Uncertainty quantification via codimension one domain partitioning and a new concentration inequality

Tim Sullivan, Ufuk Topcu, Mike McKerns & Houman Owhadi
tjs@..., utopcu@..., mmckerns@... & owhadi@caltech.edu

California Institute of Technology, Pasadena, California 91125, USA

SIAM 2010 — Minisymposium on Numerical Methods for Stochastic Computation and Uncertainty Quantification
Pittsburgh, Pennsylvania, USA
12–16 July 2010
1 Introduction
   - Aim: Certification
   - McDiarmid’s Inequality

2 McDiarmid’s Inequality with Partitioning
   - Partitioning
   - Error Bounds and Convergence
   - Codimension-One Recursive Partitioning Using Subdiameters
   - Hypervelocity Impact Example

3 Confidence Bounds and Half-Spaces
   - Empirical Bounds
   - Confidence Bounds using Half-Spaces

4 Conclusions and Outlook
   - Conclusions
   - Outlook
   - Bibliography & Acknowledgements
Introduction: Certification

**Aim**

We approach uncertainty quantification from the point of view of the certification problem: we want good (rigorous and sharp) upper bounds on

$$\mu[f(X) \leq \theta],$$

where

- $f : \mathcal{X} \rightarrow \mathbb{R}$ is a system / response function of interest;
- $X : \Omega \rightarrow \mathcal{X}$ represents the random inputs of $f$, with law $\mu$;
- $\theta \in \mathbb{R}$ is some threshold for failure.

We do this so that we (hopefully) rigorously guarantee that

$$\mu[f(X) \leq \theta] \leq \epsilon,$$

where $\epsilon \in [0, 1]$ is a maximum acceptable probability of failure.
Introduction: Monte Carlo

Why not simply certify using Monte Carlo sampling?

**Quantitative Reasons**

For systems with small failure probability $p$, certification will take of the order of $p^{-2} \log p^{-1}$ samples (evaluations of $f$), which may be more expensive than the available resources permit.

**Qualitative Reasons**

Monte Carlo certification does not distinguish between the aleatoric uncertainty in the inputs $X$ and the input parameter sensitivity of $f$. In the language of QMU (quantification of margins and uncertainties), it may be desirable to quantify margins (e.g. mean performance) and uncertainties (system sensitivity) separately.
McDiarmid’s Inequality

Definition
For any function $f : \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ and $i \in \{1, \ldots, n\}$, the \textit{i}th McDiarmid subdiameter of $f$ is defined by

$$D_i[f] := \sup \{ |f(x) - f(x')| \mid x_j = x'_j \in \mathcal{X}_j \text{ for } j \neq i \};$$

the \textit{McDiarmid diameter} of $f$ is $D[f] := \left( \sum_{i=1}^{n} D_i[f]^2 \right)^{1/2}$.

Theorem (McDiarmid 1989)
For every product measure $\mu$ on $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ such that $\mathbb{E}[|f|]$ is finite, and for every $r > 0$,

$$\mu[f - \mathbb{E}[f] \geq r] \text{ and } \mu[f - \mathbb{E}[f] \geq -r] \leq \exp \left(-\frac{2r^2}{D[f]^2}\right).$$
Certification using McDiarmid’s Inequality

McDiarmid’s inequality implies that

$$ \mu[f \leq \theta] \leq \exp \left( -\frac{2(\mathbb{E}[f] - \theta)^2}{\mathcal{D}[f]^2} \right). $$

This provides a rigorous certification criterion in terms of the performance margin \( (\mathbb{E}[f] - \theta)_+ \) and the McDiarmid diameter \( \mathcal{D}[f] \): the system is certified as safe if

$$ \exp \left( -\frac{2(\mathbb{E}[f] - \theta)^2}{\mathcal{D}[f]^2} \right) \leq \epsilon. $$

Application of McDiarmid’s inequality is not an ideal method:

- determination of \( \mathcal{D}[f] \) requires a (potentially expensive) global optimization;
- \( \mathcal{D}[f] \) is a global sensitivity measure — because of this, McDiarmid’s inequality is often not sharp.
McDiarmid’s Inequality is Not Sharp

Exact probability of failure if $\mu = \text{uniform}$: $\mu[f \leq \frac{1}{4}] = \frac{5}{12} \approx 0.42$

McDiarmid’s bound: $\mu[f \leq \frac{1}{4}] \leq e^{-1/8} \approx 0.88$
McDiarmid’s Inequality is Not Sharp

Exact probability of failure if $\mu = \text{uniform}$: $\mu[f \leq \frac{1}{4}] = \frac{5}{12} \approx 0.42$

McDiarmid’s bound: $\mu[f \leq \frac{1}{4}] \leq e^{-1/8} \approx 0.88$

McDiarmid’s bound on each third: $\mu[f \leq \frac{1}{4}] \leq \frac{1}{3}(0 + e^{-1/8} + 1) \approx 0.63$
McDiarmid’s Inequality with Partitioning

Let $\mathcal{P}$ be a finite or countable partition of $\mathcal{X}$ into pairwise-disjoint measurable rectangles, and let $\mu$ be any product measure on $\mathcal{X}$ for which $\mathbb{E}_\mu[|f|]$ is finite. Then

$$\mu[f \leq \theta] = \sum_{A \in \mathcal{P}} \mu([f \leq \theta] \cap A)$$

$$= \sum_{A \in \mathcal{P}} \mu(A) \mu[f \leq \theta | A]$$

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

$$=: \mu_P[f \leq \theta].$$
Error Bound

**Proposition (Error bound)**

Let $f : \mathcal{X} \to \mathbb{R}$ be measurable and let $\mathcal{P}$ be a partition of $\mathcal{X}$. Then, for every $\varepsilon > 0$, and for all sufficiently small $\delta > 0$,

$$
0 \leq \overline{\mu}_{\mathcal{P}}[f \leq \theta] - \mu[f \leq \theta] < \varepsilon + \sup_{A \in \mathcal{P}_\delta} \exp \left( -\frac{2 \left( \delta - \sum_{j=1}^{n} D_j[f|A] \right)^2}{D[f|A]^2} \right),
$$

where

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

I.e. the amount by which $\overline{\mu}_{\mathcal{P}}[f \leq \theta]$ is an over-estimate of the probability of failure is controlled by the McDiarmid subdiameters (not the metric diameter) of those $A \in \mathcal{P}$ on which $f$ exceeds the threshold for success by more than $\delta$ somewhere in $A$. 

---

Sullivan & al. (Caltech) UQ via Codimension 1 Partitioning SIAM UQ 2010 9 / 30
Partitioning Algorithms

For simplicity, restrict attention to parameter spaces that are compact boxes in $\mathbb{R}^n$:

$$X = [a_1, b_1] \times \cdots \times [a_n, b_n].$$

How can one efficiently construct a partition $\mathcal{P}$ of $X$ for which $\mu_{\mathcal{P}}[f \leq \theta]$ is nearly $\mu[f \leq \theta]$?

**Naïve Method**

Construct a sequence $(\mathcal{P}(k))_{k \in \mathbb{N}}$ by bisecting each box $A \in \mathcal{P}(k)$ in each of the $n$ coordinate directions to produce the boxes of $\mathcal{P}(k+1)$.

The naïve method is strongly affected by the curse of dimension: there are $2^n$ new boxes with each iteration. Therefore, we propose an algorithm in which the McDiarmid subdiameters are used as sensitivity indices to guide a codimension-one recursive partitioning scheme.
Recursively define a sequence of partitions \((\mathcal{P}(k))_{k \in \mathbb{N}}\) as follows: for each \(A \in \mathcal{P}(k)\),

1. if \(A \in \mathcal{P}(k)\) satisfies \(\inf_{x \in A} f(x) > \theta\) (i.e. \(f\) always succeeds on \(A\)), then include \(A\) in \(\mathcal{P}(k + 1)\) as it is;
2. if \(A \in \mathcal{P}(k)\) satisfies \(\sup_{x \in A} f(x) \leq \theta\) (i.e. \(f\) always fails on \(A\)), then include \(A\) in \(\mathcal{P}(k + 1)\) as it is;
3. otherwise,
   1. determine \(j \in \{1, \ldots, n\}\) such that \(D_j[f|A]\) is maximal (choose one such \(j\) arbitrarily if there are multiple maximizers);
   2. set \(c(A) := \int_A x \, dx\), the geometric centre of \(A\);
   3. bisect \(A\) by a hyperplane of codimension one (i.e. of dimension \(n - 1\)) through \(c(A)\) and normal to \(\hat{e}_j\), the unit vector in the \(j^{th}\) coordinate direction;
   4. include in \(\mathcal{P}(k + 1)\) the two subsets of \(A\) so generated, but not the original set \(A\); the two new sets are called the \textit{children} of \(A\).
McDiarmid’s Inequality with Partitioning

Codimension-One Recursive Partitioning Using Subdiameters

\[ f \leq \theta \] (failure)

\[ f > \theta \] (success)

Sullivan & al. (Caltech)
CORPUS Convergence Theorem

**Theorem**

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

$$ \mu[f \leq \theta] = \lim_{k \to \infty} \mu_{\mathcal{P}(k)}[f \leq \theta]. $$

**Sketch of Proof**

Need to show that for any initial box $A$, every generation-$g$ child $A'$ of $A$ with $g$ sufficiently large must satisfy one of the following:

$$ \mathcal{D}_j[f|A'] \leq \frac{1}{2} \mathcal{D}_j[f|A] \text{ for all } j = 1, \ldots, n, \text{ or } $$

$$ \sup_{x \in A'} f(x) \leq \theta \text{ or } \inf_{x \in A'} f(x) > \theta. $$
Hypervelocity Impact

Figure: Caltech’s Small Particle Hypervelocity Impact Range (SPHIR): a two-stage light gas gun that launches 1–50 mg projectiles at speeds of 2–10 km · s$^{-1}$. 
Hypervelocity Impact

Figure: Caltech’s Small Particle Hypervelocity Impact Range (SPHIR): a two-stage light gas gun that launches 1–50 mg projectiles at speeds of 2–10 km · s⁻¹.
Hypervelocity Impact: Surrogate Model

Experimentally-derived deterministic surrogate model for the perforation area (in mm²):

- plate thickness $h \in [1.52, 2.67]$ mm;
- impact obliquity $\alpha \in [0, \frac{\pi}{6}]$;
- impact speed $v \in [2.1, 2.8]$ km · s⁻¹.

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

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

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

is called the ballistic limit, the impact speed below which no perforation occurs. The failure event is non-perforation, i.e. $[f = 0] \equiv [f \leq 0]$. 
**Hypervelocity Impact: Effect of Partitioning**

$$\log \mu_\mathcal{P}[f = 0]$$

![Graph showing the effect of partitioning on failure probability](image)

**Figure:** In **blue**, the $\mu_\mathcal{P}$ upper bound on the failure probability versus the number of boxes $\#\mathcal{P}$ used by the CORPUS algorithm. In **green**, the corresponding upper bound obtained if all boxes are subdivided, instead of just those on which $f$ both succeeds and fails. In **red**, the exact probability of failure.
Confidence in Empirical Bounds

Suppose that we are given a partition \( \mathcal{P} = A_1 \cup \cdots \cup A_K \) for which we know \( \mu(A_k) \) and \( \mathcal{D}[f|A_k] \) for each \( k = 1, \ldots, K \), but our knowledge of the local mean performance \( \mathbb{E}[f|A_k] \) comes from \( m_k \) empirical samples:

\[
\mathbb{E}[f|A_k] \sim \langle f|A_k \rangle := \frac{1}{m_k} \sum_{j=1}^{m_k} f(X^{(j)}).
\]

It is not true that

\[
\mu[f \leq \theta] \leq \sum_{k=1}^{K} \mu(A_k) \exp \left( -\frac{2(\langle f|A_k \rangle - \theta)^2}{\mathcal{D}[f|A_k]^2} \right)
\]

however, it may be true, with acceptably high probability, that

\[
\mu[f \leq \theta] \leq \sum_{k=1}^{K} \mu(A_k) \exp \left( -\frac{2(\langle f|A_k \rangle - \alpha_k - \theta)^2}{\mathcal{D}[f|A_k]^2} \right),
\]

where \( \alpha_k > 0 \) are suitable margin hits.
McDiarmid’s Inequality with an Empirical Mean

**Theorem**

Let $X^{(1)}, \ldots, X^{(m)}$ be $m$ independent $\mu$-distributed samples of $\mathcal{X}$ and let

$$
\langle f \rangle := \frac{1}{m} \sum_{j=1}^{m} f(X^{(j)})
$$

be the associated empirical mean of $f$. Then, for every $\varepsilon > 0$, with $\mu$-probability at least $1 - \varepsilon$ on the $m$ samples,

$$
\mu[f \leq \theta] \leq \exp \left( -\frac{2(\langle f \rangle - \alpha - \theta)^2}{\mathcal{D}[f]^2} \right),
$$

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

Partitioned McDiarmid’s Inequality with Empirical Means

Given $\alpha = (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K$, let

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

We seek a bound

$$\mu[H_\alpha(Y)] \leq H_\alpha(-\alpha) \leq ??? \equiv \mu_{\mathcal{P}}[f \leq \theta]$$

where

$$Y_k := \mathbb{E}[f|A_k] - \langle f | A_k \rangle.$$ 

Note that each $Y_k$ is a real-valued random variable that concentrates about its mean, 0: for any $r > 0$,

$$\mu[Y_k \geq r] \text{ and } \mu[Y_k \leq -r] \leq \exp \left(-\frac{2m_k r^2}{D[f|A_k]^2}\right).$$
Level Sets of $H_\alpha$

Figure: 20 equally-spaced contours of $H_\alpha$, which increases from 0 in the bottom-left to 1 in the top-right. $H_\alpha$ is increasing and sublevels of small enough values are convex.
Bounds Using Orthants

Since $H_{\alpha}$ is increasing in each of its $K$ arguments and the $K$ random variables $\langle f | A_k \rangle$ are independent, one bound on $\mu[H_{\alpha}(Y) \leq H_{\alpha}(-\alpha)]$ is provided as follows: fix $\varepsilon > 0$, choose any $\varepsilon_1, \ldots, \varepsilon_K > 0$ such that

$$1 - \varepsilon = \prod_{k=1}^{K} (1 - \varepsilon_k),$$

and set

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

Then

$$\mu[H_{\alpha}(Y) \geq H_{\alpha}(-\alpha)] \geq \prod_{k=1}^{K} \mu[Y_k \geq -\alpha_k] \geq \prod_{k=1}^{K} (1 - \varepsilon_k) = 1 - \varepsilon.$$
The Problem with Orthants. . .

- The problem with the bound on the previous slide is that for even moderately large $K$, $\varepsilon_k$ must be tiny in order to make $\varepsilon$ small enough. It then follows that $m_k$ must be large in order to make the margin hit $\alpha_k$ acceptably small.

- Geometrically, this can be seen as a consequence of using $K$-dimensional orthants to estimate the measure of a set: viewed from their vertices, high-dimensional orthants look very “narrow”.

- Half-spaces are much better, dimensionally speaking, since they always fill half the “field of view”.
A Bound on the Measure of a Half-Space

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

$$H_{p,\nu} = \{ y \in \mathbb{R}^K \mid \nu \cdot y \leq \nu \cdot p \}.$$

Since $\mathbb{E}[\nu \cdot Y] = 0$, application of McDiarmid’s inequality yields that

$$\mu[Y \in H_{p,\nu}] \leq \exp \left( -2(\nu \cdot p)^2 / \sum_{k=1}^{K} \frac{|\nu_k|^2}{m_k} \mathcal{D}[f|A_k]^2 \right).$$

Hence, for any $S \subseteq \mathbb{R}^K$,

$$\mu[Y \in S] \leq \inf \left\{ \exp \left( - \frac{2(\nu \cdot p)^2}{\sum_{k=1}^{K} \frac{|\nu_k|^2}{m_k} \mathcal{D}[f|A_k]^2} \right) \mid p \in \mathbb{R}^K \text{ and } \nu \in \mathbb{R}^K \text{ such that } S \subseteq H_{p,\nu} \right\}.$$
Consequences for $H_\alpha$

Suppose it is known a priori that $H_\alpha(-\alpha)$ is small enough that the sublevel set $H^{-1}_\alpha([0, H_\alpha(-\alpha)])$ is convex. Then, applying the inequality from the previous slide with $p = -\alpha$ and $\nu = \nabla H_\alpha(-\alpha)$ yields that

$$\mu[H_\alpha(Y) \leq H_\alpha(-\alpha)] \leq \exp \left( -\frac{2(\nabla H_\alpha(-\alpha) \cdot \alpha)^2}{\sum_{k=1}^{K} \frac{|\partial_k H_\alpha(-\alpha)|^2}{m_k} \mathcal{D}[f|A_k]^2} \right).$$

Note:

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

Note also that, by assumption, $\mathbb{E}[f|A_k]$ is unknown, so in practice one takes a supremum over known ranges of values for $\mathbb{E}[f|A_k]$. 
<table>
<thead>
<tr>
<th>$K = 2$, $m_1 = m_2 = 5$</th>
<th>$\varepsilon_1 = \varepsilon_2 = 1%$</th>
</tr>
</thead>
</table>

**Upper bounds on the probability of failure**  
*(i.e. non-perforation, $[f = 0]$)*

<table>
<thead>
<tr>
<th>If inputs uniformly dist.</th>
<th>3.7%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact local means: $\mu_P[f = 0]$</td>
<td>33%</td>
</tr>
<tr>
<td>Empirical local means: $H_\alpha(Y)$</td>
<td>54%</td>
</tr>
</tbody>
</table>

**Confidence levels**  
*(i.e. upper bounds on $\mu[H_\alpha(Y) \leq \mu_P[f = 0]]$)*

<table>
<thead>
<tr>
<th>Orthant method</th>
<th>2%</th>
</tr>
</thead>
</table>

**Half-space method**

<table>
<thead>
<tr>
<th>Means known exactly</th>
<th>0.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means known to within $\pm 1\text{ mm}^2$</td>
<td>0.9%</td>
</tr>
<tr>
<td>Means known to within $\pm 5\text{ mm}^2$</td>
<td>1.0%</td>
</tr>
<tr>
<td>$K = 2$, $m_1 = m_2 = 5$</td>
<td>$\varepsilon_1 = \varepsilon_2 = 1%$</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>

**Upper bounds on the probability of failure**

(*i.e. non-perforation, $[f = 0]$*)

<table>
<thead>
<tr>
<th>If inputs uniformly dist.</th>
<th>3.7%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact local means: $\mu_P[f = 0]$</td>
<td>33%</td>
</tr>
<tr>
<td>Empirical local means: $H_\alpha(Y)$</td>
<td>54%</td>
</tr>
</tbody>
</table>

**Confidence levels**

(*i.e. upper bounds on $\mu[H_\alpha(Y) \leq \mu_P[f = 0]]$*)

<table>
<thead>
<tr>
<th>Orthant method</th>
<th>2%</th>
<th>0.20%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Half-space method</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means known exactly</td>
<td>0.9%</td>
<td>0.09%</td>
</tr>
<tr>
<td>Means known to within $\pm 1\ mm^2$</td>
<td>0.9%</td>
<td>0.09%</td>
</tr>
<tr>
<td>Means known to within $\pm 5\ mm^2$</td>
<td>1.0%</td>
<td>0.10%</td>
</tr>
</tbody>
</table>
Scaling of Confidence Levels with $K$

This example leads us to consider the very different scaling properties of the orthant and half-space methods, provided sample sizes $m_1, \ldots, m_K$ are chosen appropriately.

**Proposition**

Suppose that the same level of confidence $1 - \varepsilon_0$ is required for each local mean $\mathbb{E}[f|A_k]$, $k = 1, \ldots, K$. Choose sample sizes $m_k$ such that

$$\sqrt{m_k} \propto \partial_k H_\alpha(-\alpha) D[f|A_k].$$

Then the confidence levels for $H_\alpha$ are given by:

**half-space method:** $\mu[H_\alpha(Y) \geq H_\alpha(-\alpha)] \geq 1 - \varepsilon_0^K$;

**orthant method:** $\mu[H_\alpha(Y) \geq H_\alpha(-\alpha)] \geq (1 - \varepsilon_0)^K$. 
The use of half-spaces exploits the fact that, in a probability normed vector space $\mathcal{V}$, a convex set $C$ that does not contain the centre of mass has small measure — exponentially small with respect to its distance from the centre of mass.

Hence, a quasiconvex function $f$ on $\mathcal{V}$ is unlikely to assume values below its value at the centre of mass.

This differs from concentration/deviation results in the literature in two ways:

1. there are no smoothness assumptions on $f$;
2. the result is a one-sided concentration about the value of $f$ at the centre of mass, not about $\mathbb{E}[f]$.

We believe that results of this type indicate a deeper connection between concentration-of-measure phenomena and large deviations principles.
Conclusions

In situations where failure is a rare event but McDiarmid diameters can be computed:

- McDiarmid’s inequality offers a rigorous upper bound on the probability of failure (certification criterion);
- partitioning offers a way to obtain arbitrarily sharp upper bounds on the probability of failure, at the cost of further diameter calculations;
- this can be done in ways that avoid the naïve curse of dimension;
- half-space methods provide confidence bounds in which high-cardinality partitions are a help, not a hindrance.
Outlook

- It is not necessary to assume that the partition elements are rectangles: in the non-rectangular situation, resort to martingale inequalities.
- The $\mu$ and $f$ to which CORPUS is applied may be surrogates for the real $\mu'$ and $f'$ on which the probability of failure upper bound will be calculated (perhaps using sampling) — can the approximation error be controlled?
- Does it make sense to ask for the “optimal” partition of a given cardinality? of a given mesh size?
- How can these methods be extended to handle noisy / imperfectly observed response functions $f$?
Bibliography & Acknowledgements


Portions of this work were developed as part of the Predictive Science Academic Alliance Program, funded by United States Department of Energy contract number DE-FC52-08NA28613.