∞-Variate Linear Problems

G. W. Wasilkowski
Department of Computer Science
University of Kentucky

In this presentation:

Information-Based Complexity approach, i.e., study of the worst case complexity with respect to a whole normed space $\mathcal{F}$ of functions

In this presentation:

Information-Based Complexity approach, i.e., study of the worst case complexity with respect to a whole normed space $\mathcal{F}$ of functions

There is the **Curse of Dimensionality** for **isotropic spaces**. To brake this curse, Non-isotropic spaces are needed.

In this presentation:

Information-Based Complexity approach, i.e., study of the worst case complexity with respect to a whole normed space $\mathcal{F}$ of functions

There is the **Curse of Dimensionality** for **isotropic spaces**. To brake this curse, Non-isotropic spaces are needed.

Weighted Spaces are good candidates. Introduced by [Sloan and Woźniakowski 1998], assign different importance to different variables.

Motivating Example

Compute expectation $\mathbb{E}(g(\boldsymbol{X}(t_0)))$ for

stochastic process $\boldsymbol{X}(t) = \sum_{j=1}^{\infty} x_j \cdot \xi_j(t)$,

where x_j i.i.d. random variables w.r.t. μ

and $\xi_j(t) \to 0$ fast.

Motivating Example

Compute expectation $\mathbb{E}(g(\boldsymbol{X}(t_0)))$ for

stochastic process $\boldsymbol{X}(t) = \sum_{j=1}^{\infty} x_j \cdot \xi_j(t)$,

where x_j i.i.d. random variables w.r.t. μ

and $\xi_j(t) \to 0$ fast.

Equivalent to computing an ∞-variate integral

$$\int_{\mathbb{R}^{\mathbb{N}}} f(\boldsymbol{x})\, \mathrm{d}\mu^{\mathbb{N}}(\boldsymbol{x}) \quad \text{for} \quad f(\boldsymbol{x}) = g\left(\sum_{j=1}^{\infty} x_j \cdot \xi_j(t_0)\right)$$

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.

This leads to **WEIGHTED SPACES**

of ∞-**VARIATE FUNCTIONS**

Initial complexity attempts: [W. and Woźniakowski 96 and 00], [Plaskota, W. and Woźniakowski 00], and [Hickernell and Wang 01]

[Hickernell, Müller-Gronbach, Niu, Ritter 10]

Since 2010, many papers have been written by, e.g.,

J. Baldeaux, A. Cohen, J. Creutzig, W. Dahmen, S. Dereich, R. DeVore, J. Dick, D. Dũng, A. Gilbert, M. Gnewuch, M. Griebel, M. G. Gunzburger, M. Hefter, S. Heinrich, F. J. Hickernell, A. Hinrichs, Z. Kacewicz, P. Kritzer, F. Y. Kuo, S. Mayer, T. Müller-Gronbach, P. Morkisz, B. Niu, J. Nichols, D. Nuyens, F. Pillichshammer, L. Plaskota, P. Przybylowicz, K. Ritter, Ch. Schwab, I. H. Sloan, H. Woźniakowski

However, majority of the papers dealing with:
Hilbert spaces,
integration problem,
and **too simplistic cost model**

In this presentation:
Banach or just **normed spaces**,
linear tensor-product problems,
and **more realistic cost model**

In our model, $f(\boldsymbol{x})$ can be evaluated for $\boldsymbol{x}$ with **only finitely many** $x_j \neq 0$

$$f(\boldsymbol{x}_{\mathfrak{w}}) = g\left(\sum_{j\in\mathfrak{w}} x_j\,\xi_j(t_0)\right) \quad \text{for **finite** } \mathfrak{w}\subset\mathbb{N}$$

Cost of obtaining the sample $f(\boldsymbol{x}_{\mathfrak{w}})$ should depend on the size of $\mathfrak{w}$, i.e.

$$\mathbf{cost}(f(\boldsymbol{x}_{\mathfrak{w}})) = \$(|\mathfrak{w}|)$$

Positive and sharp results for

$$\Omega(k) = \$(k) = e^{O(k)}$$

Notation:

Domain: $D^{\mathbb{N}}$ set of sequences $\boldsymbol{x} = (x_1, x_2, \dots)$
with $x_j \in D$; e.g., $D = [0,1]$, $\mathbb{R}_+$ or $\mathbb{R}$

$\mathfrak{w}$ finite subsets of $\mathbb{N}_+$
listing the "variables in action", e.g.,
given $\boldsymbol{x} = (x_1, x_2, \dots)$, $\boldsymbol{x}_{\mathfrak{w}} = (x_j : j \in \mathfrak{w})$

$$[\boldsymbol{x}_{\mathfrak{w}}; \mathbf{0}] = (y_1, y_2, \dots) \quad \text{with} \quad y_j = \begin{cases} x_j & \text{if } j \in \mathfrak{w}, \\ 0 & \text{if } j \notin \mathfrak{w} \end{cases}$$

For $p \in [1, \infty]$ we use $p^* = \frac{p}{p-1}$

Weighted Spaces $\mathcal{F}_\gamma$

Any $f \in \mathcal{F}_\gamma$ has **decomposition**: $f = \sum_{\mathfrak{w}} f_{\mathfrak{w}}$ such that

$f_{\mathfrak{w}}$ depends only on $\boldsymbol{x}_{\mathfrak{w}}$,

$$f_{\mathfrak{w}} \in F_{\mathfrak{w}},$$

where $F_{\mathfrak{w}}$ is a Banach (or just normed) space

and $F_{\mathfrak{w}} \cap F_{\mathfrak{v}} = \{0\}$ if $\mathfrak{w} \neq \mathfrak{v}$

It is **Anchored** if

$f_{\mathfrak{w}}(\boldsymbol{x}) = 0$ when $x_j = 0$ for $j \in \mathfrak{w}$

The space $\mathcal{F}_\gamma$ is endowed with

$$\|f\|_{\mathcal{F}} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \|f_{\mathfrak{w}}\|_{F_{\mathfrak{w}}}^p\right)^{1/p}$$

Here $p \in [1, \infty]$ and $\gamma_{\mathfrak{w}} \geq 0$ are **weights**

Product Weights [Sloan and Woźniakowski 1998]:

$$\gamma_{\mathfrak{w}} = c \prod_{j \in \mathfrak{w}} j^{-\beta}$$

Product Order Depedent Weights [Kuo, Schwab, and Sloan 14]:

$$\gamma_{\mathfrak{w}} = (|\mathfrak{w}|!)^{\alpha} \prod_{j \in \mathfrak{w}} j^{-\beta} \quad \beta > \max(\alpha, 1/p^*)$$

Problem to Approximate

Approximate $\mathcal{S}(f)$, where $\mathcal{S} : \mathcal{F}_\gamma \to \mathcal{G}$

is a **linear** operator into a normed space $\mathcal{G}$

Example:

$$\begin{aligned}\mathcal{S}(f) &= \int_{D^{\mathbb{N}}} f\,\rho^{\mathbb{N}} \\ &= \lim_{d\to\infty} \int_{D^d} f(x_1,\dots,x_d,0,\dots)\prod_{j=1}^{d}\rho(x_j)\,\mathrm{d}(x_1,\dots,x_d)\end{aligned}$$

where ρ is a **probability density** on D

ASSUMPTION 1: For every $\mathfrak{w}$,

$$\|\mathcal{S}_{|F_{\mathfrak{w}}}\|_{F_{\mathfrak{w}}} \le C_1^{|\mathfrak{w}|}$$

ASSUMPTION 2: **Continuity of $\mathcal{S}$**

$$\|\mathcal{S}\|_{\mathcal{F}_\gamma} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{p^*} \|\mathcal{S}_{|F_{\mathfrak{w}}}\|_{F_{\mathfrak{w}}}^{p^*}\right)^{1/p^*} \le \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{p^*} C_1^{|\mathfrak{w}|\,p^*}\right)^{1/p^*} < \infty$$

For tensor product problems: $\|\mathcal{S}_{|F_{\mathfrak{w}}}\|_{F_{\mathfrak{w}}} = C_1^{|\mathfrak{w}|}$ with $C_1 = \|\mathcal{S}_{|F_1}\|_{F_1}$

and for product weights: $\|\mathcal{S}\|_{\mathcal{F}_\gamma} = \left(\prod_{j=1}^\infty \left(1 + (C_1\, j^{-\beta})^{p^*}\right)\right)^{1/p^*}$

Standard Integration Problem:

$\mathcal{F}_\gamma$ endowed with the norm

$$\|f\|_{\mathcal{F}} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \left\| f^{(\mathfrak{w})}([\cdot_{\mathfrak{w}};\mathbf{0}]) \right\|_{L_p(D^{|\mathfrak{w}|})}^p \right)^{1/p}, \quad f^{(\mathfrak{w})} = \prod_{j\in\mathfrak{w}} \frac{\partial}{\partial x_j} f$$

$D = [0,1]$ and $\rho \equiv 1$ (most often Hilbert case $p = 2$)

and $\mathcal{S} = \mathcal{I} = \int_{D^{\mathbb{N}}}$

Then

$$\gamma_{\mathfrak{w}} \simeq \prod_{j\in\mathfrak{w}} \frac{1}{|\xi_j(t_0)|^\delta}$$

and

$$C_1 = \|\mathcal{I}_{|F_1}\|_{F_1} = (1+p^*)^{-1/p^*} \quad (= 1 \quad \text{if } p^* = \infty)$$

How to Cope with So Many Variables?

Truncate the Dimension, i.e., replace

$f(\boldsymbol{x})$ by $f(x_1, \ldots, x_k, 0, 0, \ldots)$

for a "relatively" small $k = k(\text{error})$

or (even better) use

Multivariate Decomposition Method

i.e., approximate a small number of integrals

each with a small number of variables

Low Truncation Dimension

Anchored Decomposition

[Kritzer, Pillichshammer, W.2016]$\subset$ [Hinrichs, Kritzer, Pillichshammer, W.2018]

Let

$$f_k(x_1,\dots,x_k) \;=\; f(x_1,\dots,x_k,0,0,\dots)$$

$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$ **ε-truncation dimension**

is the smallest k such that

$$\|\mathcal{S}(f)-\mathcal{S}(f_k)\|_{\mathcal{G}} \;\le\; \varepsilon\,\|f\|_{\mathcal{F}_\gamma} \quad \text{for all } f\in\mathcal{F}_\gamma$$

Our concept of **Truncation Dimension**

is different than the one in Statistics.

If $\|\mathcal{S}(f)-\mathcal{S}(f_k)\|_{\mathcal{G}} \le \varepsilon\,\|f\|_{\mathcal{F}}$ and $\|\mathcal{S}(f_k)-A_k(f_k)\|_{\mathcal{G}} \le \varepsilon\,\|f\|_{\mathcal{F}}$ then

$$\|\mathcal{S}(f)-A_k(f_k)\|_{\mathcal{G}} \le 2\,\varepsilon\,\|f\|_{\mathcal{F}}$$

Hence **the smaller $\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$ the better**

THEOREM 1 *For product weights* $\gamma_{\mathfrak{w}} = \prod_{j\in\mathfrak{w}} j^{-\beta}$

$$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon) = O\left(\varepsilon^{-1/(\beta-1+1/p)}\right) \quad \textit{for } p>1,$$

and

$$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon) = O\left(\varepsilon^{-1/\beta}\right) \quad \textit{for } p=1.$$

Values of $\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$ for **Standard Integration Problem**
with $p = 1$ and $\gamma_{\mathfrak{w}} = \prod_{j \in \mathfrak{w}} j^{-\beta}$

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	
$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$	2	9	31	99	316	$\beta = 2$
$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$	2	4	9	21	46	$\beta = 3$
$\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$	1	3	5	9	17	$\beta = 4$

For instance, for the error demand $\varepsilon = 10^{-3}$ with $\beta = 4$,
only five variables instead of ∞-many!
Worst Case Error of QMC or Sparse Grids Methods is:

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

Arbitrary Decomposition

Theorem 1 holds true for

"Quasi"-Truncation Dimension defined by

$$\mathbf{q}-\mathbf{dim}^{\mathbf{trnc}}(\varepsilon) := \min\left\{k \;:\; \left\|\mathcal{S}(f)-\mathcal{S}\left(\sum_{\mathfrak{w}\subseteq 1:k} f_{\mathfrak{w}}\right)\right\|_{\mathcal{G}} \le \varepsilon\,\|f\|_{\mathcal{F}_\gamma}\right\}$$

of interest in Statistics (**ANOVA Decomposition**)

However for **non-Anchored Decomposition**

$$f(x_1,\dots,x_k,0,0,\dots) \neq \sum_{\mathfrak{w}\subseteq 1:k} f_{\mathfrak{w}}(\boldsymbol{x})$$

and small $\mathbf{q}-\mathbf{dim}^{\mathbf{trnc}}(\varepsilon)$ need not lead to efficient algorithms.

Unless..... we'll see it later.

Multivariate Decomposition Method

Introduced by [Kuo, Sloan, W., and Woźnakowski 2010]

MDM replaces

one ∞-variate integral

by

only few integrals

each with

only few variables

More precisely:

Given the **error demand** $\varepsilon > 0$,

construct an **"active set"** $\mathbf{Act}(\varepsilon)$

of subsets $\mathfrak{w}$ such that

$$\left\| \mathcal{S}\left(\sum_{\mathfrak{w} \notin \mathbf{Act}(\varepsilon)} f_{\mathfrak{w}} \right) \right\|_{\mathcal{G}} \le \frac{\varepsilon}{2} \, \|f\|_{\mathcal{F}_\gamma} \quad \text{for all } f \in \mathcal{F}.$$

Do nothing for $\mathcal{S}(f_{\mathfrak{w}})$ with $\mathfrak{w} \notin \mathbf{Act}(\varepsilon)$,

and **concentrate** on $\mathcal{S}(f_{\mathfrak{w}})$ with $\mathfrak{w} \in \mathbf{Act}(\varepsilon)$.

For $\mathfrak{w}\in\mathbf{Act}(\varepsilon)$, choose $n_{\mathfrak{w}}$ and algorithms $A_{\mathfrak{w},n_{\mathfrak{w}}}$

to approximate $\mathcal{S}(f_{\mathfrak{w}})$ such that

$$\sum_{\mathfrak{w}\in\mathbf{Act}(\varepsilon)} \|\mathcal{S}(f_{\mathfrak{w}}) - A_{\mathfrak{w},n_{\mathfrak{w}}}(f_{\mathfrak{w}})\|_{\mathcal{G}} \le \frac{\varepsilon}{2}\,\|f\|_{\mathcal{F}} \quad \text{for all } f\in\mathcal{F}.$$

The algorithms $A_{\mathfrak{w},n_{\mathfrak{w}}}$ could be $|\mathfrak{w}|$-variate

Sparse Grids Algorithms

Then the **MDM** given by

$$\mathcal{A}_\varepsilon(f) := \sum_{\mathfrak{w}\in\mathbf{Act}(\varepsilon)} A_{\mathfrak{w},n_{\mathfrak{w}}}(f_{\mathfrak{w}})$$

has the **"worst case error"** bounded by ε, i.e.,

$$\|\mathcal{S}(f)-\mathcal{A}_\varepsilon(f)\|_{\mathcal{G}} \le \varepsilon\,\|f\|_{\mathcal{F}} \quad \text{for all } f\in\mathcal{F}.$$

Then the **MDM** given by

$$\mathcal{A}_\varepsilon(f) := \sum_{\mathfrak{w} \in \mathbf{Act}(\varepsilon)} A_{\mathfrak{w}, n_{\mathfrak{w}}}(f_{\mathfrak{w}})$$

has the **"worst case error"** bounded by ε, i.e.,

$$\|\mathcal{S}(f) - \mathcal{A}_\varepsilon(f)\|_{\mathcal{G}} \le \varepsilon \, \|f\|_{\mathcal{F}} \quad \text{for all } f \in \mathcal{F}.$$

How about the **COST**?

The number of $\mathcal{S}(f_{\mathfrak{w}})$ to approximate is small:

$$\mathbf{card}(\varepsilon) := |\mathbf{Act}(\varepsilon)| = \mathcal{O}\left(\varepsilon^{-1/r}\right),$$

where r is regularity degree (later).

Each $f_{\mathfrak{w}}$ depends on only $|\mathfrak{w}|$ variables.

The largest number of variables is also small:

$$\mathbf{dim}(\mathbf{Act}(\varepsilon)) := \max\{|\mathfrak{w}| \ : \ \mathfrak{w} \in \mathbf{Act}(\varepsilon)\} = \mathcal{O}(???)$$

The number of $\mathcal{S}(f_{\mathfrak{w}})$ to approximate is small:

$$\mathbf{card}(\varepsilon) := |\mathbf{Act}(\varepsilon)| = \mathcal{O}\left(\varepsilon^{-1/r}\right),$$

where r is regularity degree (later).

Each $f_{\mathfrak{w}}$ depends on only $|\mathfrak{w}|$ variables.
The largest number of variables is also small:

$$\mathbf{dim}(\mathbf{Act}(\varepsilon)) := \max\{|\mathfrak{w}| \ : \ \mathfrak{w} \in \mathbf{Act}(\varepsilon)\} = \mathcal{O}(???)$$

Theorem 2 *For product order dependent weights, there is* $\mathbf{Act}(\varepsilon)$ *with*

$$\mathbf{dim}(\mathbf{Act}(\varepsilon)) = O\left(\frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))}\right)$$

Proof like in [Plaskota and W. 2011]

This leads to

Worst Case ε-Superposition Dimension

$$\mathbf{dim}^{\mathbf{sprp}}(\varepsilon) := \inf_{\mathbf{Act}(\varepsilon)} \mathbf{dim}(\mathbf{Act}(\varepsilon))$$

Theorem 2 provides upper bounds on $\dim^{\mathrm{sprp}}(\varepsilon)$

[Gilbert, Kuo, Nuyens, and W. 2018] Efficient implementation of **MDM** for **Product Order Dependent Weights**

[Gilbert and W. 2017] Very efficient algorithm to construct **optimal** $\mathbf{Act}(\varepsilon)$ for **Product Weights**

This yields exact value of $\dim^{\mathrm{sprp}}(\varepsilon)$

Illustration for **Standard Integration Problem:**

Values of $\dim^{\text{sprp}}(\varepsilon)$ and $\text{card}(\varepsilon)$

for $p = 1$ and $\gamma_{\mathfrak{w}} = \prod_{j\in\mathfrak{w}} j^{-\beta}$

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	
	2 , 6	3 , 22	4 , 113	4 , 534	5 , 2424	$\beta = 2$
	2 , 6	2 , 8	3 , 22	3 , 68	4 , 192	$\beta = 3$
	1 , 2	2 , 6	2 , 10	3 , 26	3 , 50	$\beta = 4$

For instance, for $\varepsilon = 10^{-3}$ with $\beta = 4$

it is sufficient to approximate

10 integrals with at most 2 variables!

Active Set $\mathrm{Act}(10^{-3})$

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

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

REMARK: We do **NOT** know $f_{\mathfrak{w}}$ terms. However, we can sample them: due to [Kuo, Sloan, W., and Woźniakowski 2010b]

$$f_{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}}) = \sum_{\mathfrak{v} \subseteq \mathfrak{w}} (-1)^{|\mathfrak{w}|-|\mathfrak{v}|} f([\boldsymbol{x}_{\mathfrak{v}}; \mathbf{0}])$$

requires

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

REMARK: We do **NOT** know $f_{\mathfrak{w}}$ terms. However, we can sample them: due to [Kuo, Sloan, W., and Woźniakowski 2010b]

$$f_{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}}) = \sum_{\mathfrak{v}\subseteq\mathfrak{w}} (-1)^{|\mathfrak{w}|-|\mathfrak{v}|} f([\boldsymbol{x}_{\mathfrak{v}};\mathbf{0}])$$

requires

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

but from [Plaskota and W. 2011]

$$2^{|\mathfrak{w}|} = O\left(\varepsilon^{\frac{-1}{\ln(\ln(1/\varepsilon))}}\right) = o\left(\varepsilon^{-\delta}\right) \quad \text{for any } \delta > 0.$$

Theorem 3: *Suppose* $F_{\mathfrak{w}}$ *are* $|\mathfrak{w}|$*-fold tensor product of* F_1 *and for* $d = 1$ *there are algorithms* A_n *with*

$$\mathbf{error}(A_n; F_1) = O(n^{-r}) \quad r - \textit{regularity degree.}$$

Then for **POD** *weights* $\gamma_{\mathfrak{w}} = (|\mathfrak{w}|!)^{\alpha} \prod_{j \in \mathfrak{w}} C/j^{-\beta}$, *there are* **MDM***ethods* $\mathcal{A}_\varepsilon$ *such that*

$$\mathbf{error}(\mathcal{A}; \mathcal{F}_\gamma) \le \varepsilon \quad \textit{and} \quad \mathbf{cost}(\mathcal{A}_\varepsilon) \le c_\delta\, \varepsilon^{-\kappa-\delta}$$

for and $\delta > 0$, *where*

$$\kappa = \max\left(\frac{1}{r}, \frac{1}{\beta - 1/p^*}\right).$$

Theorem is sharp (modulo arbitrarily small $\delta > 0$):
Using a proof technique from
[Kuo, Sloan, W., and Woźniakowski 2010]

$$\mathbf{complexity}(\varepsilon; \mathcal{F}) = \Omega\left(\varepsilon^{-\kappa}\right)$$

Hence

$$\Omega\left(\varepsilon^{-\kappa}\right) = \mathbf{complexity}(\varepsilon; \mathcal{F}_\gamma) = O\left(\varepsilon^{-\kappa-\delta}\right) \quad \forall \delta > 0.$$

Recall that $\mathbf{complexity}(\varepsilon; \mathcal{F}_\gamma)$
is the smallest cost among all algorithms
with errors $\leq \varepsilon$

Efficient implementation of MDM in

[Gilbert, Kuo, Nuyens, and W. [18]]

The same value $f(x_{\mathfrak{v}})$ might be needed for approximating integrals of $f_{\mathfrak{w}}$ for $\mathfrak{v} \subseteq \mathfrak{w} \in \mathbf{Act}(\varepsilon)$.
In the implementation, we avoid such repeated samplings.

Application to Problems on Unbounded Domains

Univariate Functions:

$F_1 = W_{1,p,\psi}(\mathbb{R}_+)$ is the Banach space of

$$f(x) = \int_{\mathbb{R}_+} h(t)\,(x-t)_+^0 \,\mathrm{d}t \quad \text{for} \quad h \in L_{p,\psi}(\mathbb{R}_+)$$

with the norm

$$\|f\|_{F_1} = \|f'\,\psi\|_{L_p(\mathbb{R}_+)}$$

Clearly

$$f(0) = 0 \quad \text{and} \quad f' = h,$$

Here

$$\psi : D \to \mathbb{R}_+ \quad \text{is measurable and positive.}$$

Property: The faster the decay $\psi(x) \to 0$ as $x \to \infty$, the larger the space F_1

Univariate Integration:

$$\mathcal{S}|_{F_1}(f) = I_1(f) = \int_{\mathbb{R}_+} f(x)\,\rho(x)\mathrm{d}x$$

for a probability density function

$$\rho : \mathbb{R}_+ \to \mathbb{R}_+$$

[Kuo, Plaskota, and W. 2016] necessary and sufficient condition for

$$\|I_1\|_{F_1} = C_1 < \infty$$

and for a **special quadratures** to have errors $\simeq n^{-1}$, i.e., $r = 1$

Multivariate Case: $F_{\mathfrak{w}}$ is $|\mathfrak{w}|$-fold tensor product of F_1, i.e

$$f_{\mathfrak{w}}(\boldsymbol{x}) = \int_{\mathbb{R}_+^{|\mathfrak{w}|}} h(\boldsymbol{t}_{\mathfrak{w}}) \prod_{j\in\mathfrak{w}} (x_j - t_j)_+^0 \mathrm{d}\boldsymbol{t}_{\mathfrak{w}}, \quad h\,\psi^{\mathfrak{w}} \in L_p(\mathbb{R}_+^{|\mathfrak{w}|})$$

That is

$$\|f_{\mathfrak{w}}\|_{F_{\mathfrak{w}}} = \left\| \psi^{\mathfrak{w}} \prod_{j\in\mathfrak{w}} \frac{\partial}{\partial x_j} f_{\mathfrak{w}} \right\|_{L_p(\mathbb{R}_+^{|\mathfrak{w}|})} = \|h\,\psi^{\mathfrak{w}}\|_{L_p(\mathbb{R}_+^{|\mathfrak{w}|})}$$

Integration

$$\mathcal{S}|_{F_{\mathfrak{w}}}(f_{\mathfrak{w}}) = I_{\mathfrak{w}}(f_{\mathfrak{w}}) = \int_{\mathbb{R}_+^{|\mathfrak{w}|}} f_{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}})\,\rho^{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}})\mathrm{d}\boldsymbol{x}_{\mathfrak{w}}.$$

Here

$$\rho^{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}}) = \prod_{j\in\mathfrak{w}} \rho(x_j) \quad \text{and} \quad \psi^{\mathfrak{w}}(\boldsymbol{x}_{\mathfrak{w}}) = \prod_{j\in\mathfrak{w}} \psi(x_j)$$

∞-variate Integration

As in General Setting

$$\mathcal{F} \ni f \quad \text{iff} \quad f = \sum_{\mathfrak{w}} f_{\mathfrak{w}} \quad \text{and} \quad \|f\|_{\mathcal{F}} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \, \|f_{\mathfrak{w}}\|_{F_{\mathfrak{w}}}^{p} \right)^{1/p} < \infty.$$

Equivalently

$$\|f\|_{\mathcal{F}} = \left[\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \left\| \psi^{\mathfrak{w}} \, f^{(\mathfrak{w})}([\cdot_{\mathfrak{w}};\mathbf{0}]) \right\|_{L_p(\mathbb{R}^{\mathfrak{w}})}^{p} \right]^{1/p^*} .$$

Integration

$$\mathcal{S}(f) = \mathcal{I}(f) = \int_{\mathbb{R}_+^{\mathbb{N}}} f \, \rho^{\mathbb{N}} = \sum_{\mathfrak{w}} I_{\mathfrak{w}}(f_{\mathfrak{w}})$$

REMARK: For unbounded domains D and some spaces $\mathcal{F}$, function evaluation could be discontinuous functional. That is, for some $f \in \mathcal{F}$ and some $\boldsymbol{x} \in D^{\mathbb{N}}$ we do not have point-wise convergence of

$$\sum_{\mathfrak{w}} f_{\mathfrak{w}}(\boldsymbol{x})$$

REMEDY 1: Due to [Gnewuch, Mayer, and Ritter 2014]:
Suppose that $F_{\mathfrak{w}}$ are $|\mathfrak{w}|$-fold tensor products of a reproducing kernel Hilbert space $F_1 = H(k)$ with kernel k. Then under suitable assumption on the kernel k, there is a subset $\overline{D^{\mathbb{N}}} \subset D^{\mathbb{N}}$ such that $\overline{D^{\mathbb{N}}}$ has measure 1, and function sampling $f(\boldsymbol{x})$ is continuous for every $\boldsymbol{x} \in \overline{D^{\mathbb{N}}}$.

$\overline{D^{\mathbb{N}}}$ has a complicated structure
and the restriction to Hilbert spaces is essential.

REMEDY 2: We treat $\mathcal{F}$ as a Banach space of sequences

$$f = (f_{\mathfrak{w}})_{\mathfrak{w} \subset \mathbb{N}} \quad \text{with finite} \quad \|f\|_{\mathcal{F}} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \, \|f_{\mathfrak{w}}\|_{F_{\mathfrak{w}}}^{p} \right)^{1/p}$$

We only need to assume continuity of $\mathcal{I}$: $\quad \|\mathcal{I}\| < \infty$

Sampling at $\boldsymbol{x}$ with only finitely many nonzero x_j
and cubatures that use only such points
are well defined and continuous.

It follows from **Theorem 3** that the corresponding **MDM** have

$$\mathrm{error}(\mathcal{A}_\varepsilon; \mathcal{F}) \le \varepsilon \quad \text{and} \quad \mathrm{cost}(\mathcal{A}_\varepsilon) = O\left(\varepsilon^{-\max\left(1\,,\,\frac{1}{\beta - 1/p^*}\right) - \delta} \right)$$

How About ANOVA Spaces?

$$f \in \mathcal{F}^{\mathrm{AVOVA}} \quad \text{iff} \quad f(\boldsymbol{x}) = \sum_{\mathfrak{w}} f_{\mathfrak{w},\mathrm{A}}(\boldsymbol{x}_{\mathfrak{w}})$$

with

$$\int_D f_{\mathfrak{w},\mathrm{A}}(\boldsymbol{x}_{\mathfrak{w}})\,\mathrm{d}x_j = 0 \qquad \text{for any } j \in \mathfrak{w}$$

and, as before,

$$\|f\|_{\mathcal{F}^{\mathrm{ANOVA}}} = \left(\sum_{\mathfrak{w}} \gamma_{\mathfrak{w}}^{-p} \, \|f_{\mathfrak{w},\mathrm{A}}^{(\mathfrak{w})}\|_{L_p}^p \right)^{1/p} < \infty$$

ANOVA decomposition terms $f_{\mathfrak{w},\mathrm{A}}$
cannot be sampled, i.e.,
low truncation dimension and MDM might
not be applicable.

Even worse: the 'easiest' (constant) term
is **NOT known**;
and it is the integral we want to approximate:

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

HOWEVER

If the spaces are **EQUIVALENT**, then
efficient algorithms for anchored spaces
are also efficient for ANOVA spaces

Small truncation/superposition dimension for anchored spaces
are also small for ANOVA spaces.

This motivated the study of

Equivalence of anchored and ANOVA Spaces

For Product Weights $\gamma_{\mathfrak{w}} = \prod_{j\in\mathfrak{w}} j^{-\beta}$

$$\mathcal{F} = \mathcal{F}^{\mathrm{ANOVA}} \quad \text{as sets.}$$

For the imbedding $\imath : \mathcal{F} \hookrightarrow \mathcal{F}^{\mathrm{ANOVA}}$ we have

$$\|\imath\| = \|\imath^{-1}\| \le \prod_{j=1}^{\infty} \left(1 + j^{-\beta}\right)$$

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

General Information about f:

$$L_1(f), L_2(f), \ldots, L_n(f), \qquad L_j \in \mathcal{F}^*$$

Bayesian Approach:

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

$$\int_{\mathcal{F}} \|\mathcal{S}(f) - Alg(L_1(f), \ldots, L_n(f))\|_{\mathcal{G}}^2 \, \mathrm{PROB}(\mathrm{d}f)$$

or in **Probabilistic Case Setting** introduced in [Wasilkowski 1986]

THANK YOU FOR THE ATTENTION