Measures. *The Annals of Probability* **27** 1903–1921. MR1742893

[9] BOLIN, D. (2014). Spatial Matérn Fields Driven by Non-Gaussian Noise. *Scandinavian Journal of Statistics* **41** 557–579. MR3249417

[10] BOLIN, D. and WALLIN, J. (2019). Multivariate Type G Matérn Stochastic Partial Differential Equation Random Fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).*

[11] BRIDGES, R., GRUBER, A., FELDER, C., VERMA, M. and HOFF, C. (2019). Active Manifolds: A Non-Linear Analogue to Active Subspaces. In *International Conference on Machine Learning* 764–772.

[12] BUCHHOLZ, R. H. (1992). Perfect Pyramids. *Bulletin of the Australian Mathematical Society* **45** 353–368. MR1165142

[13] BUDNINSKIY, M., YIN, G., FENG, L., TONG, Y. and DESBRUN, M. (2019). Parallel Transport Unfolding: A Connection-Based Manifold Learning Approach. *SIAM Journal on Applied Algebra and Geometry* **3** 266–291. MR3948233

[14] CHEN, L. H. (1982). An Inequality for the Multivariate Normal Distribution. *Journal of Multivariate Analysis* **12** 306–315. MR0661566

[15] COLEMAN, K. D., LEWIS, A., SMITH, R. C., WILLIAMS, B., MORRIS, M. and KHUWAILEH, B. (2019). Gradient-Free Construction of Active Subspaces for Dimension Reduction in Complex Models with Applications to Neutronics. *SIAM/ASA Journal on Uncertainty Quantification* **7** 117–142. MR3900802

[16] CONSTANTINE, P. and GLEICH, D. (2014). Computing Active Subspaces with Monte Carlo. arXiv preprint arXiv:1408.0545.

[17] CONSTANTINE, P. G. (2015). *Active Subspaces. SIAM Spotlights* **2**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA Emerg-

ing Ideas for Dimension Reduction in Parameter Studies. MR3486165

[18] Constantine, P. G., Dow, E. and Wang, Q. (2014). Active Subspace
Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM
J. Sci. Comput.* **36** A1500–A1524. MR3233940

[19] Constantine, P. G., Eftekhari, A., Hokanson, J. and Ward, R. A.
(2017). A Near-Stationary Subspace for Ridge Approximation. *Computer
Methods in Applied Mechanics and Engineering* **326** 402–421. MR3709198

[20] Constantine, P. G., Kent, C. and Bui-Thanh, T. (2016). Accelerating
Markov chain Monte Carlo with Active Subspaces. *SIAM J. Sci. Comput.*
38 A2779–A2805. MR3543164

[21] Cui, T., Martin, J., Marzouk, Y. M., Solonen, A. and Spantini, A.
(2014). Likelihood-Informed Dimension Reduction for Nonlinear Inverse
Problems. *Inverse Problems* **30** 114015, 28. MR3274599

[22] Diaz, P., Constantine, P., Kalmbach, K., Jones, E. and
Pankavich, S. (2018). A Modified SEIR Model for the Spread of Ebola in
Western Africa and Metrics for Resource Allocation. *Applied Mathematics
and Computation* **324** 141–155. MR3743663

[23] Eberlein, E. (2001). Application of Generalized Hyperbolic Lévy Motions
to Finance. In *Lévy processes* 319–336. Springer. MR1833703

[24] Glaws, A. and Constantine, P. G. (2018). A Lanczos-Stieltjes Method
for One-Dimensional Ridge Function Approximation and Integration. arXiv
preprint arXiv:1808.02095.

[25] Hokanson, J. M. and Constantine, P. G. (2018). Data-Driven Poly-
nomial Ridge Approximation Using Variable Projection. *SIAM Journal on
Scientific Computing* **40** A1566–A1589. MR3809540

[26] Jorgensen, B. (2012). *Statistical Properties of the Generalized Inverse
Gaussian Distribution* **9**. Springer Science & Business Media. MR0648107

[27] Kinsey, L. C. and Moore, T. E. (2006). *Symmetry, Shape and Space:
An Introduction to Mathematics Through Geometry.* Springer Science &
Business Media.

[28] Klenke, A. (2013). *Probability Theory: A Comprehensive Course.*
Springer Science & Business Media. MR3112259

[29] Kotz, S., Kozubowski, T. and Podgorski, K. (2012). *The Laplace Dis-
tribution and Generalizations: A Revisit with Applications to Communica-
tions, Economics, Engineering, and Finance.* Springer Science & Business
Media. MR1935481

[30] Lee, Y. T. and Vempala, S. S. (2018). The Kannan-Lovász-Simonovits
Conjecture. arXiv preprint arXiv:1807.03465.

[31] Leon, L. S., Smith, R. C., Oates, W. S. and Miles, P. (2017). Identifi-
ability and Active Subspace Analysis for a Polydomain Ferroelectric Phase
Field Model. In *ASME 2017 Conference on Smart Materials, Adaptive
Structures and Intelligent Systems* V002T03A022–V002T03A022. Ameri-
can Society of Mechanical Engineers.

[32] Lin, T. I. and Lee, J. C. (2007). Bayesian Analysis of Hierarchical Linear
Mixed Modeling Using the Multivariate *t* Distribution. *Journal of Statisti-
cal Planning and Inference* **137** 484–495. MR2298952

[33] MORAWIEC, A. (2003). *Orientations and Rotations.* Springer. MR2143690

[34] PINKUS, A. (2015). *Ridge Functions* **205**. Cambridge University Press. MR3559591

[35] STUART, A. M. (2010). Inverse Problems: A Bayesian Perspective. *Acta Numerica* **19** 451–559. MR2652785

[36] TEIXEIRA PARENTE, M., MATTIS, S., GUPTA, S., DEUSNER, C. and WOHLMUTH, B. (2019). Efficient Parameter Estimation for a Methane Hydrate Model with Active Subspaces. *Computational Geosciences* **23** 355–372.

[37] TEIXEIRA PARENTE, M., BITTNER, D., MATTIS, S. A., CHIOGNA, G. and WOHLMUTH, B. (2019). Bayesian Calibration and Sensitivity Analysis for a Karst Aquifer Model Using Active Subspaces. *Water Resources Research* **55** 7086–7107.

[38] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290** 2319–2323.

[39] TEZZELE, M., BALLARIN, F. and ROZZA, G. (2018). Combined Parameter and Model Reduction of Cardiovascular Problems by Means of Active Subspaces and POD-Galerkin Methods. In *Mathematical and Numerical Modeling of the Cardiovascular System and Applications* 185–207. Springer. MR3887546

[40] TRIPATHY, R., BILIONIS, I. and GONZALEZ, M. (2016). Gaussian Processes with Built-In Dimensionality Reduction: Applications to High-Dimensional Uncertainty Propagation. *Journal of Computational Physics* **321** 191–223. MR3527564

[41] VAN HANDEL, R. (2016). Probability in High Dimension Technical Report, Princeton University.

[42] WALLIN, J. and BOLIN, D. (2015). Geostatistical Modelling Using Non-Gaussian Matérn Fields. *Scandinavian Journal of Statistics* **42** 872–890. MR3391697

[43] WOLFRAM RESEARCH, INC. Mathematica, Version 11.3. Champaign, IL, 2018.

[44] YU, Y. (2017). On Normal Variance-Mean Mixtures. *Statistics & Probability Letters* **121** 45–50. MR3575408

[45] ZAHM, O., CONSTANTINE, P., PRIEUR, C. and MARZOUK, Y. (2018). Gradient-Based Dimension Reduction of Multivariate Vector-Valued Functions. arXiv preprint arXiv:1801.07922.

[46] ZAHM, O., CUI, T., LAW, K., SPANTINI, A. and MARZOUK, Y. (2018). Certified Dimension Reduction in Nonlinear Bayesian Inverse Problems. arXiv preprint arXiv:1807.03712v2.

[47] ZHANG, P., QIU, Z., FU, Y. and SONG, P. X.-K. (2009). Robust Transformation Mixed-Effects Models for Longitudinal Continuous Proportional Data. *Canadian Journal of Statistics* **37** 266–281. MR2531831