Uncertainty quantification via codimension-one partitioning

T. J. Sullivan[1,*,†], U. Topcu[2], M. McKerns[3] and H. Owhadi[4]

[1]*Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, U.S.A.*
[2]*Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, U.S.A.*
[3]*Materials Science, California Institute of Technology, Pasadena, CA 91125, U.S.A.*
[4]*Applied and Computational Mathematics and Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, U.S.A.*

SUMMARY

We consider uncertainty quantification in the context of certification, i.e. showing that the probability of some 'failure' event is acceptably small. In this paper, we derive a new method for rigorous uncertainty quantification and conservative certification by combining McDiarmid's inequality with input domain partitioning and a new concentration-of-measure inequality. We show that arbitrarily sharp upper bounds on the probability of failure can be obtained by partitioning the input parameter space appropriately; in contrast, the bound provided by McDiarmid's inequality is usually not sharp. We prove an error estimate for the method (Proposition 3.2); we define a codimension-one recursive partitioning scheme and prove its convergence properties (Theorem 4.1); finally, we apply a new concentration-of-measure inequality to give confidence levels when empirical means are used in place of exact ones (Section 5). Copyright © 2010 John Wiley & Sons, Ltd.

1. INTRODUCTION

In [1], it was proposed that McDiarmid's inequality [2] be used in the context of certification to provide rigorous upper bounds on the failure probability of a real-valued system of interest, f, the response of which depends on a collection of independent random inputs. McDiarmid's inequality has the advantage of providing an upper bound in terms of only three quantities: the mean response of the system, $\mathbb{E}[f]$; the failure threshold, θ; and a measure of system spread known as the McDiarmid diameter, $\mathscr{D}[f]$. McDiarmid's inequality implies that, for any product probability measure μ on the inputs of f

$$\mu[f \leqslant \theta] \leqslant \exp\left(-\frac{2(\mathbb{E}[f] - \theta)_+^2}{\mathscr{D}[f]^2}\right)$$

see Section 2 for the definitions and notation in use. One disadvantage of McDiarmid's inequality is that it is a worst-case estimate that takes a global view of the response function: even if the response function exhibits large plateaus of 'success' with only small localized regions of failure, McDiarmid's inequality is unable to use this to any advantage. In this paper, we propose a

---

generalization of McDiarmid's inequality that, by a simple partitioning argument, takes a more local view and thereby produces sharper upper bounds. We also propose a partitioning algorithm that uses McDiarmid diameters to generate sequences of partitions that are associated with arbitrarily sharp upper bounds. This codimension-one partitioning algorithm is well suited for the study of response functions that have high-dimensional input parameter spaces.

In many applications, interest lies in certifying that the probability of failure $\mu[f \leqslant \theta]$ is acceptably small [3]. A straightforward approach to certification is to bound the failure probability empirically, i.e. solely through experimental testing, which amounts to a Monte Carlo approach. The problem with this approach is that the number of experiments required to certify a system with probability of failure of order p up to a statistical error of order ε requires of the order of $(1/p^2)\log(1/\varepsilon)$ experiments, which can be prohibitively expensive if the tests are costly and the failure probability is small; furthermore, the probability-of-failure estimate so obtained may not actually be an upper bound. Indeed, there is a growing awareness that rare failure events require careful study and treatment [4]. It is worth noting that, although this paper focusses only on certification, uncertainty quantification is not limited to certification alone: there is also considerable interest in studying the propagation of uncertainties through models and systems; for some systems, polynomial chaos expansions are well adapted for this task [5, 6].

In contrast to Monte Carlo methods, McDiarmid's inequality offers a conservative certification criterion, and the number of experiments required to bound the probability of failure using McDiarmid's inequality is independent of p and depends only on the number of experiments required to determine the mean system performance and McDiarmid diameter (up to acceptable error tolerances). In this sense, the partitioning method of this paper occupies a middle ground between McDiarmid's inequality and Monte Carlo methods, in that its application to situations involving sample data (Section 5) relies on having enough samples to estimate the probabilities of subsets of the input parameter space that are significantly larger than the failure region $[f \leqslant \theta]$, as well as mean performances and diameters on those sets, but does not require extensive sampling of the failure region itself.

Our method makes extensive use of McDiarmid subdiameters and does so in three ways: they are used as sensitivity indices; they are used to provide upper bounds on the probability of failure; and they are used to generate refined partitions of the parameter space and hence sharper upper bounds on the probability of failure without succumbing to the naïve curse of dimension. Insofar as the McDiarmid subdiameters are used as sensitivity indices, this paper forms part of the literature on sensitivity analysis, including Sobol' indices [7, 8], Jansen indices [9], Fourier amplitude sensitivity tests [10–12], and random balance designs [13]. McDiarmid subdiameters have wider applicability than sensitivity analysis, however, since they give rise to rigorous upper and lower bounds on the probabilities of certain sets via concentration-of-measure inequalities like that of McDiarmid. Once formulated, such inequalities can then be re-phrased to provide certification in the context of hypothesis testing; see, e.g. [14].

In Section 5, we consider the problem of replacing exact mean performances by empirically sampled ones. In this situation, the upper bound on the probability of failure is a function of the random empirical observations; therefore, it is important to know what level of confidence can be ascribed to this probabilistic upper bound. This is usually done by allowing for some margin of uncertainty in the mean performance. Intuitively, one might expect that the overall level of confidence is exponentially decreasing with respect to the cardinality of the partition used; however, application of a new concentration inequality (fully treated in (Sullivan and Owhadi, in preparation)) yields a surprising result that the confidence level is exponentially *increasing* with respect to the cardinality of the partition.

2. NOTATION AND BACKGROUND

Throughout, $(\mathscr{X}_j, d_j, \mu_j)$ will be a metric probability space for $j = 1, \ldots, n$. $\mathscr{X} := \mathscr{X}_1 \times \cdots \times \mathscr{X}_n$ will denote the Cartesian product space, equipped with the product probability measure $\mu := \mu_1 \otimes \cdots \otimes \mu_n$

and the ℓ^1 metric $d(x,y) := \sum_{i=1}^{n} d_i(x_i, y_i)$. An alternative point of view is that of a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and independent coordinate processes $X_j : \Omega \to \mathscr{X}_j$ with law $(X_j)_* \mathbb{P} = \mu_j$, so that $X := (X_1, \ldots, X_n) : \Omega \to \mathscr{X}$ has the joint law $X_* \mathbb{P} = \mu$. $\mathbb{E}[\cdot]$ denotes the expectation operator with respect to μ.

For a subset $A \subseteq \mathscr{X}$, $d(A)$ will denote its diameter with respect to the metric d, i.e.

$$d(A) := \sup\{d(x,y) | x, y \in A\} \tag{1}$$

and $d_j(A)$ will denote the diameter of A with respect to the metric d_j, i.e. the diameter of its projection onto the space $\mathscr{X}_j$.

The prototypical situation is that $\mathscr{X}_j = [a_j, b_j] \subseteq \mathbb{R}$, that d_j is the Euclidean metric, and that μ_j is normalized Lebesgue (uniform) measure on the interval $[a_j, b_j]$. A slight generalization is to take μ_j to be a probability measure on $[a_j, b_j]$ that is absolutely continuous (i.e. has a density) with respect to the Lebesgue measure. The compactness of $\mathscr{X}$ is useful for the partitioning algorithm that will be discussed in Section 4, but is not required for the general theory.

For a function $f : \mathscr{X} \to \mathbb{R}$, diam$[f]$ denotes the *diameter of the image of* f as a subset of $\mathbb{R}$, i.e.

$$\mathrm{diam}[f] := \sup_{x, y \in \mathscr{X}} |f(x) - f(y)| = \sup_{x \in \mathscr{X}} f(x) - \inf_{x \in \mathscr{X}} f(x). \tag{2}$$

$\mathscr{D}_j[f]$ denotes the *McDiarmid subdiameter* of f in the jth coordinate direction:

$$\mathscr{D}_j[f] := \sup\left\{ |f(x) - f(y)| \;\middle|\; \begin{array}{l} x = (x_1, \ldots, x_n) \in \mathscr{X}, \\ y = (y_1, \ldots, y_n) \in \mathscr{X}, \\ x_k = y_k \in \mathscr{X}_k \text{ for every } k \neq j \end{array} \right\}. \tag{3}$$

Note that $\mathscr{D}_j[\cdot]$ is a semi-norm on the space of bounded functions $f : \mathscr{X} \to \mathbb{R}$. $\mathscr{D}_j[f]$ is equivalently defined by

$$\mathscr{D}_j[f] := \sup\{\mathrm{diam}[f(x_1, \ldots, x_{j-1}, \cdot, x_{j+1}, \ldots, x_n)] | x_k \in \mathscr{X}_k \text{ for } k \neq j\}. \tag{4}$$

$\mathscr{D}[f]$ denotes the *McDiarmid diameter* of f:

$$\mathscr{D}[f] := \left(\sum_{j=1}^{n} \mathscr{D}_j[f]^2 \right)^{1/2}. \tag{5}$$

Note that

$$\mathscr{D}[f] \leqslant \mathrm{diam}[f] \leqslant \sum_{j=1}^{n} \mathscr{D}_j[f] \tag{6}$$

with equality throughout when $n = 1$. If n is large, then typically it holds that $\mathscr{D}[f] \ll \mathrm{diam}[f]$; heuristically, this is analogous to the observation that, for large n, the unit Euclidean n-ball occupies little volume in the unit n-cube.

McDiarmid's inequality is one of a family of results known as *concentration-of-measure inequalities* [15, 16]. The concentration of measure is based on a simple but non-trivial observation originally due to Lévy [17]: product probability measures in high-dimensional spaces tend to be very concentrated about their mean: functions of many independent variables with small sensitivity to each individual input are nearly constant. Concentration-of-measure inequalities quantify this concentration as being exponentially small in the distance from the mean (or median) value. In the case of McDiarmid's inequality, the concentration rate is controlled by the McDiarmid diameter of the function.

Theorem 2.1 (McDiarmid's inequality: [2])
If $f : \mathscr{X} \to \mathbb{R}$ is absolutely integrable with respect to μ (that is, if $\mathbb{E}[|f|]$ is finite), then, for any $r > 0$,

$$\mu[f - \mathbb{E}[f] \geqslant r] \leqslant \exp\left(-\frac{2r^2}{\mathscr{D}[f]^2} \right) \tag{7a}$$

and

$$\mu[f - \mathbb{E}[f] \leqslant -r] \leqslant \exp\left(-\frac{2r^2}{\mathscr{D}[f]^2}\right). \tag{7b}$$

Given a system of interest modeled by a real-valued function $f: \mathscr{X} \to \mathbb{R}$, $\theta \in \mathbb{R}$ will denote a possible value of f that is considered to be a *failure threshold*: the event $[f \leqslant \theta]$ represents the failure of the system f, and the complementary event $[f > \theta]$ represents the success of the system f. Under the assumption that the random inputs of f (i.e. the coordinate processes $X_1, \ldots, X_n$) are independent (see Section 6.1 for remarks on the dependent case), McDiarmid's inequality implies that the μ-probability of failure is bounded as follows:

$$\mu[f \leqslant \theta] \leqslant \exp\left(-\frac{2(\mathbb{E}[f] - \theta)_+^2}{\mathscr{D}[f]^2}\right), \tag{8}$$

where, for $t \in \mathbb{R}$, $t_+ := \max\{0, t\}$ and $t_- := |\min\{0, t\}|$. Note that the sensible convention for $0/0$ in the exponent is $0/0 = 0$, since a constant function f with value less than θ has $\mu[f \leqslant \theta] = 1 = \mathrm{e}^{-0}$; on the other hand, the sensible convention for $t/0$ in the exponent, with $t > 0$, is $t/0 = +\infty$, since a constant function f with value greater than θ has $\mu[f \leqslant \theta] = 0 = \mathrm{e}^{-\infty}$.

A *rectangular partition* of $\mathscr{X}$ is a finite or countably infinite collection $\mathscr{A}$ of pairwise-disjoint subsets of $\mathscr{X}$, such that

$$\mathscr{X} = \biguplus_{A \in \mathscr{A}} A$$

and such that each $A \in \mathscr{A}$ is a product of measurable subsets of the factor spaces: for each $A \in \mathscr{A}$, there exist measurable sets $A_j \subseteq \mathscr{X}_j$ such that $A = A_1 \times \cdots \times A_n$. The elements of a rectangular partition will also be referred to as *boxes*. See Section 6.1 for remarks on non-rectangular partitions. The *fineness* (or *mesh*) of a rectangular partition $\mathscr{A}$ is defined to be $\sup_{A \in \mathscr{A}} d(A)$.

3. PARTITIONING

Let $\mathscr{A}$ be a not-necessarily-rectangular partition of $\mathscr{X}$. (Recall that all partitions in this paper are assumed to be finite or countably infinite.) Observe that, by pairwise disjointness and using the definition of conditional probability,

$$\mu[f \leqslant \theta] = \sum_{A \in \mathscr{A}} \mu([f \leqslant \theta] \cap A) \tag{9a}$$

$$= \sum_{A \in \mathscr{A}} \mu(A) \mu[f \leqslant \theta | A]. \tag{9b}$$

In the case of a rectangular partition, McDiarmid's inequality (8) implies that

$$\mu[f \leqslant \theta] \leqslant \overline{\mu_{\mathscr{A}}}[f \leqslant \theta], \tag{10}$$

where

$$\overline{\mu_{\mathscr{A}}}[f \leqslant \theta] := \sum_{A \in \mathscr{A}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2}\right). \tag{11}$$

In writing this expression, we take advantage of a convenient abuse of notation: according to the reader's preference, '$\mathbb{E}[f|A]$' can be read either as the conditional expectation of f with respect to (the σ-algebra generated by) A or simply as the mean value with respect to $\mu(\cdot|A)$ of the restricted function $f|A: A \to \mathbb{R}$; the same applies for the restricted McDiarmid diameter $\mathscr{D}[f|A]$.

By the above argument, $\overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$ is always an upper bound for $\mu[f \leqslant \theta]$, but this can be improved upon: if $\inf_{x \in A} f(x) > \theta$ for some $A \in \mathscr{A}$, then

$$0 = \mu[f \leqslant \theta | A] \leqslant \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2}\right).$$

Therefore, define

$$\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta] := \sum_{A \in \mathscr{A}} \mu(A)\mathbb{1}_{(-\infty, \theta]}(\inf f(A)) \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2}\right) \qquad (12a)$$

$$= \sum_{\substack{A \in \mathscr{A} \\ \inf f(A) \leqslant \theta}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2}\right). \qquad (12b)$$

That is, $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$ is the same as $\overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$ except that, for those $A \in \mathscr{A}$ on which the local probability of failure $\mu[f \leqslant \theta | A]$ is known to be zero, zero is used as an upper bound instead of the McDiarmid exponential.

It is clear that, for any partition $\mathscr{A}$ of $\mathscr{X}$, it holds true that

$$0 \leqslant \mu[f \leqslant \theta] \leqslant \widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta] \leqslant \overline{\mu_{\mathscr{A}}}[f \leqslant \theta] \leqslant 1. \qquad (13)$$

Example 3.1
For a simple example of the improvements afforded by the partitioning method, consider $\mathscr{X} := [0, 1] \subseteq \mathbb{R}$, $\mu =$ uniform measure, and the piecewise-affine function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} 1, & 0 \leqslant x \leqslant \frac{1}{3}, \\ 2 - 3x, & \frac{1}{3} \leqslant x \leqslant \frac{2}{3}, \\ 0, & \frac{2}{3} \leqslant x \leqslant 1. \end{cases}$$

Let the failure region be $(-\infty, \frac{1}{4}]$. The exact probability of failure can be calculated explicitly as $\mu[f \leqslant \frac{1}{4}] = \frac{5}{12} \approx 0.4167$. Global application of McDiarmid's inequality yields an upper bound that is far from being sharp: since $\mathbb{E}[f] = \frac{1}{2}$ and $\mathscr{D}[f] = 1$, McDiarmid's inequality (8) implies that

$$\mu[f \leqslant \tfrac{1}{4}] \leqslant \exp\left(-\frac{2\left(\frac{1}{4}\right)^2}{1^2}\right) \approx 0.8825.$$

On the other hand, taking $A_1 := [0, \frac{1}{3})$, $A_2 := [\frac{1}{3}, \frac{2}{3})$, $A_3 := [\frac{2}{3}, 1]$ and applying (11) yields that

$$\mu\left[f \leqslant \tfrac{1}{4}\right] \leqslant \sum_{i=1}^3 \mu(A_i) \exp\left(-\frac{2\left(\mathbb{E}[f|A_i] - \frac{1}{4}\right)_+^2}{\mathscr{D}[f|A_i]^2}\right)$$

$$= \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot e^{-1/8} + \frac{1}{3} \cdot 1$$

$$\approx 0.6275,$$

which is a sharper bound than that supplied by applying McDiarmid's inequality (8) globally, i.e. to the whole of $\mathscr{X}$ instead of the partition $\mathscr{A} = \{A_1, A_2, A_3\}$.

Application of this partitioning method assumes knowledge and free exercise of the functions μ and f: μ is needed to calculate $\mu(A)$, f is needed to calculate $\mathscr{D}[f|A]$, and both are needed to

calculate $\mathbb{E}[f|A]$. See Section 5 for a discussion of what can be done if this assumption does not hold and sample data are used to determine, say, $\mathbb{E}[f|A]$.

It is natural to ask whether or not $\mu[f \leqslant \theta]$ is well approximated by the bounds $\overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$ and $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$. For example, does there exist a sequence of partitions $(\mathscr{A}(k))_{k \in \mathbb{N}_0}$ such that

$$\mu[f \leqslant \theta] = \lim_{k \to \infty} \overline{\mu_{\mathscr{A}(k)}}[f \leqslant \theta]? \tag{14}$$

If not, does there at least exist a sequence of partitions $(\mathscr{A}(k))_{k \in \mathbb{N}_0}$ such that

$$\mu[f \leqslant \theta] = \lim_{k \to \infty} \widetilde{\mu_{\mathscr{A}(k)}}[f \leqslant \theta]? \tag{15}$$

In Section 4, algorithms for constructing such sequences will be defined and examined. It seems intuitively reasonable that a sequence of partitions $(\mathscr{A}(k))_{k \in \mathbb{N}_0}$ should satisfy (14) or (15) if the fineness of $\mathscr{A}(k)$ tends to zero in an appropriate way. The following result puts this intuition on a firm footing. In fact, the error is controlled not by the fineness of the partition but instead by the local subdiameters $\mathscr{D}_j[f|A]$ on those sets $A \in \mathscr{A}$ on which f succeeds somewhere. To make this precise, we introduce notation for these sets: given a partition $\mathscr{A}$ and $\delta > 0$, define

$$\mathscr{A}_{\delta} := \{A \in \mathscr{A} \mid f(A) \cap (\theta + \delta, +\infty) \neq \emptyset\},$$

$$\mathscr{A}_{\delta}^{+} := \{A \in \mathscr{A} \mid f(A) \subseteq (\theta + \delta, +\infty)\} \subseteq \mathscr{A}_{\delta}.$$

That is, an element A of a partition $\mathscr{A}$ is an element of $\mathscr{A}_{\delta}$ if the response function f exceeds the threshold for success by more than δ *somewhere* in A; A is an element of $\mathscr{A}_{\delta}^{+}$ if the response function f exceeds the threshold for success by more than δ *everywhere* in A.

Proposition 3.2 (error bound)
Let $f : \mathscr{X} \to \mathbb{R}$ be measurable and let $\mathscr{A}$ be a partition of $\mathscr{X}$. Then, for every $\varepsilon > 0$, and for all sufficiently small $\delta > 0$,

$$0 \leqslant \overline{\mu_{\mathscr{A}}}[f \leqslant \theta] - \mu[f \leqslant \theta] < \varepsilon + \sup_{A \in \mathscr{A}_{\delta}} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^{n} \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right). \tag{16}$$

In addition, for every $\varepsilon > 0$, and for all sufficiently small $\delta > 0$,

$$0 \leqslant \widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta] - \mu[f \leqslant \theta] < \varepsilon + \sup_{A \in \mathscr{A}_{\delta} \setminus \mathscr{A}_{\delta}^{+}} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^{n} \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right). \tag{17}$$

The proof of Proposition 3.2 is in the Appendix. The only difference between the inequality for $\overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$ and the inequality for $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$ is that the second inequality involves a supremum over a smaller subset of $\mathscr{A}$: only the sets $A \in \mathscr{A}$ that are near the frontier of the failure region contribute to the error for $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$.

The following corollary of Proposition 3.2 for sequences of suitably fine partitions is immediate:

Corollary 3.3 (sufficient condition for convergence)
Let $f : \mathscr{X} \to \mathbb{R}$ be measurable. If $(\mathscr{A}(k))_{k \in \mathbb{N}_0}$ is a sequence of partitions of $\mathscr{X}$ such that, for all sufficiently small $\delta > 0$,

$$\lim_{k \to \infty} \sup_{\substack{A \in \mathscr{A}(k)_{\delta} \\ 1 \leqslant j \leqslant n}} \mathscr{D}_j[f|A] = 0 \tag{18}$$

then $\overline{\mu_{\mathscr{A}(k)}}[f\leqslant\theta]\to\mu[f\leqslant\theta]$ from above as $k\to\infty$. Similarly, if $(\mathscr{A}(k))_{k\in\mathbb{N}}$ is a sequence of partitions of $\mathscr{X}$ such that, for all sufficiently small $\delta>0$,

$$\lim_{k\to\infty}\sup_{\substack{A\in\mathscr{A}(k)_\delta\setminus\mathscr{A}(k)_\delta^+\\1\leqslant j\leqslant n}}\mathscr{D}_j[f|A]=0 \tag{19}$$

then $\widetilde{\mu_{\mathscr{A}(k)}}[f\leqslant\theta]\to\mu[f\leqslant\theta]$ from above as $k\to\infty$.

If f is continuous, then the hypotheses of Proposition 3.2 or Corollary 3.3 can be verified using the fineness of the partition $\mathscr{A}$ and the modulus of continuity ϖ_f of f defined by

$$\varpi_f(\delta):=\sup\{|f(x)-f(y)||x,y\in\mathscr{X},\mathrm{d}(x,y)\leqslant\delta\} \tag{20}$$

through the simple inequality that $\mathscr{D}_j[f|A]\leqslant\varpi_f(d_j(A))$ whenever A is a rectangle. If f is strongly discontinuous, then the fineness of $\mathscr{A}$, even near the frontier of the failure region, is of little use in constructing an error bound. Consider $f=\mathbb{1}_{\mathbb{Q}\cap[0,1]}:[0,1]\to\mathbb{R}$ with failure region $[f=0]$ as a simple pathological example: let $\mathscr{A}$ be any partition of $[0,1]$ into pairwise-disjoint intervals of non-zero length; then, for every $A\in\mathscr{A}$, it holds that $\mathscr{D}[f|A]=1$ and $\mathbb{E}[f|A]=1$, which yields the upper bound

$$\overline{\mu_{\mathscr{A}}}[f=0]=\mathrm{e}^{-2}>0=\mu[f=0].$$

Note, however, that the continuity of f is not a prerequisite for convergence. For example, the CORPUS algorithm of the following section produces sequences of partitions that satisfy the hypotheses of Corollary 3.3 even for discontinuous f, so long as the discontinuities of f are 'topologically respectable' (e.g. a rectifiable $(n-1)$-dimensional subset of $\mathscr{X}$).

It is worthwhile to review the relative advantages of the Monte Carlo, McDiarmid and partitioned McDiarmid methods are in order, since no one method is the best under all circumstances. As previously noted, a weakness of Monte Carlo methods is that they are extremely expensive in terms of number of trials (evaluations of f) if the aim is to certify that the probability of failure is very small relative to 1: to certify that the probability of failure is at most p, up to a statistical error of order ε, requires of the order of $(1/p^2)\log(1/\varepsilon)$ experiments; conversely, Monte Carlo methods are appropriate if p is not too small. In contrast, the number of experiments required to evaluate McDiarmid's bound does not depend upon the rarity of failure: instead, it depends largely on the regularity of f and the dimension of $\mathscr{X}$, since these factors govern the ease with which the optimization calculation of $\mathscr{D}[f]$ can be performed. The McDiarmid approach offers a qualitative advantage over the Monte Carlo approach in that it explicitly separates the performance margin $(\mathbb{E}[f]-\theta)_+$ from the system sensitivity $\mathscr{D}[f]$: this separation of rôles may be desirable in some settings. In particular, the exercise of calculating the subdiameters $\mathscr{D}_i[f]$ provides useful information about the behavior of f. The main reason why the partitioned McDiarmid approach might be preferred over the global McDiramid inequality is that the McDiarmid diameter cannot tell the difference between a localized cliff/spike in the graph of f and a more evenly spread out variation; the partitioned McDiarmid approach overcomes this deficiency and provides arbitrarily sharp bounds on the probability of failure, but at the increased computational cost of having to determine multiple local McDiarmid diameters.

4. CODIMENSION-ONE PARTITIONING ALGORITHMS

A simple method for constructing a sequence of partitions that satisfy the hypotheses of Corollary 3.3 is to bisect all the sets of $\mathscr{A}(k)$ in all n coordinate directions to produce the sets of $\mathscr{A}(k+1)$. This naïve algorithm makes no use of any information about the response function f and is strongly affected by the curse of dimension: in each iteration of the naïve algorithm, each set $A\in\mathscr{A}(k)$ gives rise to 2^n child sets in $\mathscr{A}(k+1)$.

For simplicity, we now restrict ourselves to the prototypical situation in which each $\mathscr{X}_j:=[a_j,b_j]$ is a compact interval in $\mathbb{R}$, and hence $\mathscr{X}:=\prod_{j=1}^n[a_j,b_j]$ is a compact rectangular box in $\mathbb{R}^n$.
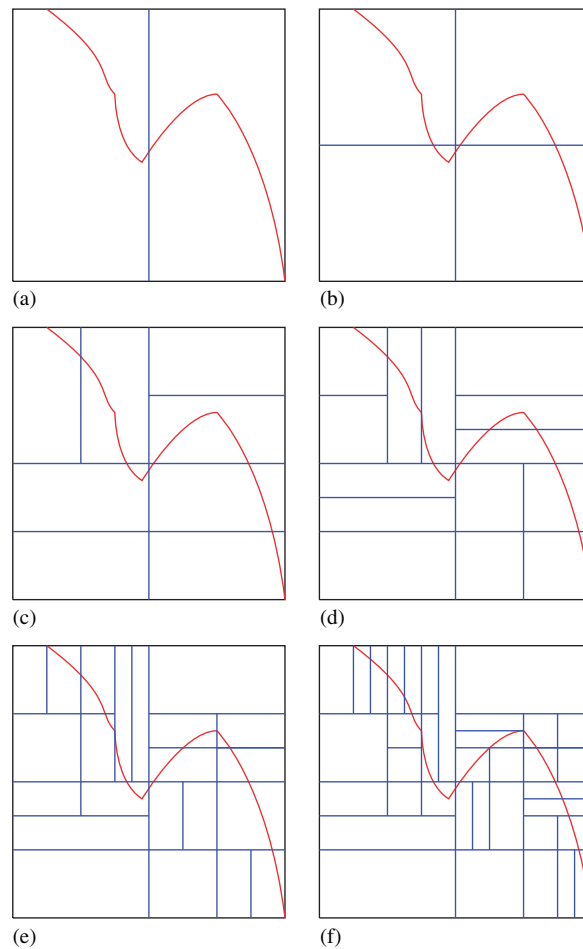
Figure 1. A schematic illustration of six iterations of the CORPUS algorithm. The curve shows the frontier of the failure region; the lines show the boundaries of the sets forming the partition $\mathscr{A}(k)$. The trivial initial partition $\mathscr{A}(0)=\{\mathscr{X}\}$ is not shown. Note how the sets of small metric diameter concentrate around the frontier of the failure region: (a) $\mathscr{A}(1)$: 1 new child set; (b) $\mathscr{A}(2)$: 2 new child sets; (c) $\mathscr{A}(3)$: 4 new child sets; (d) $\mathscr{A}(4)$: 6 new child sets; (e) $\mathscr{A}(5)$: 9 new child sets; and (f) $\mathscr{A}(6)$: 14 new child sets.

In this section, we propose a recursive algorithm for the production of partitions $\mathscr{A}$ such that (under appropriate hypotheses on f) the upper bound $\widetilde{\mu}_{\mathscr{A}}[f \leqslant \theta]$ is close to the true probability of failure $\mu[f \leqslant \theta]$. We refer to this algorithm as CORPUS (Codimension-One Recursive Partitioning Using Subdiameters). See Figure 1 for a schematic illustration of six iterations of the CORPUS algorithm. As can be seen from that figure, the CORPUS algorithm produces partitions with a binary tree structure similar to structures produced by branch-and-bound methods [18] and tree-structured density estimation techniques [19]. The most obvious difference between the naïve approach and the CORPUS algorithm is that the CORPUS algorithm bisects each $A \in \mathscr{A}(k)$ in at most one coordinate direction, whereas the naïve algorithm bisects each $A \in \mathscr{A}(k)$ in all n coordinate directions

CORPUS Algorithm

(1) Let $\mathscr{A}(0)$ be any rectangular partition of $\mathscr{X}$. Usually, $\mathscr{A}(0)=\{\mathscr{X}\}$, but this is not a necessary assumption. (In particular, $\mathscr{A}(0)$ could be given by heuristic means, by an expert judgement, or by the output of another partitioning algorithm.)

(2) Given $\mathscr{A}(k)$, recursively define $\mathscr{A}(k+1)$ as follows: for each $A \in \mathscr{A}(k)$,

  (a) if $A \in \mathscr{A}(k)$ satisfies $\inf_{x \in A} f(x) > \theta$ (i.e. f always succeeds on A), then include A in $\mathscr{A}(k+1)$ as it is,

(b) if $A \in \mathscr{A}(k)$ satisfies $\sup_{x \in A} f(x) \leqslant \theta$ (i.e. f always fails on A), then include A in $\mathscr{A}(k+1)$ as it is,

(c) otherwise,

   (i) determine $j \in \{1, \ldots, n\}$ such that $\mathscr{D}_j[f|A]$ is maximal (choose one such j arbitrarily if there are multiple maximizers),
   (ii) set $c(A) := \int_A x \, \mathrm{d}x$, the geometric center of A;
   (iii) bisect A by a hyperplane of codimension one (i.e. of dimension $n-1$) through $c(A)$ and normal to $\widehat{e}_j$, the unit vector in the jth coordinate direction,
   (iv) include in $\mathscr{A}(k+1)$ the two subsets of A so generated, but not the original set A; the two new sets are called the *children* or *child sets* of A.

(3) If $\mathscr{A}(k) = \mathscr{A}(k+1)$, or some other termination criterion (e.g. maximum number of boxes, minimum box size) is met, then the algorithm terminates.

Before proving the convergence properties of the CORPUS algorithm, a few remarks are in order.

First, observe that CORPUS delays the onset of the curse of dimension that faces the naïve partitioning algorithm since it uses the subdiameters as measures of parameter sensitivity for f. At least initially, only the most sensitive input parameter ranges will be subject to partitioning. The number of boxes (partition elements) is much better controlled under CORPUS (in which case $\#\mathscr{A}(k) \leqslant 2^k \#\mathscr{A}(0)$) than the naïve algorithm (in which case $\#\mathscr{A}(k) \leqslant 2^{nk} \#\mathscr{A}(0)$).

Second, note that the algorithm pre-supposes that it is feasible to compute the infimum, supremum and subdiameters of f over each A, and the geometric center of each A. Note, however, that neither expectations nor the probability masses of sets $A \in \mathscr{A}(k)$ need to be computed: those calculations only need to be performed if one wishes to keep track of the upper bounds $\overline{\mu_{\mathscr{A}(k)}}[f \leqslant \theta]$ and $\widetilde{\mu_{\mathscr{A}(k)}}[f \leqslant \theta]$ 'on the fly' during the execution of the algorithm. For example, if computational resources are limited, it may be desirable to split a set $A \in \mathscr{A}(k)$ to form two child sets in $\mathscr{A}(k+1)$ only if the contribution

$$\mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2}\right)$$

to the probability-of-failure bound is large relative to the corresponding contributions from the other sets $A' \in \mathscr{A}(k)$; to make such comparisons, probability masses and expectations must be computed (or must be known *a priori*).

Third, there is an opportunity for computational parallelization: since the calculations performed on disjoint sets $A, A' \in \mathscr{A}(k)$ to determine the next partition $\mathscr{A}(k+1)$ have no dependency on one another, they can be performed in parallel. This may be of particular benefit if the function f is not easy or quick to evaluate.

The CORPUS algorithm satisfies the following convergence theorem, the proof of which is in the Appendix:

Theorem 4.1
For every bounded $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_n \subseteq \mathbb{R}^n$ and every continuous $f : \mathscr{X} \to \mathbb{R}$, the CORPUS algorithm generates a sequence of partitions $(\mathscr{A}(k))_{k \in \mathbb{N}_0}$ such that

$$\mu[f \leqslant \theta] = \lim_{k \to \infty} \widetilde{\mu_{\mathscr{A}(k)}}[f \leqslant \theta]. \tag{21}$$

Remark 4.2
Theorem 4.1 also holds for a barycentric variant of the CORPUS algorithm in which boxes are subdivided through their barycenter $\int_A x \, \mathrm{d}\mu(x)$, provided that the measure μ has the property that the barycenter of any box (product of intervals) $A \subseteq \mathscr{X}$ cannot be arbitrarily close to its frontier. That is, it is required that there exists $C(\mu) > 0$ such that, for any box $A \subseteq \mathscr{X}$,

$$d_j\left(\int_A x_j \, \mathrm{d}\mu(x), \int_A x_j \, \mathrm{d}x\right) \leqslant \left(\frac{1}{2} - C(\mu)\right) d_j(A). \tag{22}$$
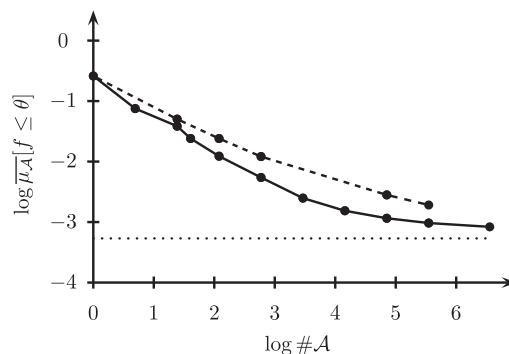
Figure 2. A log–log plot of the $\overline{\mu_{\mathscr{A}}}$ upper bound on the failure probability versus the number of boxes #$\mathscr{A}$ used by the CORPUS algorithm in Example 4.3. Dashed, the corresponding upper bound obtained if all boxes are subdivided, instead of just those on which f both succeeds and fails. Dotted, the exact probability of failure.

Thus, for example, barycentric CORPUS will satisfy (21) if μ has a density with respect to Lebesgue measure that is bounded away from 0 and ∞, whereas it may fail to converge if μ has an atomic point mass. Note that barycentric CORPUS is more computationally expensive than the original CORPUS algorithm, since calculation of barycenter $\int_A x \, d\mu(x)$ requires access to the input distribution μ and may be subject to sampling error, whereas the geometric center of a rectangular box is easy to calculate.

Example 4.3 (*hypervelocity impact model*)
One scenario to which the partitioned version of McDiarmid's inequality and the CORPUS algorithm can be applied is the study of hypervelocity impact.[‡] The response function of study is the perforation area caused in a steel plate of thickness h by the impact of a steel ball that impacts at obliquity α from the plate normal and at speed v; the perforation area is measured using a line of sight normal to the plate, not a line of sight parallel to the impact direction. A first approximation to the perforation area $f(h, \alpha, v)$ (in mm^2) in the range

$$h \in [1.52, 2.67]\,\text{mm}, \quad \alpha \in [0, \tfrac{\pi}{6}], \quad v \in [2.1, 2.8]\,\text{km}\,\text{s}^{-1}$$

is given by the experimentally derived surrogate

$$f(h, \alpha, v) := 10.396 \left(\left(\frac{h}{1.778} \right)^{0.476} (\cos \alpha)^{1.028} \tanh \left(\frac{v}{v_{\text{bl}}} - 1 \right) \right)_{+}^{0.468}. \tag{23}$$

The quantity $v_{\text{bl}}(h, \alpha)$ given by

$$v_{\text{bl}}(h, \alpha) := 0.579 \left(\frac{h}{(\cos \alpha)^{0.448}} \right)^{1.400} \tag{24}$$

is called the *ballistic limit*, the impact speed below which no perforation occurs. The failure event is non-perforation, i.e. $[f = 0] \equiv [f \leqslant 0]$.

For simplicity and to provide some explicit quantities for comparison, suppose that h, α and v are uniformly distributed in their ranges. Under this assumption, the probability of non-perforation for the model f is approximately 3.8%. In this situation, the mean perforation area is $6.577\,\text{mm}^2$ and $\mathscr{D}[f] = 12.151\,\text{mm}^2$; therefore, McDiarmid's inequality (8) yields an upper bound of 55.7%. The probability-of-failure upper bounds obtained by the CORPUS algorithm in this case are illustrated in Figure 2.

One (empirical) observation on this example is that the improvement of the CORPUS algorithm upon the usual McDiarmid inequality is of order $1/\sqrt{\#\mathscr{A}}$:

$$\overline{\mu_{\mathscr{A}}}[f=0]-\mu[f=0]\approx\frac{\overline{\mu_{\{\mathscr{X}\}}}[f=0]-\mu[f=0]}{\sqrt{\#\mathscr{A}}}. \tag{25}$$

It is to be expected that, in general, the rate of convergence of $\overline{\mu_{\mathscr{A}}}[f\leqslant\theta]$ to $\mu[f\leqslant\theta]$ will be dependent upon the regularity of f and the failure region $[f\leqslant\theta]\subseteq\mathscr{X}$.

Recall that, in order to ensure that $\overline{\mu_{\mathscr{A}(k)}}[f\leqslant\theta]$ converges to $\mu[f\leqslant\theta]$ as the iteration number $k\to\infty$, it is sufficient to bisect those boxes that 'straddle' the failure frontier. Figure 2 shows that subdividing only those boxes—as opposed to all boxes—does not provide a significant improvement in the upper bound on the probability of failure. More precisely, although subdividing only the boxes that 'straddle' the frontier does provide a sharper upper bound for a given total number of boxes, $\#\mathscr{A}$, the improved sharpness is only significant when $\#\mathscr{A}$ is of order e^3 to e^6 (20–400 boxes), and this is much larger than may be practical for computationally intensive response functions f. It is worth bearing in mind, however, that there may be a significant reduction in the total *computational* cost associated to not having to evaluate f again on boxes that have been established to be pure failure/success regions.

5. SAMPLE DATA

In practice, it may be the case that the true response function and probability distribution on input parameters can be accessed only through experimental sampling. In particular, sampling may be required to approximately determine all or some of the quantities $\mu(A)$, $\mathscr{D}[f|A]$ and $\mathbb{E}[f|A]$ as used in (11). For the purposes of this section, assume that $\mathscr{A}=\{A_k\}_{k=1}^{K}$ is a *finite* rectangular partition of $\mathscr{X}$.

Suppose that a finite rectangular partition $\mathscr{A}$ of $\mathscr{X}$ is given, but that our knowledge of the measure μ and response function f comes from $m\in\mathbb{N}$ independent μ-distributed sample points $X^{(1)}, \ldots, X^{(m)}$ in $\mathscr{X}$ and the corresponding values $f(X^{(1)}), \ldots, f(X^{(m)})$ in $\mathbb{R}$; m_k denotes the number of sample points that lie in A_k. In general, the inequality in (11) does not hold if the quantities $\mu(A_k)$, $\mathbb{E}[f|A_k]$ and $\mathscr{D}[f|A_k]$ are replaced by empirical estimates calculated using the m samples. Therefore, our interest lies in determining what corrections must be made to the empirical analogue of (11) to obtain a bound on the failure probability $\mu[f\leqslant\theta]$.

For simplicity, the following discussion focuses on how to replace exact local means $\mathbb{E}[f|A]$ by sampled ones $\widehat{\mathbb{E}}[f|A]$ in (11). The estimation of the probability masses $\mu(A_k)$ can be treated similarly, but with one major *caveat*: accurate estimation of a small $\mu(A_k)$ suffers from the same difficulties as Monte Carlo estimation of a small $\mu[f\leqslant\theta]$. It will be assumed that the local McDiarmid diameters $\mathscr{D}[f|A_k]$ are known (or at least bounded from above), perhaps through sampling directed by an optimization algorithm. If K, the number of boxes in the partition, is large, then it may be too expensive to determine $\mathscr{D}[f|A_k]$ accurately for every $k\in\{1, \ldots, K\}$. However, McDiarmid diameters are monotone with respect to the domain of the function, i.e.

$$A\subseteq B\Rightarrow\mathscr{D}[f|A]\leqslant\mathscr{D}[f|B]$$

and so the bound $\mathscr{D}[f|A_k]\leqslant\mathscr{D}[f]$ always holds; this bound can be used to provide an upper estimate of (11) that is still an improvement on the global McDiarmid inequality.

As an introductory case, consider the trivial partition (i.e. $K=1$) and m independent μ-distributed samples, as in [1, Section 2.3]. Application of McDiarmid's inequality to the mn independent random variables

$$\{X_i^{(j)}|i=1, \ldots, n \text{ and } j=1, \ldots, m\}$$

yields that, for any given $\varepsilon > 0$, with μ-probability at least $1 - \varepsilon$ on the m samples

$$\mu[f \leqslant \theta] \leqslant \exp\left(-\frac{2(\widehat{\mathbb{E}}[f] - \alpha - \theta)_+^2}{\mathscr{D}[f]^2}\right),\tag{26}$$

where the *empirical mean* $\widehat{\mathbb{E}}[f]$ and the *confidence shift* α are defined by

$$\widehat{\mathbb{E}}[f] := \frac{1}{m}\sum_{j=1}^{m} f(X^{(j)}),\tag{27}$$

$$\alpha := \mathscr{D}[f]\sqrt{\frac{\log(1/\varepsilon)}{2m}}.\tag{28}$$

The situation of a non-trivial partition (i.e. $K > 1$) can be treated using similar methods. There are many powerful concentration results for empirical processes in the literature: [20–23] are significant contributions. We will outline two confidence-shift-type methods in the style of (26) that can be applied to the present problem. Given a finite rectangular partition $\mathscr{A} = \{A_k\}_{k=1}^{K}$ and $\alpha \in \mathbb{R}^K$, define $H_\alpha : \mathbb{R}^K \to [0, 1]$ by

$$H_\alpha(y) := \sum_{k=1}^{K} \mu(A_k)\exp\left(-\frac{2(\mathbb{E}[f|A_k] - y_k - \alpha_k - \theta)_+^2}{\mathscr{D}[f|A_k]^2}\right).\tag{29}$$

Note that $H_\alpha(-\alpha) = \overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$, which is known to be an upper bound on the probability of failure $\mu[f \leqslant \theta]$. Let $Y_k := \mathbb{E}[f|A_k] - \widehat{\mathbb{E}}[f|A_k]$, let $Y := (Y_1, \ldots, Y_K)$, and consider the random variable $H_\alpha(Y)$. The two methods that will follow both attempts to answer the following question: with what probability is $H_\alpha(Y)$ an upper bound for $\mu[f \leqslant \theta]$? Note that

$$H_\alpha((\mathbb{E}[f|A_k] - \widehat{\mathbb{E}}[f|A_k])_{k=1}^{K}) = \sum_{k=1}^{K} \mu(A_k)\exp\left(-\frac{2(\widehat{\mathbb{E}}[f|A_k] - \alpha_k - \theta)_+^2}{\mathscr{D}[f|A_k]^2}\right)$$

and so $H_\alpha(Y)$ is not a function of the exact means.

See Figure 3 for a contour plot of $H_\alpha : \mathbb{R}^2 \to [0, 1]$ in the case $K = 2$. Observe that H_α is a piecewise-smooth, increasing but non-convex function of its arguments. Since $H_\alpha(-\alpha) = \overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$, the aim is to bound from above the probability that $H_\alpha(Y)$ fails to be an upper bound for $\mu[f \leqslant \theta]$; to do this, it is sufficient to bound

$$\mu[H_\alpha(Y) \leqslant H_\alpha(-\alpha)]\tag{30}$$

from above.

The methods that follow will exploit the fact that each Y_k is a centered (mean zero) real-valued random variable with exponentially small tails: application of McDiarmid's inequality (7a) yields that, for $r > 0$,

$$\mu[Y_k \geqslant r] \leqslant \exp\left(-\frac{2m_k r^2}{\mathscr{D}[f|A_k]^2}\right)\tag{31}$$

and similarly for $\mu[Y_k \leqslant r]$ if $r < 0$. We now give two examples of how (31) can be translated into concentration-of-measure results for $H_\alpha(Y)$, and hence an upper bound on (30).

Example 5.1 (confidence inequalities using (26))
One upper bound on (30) is given by unconditioned global sampling and K independent applications of (26). That is, consider $m \in \mathbb{N}$ independent μ-distributed samples of the space $\mathscr{X}$. Then, for any $\varepsilon > 0$,

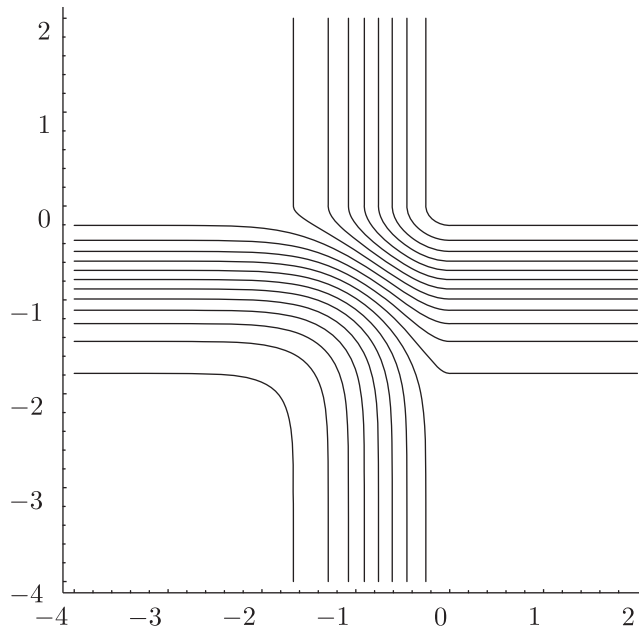$$\mu[H_\alpha(-\alpha) \leqslant H_\alpha(Y)] \geqslant (1 - \varepsilon)^K.\tag{32}$$

Figure 3. 20 equally spaced contours of the empirical upper bound function $H_\alpha(y_1, y_2)$ as defined by (29), with $\mu(A_1)=0.4$, $\mu(A_2)=0.6$. H_α increases from very nearly 0 in the bottom-left to exactly 1 in the top-right. Note that the sublevel set $H_\alpha^{-1}([0,t])$ is convex for small enough t, but is non-convex for $t > \min_{1 \leqslant k \leqslant K} \mu(A_k)$. The coordinate on axis i, for $i=1, 2$, is $(\mathbb{E}[f|A_i] - y_i - \alpha_i - \theta)/\mathscr{D}[f|A_i]$.

where $\alpha = (\alpha_1, \ldots, \alpha_K)$ is the random variable defined in terms of $m_k := m\widehat{\mu}_m(A_k)$ by

$$\alpha_k := \mathscr{D}[f|A_k]\sqrt{\frac{\log(1/\varepsilon)}{2m_k}}.$$

It is not always the case that m_k (and hence α_k) is a random variable over which there is no control; it may be possible to choose to draw m_k independent samples from the conditional distribution $\mu(\cdot|A_k)$, i.e. to target the experiments according to the partition $\mathscr{A}$. In this case, choose $m_1, \ldots, m_K \in \mathbb{N}$ and $\varepsilon > 0$, choose any $\varepsilon_1, \ldots, \varepsilon_K > 0$ such that $1 - \varepsilon = \prod_{k=1}^K (1 - \varepsilon_k)$, and set

$$\alpha_k := \mathscr{D}[f|A_k]\sqrt{\frac{\log(1/\varepsilon_k)}{2m_k}}. \tag{33}$$

(Note that, once m_k and ε_k are chosen, α_k is a deterministic object.) Then

$$\mu[H_\alpha(-\alpha) \leqslant H_\alpha(Y)] \geqslant 1 - \varepsilon = \prod_{k=1}^K (1 - \varepsilon_k). \tag{34}$$

This inequality is not a significant improvement on (32) except in that it allows for conditioning of the samples according to the partition $\mathscr{A}$. Both inequalities suffer from the problem that, if K is large, then many samples will be required in order to make the confidence shifts α_k acceptably small: otherwise, α will be so large that the result of the calculation is likely to be the trivial upper bound $\mu[f \leqslant \theta] \leqslant H_\alpha(-\alpha) = 1$.

Geometrically speaking, both the inequalities in the previous example amount to bounding the measure of the superlevel set $\{y \in \mathbb{R}^K | H_\alpha(y) > H_\alpha(-\alpha)\}$ from below by that of a K-dimensional positive orthant (a product of semi-infinite intervals $(-\alpha_k, +\infty)$). A better method is to bound

the measure of $[H_\alpha > H_\alpha(-\alpha)]$ from below by that of a half-space, if indeed it contains such a half-space. The measure of a half-space is easy to bound and is not adversely affected by K being large. The intuitive explanation for this is that, in a high-dimensional space, a cuboid/orthant appears to be very narrow when viewed from one of its corners, whereas a half-space always fills half of one's field of view.

Denote by $\mathbb{H}_{p,v}$ the closed half-space in $\mathbb{R}^K$ that has p on its boundary and v as an outward-pointing normal:

$$\mathbb{H}_{p,v} = \{y \in \mathbb{R}^K \,|\, v \cdot y \leqslant v \cdot p\}. \tag{35}$$

Since $\mathbb{E}[v \cdot Y] = 0$, application of McDiarmid's inequality (7b) yields that

$$\mu[Y \in \mathbb{H}_{p,v}] \leqslant \exp\left(-2(v \cdot p)_-^2 \,\Big/\, \sum_{k=1}^K \frac{|v_k|^2}{m_k} \mathscr{D}[f|A_k]^2\right). \tag{36}$$

(Recall that, for $t \in \mathbb{R}$, $t_- := |\min\{0,t\}| \geqslant 0$.) It follows that, for any $S \subseteq \mathbb{R}^K$,

$$\mu[Y \in S] \leqslant \inf\left\{ \exp\left(-\frac{2(v \cdot p)_-^2}{\sum_{k=1}^K \frac{|v_k|^2}{m_k} \mathscr{D}[f|A_k]^2}\right) \,\middle|\, \begin{array}{l} p \in \mathbb{R}^K \text{ and } v \in \mathbb{R}^K \\ \text{such that } S \subseteq \mathbb{H}_{p,v} \end{array} \right\}. \tag{37}$$

Concentration-of-measure inequalities that arise from half-space considerations in this way appear to be new to the literature and are a topic of the current research (Sullivan and Owhadi, in preparation). In the following example, inequality (36) is applied with suitable choices of p and v.

Example 5.2 (confidence inequalities using (36))
Suppose it is known *a priori* that $H_\alpha(-\alpha)$ is small enough that the sublevel set $H_\alpha^{-1}([0, H_\alpha(-\alpha)])$ is convex. Under this convexity assumption, (36) can be applied with $p = -\alpha$ and $v = \nabla H_\alpha(-\alpha)$. Note that the partial derivatives of H_α at $-\alpha$ do not depend on α and are given by

$$\partial_k H_\alpha(-\alpha) = \frac{4\mu(A_k)(\mathbb{E}[f|A_k] - \theta)_+}{\mathscr{D}[f|A_k]^2} \exp\left(-\frac{2(\mathbb{E}[f|A_k] - \theta)_+^2}{\mathscr{D}[f|A_k]^2}\right) \geqslant 0. \tag{38}$$

Hence, by (36) with $p = -\alpha$ and $v = \nabla H_\alpha(-\alpha)$,

$$\mu[H_\alpha(Y) \leqslant H_\alpha(-\alpha)] \leqslant \exp\left(-\frac{2(\nabla H_\alpha(-\alpha) \cdot \alpha)_+^2}{\sum_{k=1}^K |\partial_k H_\alpha(-\alpha)|^2 \mathscr{D}[f|A_k]^2/m_k}\right). \tag{39}$$

(If the partial derivative $\partial_k H_\alpha(-\alpha)$ is undefined in the strong sense, i.e. if $\mathbb{E}[f|A_k] = \theta$, then, since H_α is increasing with respect to each of its K arguments, it makes sense to observe the convention that $\partial_k H_\alpha(-\alpha) = 0$. This convention maintains the validity of (39). In particular, if $\nabla H_\alpha(-\alpha) = 0$ either in the strong sense or by convention, then the right-hand side of (39) is the trivial upper bound 1.)

In practice, $\mathbb{E}[f|A_k]$ is not known exactly, and so the gradient $\nabla H_\alpha(-\alpha)$ and the bound (39) cannot be exactly evaluated. Therefore, it is necessary to take a worst-case estimate with respect to whatever information about $\mathbb{E}[f|A_k]$ is known. Therefore, as a 'place-holder' for the partial derivative $\partial_k H_\alpha(-\alpha)$, define

$$\Delta_k(E_k) := \frac{4\mu(A_k)(E_k - \theta)_+}{\mathscr{D}[f|A_k]^2} \exp\left(-\frac{2(E_k - \theta)_+^2}{\mathscr{D}[f|A_k]^2}\right) \tag{40}$$

that is, $\Delta_k(E_k) = \partial_k H_\alpha(-\alpha)$ if it is known that $\mathbb{E}[f|A_k] = E_k$. Suppose that it is given that the exact means satisfy $(\mathbb{E}[f|A_k])_{k=1}^K \in \mathscr{E} \subseteq \mathbb{R}^K$. A worst-case confidence bound given this information is obtained by optimizing over $E = (E_1, \ldots, E_K) \in \mathscr{E}$, yielding the following result:

Proposition 5.3
Suppose it is known that $H_\alpha(-\alpha)$ is small enough that the sublevel set $H_\alpha^{-1}([0, H_\alpha(-\alpha)])$ is convex and that $(\mathbb{E}[f|A_k])_{k=1}^K \in \mathscr{E} \subseteq \mathbb{R}^K$. Then

$$\mu[H_\alpha(Y) \leqslant H_\alpha(-\alpha)] \leqslant \sup_{E \in \mathscr{E}} \exp\left(-\frac{2\left(\sum_{k=1}^K \Delta_k(E_k)\alpha_k\right)^2}{\sum_{k=1}^K \Delta_k(E_k)^2 \mathscr{D}[f|A_k]^2/m_k} \right) \tag{41a}$$

$$= \sup_{E \in \mathscr{E}} \exp\left(-\frac{\left(\sum_{k=1}^K \Delta_k(E_k)\mathscr{D}[f|A_k]/\sqrt{m_k}\right)^2}{\sum_{k=1}^K \Delta_k(E_k)^2 \mathscr{D}[f|A_k]^2/m_k} \right). \tag{41b}$$

The heuristic behind the half-space method is that if one of the summands in (29) is an underestimate and K is large, then it is easy for the other $K-1$ summands to collectively accrue enough overestimation that the initial underestimate is canceled out. However, this correction effect is unlikely to hold if the error occurs in a box with much larger μ-mass than all the others: in this case, it is very unlikely that the other boxes' contributions will correct the error coming from the large box. In this situation, when there is a great disparity in the masses $\mu(A_k)$, the sublevel set

$$\{y \in \mathbb{R}^K \,|\, H_\alpha^{-1}(y) \leqslant H_\alpha(-\alpha)\}$$

is non-convex and its convex hull contains $0 \in \mathbb{R}^K$, and so the half-space method yields a trivial bound.

Remark 5.4
The scaling properties of (34) versus (39) are significantly different. Suppose that the same level of confidence $1 - \varepsilon_k \equiv 1 - \varepsilon_0$ is imposed for each of the partition elements $A_1, \ldots, A_K$ and that $\alpha = (\alpha_1, \ldots, \alpha_K)$ is defined by (33). The level of confidence in the empirical upper bound $H_\alpha(Y)$ as given by the orthant method (34) is $(1-\varepsilon_0)^K$. On the other hand, the half-space method (39) yields a level of confidence of

$$1 - \exp\left(-\frac{\left(\sum_{k=1}^K \partial_k H_\alpha(-\alpha)\mathscr{D}[f|A_k]/\sqrt{m_k}\right)^2}{\sum_{k=1}^K \left(\partial_k H_\alpha(-\alpha)\mathscr{D}[f|A_k]\right)^2/m_k} \log\frac{1}{\varepsilon_0} \right)$$

$$= 1 - \varepsilon_0^{\left(\sum_{k=1}^K \partial_k H_\alpha(-\alpha)\mathscr{D}[f|A_k]/\sqrt{m_k}\right)^2 / \sum_{k=1}^K (\partial_k H_\alpha(-\alpha)\mathscr{D}[f|A_k])^2/m_k}.$$

This expression is maximized when all the summands are equal, and is then equal to $1 - \varepsilon_0^K \gg (1-\varepsilon_0)^K$; that is, when K is large, the level of confidence that the half-space method provides is much greater than that provided by the orthant method. It is also worth noting that this argument gives a prescription for the optimal number of samples m_k to take in box A_k. Heuristically, we should take $m = (m_1, \ldots, m_K) \in \mathbb{N}^K$ such that the ratio

$$\frac{m_k}{(\partial_k H_\alpha(-\alpha)\mathscr{D}[f|A_k])^2}$$

is independent of k. More precisely, and taking account of the fact that the exact local means $\mathbb{E}[f|A_k]$ are only known *a priori* to lie in some region $\mathscr{E} \subseteq \mathbb{R}^K$, given a total number $m_{\text{total}} \in \mathbb{N}$

of available samples, we seek a solution $m = (m_1, \ldots, m_K) \in \mathbb{N}^K$ to the constrained optimization problem

$$\text{minimize:} \quad \sup_{\substack{E = (E_1, \ldots, E_K) \in \mathscr{E} \\ 1 \leqslant k < \ell \leqslant K}} \left| \log \left(\frac{m_k}{\Delta_k(E_k)^2 \mathscr{D}[f|A_k]^2} \bigg/ \frac{m_\ell}{\Delta_\ell(E_\ell)^2 \mathscr{D}[f|A_\ell]^2} \right) \right|$$

$$\text{subject to:} \quad m = (m_1, \ldots, m_K) \in \mathbb{N}^K,$$

$$\sum_{k=1}^{K} m_k = m_{\text{total}}.$$

In comparison with orthant methods, the discussion above shows how half-space considerations can give a greater level of confidence in an empirically-derived upper bound on the probability of failure. One could equally well adopt the dual point of view that half-space considerations allow for a sharper (lower) empirically-derived upper bound on the probability of failure with the same prescribed level of confidence.

To conclude this section and show the relative strengths of the two methods outlined above, we return to the earlier example of the hypervelocity impact surrogate (Example 4.3).

Example 5.5 (Example 4.3 continued)
Consider the 8-element partition $\mathscr{A}$ obtained by three iterations of the CORPUS algorithm. Suppose that, for $k = 1, \ldots, 8$, each box $A_k \in \mathscr{A}$ is sampled with $m_k := 5$ independent $\mu(\cdot|A_k)$-distributed samples. Set $\varepsilon_k := 1\%$ for $k = 1, \ldots, 8$ and calculate the confidence shifts α_k as given by (33). In this case

$$\overline{\mu_{\mathscr{A}}}[f = 0] \equiv H_\alpha(-\alpha) = 1.48 \times 10^{-1},$$

whereas the above specifications generate an empirical upper bound of

$$H_\alpha(Y) = 3.14 \times 10^{-1}.$$

The positive orthant method in the form of (34) yields that

$$\mu[H_\alpha(Y) \leqslant H_\alpha(-\alpha)] \leqslant 7.72 \times 10^{-2}.$$

Since it always holds that $H_\alpha(-\alpha) \geqslant \mu[f = 0]$, it follows that we have a confidence level of at least 92.2% that the true probability of failure is less than the empirical upper bound $H_\alpha(Y)$.

Unfortunately, it has not been possible to directly compare the orthant and half-space methods (i.e. (34) versus (41b)) for $K = 8$ since this would require that $\{y \in \mathbb{R}^8 | H_\alpha(y) \leqslant 0.148\}$ be convex—however, it is not convex, and its convex hull includes the origin, and so the half-space method yields a trivial bound. However, a direct comparison is possible in the case $K = 2$. Table I gives upper bounds on the probability of failure and upper bounds on the probability that $H_\alpha(Y) \leqslant H_\alpha(-\alpha)$, assuming that the local means $\mathbb{E}[f|A_k]$ are known *a priori* to within $\pm\delta$; that is, we take

$$\mathscr{E}_\delta := \prod_{k=1}^{K} [\mathbb{E}[f|A_k] - \delta, \mathbb{E}[f|A_k] + \delta]$$

and calculate the bound (41b).

The results in Table I illustrate another benefit of using partitioning rather than McDiarmid's inequality alone: for sufficiently large K, concentration-of-measure effects ensure that even large uncertainty in the true mean performance of the system need not result in low levels of confidence in empirically derived upper bounds on $\mu[f \leqslant \theta]$.

Table I. Upper bounds on the probability of failure (non-perforation) $\mu[f=0]$ for the hypervelocity impact example (Example 4.3), together with associated levels of confidence. Throughout, $K=2$, $m_1=m_2=5$, and all numerical data are given to 3 significant figures. Of particular note is the final line: knowing the mean performance to within $\pm 5\,\mathrm{mm}^2$ (which is of the same order as the mean perforation area of $6.58\,\mathrm{mm}^2$) is sufficient to give confidence levels twice as good as the orthant method.

	$\varepsilon_1=\varepsilon_2=10^{-2}$	$\varepsilon_1=\varepsilon_2=10^{-3}$
Upper bounds on the probability of failure		
(i.e. non-perforation, $[f=0]$)		
Exact probability of failure: $\mu[f=0]$	3.78×10^{-2}	
Exact local means: $\overline{\mu_{\mathscr{A}}}[f=0]$	3.25×10^{-1}	
Empirical local means: $H_\alpha(Y)$	5.41×10^{-1}	5.78×10^{-1}
Confidence levels		
(i.e. upper bounds on $\mu[H_\alpha(Y)\leqslant\overline{\mu_{\mathscr{A}}}[f\leqslant\theta]]$)		
Orthant method (34)	2.00×10^{-2}	2.00×10^{-3}
Half-space method (41b)		
$\delta=0$	9.41×10^{-3}	9.13×10^{-4}
$\delta=1$	9.89×10^{-3}	9.83×10^{-4}
$\delta=5$	1.00×10^{-2}	1.00×10^{-3}

6. OUTLOOK

6.1. Martingale and other concentration inequalities

It is not necessary to assume that $\mathscr{X}$ is a product space, that the measure μ on $\mathscr{X}$ is a product measure (i.e. that the random inputs $X_1,\ldots,X_n$ of f are independent) and that the measurable sets $A\in\mathscr{A}$ into which $\mathscr{X}$ is decomposed are themselves products (rectangular boxes). However, the notation and intuition are significantly easier in the rectangular/independent case. In the general case, consider a filtration

$$\sigma(\emptyset)\subseteq\mathscr{F}_{A,0}\subseteq\mathscr{F}_{A,1}\subseteq\cdots\subseteq\mathscr{F}_{A,m}\subseteq\mathscr{B}(A)$$

of the Borel σ-algebra $\mathscr{B}(A)$ and consider the *martingale diameter* defined by

$$\mathscr{D}[f|A]^2:=\sum_{j=1}^{m}\mathrm{diam}(\mathbb{E}[f|\mathscr{F}_{A,j}]-\mathbb{E}[f|\mathscr{F}_{A,j-1}])^2.$$

The result that corresponds to McDiarmid's inequality in this case is the so-called *martingale inequality*: for any $r>0$, with μ-probability one,

$$\mu[\mathbb{E}[f|\mathscr{F}_{A,m}]-\mathbb{E}[f|\mathscr{F}_{A,0}]\geqslant r|\mathscr{F}_{A,0}]\leqslant\exp\left(-\frac{2r^2}{\mathscr{D}[f|A]^2}\right)$$

and

$$\mu[\mathbb{E}[f|\mathscr{F}_{A,m}]-\mathbb{E}[f|\mathscr{F}_{A,0}]\leqslant -r|\mathscr{F}_{A,0}]\leqslant\exp\left(-\frac{2r^2}{\mathscr{D}[f|A]^2}\right).$$

See, for example, [2, Corollary 6.9] [24, Theorem 2.3] for details and proofs. In the case of a not-necessarily-rectangular partition $\mathscr{A}$ of $\mathscr{X}$, with $\mathscr{F}_{A,0}=\sigma(\emptyset)$ and $\mathscr{F}_{A,m}=\mathscr{B}(A)$, it follows that

$$\mathbb{E}[f|\mathscr{F}_{A,0}]=\mathbb{E}[f|A]\quad\text{and}\quad\mathbb{E}[f|\mathscr{F}_{A,m}]=f$$

thereby yielding the almost sure bound

$$\mu[f \leqslant \theta] \leqslant \sum_{A \in \mathscr{A}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A] - \theta)^2}{\mathscr{D}[f|A]^2}\right),$$

where $\mathscr{D}[f|A]$ is now interpreted as a martingale diameter instead of a McDiarmid diameter.

There are many other concentration-of-measure inequalities that can be used to bound the terms $\mu[f \leqslant \theta|A]$ in

$$\mu[f \leqslant \theta] = \sum_{A \in \mathscr{A}} \mu(A) \mu[f \leqslant \theta|A].$$

Aside from those derived using martingale methods, there are also concentration-of-measure inequalities produced using the *entropy method*: this method has its roots in the theory of logarithmic Sobolev inequalities for Gaussian measures [25]; more recent results include [26, 27]. We see no great theoretical obstacles to the combination of such methods with the partitioning scheme. As used in the methods of this paper, the chief advantages of McDiarmid's inequality are its notational simplicity and that the concentration rate is controlled by the subdiameters $\mathscr{D}_j[f|A]$, which measure parameter sensitivity and can be easily used in algorithms such as CORPUS.

6.2. Multi-fidelity modeling

Although the partitioning method applies to the situation in which the true response function and probability distribution on input parameters can be accessed only through experimental sampling, the CORPUS algorithm does not directly apply to this situation. However, CORPUS can be applied indirectly through the use of surrogate response functions and probability distributions.

More precisely, suppose that the response function and probability distribution are actually unknowns G and $\mathbb{P}$. If μ (resp. f) is any approximation to $\mathbb{P}$ (resp. G) for which the CORPUS algorithm is practical to implement, then CORPUS may be applied to generate a partition $\mathscr{A}$ that is well adapted to the pair (μ, f) and on which the sampled version of the upper bound $\overline{\mathbb{P}_{\mathscr{A}}}[G \leqslant \theta]$ can be calculated as in Section 5. The two main *caveats* are that the approximation

$$\overline{\mathbb{P}_{\mathscr{A}}}[G \leqslant \theta] \approx \overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$$

is uncontrolled, and hence $\mathscr{A}$ may be a poor partition for $(\mathbb{P}, G)$ if μ (resp. f) is not good approximation to $\mathbb{P}$ (resp. G); and that sampling to estimate a small mass $\mathbb{P}(A)$ suffers from the same $\mathbb{P}(A)^{-2} \log 1/\varepsilon$ problems as naïve Monte Carlo methods.

The use of approximations provides an opportunity for an additional layer of modeling. Since the subdiameters are semi-norms, they satisfy the triangle inequality. In this case, that means that if ϕ is any model for f (perhaps a computational scheme that is much cheaper and easier to exercise than f itself), then

$$\mathscr{D}[f|A] \leqslant \mathscr{D}[\phi|A] + \mathscr{D}[f - \phi|A].$$

If the model ϕ is good — at least on some subset A of the input parameter space $\mathscr{X}$ — then $\mathscr{D}[f - \phi|A]$ may be small. Of particular note is that ϕ may depend on A. That is, regions where the behavior of f is very simple may be paired with a very coarse model $\phi_A: A \to \mathbb{R}$; regions where the behavior of f is complicated may be paired with a finer model ϕ_A; such multi-scale and multi-fidelity modeling offers significant advantages in terms of computational efficiency. In this case, the probability-of-failure bound for f takes the form

$$\mu[f \leqslant \theta] \leqslant \sum_{A \in \mathscr{A}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[\phi_A|A] + \mathbb{E}[f - \phi_A|A] - \theta)_+^2}{(\mathscr{D}[\phi_A|A] + \mathscr{D}[f - \phi_A|A])^2}\right)$$

with similar confidence corrections to those in Section 5 if μ and f are known only through samples. It is expected that this approach will be particularly useful in multi-scale simulation efforts; see e.g. [28–31].

6.3. Optimal partitions

In practice, partitions of high cardinality will be impractical to work with, since they will require too many experimental samples to implement the methods of Section 5, especially if some partition elements $A \in \mathscr{A}$ have small measure. Therefore, one problem for future research is the determination of *optimal* partitions of a given cardinality. That is, given μ, f, θ and N, find

$$\mathscr{A} \in \arg\min\{\overline{\mu_{\mathscr{A}}}[f \leqslant \theta] | \#\mathscr{A} \leqslant N\}$$

or

$$\mathscr{A} \in \arg\min\{\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta] | \#\mathscr{A} \leqslant N\}.$$

It may also be of interest to add additional feasibility constraints on the partitions $\mathscr{A}$: for example, that $\mu(A) \geqslant \mu_{\min}$ or $d(A) \geqslant \delta_{\min}$ for all $A \in \mathscr{A}$, where $\mu_{\min}, \delta_{\min} > 0$ are parameters determined by the user or by practical considerations.

6.4. Partitions based on sample data

The discussions in this paper have assumed that a partition $\mathscr{A}$ is given before any sampling begins, whether constructed by an expert or using an algorithm such as CORPUS. In practice, it may be necessary to construct a partition based solely on sample data. One problem here is that samples/experiments must be used carefully since they may have been strongly conditioned (i.e. not been drawn from the unbiased global distribution on input parameters). The adaptation of CORPUS to this situation will be the topic of the future work.

7. CONCLUSIONS

As noted in the introduction, McDiarmid's inequality provides an upper bound on the probability of failure of a real-valued system of interest that is notationally simple but often not sharp. Via an elementary partitioning argument, we arrive at the alternative upper bound (11). As Proposition 3.2 shows, the error in (11), i.e. the amount by which (11) is an overestimate of the true probability of failure, is controlled by the local McDiarmid subdiameters $\mathscr{D}_j[f|A]$, $j \in \{1, \ldots, n\}$, $A \in \mathscr{A}$ and not by the partition fineness $\max_{A \in \mathscr{A}} d(A)$. Moreover, these McDiarmid subdiameters can be used as sensitivity indices to guide a recursive partitioning scheme (the CORPUS algorithm). The CORPUS algorithm partitions the parameter space according to a binary tree in a way that avoids the naïve curse of dimension and is reminiscent of adaptive mesh refinement, branch-and-bound, and tree-structured density estimation methods.

Given a partition of the input parameter space, whether generated by CORPUS or otherwise, sampling techniques can be used to evaluate an empirical upper bound on the failure probability of interest, and the level of confidence in this empirical upper bound can be quantified. Well-established concentration-of-measure inequalities such as Hoeffding's and McDiarmid's inequalities can be used to provide these confidence bounds; in certain cases, though, it is possible to take advantage of the geometry of the region in sample space where the empirical estimate is too small and use half-space-based techniques that yield a higher level of confidence in a given empirical bound. These half-space-based concentration-of-measure inequalities appear to be novel and interesting objects of study in their own right.

APPENDIX A: PROOFS

Proof of Proposition 3.2

First observe that, by the basic monotonicity and continuity properties of measures, for any $\delta > 0$, $\mu[f \leqslant \theta] \leqslant \mu[f \leqslant \theta + \delta]$ and

$$\mu[f \leqslant \theta] = \mu \left(\bigcap_{\delta > 0} [f \leqslant \theta + \delta] \right) = \lim_{\delta \to 0+} \mu[f \leqslant \theta + \delta]. \tag{A1}$$

The measure of the set $[f \leqslant \theta + \delta]$ can be bounded from below by the measure of those $A \in \mathscr{A}$ that are contained within $f^{-1}((-\infty, \theta + \delta])$, yielding

$$\mu[f \leqslant \theta + \delta] \geqslant \sum_{\substack{A \in \mathscr{A} \\ f(A) \subseteq (-\infty, \theta + \delta]}} \mu(A)$$

$$= \sum_{A \in \mathscr{A} \setminus \mathscr{A}_\delta} \mu(A)$$

$$\geqslant \sum_{A \in \mathscr{A} \setminus \mathscr{A}_\delta} \mu(A) \exp \left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2} \right).$$

A bound on the contributions arising from the sets $A \in \mathscr{A}$ that are not included in this sum—i.e. $A \in \mathscr{A}_\delta$—is now required.

For each $A \in \mathscr{A}$, we have the conditional analogue of (6):

$$\mathscr{D}[f|A] \leqslant \operatorname{diam}[f|A] \leqslant \sum_{j=1}^n \mathscr{D}_j[f|A]. \tag{A2}$$

Note that

$$A \in \mathscr{A}_\delta \iff f(A) \cap [\theta + \delta, +\infty) \neq \emptyset$$

$$\Rightarrow f(A) \subseteq \left[\theta + \delta - \sum_{j=1}^n \mathscr{D}_j[f|A], +\infty \right) \quad \text{by (A2)}$$

$$\Rightarrow \mathbb{E}[f|A] \geqslant \theta + \delta - \sum_{j=1}^n \mathscr{D}_j[f|A]$$

$$\iff \mathbb{E}[f|A] - \theta \geqslant \delta - \sum_{j=1}^n \mathscr{D}_j[f|A].$$

Therefore,

$$\sum_{A \in \mathscr{A}_\delta} \mu(A) \exp \left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2} \right) \leqslant \sum_{A \in \mathscr{A}_\delta} \mu(A) \exp \left(-\frac{2(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A])_+^2}{\mathscr{D}[f|A]^2} \right)$$

$$\leqslant \sup_{A \in \mathscr{A}_\delta} \exp \left(-\frac{2(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A])_+^2}{\mathscr{D}[f|A]^2} \right)$$

since $\mu(\biguplus_{A \in \mathscr{A}_\delta} A) \leqslant \mu(\mathscr{X}) = 1$. Similarly,

$$\sum_{A \in \mathscr{A}_\delta \setminus \mathscr{A}_\delta^+} \mu(A) \exp \left(-\frac{2(\mathbb{E}[f|A] - \theta)_+^2}{\mathscr{D}[f|A]^2} \right) \leqslant \sup_{A \in \mathscr{A}_\delta \setminus \mathscr{A}_\delta^+} \exp \left(-\frac{2(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A])_+^2}{\mathscr{D}[f|A]^2} \right).$$

Now fix any $\varepsilon > 0$; by (A1), there exists $\delta(\varepsilon) > 0$ such that

$$0 < \delta < \delta(\varepsilon) \Rightarrow 0 \leqslant \mu[f \leqslant \theta + \delta] - \mu[f \leqslant \theta] < \varepsilon.$$

Hence, for any such δ,

$$\overline{\mu_{\mathscr{A}}}[f \leqslant \theta] - \mu[f \leqslant \theta] = \sum_{A \in \mathscr{A}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A]-\theta)_+^2}{\mathscr{D}[f|A]^2}\right) - \mu[f \leqslant \theta]$$

$$\leqslant \mu[f \leqslant \theta+\delta] + \sup_{A \in \mathscr{A}_\delta} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right) - \mu[f \leqslant \theta]$$

$$< \varepsilon + \sup_{A \in \mathscr{A}_\delta} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right).$$

This establishes the claim for the bound $\overline{\mu_{\mathscr{A}}}[f \leqslant \theta]$. For the bound $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$, note that

$$A \in \mathscr{A}_\delta^+ \text{ for some } \delta > 0 \Rightarrow \inf_{x \in A} f(x) \geqslant \theta + \delta > \theta \Rightarrow \mu[f \leqslant \theta|A] = 0$$

and so $\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta]$ makes no error on such sets A. Therefore,

$$\widetilde{\mu_{\mathscr{A}}}[f \leqslant \theta] - \mu[f \leqslant \theta] = \sum_{\substack{A \in \mathscr{A} \\ \inf_A f \leqslant \theta}} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A]-\theta)_+^2}{\mathscr{D}[f|A]^2}\right) - \mu[f \leqslant \theta]$$

$$\leqslant \sum_{A \in \mathscr{A} \setminus \mathscr{A}_\delta^+} \mu(A) \exp\left(-\frac{2(\mathbb{E}[f|A]-\theta)_+^2}{\mathscr{D}[f|A]^2}\right) - \mu[f \leqslant \theta] \quad \text{(for small enough } \delta > 0)$$

$$\leqslant \mu[f \leqslant \theta+\delta] + \sup_{A \in \mathscr{A}_\delta \setminus \mathscr{A}_\delta^+} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right) - \mu[f \leqslant \theta]$$

$$< \varepsilon + \sup_{A \in \mathscr{A}_\delta \setminus \mathscr{A}_\delta^+} \exp\left(-\frac{2\left(\delta - \sum_{j=1}^n \mathscr{D}_j[f|A]\right)_+^2}{\mathscr{D}[f|A]^2}\right). \qquad \square$$

Proof of Theorem 4.1
Recall that, for any function $f: \mathscr{X} \to \mathbb{R}$, the *modulus of continuity* of f is the increasing function $\varpi_f: (0, +\infty) \to [0, +\infty]$ defined by

$$\varpi_f(\delta) := \sup\{|f(x) - f(y)| \,|\, x, y \in \mathscr{X}, \mathrm{d}(x,y) \leqslant \delta\}$$

$$= \sup\{\mathrm{diam}[f|B] | B \subseteq \mathscr{X}, d(B) \leqslant \delta\}.$$

It is readily verified that f is continuous if, and only if, $\lim_{\delta \to 0} \varpi_f(\delta) = 0$. Hence, for continuous f, we have that

$$\mathscr{D}_j[f|A] \leqslant \varpi_f(d_j(A)) \to 0 \quad \text{as } d_j(A) \to 0.$$

We introduce some further notation for child sets. For any rectangular set $A \subseteq \mathscr{X}$, let A_{j+} and A_{j-} be the two child sets obtained by partitioning A (though its mid-point) by an $(n-1)$-dimensional subspace normal to $\widehat{e}_j$. By extension, if J is a finite string of the form $J = j_1 \pm j_2 \pm \cdots j_k \pm$, then A_J denotes the set obtained by partitioning A by a hyperplane normal to $\widehat{e}_{j_1}$, taking the positive/negative half, then partitioning the result by a hyperplane normal to $\widehat{e}_{j_2}$, and so on. If J is an infinite string, then $J|k$ denotes the substring consisting of the first k terms.

We claim that any set A_J for which the CORPUS-generated string $J = j_1 \pm j_2 \pm \cdots j_k \pm$ is sufficiently long must have

$$\mathcal{D}_j[f|A_J] \leqslant \tfrac{1}{2}\mathcal{D}_j[f|A] \quad \text{for all } j = 1, \ldots, n$$

or else have

$$\sup_{x \in A_J} f(x) \leqslant \theta \quad \text{or} \quad \inf_{x \in A_J} f(x) > \theta.$$

For a contradiction, suppose not. Then, for some 'bad' index $\xi \in \{1, \ldots, n\}$, there exists an infinitely long string $J = j_1 \pm j_2 \pm \cdots j_k \pm \cdots$ that results from of application of CORPUS but

$$\mathcal{D}_\xi[f|A_{J|k}] > \tfrac{1}{2}\mathcal{D}_\xi[f|A] \quad \text{for all } k \in \mathbb{N}. \tag{A3}$$

In any such string, at least one index $\eta \in \{1, \ldots, n\} \setminus \{\xi\}$ must appear infinitely many times; each occurrence of η halves the d_η-diameter of the resulting child sets (in the case of μ satisfying the barycenter condition (22), the diameter is reduced by a factor of $1 - C(\mu) < 1$). Therefore, there exists $k \in \mathbb{N}$ large enough that

$$\mathcal{D}_\eta[f|A_{J|k}] \leqslant \varpi_f(d_\eta(A_{J|k})) \leqslant \tfrac{1}{2}\mathcal{D}_\xi[f|A].$$

Suppose that $j_{k'}$ is the next occurrence of η after $J|k$. Then, since CORPUS bisects normal to $\widehat{e}_j$ only when the jth subdiameter is maximal, for every $j \in \{1, \ldots, n\}$,

$$\mathcal{D}_j[f|A_{J|k'}] \leqslant \mathcal{D}_\eta[f|A_{J|k'}] \leqslant \tfrac{1}{2}\mathcal{D}_\xi[f|A].$$

In particular, taking $j = \xi$ in the above and applying (A3) yields

$$\tfrac{1}{2}\mathcal{D}_\xi[f|A] < \mathcal{D}_\xi[f|A_{J|k'}] \leqslant \mathcal{D}_\eta[f|A_{J|k'}] \leqslant \tfrac{1}{2}\mathcal{D}_\xi[f|A],$$

which is a contradiction. Hence, no such ξ can exist.

Therefore, given any $\varepsilon > 0$, after sufficiently many iterations $k \in \mathbb{N}$, all the sets $A \in \mathscr{A}(k)$ are sets on which f always fails, or always succeeds, or else satisfy

$$\max_{1 \leqslant j \leqslant n} \mathcal{D}_j[f|A] < \varepsilon.$$

The claimed convergence now follows from Corollary 3.3. □

REFERENCES

1. Lucas LJ, Owhadi H, Ortiz M. Rigorous verification, validation, uncertainty quantification and certification through concentration-of-measure inequalities. *Computer Methods in Applied Mechanics and Engineering* 2008; **197**(51–52):4591–4609.
2. McDiarmid C. On the method of bounded differences. *Surveys in Combinatorics, 1989 (Norwich, 1989). London Mathematical Society*, Lecture Note Series, vol. 141. Cambridge University Press: Cambridge, 1989; 148–188.
3. Oberkampf WL, Trucano TG, Hirsch C. Verification, validation and predictive capability in computational engineering and physics. *Applied Mechanics Reviews* 2004; **57**(5):345–384.

4. Babuška I, Nobile F, Tempone R. Reliability of computational science. *Numerical Methods for Partial Differential Equations* 2007; **23**(4):753–784. Available from: http://dx.doi.org/10.1002/num.20263.

5. Najm HM, Debusschere BJ, Marzouk YM, Widmer S, Le Maître OP. Uncertainty quantification in chemical systems. *International Journal for Numerical Methods in Engineering* 2009; **80**(6–7):789–814.

6. Red-Horse JR, Ghanem RG. Elements of a function analytic approach to probability. *International Journal for Numerical Methods in Engineering* 2009; **80**(6–7):689–716.

7. Saltelli A, Tarantola S. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association* 2002; **97**(459):702–709. Available from: http://dx.doi.org/10.1198/016214502388618447.

8. Sobol' IM. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment* 1993; **1**(4):407–414.

9. Jansen MJW, Rossing WAH, Daamen RA. Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In *Predictability and Nonlinear Modelling in Natural Sciences and Economics*, Grasman J, van Straaten G (eds). Kluwer Acad. Publishers: Dordrecht, 1994; 335–343.

10. Cukier RI, Fortuin CM, Shuler KE, Petschek AG, Schaibly JH. Study of the sensitivity of coupled reaction systems to uncertainties in rate coeffcients. I. Theory. *Journal of Chemical Physics* 1973; **59**:3873–3878.

11. Cukier RI, Schaibly JH, Shuler KE. Study of the sensitivity of coupled reaction systems to uncertainties in rate coeffcients. III. Analysis of the approximations. *Journal of Chemical Physics* 1978; **63**:1140–1149.

12. Saltelli A, Tarantola S, Chan KP-S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 1999; **41**(1):39–56.

13. Tarantola S, Gatelli D, Mara TA. Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering and System Safety* 2006; **91**(6):717–727.

14. Scovel C, Steinwart I. Hypothesis testing for validation and certification. *Statistics and Probability Letters* 2010. Available from: http://www.ccs3.lanl.gov/group/papers/la-ur-10-02355.pdf, submitted.

15. Ledoux M. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society: Providence, RI, 2001.

16. McDiarmid C. Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics*. Algorithms and Combinatorics, vol. 16. Springer: Berlin, 1998; 195–248.

17. Lévy P. *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino* (2nd edn). Gauthier-Villars: Paris, 1951.

18. Lawler EL, Wood DE. Branch-and-bound methods: a survey. *Operations Research* 1966; **14**:699–719.

19. Spear RC, Grieb TM, Shang N. Parameter uncertainty and interaction in complex environmental models. *Water Resources Research* 1994; **30**(11):3159–3169.

20. Hoeffding W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 1963; **58**:13–30.

21. Massart P. About the constants in Talagrand's concentration inequalities for empirical processes. *Annals of Probability* 2000; **28**(2):863–884.

22. Talagrand M. Concentration of measure and isoperimetric inequalities in product spaces. *Institut des Hautes Études Scientifiques. Publications Mathématiques* 1995; (81):73–205.

23. Talagrand M. New concentration inequalities in product spaces. *Inventiones mathematicae* 1996; **126**(3):505–563.

24. McDiarmid C. Centering sequences with bounded differences. *Combinatorics, Probability and Computing* 1997; **6**(1):79–86.

25. Gross L. Logarithmic Sobolev inequalities. *American Journal of Mathematics* 1975; **97**(4):1061–1083.

26. Boucheron S, Lugosi G, Massart P. Concentration inequalities using the entropy method. *Annals of Probability* 2003; **31**(3):1583–1614.

27. Maurer A. Concentration inequalities for functions of independent variables. *Random Structures and Algorithms* 2006; **29**(2):121–138.

28. Aytona GS, Noida WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. *Current Opinion in Structural Biology* 2007; **17**:192–198.

29. Cuitiño AM, Stainier L, Wang G, Strachan A, Cagin T, Goddard III WA, Ortiz M. A multiscale approach for modeling crystalline solids. *Journal of Computer-Aided Design* 2001; **8**:127–149.

30. DeVolder B, Glimm J, Grove JW, Kang Y, Lee Y, Pao K, Sharp DH, Ye K. Uncertainty quantification for multiscale simulations. *Journal of Fluids Engineering* 2002; **124**(1):29–41.

31. Liu WK, Karpov EG, Park HS. *Nano Mechanics and Materials*: *Theory, Multiscale Methods and Applications*. Wiley: New York, 2006.

32. McKerns M, Hung P, Aivazis M. Mystic A simple model-independent inversion framework, 2009. Available from: http://dev.danse.us/trac/mystic.