G. W. Wasilkowski

Department of Computer Science
University of Kentucky

Presentation based on papers co-authored with

A. Gilbert, M. Gnewuch, M. Hefter,
A. Hinrichs, P. Kritzer, F. Y. Kuo,
F. Pillichshammer, L. Plaskota,
K. Ritter, I. H. Sloan, H. Woźniakowski
**INFORMATION-BASED COMPLEXITY** approach very nicely explained in Houman Owhadi’s tutorial: many thanks Houman

Briefly: instead of working with specific functions, **IBC** deals with problems on whole spaces of functions and tries to determine the **complexity**, i.e., the minimal cost, and (almost) **optimal algorithms**.

This is done in various settings including: **worst case, average case and randomized settings**
"Classical" Integration Problem

Given a normed space $F$ of functions on $[0, 1]^d$

approximate $I(f) = \int_{D^d} f(\mathbf{x}) \, d\mathbf{x}$

by cubatures $Q_n(f) = \sum_{j=1}^n f(t_{n,j}) \cdot a_{n,j}$

with small error $\|I - Q_n\|$ and

(if possible) small cost
Classical Methods are Extremely Bad!!!!
even for Finitely Many Variables.

(Product) Trapezoidal $T_n$ with $n$ samples has error

$$\text{error}(T_n; d = 2) \simeq \frac{1}{n} \quad d = 2 \text{ variables}$$

$$\text{error}(T_n; d = 360) \simeq \frac{1}{n^{2/360}} \quad d = 360 \text{ variables}$$

E.g., for 360 variables, it needs

$$n \sim 20^{180} \text{ to get only 1 digit of accuracy}$$
\[ 20^{180} = 1532495540865888858358347027150309183618739122 \\
18360217600000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
For Classical (Isotropic) Spaces

One Cannot Do Better

“Curse of Dimensionality”
For Classical (Isotropic) Spaces

One **Cannot Do Better**

“Curse of Dimensionality”

To Break this Curse

**Different Spaces Are Needed**

**Weighted Spaces**

treat different variables differently
Motivating Example

Compute expectation $\mathbb{E}(g(X(t_0)))$ for stochastic process $X(t) = \sum_{j=1}^{\infty} x_j \xi_j(t)$

Equivalent to computing an integral of

$$f(x_1, x_2, \ldots) = g \left( \sum_{j=1}^{\infty} x_j \xi_j(t_0) \right)$$
In
\[ g \left( \sum_{j=1}^{\infty} x_j \xi_j(t_0) \right) \]

"importance" of \( x_j \)

is quantized by the size of \( |\xi_j(t_0)| \).

The larger \( |\xi_j(t_0)|\) the more important \( x_j \).
Although there are results for quite general spaces and problems

we present results for Integration
over a special space $\mathcal{F}$

$\mathcal{F}$ is the $\gamma$-weighted Banach space
of functions with dominating mixed derivatives
of order one bounded in $L_p$-norm
Notation:

\( \mathcal{w} \) finite subsets of \( \mathbb{N}_+ \)

listing the "variables in action"

Given \( \mathbf{x} = (x_1, x_2, \ldots) \), \( \mathbf{x}_\mathcal{w} = (x_j : j \in \mathcal{w}) \)

\[
[x_\mathcal{w}; 0] = (y_1, y_2, \ldots) \quad \text{with} \quad y_j = \begin{cases} 
    x_j & \text{if } j \in \mathcal{w}, \\
    0 & \text{if } j \notin \mathcal{w}
\end{cases}
\]

\[
f^{(\mathcal{w})} = \frac{\partial^{|\mathcal{w}|}}{\partial \mathbf{x}_\mathcal{w}} f = \prod_{j \in \mathcal{w}} \frac{\partial}{\partial x_j} f
\]
Domain: $D^N$ set of sequences $(x_j)_{j \in \mathbb{N}}$ with $x_j \in D$; for simplicity $D = [0, 1]$.

$\mathcal{F}$ the Banach space of $f : D^N \rightarrow \mathbb{R}$ endowed with the norm

$$
\| f \|_{\mathcal{F}} = \left( \sum_{\mathbf{w}} \gamma_{\mathbf{w}}^{-p} \left\| f^{(\mathbf{w})}([\mathbf{w}; 0]) \right\|_{L^p(D^{\mathbf{w}})}}^p \right)^{1/p} < \infty
$$

Here $p \in [1, \infty]$ and $\gamma_{\mathbf{w}} \geq 0$ are weights
For simplicity

**Product Weights** introduced by [Sloan and Woźniakowski 1998]:

$$\gamma_w = c \prod_{j \in w} \gamma_j \quad (\gamma_j = j^{-\beta})$$

For ‘motivating example’ we have

$$\gamma_w \approx \prod_{j \in w} |\xi_j(t_0)|^\alpha$$
Integration Problem

APPROXIMATE:

\[ \mathcal{I}(f) := \int_{D^N} f(x) \, d^N x \]

\[ = \lim_{d \to \infty} \int_{D^d} f(x_1, \ldots, x_d, 0, 0, \ldots) \, d[x_1, \ldots, x_d] \]
Integration Problem

**APPROXIMATE:**

\[ \mathcal{I}(f) := \int_{D^N} f(x) \, d^N x \]

\[ = \lim_{d \to \infty} \int_{D^d} f(x_1, \ldots, x_d, 0, 0, \ldots) \, d[x_1, \ldots, x_d] \]

We have

\[ \|\mathcal{I}\| = \left( \sum_{\mathbf{w}} \gamma_{\mathbf{w}}^{p^*} / (1 + p^*) |\mathbf{w}| \right)^{1/p^*} \]

\[ = \max_{\mathbf{w}} \gamma_{\mathbf{w}} \text{ for } p = 1 \]
ASSUMPTIONS:

(A1) \[ \|I\| < \infty \]

(A2) We can sample \( f \) at points \([x_w; 0]\)

For ‘motivating example’ we have:

\[
f([x_w; 0]) = g \left( \sum_{j \in w} x_j \xi(t_0) \right)
\]
How to Cope with So Many Variables?

(i) **Truncate the dimension**, i.e., approximate the integral of

$$f(x_1, \ldots, x_k, 0, 0, \ldots)$$

or (even better)

(ii) use

**Multivariate Decomposition Method**
Let
\[ f_k(x_1, \ldots, x_k) = f(x_1, \ldots, x_k, 0, 0, \ldots) \]

Given the **error demand** \( \varepsilon > 0 \),
\[ \text{dim}^{\text{trnc}}(\varepsilon) \text{ \( \varepsilon \)-truncation dimension} \]
the smallest \( k \) such that
\[ |\mathcal{I}(f) - \mathcal{I}(f_k)| \leq \varepsilon \|f\|_\mathcal{F} \quad \text{for all } f \in \mathcal{F} \]

Our concept of **Truncation Dimension**
is different than the one in Statistics!!!
If

$$|\mathcal{I}(f) - \mathcal{I}(f_k)| \leq \varepsilon \|f\|_\mathcal{F} \quad \text{and} \quad |\mathcal{I}(f_k) - Q_k(f_k)| \leq \varepsilon \|f\|_\mathcal{F}$$

then

$$|\mathcal{I}(f) - Q_k(f_k)| \leq 2\varepsilon \|f\|_\mathcal{F}$$

Hence

the smaller \( \dim(\varepsilon) \) the better
Special Case: \(\gamma_w = c \prod_{j \in w} j^{-\beta}\)

\[
\text{dim}^{\text{trnc}}(\varepsilon) \leq \min \left\{ \ell : \sum_{j=\ell+1} \frac{p^* + 1}{c^{p^*}} \ln\left(\frac{1}{1 - \varepsilon^{p^*}}\right) \leq \frac{\sum_{j=\ell+1} j^{-\beta p^*}}{c^{p^*}} \leq \frac{p^* + 1}{c^{p^*}} \ln\left(\frac{1}{1 - \varepsilon^{p^*}}\right) \right\}
\]

\[= O\left(\varepsilon^{-1/(\beta-1+1/p)}\right)\]

for \(p > 1\) and

\[
\text{dim}^{\text{trnc}}(\varepsilon) = \left\lceil \left(\frac{c}{\varepsilon}\right)^{1/\beta} \right\rceil - 1
\]

for \(p = 1\)
Specific Values of $\dim^{\text{trnc}}(\varepsilon)$ for $p = 1$ and $\gamma_w = \prod_{j \in w} j^{-\beta}$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$10^{-1}$</th>
<th>$10^{-2}$</th>
<th>$10^{-3}$</th>
<th>$10^{-4}$</th>
<th>$10^{-5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dim^{\text{trnc}}(\varepsilon)$</td>
<td>2</td>
<td>9</td>
<td>31</td>
<td>99</td>
<td>316</td>
</tr>
<tr>
<td>$\dim^{\text{trnc}}(\varepsilon)$</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>21</td>
<td>46</td>
</tr>
<tr>
<td>$\dim^{\text{trnc}}(\varepsilon)$</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>17</td>
</tr>
</tbody>
</table>

For instance, for the error demand $\varepsilon = 10^{-3}$ with $\beta = 4$, only five variables instead of $\infty$-many!

Worst Case Error of QMC or Sparse Grids Methods is:

$$\leq O\left(\frac{\ln^4 n}{n}\right)$$
Multivariate Decomposition Method

MDM replaces one infinite-variate integral by only few integrals each with only few variables.
Any function $f$ has the unique anchored decomposition

$$f(x) = \sum_{\mathcal{w}} f_{\mathcal{w}}(x_{\mathcal{w}}),$$

where $f_{\mathcal{w}}$ depends only on $x_j$ with $j \in \mathcal{w}$ and vanishes if $x_j = 0$.

**General Idea:**
- Select the "most important" $\mathcal{w}$'s
- Approximate Integrals of $f_{\mathcal{w}}$ only for the selected $\mathcal{w}$'s
More precisely:

Given the error demand $\varepsilon > 0$, construct an "active set" $\text{Act}(\varepsilon)$ of subsets $w$ such that

$$\left| \mathcal{I} \left( \sum_{w \notin \text{Act}(\varepsilon)} f_w \right) \right| \leq \frac{\varepsilon}{2} \| f \|_{\mathcal{F}}$$

for all $f \in \mathcal{F}$.

Do nothing for integrals of $f_w$ with $w \notin \text{Act}(\varepsilon)$
For \( \mathbf{w} \in \text{Act}(\varepsilon) \), choose \( n_{\mathbf{w}} \) and cubatures \( Q_{\mathbf{w},n_{\mathbf{w}}} \)

to approximate integrals of \( f_{\mathbf{w}} \) such that

\[
\sum_{\mathbf{w} \in \text{Act}(\varepsilon)} |\mathcal{I}(f_{\mathbf{w}}) - Q_{\mathbf{w},n_{\mathbf{w}}}(f_{\mathbf{w}})| \leq \frac{\varepsilon}{2} \| f \|_{\mathcal{F}} \quad \text{for all } f \in \mathcal{F}.
\]

The cubatures \( Q_{\mathbf{w},n_{\mathbf{w}}} \) could be QMC or Sparse Grids
Then the MDM given by

$$Q_\varepsilon(f) := \sum_{w \in \text{Act}(\varepsilon)} \sum_{m} Q_{w,m} (f_w)$$

has the "worst case error" bounded by $\varepsilon$, i.e.,

$$|I(f) - Q_\varepsilon(f)| \leq \varepsilon \|f\|_\mathcal{F}$$

for all $f \in \mathcal{F}$.

How about the COST?
The number of integrals to approximate is small:

$$\text{card}(\varepsilon) := |\text{Act}(\varepsilon)| = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$$

Each $f_w$ depends on only $|w|$ variables.

The largest number of variables is also small:

$$\text{dim}(\varepsilon) := \max \{|w| : w \in \text{Act}(\varepsilon)\} = \mathcal{O}(???)$$
The number of integrals to approximate is small:

\[ \text{card}(\varepsilon) := |\text{Act}(\varepsilon)| = \mathcal{O}\left(\frac{1}{\varepsilon}\right) \]

Each \( f_w \) depends on only \(|w|\) variables.

The largest number of variables is also small:

\[ \text{dim}(\varepsilon) := \max \{|w| : w \in \text{Act}(\varepsilon)\} = \mathcal{O}\left(\frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))}\right) \]

[Plaskota and W. 2011]

\( \text{dim}(\varepsilon) \) is like

**Superposition Dimension** in Statistics
Very efficient algorithm to construct $\text{Act}(\varepsilon)$ in [Gilbert and W. 2016]

Specific Values of $\dim(\varepsilon)$ and $\text{card}(\varepsilon)$ for $p = 1$ and $\gamma_w = \prod_{j \in w} j^{-\beta}$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$10^{-1}$</th>
<th>$10^{-2}$</th>
<th>$10^{-3}$</th>
<th>$10^{-4}$</th>
<th>$10^{-5}$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-1}$</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>26</td>
</tr>
</tbody>
</table>

For instance, for $\varepsilon = 10^{-3}$ with $\beta = 4$
it is sufficient to approximate

10 integrals with at most 2 variables!
Active Set \( \text{Act}(10^{-3}) \)

For \( \beta = 4 \)
\[ \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\} \]

For \( \beta = 3 \)
\[ \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{1, 7\}, \{1, 8\}, \{1, 9\}, \{2, 3\}, \{2, 4\}, \{1, 2, 3\}, \{1, 2, 4\} \]
For $\beta = 2$

31 of integrals with 1 variable
54 of integrals with 2 variables
26 of integrals with 3 variables
2 of integrals with 4 variables

$\emptyset$, 
{1}, ..., {31}, 
{1, 2}, ..., {1, 31}, {2, 3}, ..., {2, 15}, 
{3, 4}, ..., {3, 10}, {4, 5}, {4, 6}, {4, 7}, {5, 6}, 
{1, 2, 3}, ..., {1, 2, 15}, {1, 3, 4}, ..., {1, 3, 10}, 
{1, 4, 5}, {1, 4, 6}, {1, 4, 7}, {1, 5, 6}, {2, 3, 4}, {2, 3, 5}, 
{1, 2, 3, 4}, {1, 2, 3, 5}
REMARK:

We do **NOT** know $f_w$ terms. However, we can sample them. Indeed due to [Kuo, Sloan, W., and Woźniakowski 2010b]

$$f_w(x_w) = \sum_{v \subseteq w} (-1)^{|w| - |v|} f([x_v; 0])$$

requires

$$2^{|w|}$$ samples of $f$
**REMARK:**

We do **NOT** know $f_w$ terms. However, we can sample them. Indeed due to [Kuo, Sloan, W., and Woźniakowski 2010b]

$$f_w(x_w) = \sum_{v \subseteq w} (-1)^{|w|-|v|} f([x_v; 0])$$

requires

$$2^{|w|} \text{ samples of } f$$

but from [Plaskota and W. 2011]

$$2^{|w|} = O\left(\varepsilon^{\ln(\ln(1/\varepsilon))}\right) \text{ is small for modest } \varepsilon.$$
How About **ANOVA** Spaces?

\[ f \in \mathcal{F}^{\text{ANOVA}} \text{ iff } \]

\[ f(x) = \sum_{w} f_{w,A}(x_{w}) \]

with \( \int_{D} f_{w,A}(x_{w}) \, dx_{j} = 0 \) and

\[ \| f \|_{\mathcal{F}^{\text{ANOVA}}} = \left( \sum_{w} \gamma_{w}^{-p} \| f_{w,A}^{(w)} \|_{L_{p}}^{p} \right)^{1/p} < \infty \]
ANOVA decomposition terms $f_{w,A}$ cannot be sampled, i.e., low truncation dimension and MDM might not be applicable.

Even worse: the ‘easiest’ (constant) term is **NOT known**; it is the integral we want:

$$f_{\emptyset,A} = \mathcal{I}(f)$$
H OWEVER

If the spaces are EQUIVALENT, then efficient algorithms for anchored spaces are also efficient for ANOVA spaces. This motivated the study of...
Equivalence of anchored and ANOVA Spaces

For product weights \( \gamma_w = \prod_{j \in w} j^{-\beta} \)

\[ \mathcal{F} = \mathcal{F}^{\text{ANOVA}} \] as sets.

For the imbedding \( \iota : \mathcal{F} \hookrightarrow \mathcal{F}^{\text{ANOVA}} \) we have

\[ \| \iota \| = \| \iota^{-1} \| \leq \prod_{j=1}^{\infty} (1 + j^{-\beta}) \]

EQUIVALENCE iff \( \beta > 1 \)
Research direction initiated in [Hefter and Ritter 2014], Hilbert spaces setting $p = 2$ and product weights

[Hefter, Ritter and W. 2016]
$p \in \{1, \infty\}$ and general weights,

[Hinrichs and Schneider 2016]
$p \in (1, \infty)$,

[Gnewuch, Hefter, Hinrichs, Ritter, and W. 2016]
more general spaces,

[Kritzer, Pillichshammer, and W. 2017]
sharp lower bounds,

[Hinrichs, Kritzer, Pillichshammer, and W. 2017] most general
GENERALIZATIONS

More General Domain:
Any interval $D$ including $D = \mathbb{R}$

More General Distributions $\mu$ on $D$:
e.g., Exponential, Gaussian

More General Integrals: $\int_{\mathbb{R}^N} f(\mathbf{x}) \mu^N(\mathbf{d}x)$

Other Linear Solution Operators: $S(f) = ???$
e.g., Function Approximation, ODE’s, PDE’s

General Information about $f$:
$L_1(f), L_2(f), \ldots, L_n(f), \quad L_j \in \mathcal{F}^*$
Bayesian Approach:

Endowing $\mathcal{F}$ with **Gaussian** probability measure $\text{PROB}$ and studying **average case errors**:

$$\int_{\mathcal{F}} \| S(f) - \text{Alg}(L_1(f), \ldots, L_n(f)) \|^p_{S(\mathcal{F})} \text{PROB}(df)$$

Similar results in [W. 2014]
Comments to Houman Owhadi’s 1st talk:

[Traub, W., and Woźniakowski 1988] has a number of chapters devoted to the average, randomized and probabilistic settings for \textit{infinitely dimensional} Hilbert and Banach spaces. They are based on a number of earlier papers. Currently there are 100’s of IBC such papers, see e.g. 3 Volumes monograph:

[E. Novak and H. Woźniakowski 2008-10]
On page 16, the IBC Probabilistic Setting was attributed to H. Woźniakowski’s paper. However, as acknowledged in that paper, the results were based on some of the results of my paper: Optimal algorithms for linear problems with Gaussian measures, *Rocky Mountains J. of Math.* 1986, where IBC Probabilistic Setting has been introduced for the first time.
Comment to Ilias Bilionis’ talk:

Research that seem to be related:

IBC approach to PDE’s with random coefficients
by Ch. Shwab and his collaborators,
e.g., F.Y.Kuo, D. Nuyens, I. H. Sloan
THANK YOU FOR THE ATTENTION